4

Asymptotic Expansions of Solutions for Forward Equations

4.1 Introduction

This chapter is concerned with the analysis of the probability distributions of two-time-scale Markov chains. We aim to approximate the solution of forward equation by means of sequences of functions so that the desired accuracy is reached. As alluded to in Chapter 1, we devote our attention to nonstationary Markov chains with time-varying generators. A key feature here is time-scale separation. By introducing a small parameter $\varepsilon > 0$, the generator and hence the corresponding Markov chain have "two times," a usual running time t and a fast time t/ε. The main approach that we are using is the matched asymptotic expansions from singular perturbation theory. We first construct a sequence of functions that well approximate the solution of the forward equation when t is large enough (outside the initial layer of $O(\varepsilon)$). By adopting the notion of singular perturbation theory, this part of the approximation will be called outer expansions. We demonstrate that it is a good approximation as long as t is not in a neighborhood of 0 of the order $O(\varepsilon)$. Nevertheless, this sequence of functions does not satisfy the given initial condition and the approximation breaks down when $t \leq O(\varepsilon)$. To circumvent these difficulties, we construct another sequence of functions by magnifying the asymptotic behavior of the solution near 0 using the stretched fast time $\tau = t/\varepsilon$. Following the traditional terminology in singular perturbation theory, we call this sequence of functions initial-layer corrections (or sometimes, boundary-layer corrections). It effectively yields

corrections to the outer expansions and makes sure that the approximation is good in a neighborhood of $O(\varepsilon)$. By combining the outer expansions and the initial-layer corrections, we obtain a sequence of matched asymptotic expansions. The entire process is constructive. Our aims in this chapter include:

- Construct the outer expansions and the initial-layer corrections. This construction is often referred to as formal expansions.

- Justify the sequence of approximations obtained by deriving the desired error bounds. To achieve this, we show that (i) the outer solutions are sufficiently smooth, (ii) the initial-layer terms all decay exponentially fast, and (iii) the error is of the desired order. Thus not only is convergence of the asymptotic expansions proved, but also the error bound is obtained.

- Demonstrate that the error bounds hold uniformly. We would like to mention that in the usual singular perturbation theory, for example, in treating a linear system of differential equations, it is required that the system matrix be stable (i.e., all eigenvalues have negative real parts). In our setup, even for a homogeneous Markov chain, the generator (the system matrix in the equation) has an eigenvalue 0, so is not invertible. Thus, the stability requirement is violated. Nevertheless, using Markov properties, we are still able to obtain the desired asymptotic expansions.

Before proceeding further, we present a lemma. Let $Q(t) \in \mathbb{R}^{m \times m}$ be a generator, and let $\alpha(t)$ be a finite-state Markov chain with state space $\mathcal{M} = \{1, \ldots, m\}$ and generator $Q(t)$. Denote by

$$p(t) = (P(\alpha(t) = 1), \ldots, P(\alpha(t) = m)) \in \mathbb{R}^{1 \times m}$$

the row vector of the probability distribution of the underlying chain at time t. Then in view of Theorem 2.5, $p(\cdot)$ is a solution of the forward equation

$$\frac{dp(t)}{dt} = pQ(t) = p(t)Q(t),$$

$$p(0) = p^0 \text{ such that } p_i^0 \geq 0 \text{ for each } i, \text{ and } \sum_{i=1}^{m} p_i^0 = 1, \tag{4.1}$$

where $p^0 = (p_1^0, \ldots, p_m^0)$ and p_i^0 denotes the ith component of p^0. Therefore, studying the probability distribution is equivalent to examining the solution of (4.1). Note that the forward equation is linear, so the solution is unique. As a result, the following lemma is immediate. This lemma will prove useful in subsequent study.

Lemma 4.1. *The solution $p(t)$ of* (4.1) *satisfies the conditions*

$$0 \le p_i(t) \le 1 \ and \ \sum_{i=1}^{m} p_i(t) = 1. \qquad (4.2)$$

Remark 4.2. For the reader whose interests are mainly in differential equations, we point out that the initial condition $\sum_{i=1}^{m} p_i^0 = 1$ in (4.1) is not restrictive since if $p^0 = 0$, then $p(t) = 0$ is the only solution to (4.1). If $p_i^0 > 0$ for some i, one may divide both sides of (4.1) by $\sum_{i=1}^{m} p_i^0 \ (> 0)$ and consider $\widetilde{p}(t) = p(t)/\sum_{i=1}^{m} p_i^0$ in lieu of $p(t)$.

To achieve our goal, we first treat a simple case, namely, the case that the generator is weakly irreducible. Once this is established, we proceed to the more complex case that the generator has several weakly irreducible classes, the inclusion of absorbing states, and the inclusion of transient states.

The rest of the chapter is arranged as follows. Section 4.2 begins with the study of the situation in which the generator is weakly irreducible. Although it is a simple case, it outlines the main ideas behind the construction of asymptotic expansions. This section begins with the construction of formal expansions, proves the needed regularity, and ascertains the error estimates. Section 4.3 develops asymptotic expansions of the underlying probability distribution for the chains with recurrent states. As will be seen in the analysis to follow, extreme care must be taken to handle two-time-scale Markov chains with fast and slow components. One of the key issues is the selection of appropriate initial conditions to make the series a "matched" asymptotic expansions, in which the separable form of our asymptotic expansion appears to be advantageous compared with the two-time-scale expansions. For easy reference, a subsection is also provided as a user's guide.

Using the methods of matched asymptotic expansion, Section 4.4 extends the results to include absorbing states. It demonstrates that similar techniques can be used. We also demonstrate that the techniques and methods of Section 4.3 are rather general and can be applied to a wide variety of cases. Section 4.5 continues the study of problems involving transient states. By treating chains having recurrent states, chains including absorbing states, and chains including transient states, we are able to characterize the probability distributions of the underlying singularly perturbed chains of general cases with finite-state spaces, and hence provide comprehensive pictures through these "canonical" models.

While Sections 4.3–4.5 cover most practical concerns of interest for the finite-state-space cases, the rest of the chapter makes several remarks on Markov chains with countable-state spaces and two-time-scale diffusions. In Section 4.6.1, we extend the results to processes with countable-state spaces in which $\widetilde{Q}(t)$ is a block-diagonal matrix with infinitely many blocks each of which is finite-dimensional. Then Section 4.6.2 treats the problem in which $\widetilde{Q}(t)$ itself is an infinite-dimensional matrix. In this case,

further conditions are necessary. As in the finite-dimensional counterpart, sufficient conditions that ensure the validity of the asymptotic expansions are provided. The essential ingredients include Fredholm-alternative-like conditions and the notion of weak irreducibility. Finally, we mention related results of singularly perturbed diffusions in Section 4.7. Additional notes and remarks are given in Section 4.8.

4.2   Irreducible Case

We begin with the case concerning weakly irreducible generators. Let $Q(t) \in \mathbb{R}^{m \times m}$ be a generator, $\varepsilon > 0$ be a small parameter, and suppose that $\alpha^\varepsilon(t)$ is a finite-state Markov chain with state space $\mathcal{M} = \{1, \ldots, m\}$ generated by $Q^\varepsilon(t) = Q(t)/\varepsilon$. The row vector $p^\varepsilon(t) = (P(\alpha^\varepsilon(t) = 1), \ldots, P(\alpha^\varepsilon(t) = m)) \in \mathbb{R}^{1 \times m}$ denotes the probability distribution of the underlying chain at time t. Then by virtue of Theorem 2.5, $p^\varepsilon(\cdot)$ is a solution of the forward equation

$$\frac{dp^\varepsilon(t)}{dt} = p^\varepsilon Q^\varepsilon(t) = \frac{1}{\varepsilon} p^\varepsilon(t) Q(t),$$

$$p^\varepsilon(0) = p^0 \text{ such that } p_i^0 \geq 0 \text{ for each } i, \text{ and } \sum_{i=1}^m p_i^0 = 1, \tag{4.3}$$

where $p^0 = (p_1^0, \ldots, p_m^0)$ and p_i^0 denotes the ith component of p^0. Therefore, studying the probability distribution is equivalent to examining the solution of (4.3). Now, Lemma 4.1 continues to hold for the solution $p^\varepsilon(t)$.

As discussed in Chapters 1 and 3, the equation in (4.3) arises from various applications involving a rapidly fluctuating Markov chain governed by the generator $Q(t)/\varepsilon$. As ε gets smaller and smaller, the Markov chain fluctuates more and more rapidly. Normally, the fast-changing process $\alpha^\varepsilon(\cdot)$ in an actual system is difficult to analyze. The desired limit properties, however, provide us with an alternative. We can replace the actual process by its "average" in the system under consideration. This approach has significant practical value. A fundamental question common to numerous applications involving two-time-scale Markov chains is to understand the asymptotic properties of $p^\varepsilon(\cdot)$, namely, the limit behavior as $\varepsilon \to 0$. If $Q(t) = Q$, a constant matrix, and if Q is irreducible (see Definition 2.7), then for each $t > 0$, $p^\varepsilon(t) \to \nu$, the familiar stationary distribution. For the time-varying counterpart, it is reasonable to expect that the corresponding distribution will converge to a probability distribution that mimics the main features of the distribution of stationary chains, meanwhile preserving the time-varying nature of the nonstationary system. A candidate bearing such characteristics is the quasi-stationary distribution $\nu(t)$. Recall

that $\nu(t)$ is said to be a quasi-stationary distribution (see Definition 2.8) if $\nu(t) = (\nu_1(t), \ldots, \nu_m(t)) \geq 0$ and it satisfies the equations

$$\nu(t)Q(t) = 0 \text{ and } \sum_{i=1}^{m} \nu_i(t) = 1. \tag{4.4}$$

If $Q(t) \equiv Q$, a constant matrix, then an analytic solution of (4.3) is obtainable, since the fundamental matrix solution (see Hale [79]) takes the simple form $\exp(Qt)$; the limit behavior of $p^\varepsilon(t)$ is derivable through the solution $p^0 \exp(Qt/\varepsilon)$. For time-dependent $Q(t)$, although the fundamental matrix solution still exists, it does not have a simple form. The complex integral representation is not very informative in the asymptotic study of $p^\varepsilon(t)$, except in the case $m = 2$. In this case, $\alpha^\varepsilon(\cdot)$ is a two-state Markov chain and the constraint $p_1^\varepsilon(t) + p_2^\varepsilon(t) = 1$ reduces the current problem to a scalar one. Therefore, a closed-form solution is possible. However, such a technique cannot be generalized to $m > 2$. Let $0 < T < \infty$ be a finite real number. We divide the interval $[0, T]$ into two parts. One part is for t very close to 0 (in the range of an ε-layer), and the other is for t bounded away from 0. The behavior of $p^\varepsilon(\cdot)$ differs significantly in these two regions. Such a division led us to the utilization of the matched asymptotic expansion. Not only do we prove the convergence of $p^\varepsilon(t)$ as $\varepsilon \to 0$, but we also obtain an asymptotic series. The procedure involves constructing the regular part (outer expansion) for t to be away from 0, as well as the initial-layer corrections for small t, and to match these expansions by a proper choice of initial conditions.

In what follows, in addition to obtaining the zeroth-order approximation, i.e., the convergence of $p^\varepsilon(\cdot)$ to its quasi-stationary distribution, we derive higher-order approximations and error bounds. A consequence of the findings is that the convergence of the probability distribution and related occupation measures of the corresponding Markov chain takes place in an appropriate sense. The asymptotic properties of a suitably scaled occupation time and the corresponding central limit theorem for $\alpha^\varepsilon(\cdot)$ (based on the expansion) will be studied in Chapter 5.

4.2.1 Asymptotic Expansions

To proceed, we make the following assumptions.

(**A4.1**) Given $0 < T < \infty$, for each $t \in [0, T]$, $Q(t)$ is weakly irreducible, that is, the system of equations

$$\begin{aligned} f(t)Q(t) &= 0, \\ \sum_{i=1}^{m} f_i(t) &= 1 \end{aligned} \tag{4.5}$$

has a unique nonnegative solution.

(**A4.2**) For some n, $Q(\cdot)$ is $(n+1)$-times continuously differentiable on $[0,T]$, and $(d^{n+1}/dt^{n+1})Q(\cdot)$ is Lipschitz on $[0,T]$.

Remark 4.3. Condition (A4.2) requires that the matrix $Q(t)$ be sufficiently smooth. This is necessary for obtaining the desired asymptotic expansion. To validate the asymptotic expansion, we need to estimate the remainder term. Thus for the nth-order approximation, we need the $(n+1)$st-order smoothness.

To proceed, we first state a lemma. Its proof is in Lemma A.2 in the appendix.

Lemma 4.4. *Consider the matrix differential equation*

$$\frac{dP(s)}{ds} = P(s)A, \quad P(0) = I, \tag{4.6}$$

where $P(s) \in \mathbb{R}^{m \times m}$. Suppose $A \in \mathbb{R}^{m \times m}$ is a generator of a (homogeneous or stationary) finite-state Markov chain and is weakly irreducible. Then $P(s) \to \overline{P}$ as $s \to \infty$ and

$$\left| \exp(As) - \overline{P} \right| \leq K \exp(-\widetilde{\kappa}s) \quad \text{for some } \widetilde{\kappa} > 0, \tag{4.7}$$

where $\overline{P} = \mathbb{1}(\overline{\nu}_1, \cdots, \overline{\nu}_m) \in \mathbb{R}^{m \times m}$, and $(\overline{\nu}_1, \ldots, \overline{\nu}_m)$ is the quasi-stationary distribution of the Markov process with generator A.

Recall that $\mathbb{1} = (1, \ldots, 1)' \in \mathbb{R}^{m \times 1}$ and $(\overline{\nu}_1, \ldots, \overline{\nu}_m) \in \mathbb{R}^{1 \times m}$. Thus $\mathbb{1}(\overline{\nu}_1, \ldots, \overline{\nu}_m)$ is the usual matrix product. Recall that an $m \times m$ matrix $P(s)$ is said to be a solution of (4.6) if each row of $P(s)$ satisfies the equation. In the lemma above, if A is a constant matrix that is irreducible, then $(\overline{\nu}_1, \ldots, \overline{\nu}_m)$ becomes the familiar stationary distribution. In general, A could be time-dependent, e.g., $A = A(t)$. As shown in Lemma A.4, by assuming the existence of the solution $\nu(t)$ to (4.5), it follows that $\nu(t) \geq 0$; that is, the nonnegativity assumption is redundant. We seek asymptotic expansions of the form

$$p^\varepsilon(t) = \Phi_n^\varepsilon(t) + \Psi_n^\varepsilon\left(\frac{t}{\varepsilon}\right) + e_n^\varepsilon(t),$$

where $e_n^\varepsilon(t)$ is the remainder,

$$\Phi_n^\varepsilon(t) = \varphi_0(t) + \varepsilon\varphi_1(t) + \cdots + \varepsilon^n\varphi_n(t), \tag{4.8}$$

and

$$\Psi_n^\varepsilon\left(\frac{t}{\varepsilon}\right) = \psi_0\left(\frac{t}{\varepsilon}\right) + \varepsilon\psi_1\left(\frac{t}{\varepsilon}\right) + \cdots + \varepsilon^n\psi_n\left(\frac{t}{\varepsilon}\right), \tag{4.9}$$

with the functions $\varphi_i(\cdot)$ and $\psi_i(\cdot)$ to be determined in the sequel. We now state the main result of this section.

Theorem 4.5. *Suppose that* (A4.1) *and* (A4.2) *are satisfied. Denote the unique solution of* (4.3) *by* $p^\varepsilon(\cdot)$. *Then two sequences of functions* $\varphi_i(\cdot)$ *and* $\psi_i(\cdot)$, $0 \leq i \leq n$, *can be constructed such that*

(a) $\varphi_i(\cdot)$ *is* $(n+1-i)$-*times continuously differentiable on* $[0,T]$;

(b) *for each* i, *there is a* $\kappa_0 > 0$ *such that*

$$\left| \psi_i \left(\frac{t}{\varepsilon} \right) \right| \leq K \exp \left(-\frac{\kappa_0 t}{\varepsilon} \right);$$

(c) *the following estimate holds:*

$$\sup_{t \in [0,T]} \left| p^\varepsilon(t) - \sum_{i=0}^{n} \varepsilon^i \varphi_i(t) - \sum_{i=0}^{n} \varepsilon^i \psi_i \left(\frac{t}{\varepsilon} \right) \right| \leq K \varepsilon^{n+1}. \qquad (4.10)$$

Remark 4.6. The method described in what follows gives an explicit construction of the functions $\varphi_i(\cdot)$ and $\psi_i(\cdot)$ for $i \leq n$. Thus the proof to be presented is constructive. Our plan is first to obtain these sequences, and then validate properties (a) and (b) above and derive an error bound in (c) by showing that the remainder

$$\left| p^\varepsilon(t) - \sum_{i=0}^{n} \varepsilon^i \varphi_i(t) - \sum_{i=0}^{n} \varepsilon^i \psi_i \left(\frac{t}{\varepsilon} \right) \right|$$

is of order $O(\varepsilon^{n+1})$ uniformly in t.

It will be seen from the subsequent development that $\varphi_0(t)$ is equal to the quasi-stationary distribution, that is, $\varphi_0(t) = \nu(t)$. In particular, if $n = 0$ in the above theorem, we have the following result.

Corollary 4.7. *Suppose* $Q(\cdot)$ *is continuously differentiable on* $[0,T]$, *which satisfies* (A4.1), *and* $(d/dt)Q(\cdot)$ *is Lipschitz on* $[0,T]$. *Then for all* $t > 0$,

$$\lim_{\varepsilon \to 0} p^\varepsilon(t) = \nu(t) = \varphi_0(t), \qquad (4.11)$$

i.e., $p^\varepsilon(\cdot)$ *converges to the quasi-stationary distribution.*

Remark 4.8. The theorem manifests the convergence of $p^\varepsilon(\cdot)$ to $\varphi_0(\cdot)$, as well as the rate of convergence. In addition to the zeroth-order approximation, we have the first-order approximation, the second-order approximation, and so on. In fact, the difference $p^\varepsilon(\cdot) - \varphi_0(\cdot)$ is characterized by the initial-layer term $\psi_0(\cdot)$ and the associated error bound.

If the initial condition is chosen to be exactly equal to $p^0 = \varphi_0(0)$, then in the expansion, the zeroth-order initial layer $\psi_0(\cdot)$ will vanish. This cannot be expected in general, however. Even if $\psi_0(\cdot) = 0$, the rest of the initial-layer terms $\psi_i(\cdot)$, $i \geq 1$ will still be there.

To proceed, we define an operator $\mathcal{L}^\varepsilon$ by

$$\mathcal{L}^\varepsilon f = \varepsilon \frac{df}{dt} - fQ, \tag{4.12}$$

for any smooth row-vector-valued function $f(\cdot)$. Then $\mathcal{L}^\varepsilon f = 0$ iff f is a solution to the differential equation in (4.3). The proof of Theorem 4.5 is divided into the following steps.

1. Construct the asymptotic series, i.e., find $\varphi_i(\cdot)$ and $\psi_i(\cdot)$, for $i \leq n$. For the purpose of evaluating the remainder, we need to calculate two extra terms $\varphi_{n+1}(\cdot)$ and $\psi_{n+1}(\cdot)$. This will become clear when we carry out the error analysis.

2. Obtain the regularity of $\varphi_i(\cdot)$ and $\psi_i(\cdot)$ by proving that $\varphi_i(\cdot)$ is $(n + 1 - i)$-times continuously differentiable on $[0, T]$ and that $\psi_i(\cdot)$ decays exponentially fast.

3. Carry out the error analysis and justify that the remainder has the desired property.

4.2.2   Outer Expansion

We begin with the construction of $\Phi_n^\varepsilon(\cdot)$ in the asymptotic expansion. We call it the outer expansion or the regular part of expansion. Consider the differential equation

$$\mathcal{L}^\varepsilon \Phi_{n+1}^\varepsilon = 0$$

where $\mathcal{L}^\varepsilon$ is given by (4.12).

By equating the coefficients of ε^k, for $k = 1, \ldots, n + 1$, we obtain

$$\begin{aligned}
\varepsilon^0 : \quad & \varphi_0(t)Q(t) = 0, \\[2mm]
\varepsilon^1 : \quad & \varphi_1(t)Q(t) = \frac{d\varphi_0(t)}{dt}, \\[2mm]
& \cdots \\[2mm]
\varepsilon^k : \quad & \varphi_k(t)Q(t) = \frac{d\varphi_{k-1}(t)}{dt}, \quad \text{for } k = 1, \ldots, n + 1.
\end{aligned} \tag{4.13}$$

Remark 4.9. First, one has to make sure that the equations above have solutions, that is, a consistency condition needs to be verified. For each $t \in [0, T]$, denote the null space of $Q(t)$ by $N(Q(t))$. Note that the irreducibility of $Q(t)$ implies that

$$\text{rank}(Q(t)) = m - 1,$$

thus
$$\dim(N(Q(t))) = 1.$$

It is easily seen that $N(Q(t))$ is spanned by the vector $\mathbb{1}$. By virtue of the Fredholm alternative (see Corollary A.38), the second equation in (4.13) has a solution only if its right-hand side, namely, $(d/dt)\varphi_0(t)$ is orthogonal to $N(Q(t))$. Since $N(Q(t))$ is spanned by $\mathbb{1}$,

$$\varphi_0(t)\mathbb{1} = 1$$

and

$$\frac{d\varphi_0(t)}{dt}\mathbb{1} = \frac{d(\varphi_0(t)\mathbb{1})}{dt} = 0,$$

the orthogonality is easily verified. Similar arguments hold for the rest of the equations. The consistency in fact is rather crucial. Without such a condition, one would not be able to solve the equations in (4.13). This point will be made again when we deal with weak and strong interaction models in Section 4.3.

Recall that the components of $p^\varepsilon(\cdot)$ are probabilities (see (4.2)). In what follows, we show that all these $\varphi_i(\cdot)$ can be determined by (4.13) and (4.2).

Note that $\mathrm{rank}(Q(t)) = m - 1$. Thus $Q(t)$ is singular, and each equation in (4.13) is not uniquely solvable. For example, the first equation (4.13) cannot be solved uniquely. Nevertheless, this equation together with the constraint $\sum_{i=1}^{m} \varphi_0^i(t) = 1$ leads to a unique solution, namely, the quasi-stationary distribution.

In fact, a direct consequence of (A4.3) and (A4.4) is that the weak irreducibility of $Q(t)$ is uniform in the sense that for any $t \in [0, T]$, if *any* column of $Q(t)$ is replaced by $\mathbb{1} \in \mathbb{R}^{m \times 1}$, the resulting determinant $\Delta(t)$ satisfies $|\Delta(t)| > 0$, since (4.5) has only one solution, and $\sum_{j=1}^{m} q_{ij}(t) = 0$ for each $i = 1, \ldots, m$. Moreover, there is a number $c > 0$ such that $|\Delta(t)| \geq c > 0$. Thus, in view of the uniform continuity of $Q(t)$, $|\Delta(t)| \geq c > 0$ on $[0, T]$. We can replace any equation in the first m equations of the system $\varphi_0(t)Q(t) = 0$ by the equation $\sum_{i=1}^{m} \varphi_0^i(t) = 1$. The corresponding determinant $\Delta(t)$ of the resulting coefficient matrix satisfies $|\Delta(t)| \geq c > 0$, for some $c > 0$ and all $t \in [0, T]$. To illustrate, we may suppose without loss of generality that the mth equation is the one that can be replaced. Then we have

$$
\begin{aligned}
q_{11}(t)\varphi_0^1(t) + \cdots + q_{m1}(t)\varphi_0^m(t) &= 0, \\
q_{12}(t)\varphi_0^1(t) + \cdots + q_{m2}(t)\varphi_0^m(t) &= 0, \\
\cdots & \\
q_{1,m-1}(t)\varphi_0^1(t) + \cdots + q_{m,m-1}(t)\varphi_0^m(t) &= 0, \\
\varphi_0^1(t) + \cdots + \varphi_0^m(t) &= 1.
\end{aligned}
\tag{4.14}
$$

The determinant of the coefficient matrix in (4.14) is

$$
\Delta(t) =
\begin{vmatrix}
q_{11}(t) & q_{21}(t) & \cdots & q_{m1}(t) \\
q_{12}(t) & q_{22}(t) & \cdots & q_{m2}(t) \\
\vdots & \vdots & \cdots & \vdots \\
q_{1,m-1}(t) & q_{2,m-1}(t) & \cdots & q_{m,m-1}(t) \\
1 & 1 & \cdots & 1
\end{vmatrix}
\tag{4.15}
$$

and satisfies $|\Delta(t)| \geq c > 0$. Now by Cramer's rule, for each $0 \leq i \leq m$,

$$
\varphi_0^i(t) = \frac{1}{\Delta(t)}
\begin{vmatrix}
q_{11}(t) & \cdots & 0 & \cdots & q_{m1}(t) \\
q_{12}(t) & \cdots & 0 & \cdots & q_{m2}(t) \\
\vdots & \cdots & \vdots & \cdots & \vdots \\
q_{1,m-1}(t) & \cdots & 0 & \cdots & q_{m,m-1}(t) \\
1 & \cdots & \underbrace{1}_{i\text{th column}} & \cdots & 1
\end{vmatrix},
$$

that is, the ith column of $\Delta(t)$ in (4.15) is replaced by $(0, \ldots, 0, 1)' \in \mathbb{R}^{m \times 1}$. By the assumption of $Q(\cdot)$, it is plain that $\varphi_0(\cdot)$ is $(n+1)$-times continuously differentiable on $[0, T]$.

The foregoing method can be used to solve other equations in (4.13) analogously. Owing to the smoothness of $\varphi_0(\cdot)$, $(d/dt)\varphi_0(t)$ exists, and we can proceed to obtain $\varphi_1(\cdot)$. Repeat the procedure above, and continue inductively. For each $k \geq 1$,

$$
\begin{aligned}
\sum_{i=1}^{m} \varphi_k^i(t) q_{ij}(t) &= \frac{d\varphi_{k-1}^j(t)}{dt} \text{ for } j = 1, \ldots, m, \\
\sum_{i=1}^{m} \varphi_k^i(t) &= 0.
\end{aligned}
\tag{4.16}
$$

Note that $\varphi_{k-1}^j(\cdot)$ has been found so $(d/dt)\varphi_{k-1}^j(t)$ is a known function. After a suitable replacement of one of the first m equations by the last equation in (4.16), the determinant $\Delta(t)$ of the resulting coefficient matrix satisfies $|\Delta(t)| \geq c > 0$. We obtain for each $0 \leq i \leq m$,

$$
\varphi_k^i(t) = \frac{1}{\Delta(t)}
\begin{vmatrix}
q_{11}(t) & \cdots & \dfrac{d\varphi_{k-1}^1(t)}{dt} & \cdots & q_{m1}(t) \\
q_{12}(t) & \cdots & \dfrac{d\varphi_{k-1}^2(t)}{dt} & \cdots & q_{m2}(t) \\
\vdots & \cdots & \vdots & \cdots & \vdots \\
q_{1,m-1}(t) & \cdots & \dfrac{d\varphi_{k-1}^{m-1}(t)}{dt} & \cdots & q_{m,m-1}(t) \\
1 & \cdots & \underbrace{0}_{i\text{th column}} & \cdots & 1
\end{vmatrix}.
$$

Hence $\varphi_k(\cdot)$ is $(n+1-k)$-times continuously differentiable on $[0, T]$. Thus we have constructed a sequence of functions $\varphi_k(t)$ that are $(n+1-k)$-times continuously differentiable on $[0, T]$ for $k = 0, 1, \ldots, n + 1$.

Remark 4.10. The method used above is convenient for computational purposes. An alternative way of obtaining the sequence $\varphi_k(t)$ is as follows. For example, to solve

$$\varphi_0(t)Q(t) = 0, \quad \sum_{j=1}^{m} \varphi_0^j(t) = 1,$$

define $Q_c(t) = (\mathbb{1} \vdots Q(t)) \in \mathbb{R}^{m \times (m+1)}$. Then the equation above can be written as

$$\varphi_0(t)Q_c(t) = (1, 0, \ldots, 0).$$

Note that $Q_c(t)Q_c'(t)$ has full rank m owing to weak irreducibility. Thus the solution of the equation is

$$\varphi_0(t) = (1, 0, \ldots, 0)Q_c'(t)[Q_c(t)Q_c'(t)]^{-1}.$$

We can obtain all other $\varphi_k(t)$ for $k = 1, \ldots, n + 1$, similarly.

The regular part $\Phi_n^\varepsilon(\cdot)$ is a good approximation to $p^\varepsilon(\cdot)$ when t is bounded away from 0. When t approaches 0, an initial layer (or a boundary layer) develops and the approximation breaks down. To accommodate this situation, an initial-layer correction, i.e., a sequence of functions $\psi_k(t/\varepsilon)$ for $k = 0, 1, \ldots, n + 1$ needs to be constructed.

4.2.3   Initial-Layer Correction

This section is on the construction of the initial-layer terms. The presentation consists of two parts. We obtain the sequence $\{\psi_k(\cdot)\}$ in the first subsection, and derive the exponential decay property in the second subsection.

Construction of $\psi_k(\cdot)$. Following usual practice in singular perturbation theory, define the stretched (or rescaled) time variable by

$$\tau = \frac{t}{\varepsilon}. \tag{4.17}$$

Note that $\tau \to \infty$ as $\varepsilon \to 0$ for any given $t > 0$.

Consider the differential equation

$$\mathcal{L}^\varepsilon \Psi_{n+1}^\varepsilon = \sum_{i=0}^{n+1} \varepsilon^i \mathcal{L}^\varepsilon \psi_i = 0.$$

Using the stretched time variable τ, we arrive at

$$\frac{d\Psi^{\varepsilon}_{n+1}(\tau)}{d\tau} = \Psi^{\varepsilon}_{n+1}(\tau)Q(\varepsilon\tau).$$

Owing to the smoothness of $Q(\cdot)$, a truncated Taylor expansion about $\tau = 0$ leads to

$$Q(t) = Q(\varepsilon\tau) = \sum_{i=0}^{n+1} \frac{(\varepsilon\tau)^i}{i!} \frac{d^i Q(0)}{dt^i} + R_{n+1}(\varepsilon\tau),$$

where

$$R_{n+1}(t) = \frac{t^{n+1}}{(n+1)!} \left(\frac{d^{n+1}Q(\xi)}{dt^{n+1}} - \frac{d^{n+1}Q(0)}{dt^{n+1}} \right),$$

for some $0 < \xi < t$. In view of (A4.2),

$$R_{n+1}(t) = O(t^{n+2}) \text{ uniformly in } t \in [0, T].$$

Drop the term $R_{n+1}(t)$ and use the first $n + 2$ terms to get

$$\frac{d\Psi^{\varepsilon}_{n+1}(\tau)}{d\tau} = \Psi^{\varepsilon}_{n+1}(\tau) \left(\sum_{i=0}^{n+1} \frac{(\varepsilon\tau)^i}{i!} \frac{d^i Q(0)}{dt^i} \right).$$

Similar to the previous section, for $k = 1, \ldots, n + 1$, equating coefficients of ε^k, we have

$$\varepsilon^0 : \quad \frac{d\psi_0(\tau)}{d\tau} = \psi_0(\tau)Q(0),$$

$$\varepsilon^1 : \quad \frac{d\psi_1(\tau)}{d\tau} = \psi_1(\tau)Q(0) + \tau\psi_0(\tau)\frac{dQ(0)}{dt}, \tag{4.18}$$

$$\cdots$$

$$\varepsilon^k : \quad \frac{d\psi_k(\tau)}{d\tau} = \psi_k(\tau)Q(0) + r_k(\tau),$$

where $r_k(\tau)$ is a function having the form

$$r_k(\tau) = \frac{\tau^k}{k!}\psi_0(\tau)\frac{d^k Q(0)}{dt^k} + \cdots + \tau\psi_{k-1}(\tau)\frac{dQ(0)}{dt}$$
$$= \sum_{i=1}^{k} \frac{\tau^i}{i!}\psi_{k-i}(\tau)\frac{d^i Q(0)}{dt^i}. \tag{4.19}$$

These equations together with appropriate initial conditions allow us to determine the $\psi_k(\cdot)$'s. For constructing $\varphi_k(\cdot)$, a number of algebraic equations are solved, whereas when determining ψ_k, one has to solve a number of differential equations instead. Two points are worth mentioning in

connection with (4.18). First the time-varying differential equation is replaced by one with constant coefficients; the solution thus can be written explicitly. The second point is on the selection of the initial conditions for $\psi_k(\cdot)$, with $k = 0, 1, \ldots, n + 1$. We choose the initial conditions so that the initial data of the asymptotic expansion will "match" that of the differential equation (4.3). To be more specific,

$$\varphi_0(0) + \psi_0(0) = p^0, \text{ and}$$

$$\varphi_k(0) + \psi_k(0) = 0 \text{ for } k = 1, 2, \ldots, n + 1.$$

Corresponding to ε^0, solving

$$\frac{d\psi_0(\tau)}{d\tau} = \psi_0(\tau)Q(0),$$

$$\psi_0(0) = p^0 - \varphi_0(0),$$

where p^0 is the initial data given in (4.3), one has

$$\psi_0(\tau) = (p^0 - \varphi_0(0)) \exp\left(Q(0)\tau\right). \qquad (4.20)$$

Continuing in this fashion, for $k = 1, \ldots, n + 1$, we obtain

$$\frac{d\psi_k(\tau)}{d\tau} = \psi_k(\tau)Q(0) + r_k(\tau),$$

$$\psi_k(0) = -\varphi_k(0).$$

In the equations above, we purposely separated $Q(0)$ from the term $r_k(\tau)$. As a result, the equations are linear systems with a constant matrix $Q(0)$ and time-varying forcing terms. This is useful for our subsequent investigation.

For $k = 1, 2, \ldots$, the solutions are given by

$$\psi_k(\tau) = -\varphi_k(0) \exp(Q(0)\tau) \qquad (4.21)$$
$$+ \int_0^\tau r_k(s) \exp\left(Q(0)(\tau - s)\right) ds.$$

The construction of $\psi_k(\cdot)$ for $k = 0, 1, \ldots, n+1$, and hence the construction of the asymptotic series is complete.

4.2.4   Exponential Decay of $\psi_k(\cdot)$

This subsection concerns the exponential decay of $\psi_k(\cdot)$. At first glance, it seems to be troublesome since $Q(0)$ has a zero eigenvalue. Nevertheless, probabilistic argument helps us to derive the desired property. Two key points in the proof below are the utilization of orthogonality and repeated application of the approximation of $\exp(Q(0)\tau)$ in Lemma 4.4.

By virtue of Assumption (A4.1), the finite-state Markov chain generated by $Q(0)$ is weakly irreducible. Identifying $Q(0)$ with the matrix A in Lemma 4.4 yields that

$$\exp(Q(0)\tau) \to \overline{P} \text{ as } \tau \to \infty,$$

where $\overline{P} = \mathbb{1}\overline{\nu}$, and $\overline{\nu} = (\overline{\nu}_1, \ldots, \overline{\nu}_m)$ is the quasi-stationary distribution corresponding to the constant matrix $Q(0)$.

Proposition 4.11. *Under the conditions of Theorem 4.5, for each $0 \leq k \leq n+1$, there exist a nonnegative real polynomial $c_{2k}(\tau)$ of degree $2k$ and a positive number $\kappa_{0,0} > 0$ such that*

$$|\psi_k(\tau)| \leq c_{2k}(\tau) \exp(-\kappa_{0,0}\tau). \tag{4.22}$$

Proof: First of all, note that

$$\sum_{i=1}^{m} p_i^0 = 1 \text{ and } \sum_{i=1}^{m} \varphi_0^i(0) = 1.$$

It follows that

$$\sum_{i=1}^{m} \psi_0^i(0) = \sum_{i=1}^{m} p_i^0 - \sum_{i=1}^{m} \varphi_0^i(0) = 0.$$

That is, $\psi_0(0)$ is orthogonal to $\mathbb{1}$. Consequently, $\psi_0(0)\overline{P} = 0$ and by virtue of Lemma 4.4 (with $A = Q(0)$), for some $\kappa_{0,0} := \widetilde{\kappa} > 0$,

$$
\begin{aligned}
|\psi_0(\tau)| &= |\psi_0(0)\exp(Q(0)\tau)| \\
&\leq |\psi_0(0)\overline{P}| + |\psi_0(0)(\exp(Q(0)\tau) - \overline{P})| \\
&= |\psi_0(0)(\exp(Q(0)\tau) - \overline{P})| \leq K\exp(-\kappa_{0,0}\tau).
\end{aligned} \tag{4.23}
$$

Note that

$$Q(t)\mathbb{1} = 0 \text{ for all } t \geq 0.$$

Differentiating this equation repeatedly leads to

$$\frac{d^k Q(t)}{dt^k}\mathbb{1} = \frac{d^k(Q(t)\mathbb{1})}{dt^k} = 0.$$

Hence, it follows that

$$\frac{d^k Q(0)}{dt^k}\mathbb{1} = 0 \quad \text{and} \quad \frac{d^k Q(0)}{dt^k}\overline{P} = 0,$$

for each $0 \le k \le n+1$. Owing to Lemma 4.4 and (4.21),

$$|\psi_1(\tau)| \le |\varphi_1(0)\exp(Q(0)\tau)|$$

$$+ \left| \int_0^\tau \psi_0(s)\frac{dQ(0)}{dt}\left(\overline{P} + \left(\exp(Q(0)(\tau - s)) - \overline{P}\right)\right) s\, ds \right|$$

$$\le K\exp(-\kappa_{0,0}\tau)$$

$$+ \int_0^\tau |\psi_0(s)|\left|\frac{dQ(0)}{dt}\left(\exp(Q(0)(\tau - s)) - \overline{P}\right)\right| s\, ds$$

$$\le K\exp(-\kappa_{0,0}\tau) + K\int_0^\tau \exp(-\kappa_{0,0}s)\exp(-\kappa_{0,0}(\tau - s))s\, ds$$

$$\le K\exp(-\kappa_{0,0}\tau) + K\tau^2\exp(-\kappa_{0,0}\tau) \le c_2(\tau)\exp(-\kappa_{0,0}\tau),$$

for some nonnegative polynomial $c_2(\tau)$ of degree 2.

Note that $r_k(s)$ is orthogonal to $\overline{P}$. By induction, for any k with $k = 1, \ldots, n+1$,

$$|\psi_k(\tau)|$$

$$\le |\varphi_k(0)\exp(Q(0)\tau)| + \int_0^\tau \left|r_k(s)\left(\exp(Q(0)(\tau - s)) - \overline{P}\right)\right| ds$$

$$\le K\exp(-\kappa_{0,0}\tau) + \sum_{i=1}^k \frac{1}{i!}\int_0^\tau s^i |\psi_{k-i}(s)|$$

$$\times \left|\frac{d^i Q(0)}{dt^i}\left(\exp(Q(0)(\tau - s)) - \overline{P}\right)\right| ds$$

$$\le K\exp(-\kappa_{0,0}\tau) + K\sum_{i=1}^{2k-1}\int_0^\tau s^i \exp(-\kappa_{0,0}\tau)ds$$

$$\le K\exp(-\kappa_{0,0}\tau) + K\sum_{i=1}^{2k}\tau^i \exp(-\kappa_{0,0}\tau) \le c_{2k}(\tau)\exp(-\kappa_{0,0}\tau),$$

where $c_{2k}(\tau)$ is a nonnegative polynomial of degree $2k$. This completes the proof of the proposition. $\qquad\square$

Since n is a finite integer, the growth of $c_{2k}(\tau)$ for $0 \le k \le n+1$ is much slower than exponential. Thus the following corollary is in force.

Corollary 4.12. *For each $0 \leq k \leq n + 1$, with $\kappa_{0,0}$ given in Proposition 4.11,*

$$|\psi_k(\tau)| \leq K \exp\left(-\kappa_0 \tau\right), \text{ for some } \kappa_0 \text{ with } 0 < \kappa_0 < \kappa_{0,0}.$$

4.2.5 Asymptotic Validation

Recall that $\mathcal{L}^\varepsilon f = \varepsilon(d/dt)f - fQ$. Then we have the following lemma.

Lemma 4.13. *Suppose that for some $0 \leq k \leq n + 1$,*

$$\sup_{t \in [0,T]} |\mathcal{L}^\varepsilon v^\varepsilon(t)| = O\left(\varepsilon^{k+1}\right) \text{ and } v^\varepsilon(0) = 0.$$

Then

$$\sup_{t \in [0,T]} |v^\varepsilon(t)| = O\left(\varepsilon^k\right).$$

Proof: Let $\eta^\varepsilon(\cdot)$ be a function satisfying $\sup_{t \in [0,T]} |\eta^\varepsilon(t)| = O\left(\varepsilon^{k+1}\right)$. Consider the differential equation

$$\mathcal{L}^\varepsilon v^\varepsilon(t) = \eta^\varepsilon(t), \tag{4.24}$$
$$v^\varepsilon(0) = 0.$$

Then the solution of (4.24) is given by

$$v^\varepsilon(t) = \frac{1}{\varepsilon} \int_0^t \eta^\varepsilon(s) X^\varepsilon(t, s) ds,$$

where $X^\varepsilon(t, s)$ is a principal matrix solution. Recall that (see Hale [79, p. 80]) a fundamental matrix solution of the differential equation is an invertible matrix each row of which is a solution of the equation; a principal matrix solution is a fundamental matrix solution with initial value the identity matrix. In view of Lemma 4.1,

$$|X^\varepsilon(t, s)| \leq K \quad \text{for all } t, s \in [0, T].$$

Therefore, we have the inequalities

$$\sup_{t \in [0,T]} |v^\varepsilon(t)| \leq \frac{K}{\varepsilon} \sup_{t \in [0,T]} \int_0^t |\eta^\varepsilon(s)| ds \leq K\varepsilon^k.$$

The proof of the lemma is thus complete. $\qquad\square$

Recall that the vector-valued "error" or remainder $e_n^\varepsilon(t)$ is defined by

$$e_n^\varepsilon(t) = p^\varepsilon(t) - \sum_{i=0}^{n} \varepsilon^i \varphi_i(t) - \sum_{i=0}^{n} \varepsilon^i \psi_i\left(\frac{t}{\varepsilon}\right), \qquad (4.25)$$

where $p^\varepsilon(\cdot)$ is the solution of (4.3), and $\varphi_i(\cdot)$ and $\psi_i(\cdot)$ are constructed in (4.13) and (4.18). It remains to show that $e_n^\varepsilon(t) = O\left(\varepsilon^{n+1}\right)$. To do so, we utilize Lemma 4.13 as a bridge. It should be pointed out, however, that to get the correct order for the remainder, a trick involving "back up one step" is needed. The details follow.

Proposition 4.14. *Assume* (A4.1) *and* (A4.2), *for each* $k = 0, \ldots, n$,

$$\sup_{t \in [0,T]} |e_k^\varepsilon(t)| = O(\varepsilon^{k+1}).$$

Proof: We begin with

$$e_1^\varepsilon(t) = p^\varepsilon(t) - \varphi_0(t) - \varepsilon\varphi_1(t) - \psi_0\left(\frac{t}{\varepsilon}\right) - \varepsilon\psi_1\left(\frac{t}{\varepsilon}\right). \qquad (4.26)$$

We will use the exponential decay property given in $\psi_i(\tau)$ Corollary 4.12. Clearly, $e_1^\varepsilon(0) = 0$, and hence the condition of Lemma 4.13 on the initial data is satisfied. By virtue of the exponential decay property of $\psi_i(\cdot)$ in conjunction with the defining equations of $\varphi_i(\cdot)$ and $\psi_i(\cdot)$,

$$
\begin{aligned}
\mathcal{L}^\varepsilon e_1^\varepsilon(t) &= -\Bigg[\varepsilon\frac{d\varphi_0(t)}{dt} - \varphi_0(t)Q(t) + \varepsilon^2\frac{d\varphi_1(t)}{dt} - \varepsilon\varphi_1(t)Q(t) \\
&\quad + \varepsilon\frac{d}{dt}\psi_0\left(\frac{t}{\varepsilon}\right) - \psi_0\left(\frac{t}{\varepsilon}\right)Q(t) + \varepsilon^2\frac{d}{dt}\psi_1\left(\frac{t}{\varepsilon}\right) \\
&\quad - \varepsilon\psi_1\left(\frac{t}{\varepsilon}\right)Q(t)\Bigg] \\
&= -\varepsilon^2\frac{d\varphi_1(t)}{dt} + \psi_0\left(\frac{t}{\varepsilon}\right)\left[Q(t) - Q(0) - t\frac{dQ(0)}{dt}\right] \\
&\quad + \varepsilon\psi_1\left(\frac{t}{\varepsilon}\right)[Q(t) - Q(0)].
\end{aligned}
$$

For the term involving $\psi_0(t/\varepsilon)$, using a Taylor expansion on $Q(t)$ yields that for some $\xi \in (0,t)$

$$\left|Q(t) - Q(0) - t\frac{dQ(0)}{dt}\right| = \left|\frac{1}{2}\left(\frac{d^2Q(\xi)}{dt^2}\right)t^2\right| \le Kt^2.$$

Owing to the exponential decay property of $\psi_i(\cdot)$, the fact that $\varphi_1(\cdot)$ is n-times continuously differentiable on $[0,T]$, and the above estimate, we have

$$|\mathcal{L}^\varepsilon e_1^\varepsilon(t)| \le K\left(\varepsilon^2 + (\varepsilon t + t^2)\exp\left(-\frac{\kappa_0 t}{\varepsilon}\right)\right).$$

Moreover, for any $k = 0, 1, 2 \ldots, n + 1$, it is easy to see that

$$t^k \exp\left(-\frac{\kappa_0 t}{\varepsilon}\right) = \varepsilon^k \left(\frac{t}{\varepsilon}\right)^k \exp\left(-\frac{\kappa_0 t}{\varepsilon}\right) \leq K \varepsilon^k. \tag{4.27}$$

This implies $\mathcal{L}^\varepsilon e_1^\varepsilon(t) = O(\varepsilon^2)$ uniformly in t. Thus, $e_1^\varepsilon(t) = O(\varepsilon)$ by virtue of Lemma 4.13 and the bound is uniform in $t \in [0, T]$.

We now go back one step to show that the zeroth-order approximation also possesses the correct error estimate, that is, $e_0^\varepsilon(t) = O(\varepsilon)$. Note that the desired order seems to be difficult to obtain directly, and as a result the back-tracking is employed.

Note that

$$e_1^\varepsilon(t) = e_0^\varepsilon(t) - \varepsilon \varphi_1(t) - \varepsilon \psi_1\left(\frac{t}{\varepsilon}\right). \tag{4.28}$$

However, the smoothness of $\varphi_1(\cdot)$ and the exponential decay of $\psi_1(\cdot)$ imply that

$$\varepsilon \varphi_1(t) + \varepsilon \psi_1\left(\frac{t}{\varepsilon}\right) = O(\varepsilon) \quad \text{uniformly in } t. \tag{4.29}$$

Thus $e_0^\varepsilon(t) = O(\varepsilon)$ uniformly in t.

Proceeding analogously, we obtain

$$
\begin{aligned}
&\mathcal{L}^\varepsilon e_{n+1}^\varepsilon \\
&= \mathcal{L}^\varepsilon \left(p^\varepsilon(t) - \sum_{i=0}^{n+1} \varepsilon^i \varphi_i(t) - \sum_{i=0}^{n+1} \varepsilon^i \psi_i\left(\frac{t}{\varepsilon}\right) \right) \\
&= -\varepsilon \left(\sum_{i=0}^{n+1} \varepsilon^i \frac{d\varphi_i(t)}{dt} + \sum_{i=0}^{n+1} \varepsilon^i \frac{d}{dt} \psi_i\left(\frac{t}{\varepsilon}\right) \right) \\
&\quad + \left(\sum_{i=0}^{n+1} \varepsilon^i \varphi_i(t) + \sum_{i=0}^{n+1} \varepsilon^i \psi_i\left(\frac{t}{\varepsilon}\right) \right) Q(t) \\
&= -\varepsilon^{n+2} \frac{d\varphi_{n+1}(t)}{dt} + \left[\sum_{i=0}^{n+1} \varepsilon^i \varphi_i(t) Q(t) - \sum_{i=0}^{n} \varepsilon^{i+1} \varphi_{i+1}(t) Q(t) \right] \\
&\quad + \sum_{i=0}^{n+1} \varepsilon^i \psi_i\left(\frac{t}{\varepsilon}\right) Q(t) - \sum_{i=0}^{n+1} \varepsilon^i \left[\psi_i\left(\frac{t}{\varepsilon}\right) Q(0) + r_i\left(\frac{t}{\varepsilon}\right) \right].
\end{aligned}
\tag{4.30}
$$

Note that the term in the fifth line above is

$$\sum_{i=0}^{n+1} \varepsilon^i \varphi_i(t) Q(t) - \sum_{i=1}^{n+1} \varepsilon^i \varphi_i(t) Q(t) = \varphi_0(t) Q(t) = 0.$$

Using (4.19), we represent $r_i(t)$ in terms of $(d^i/dt^i)Q(0)$, etc. For the term involving $\psi_0(t/\varepsilon)$, using a truncated Taylor expansion up to order $(n+1)$

for $Q(t)$, by virtue of the Lipschitz continuity of $(d^{n+1}/dt^{n+1})Q(\cdot)$, there is a $\xi \in (0,t)$ such that

$$\left| Q(t) - \sum_{i=0}^{n+1} \frac{t^i}{i!} \frac{d^i Q(0)}{dt^i} \right| = \frac{1}{(n+1)!} \left| t^{n+1} \frac{d^{n+1}Q(\xi)}{dt^{n+1}} - t^{n+1} \frac{d^{n+1}Q(0)}{dt^{n+1}} \right|$$

$$\leq K t^{n+1} \xi \leq K t^{n+2}.$$

For all the other terms involving $\psi_i(t/\varepsilon)$, for $i = 1, \ldots, n+1$ in (4.30), we proceed as in the calculation of $\mathcal{L}^\varepsilon e_1^\varepsilon$. As a result, the last two terms in (4.30) are bounded by

$$\psi_0\left(\frac{t}{\varepsilon}\right) O(t^{n+2}) + \varepsilon \psi_1\left(\frac{t}{\varepsilon}\right) O(t^{n+1}) + \cdots + \varepsilon^{n+1} \psi_{n+1}\left(\frac{t}{\varepsilon}\right) O(t),$$

which in turn leads to the bound

$$K(t^{n+2} + \varepsilon t^{n+1} + \cdots + \varepsilon^{n+1} t) \exp\left(-\frac{\kappa_0 t}{\varepsilon}\right) \leq K \varepsilon^{n+2},$$

in accordance with (4.27). Moreover, it is clear that $e_{n+1}^\varepsilon(0) = 0$. In view of the fact that $\varphi_{n+1}(\cdot)$ is continuously differentiable on $[0, T]$ and $Q(\cdot)$ is $(n+1)$-times continuously differentiable on $[0, T]$, by virtue of Lemma 4.13, we infer that $e_{n+1}^\varepsilon(t) = O(\varepsilon^{n+1})$ uniformly in t. Since

$$e_{n+1}^\varepsilon(t) = e_n^\varepsilon(t) + O(\varepsilon^{n+1}),$$

it must be that $e_n^\varepsilon(t) = O(\varepsilon^{n+1})$. The proof of Proposition 4.14 is complete, and so is the proof of Theorem 4.5. □

Remark 4.15. In the estimate given above, we actually obtained

$$\mathcal{L}^\varepsilon e_k^\varepsilon(t) = O\left(\varepsilon^{k+1} + (\varepsilon t^k + \cdots + \varepsilon^k t) \exp\left(-\frac{\kappa_0 t}{\varepsilon}\right)\right). \tag{4.31}$$

This observation will be useful when we consider the unbounded interval $[0, \infty)$.

The findings reported are very useful for further study of the limit behavior of the corresponding Markov chain problems of central limit type, which will be discussed in the next chapter. In many applications, a system is governed by a Markov chain, which consists of both slow and fast motions. An immediate question is this: Can we still develop an asymptotic series expansion? This question will be dealt with in Section 4.3.

Suppose that in lieu of (A4.2), we assume that $Q(\cdot)$ is piecewise $(n+1)$-times continuously differentiable on $[0, T]$, and $(d^{n+1}/dt^{n+1})Q(\cdot)$ is piecewise Lipschitz, that is, there is a partition of $[0, T]$, namely,

$$t_0 = 0 < t_1 < t_2 < \cdots \leq t_k = T$$

such that $Q(\cdot)$ is $(n+1)$-times continuously differentiable and (d^{n+1}/dt^{n+1}) $Q(\cdot)$ is Lipschitz on each subinterval $[t_i, t_{i+1})$. Then the result obtained still holds. In this case, in addition to the initial layers, one also has a finite number of inner-boundary layers. In each interval $[t_i, t_{i+1} - \eta]$ for $\eta > 0$, the expansion is similar to that presented in Theorem 4.5.

4.2.6   Examples

As a further illustration, we consider two examples in this section. The first example is concerned with a stationary Markov chain, i.e., $Q(t) = Q$ is a constant matrix. The second example deals with an analytically solvable case for a two-state Markov chain with nonstationary transition probabilities. Although they are simple, these examples give us insight into the asymptotic behavior of the underlying systems.

Example 4.16. Let $\alpha^\varepsilon(t)$ be an m-state Markov chain with a constant generator $Q(t) = Q$ that is irreducible. This is an analytically solvable case, with

$$p^\varepsilon(t) = p^0 \exp\left(\frac{Qt}{\varepsilon}\right).$$

Using the technique of asymptotic expansion, we obtain

$$\varphi_0(t) + \psi_0\left(\frac{t}{\varepsilon}\right) = \varphi_0 + (p^0 - \varphi_0)\exp\left(\frac{Qt}{\varepsilon}\right),$$

$$\text{with } \exp\left(\frac{Qt}{\varepsilon}\right) \to \overline{P}, \text{ as } \varepsilon \to 0,$$

where

$$\varphi_0(t) = (\nu_1, \ldots, \nu_m) \text{ and } \overline{P} = \mathbb{1}\varphi_0.$$

Note that $p^0\overline{P} = \varphi_0$, and hence

$$(p^0 - \varphi_0)\exp\left(\frac{Qt}{\varepsilon}\right) = (p^0 - \varphi_0)\left[\exp\left(\frac{Qt}{\varepsilon}\right) - \overline{P}\right].$$

Moreover,

$$\varphi_i(t) \equiv 0, \quad \psi_i\left(\frac{t}{\varepsilon}\right) \equiv 0 \text{ for } i \geq 1.$$

In this case, $\varphi_0(t) \equiv \varphi_0$, a constant vector, which is the equilibrium distribution of Q; the series terminates. Moreover, the solution consists of two terms, one of them the equilibrium distribution (the zeroth-order approximation) and the other the zeroth-order initial-layer correction term. Since φ_0 is the quasi-stationary distribution,

$$\varphi_0 Q = 0 \text{ and } \varphi_0 \exp\left(\frac{Qt}{\varepsilon}\right) = \varphi_0.$$

Hence the analytic solution and the asymptotic expansion coincide.

In particular, let Q be a two-dimensional matrix, i.e.,

$$Q = \begin{pmatrix} -\lambda & \lambda \\ \mu & -\mu \end{pmatrix}.$$

Then setting

$$y_0^\varepsilon(t) = \varphi_0(t) + \psi_0(t/\varepsilon),$$

we have

$$p_1^\varepsilon(t) = y_{0,1}^\varepsilon(t) = \frac{\mu}{\lambda + \mu} + \left(p_1^0 - \frac{\mu}{\lambda + \mu} \right) \exp\left(-\frac{(\lambda + \mu)t}{\varepsilon} \right),$$

$$p_2^\varepsilon(t) = y_{0,2}^\varepsilon(t) = \frac{\lambda}{\lambda + \mu} + \left(p_2^0 - \frac{\lambda}{\lambda + \mu} \right) \exp\left(-\frac{(\lambda + \mu)t}{\varepsilon} \right).$$

Therefore,

$$\varphi_0(t) = \left(\frac{\mu}{\lambda + \mu}, \frac{\lambda}{\lambda + \mu} \right),$$

$$\psi_0\left(\frac{t}{\varepsilon} \right) = \left(\left(p_1^0 - \frac{\mu}{\lambda + \mu} \right), \left(p_2^0 - \frac{\lambda}{\lambda + \mu} \right) \right) \exp\left(-\frac{(\lambda + \mu)t}{\varepsilon} \right),$$

$$\varphi_i(t) \equiv 0 \text{ and } \psi_i\left(\frac{t}{\varepsilon} \right) \equiv 0 \quad \text{for } i \geq 1.$$

Example 4.17. Consider a two-state Markov chain with generator

$$Q(t) = \begin{pmatrix} -\lambda(t) & \lambda(t) \\ \mu(t) & -\mu(t) \end{pmatrix}$$

where $\lambda(t) \geq 0$, $\mu(t) \geq 0$ and $\lambda(t) + \mu(t) > 0$ for each $t \in [0, T]$. Therefore $Q(\cdot)$ is weakly irreducible. For the following discussion, assume $Q(\cdot)$ to be sufficiently smooth. Although it is time-varying, a closed-form solution is obtainable. Since $p_1^\varepsilon(t) + p_2^\varepsilon(t) = 1$ for each t, (4.3) can be solved explicitly and the solution is given by

$$p_1^\varepsilon(t) = p_1^0 \exp\left(-\frac{1}{\varepsilon} \int_0^t (\lambda(s) + \mu(s))ds \right)$$

$$+ \int_0^t \frac{\mu(u)}{\varepsilon} \exp\left(-\frac{1}{\varepsilon} \int_u^t (\lambda(s) + \mu(s))ds \right) du,$$

$$p_2^\varepsilon(t) = p_2^0 \exp\left(-\frac{1}{\varepsilon} \int_0^t (\lambda(s) + \mu(s))ds \right)$$

$$+ \int_0^t \frac{\lambda(u)}{\varepsilon} \exp\left(-\frac{1}{\varepsilon} \int_u^t (\lambda(s) + \mu(s))ds \right) du.$$

Following the approach in the previous sections, we construct the first a few terms in the asymptotic expansion. By considering (4.13) together with (4.2), a system of the form

$$\lambda(t)\varphi_0^1(t) - \mu(t)\varphi_0^2(t) = 0,$$

$$\varphi_0^1(t) + \varphi_0^2(t) = 1$$

is obtained. The solution of the system yields that

$$\varphi_0(t) = \left(\frac{\mu(t)}{\lambda(t) + \mu(t)}, \frac{\lambda(t)}{\lambda(t) + \mu(t)} \right).$$

To find $\varphi_1(\cdot)$, consider

$$\lambda(t)\varphi_1^1(t) - \mu(t)\varphi_1^2(t) = \frac{\dot{\lambda}(t)\mu(t) - \dot{\mu}(t)\lambda(t)}{(\lambda(t) + \mu(t))^2},$$

$$\varphi_1^1(t) + \varphi_1^2(t) = 0,$$

where $\dot{\lambda} = (d/dt)\lambda$ and $\dot{\mu} = (d/dt)\mu$. Solving this system of equations gives us

$$\varphi_1(t) = \left(\frac{\dot{\lambda}(t)\mu(t) - \dot{\mu}(t)\lambda(t)}{(\lambda(t) + \mu(t))^3}, \frac{\lambda(t)\dot{\mu}(t) - \mu(t)\dot{\lambda}(t)}{(\lambda(t) + \mu(t))^3} \right).$$

To get the inner expansion, consider the differential equation

$$\frac{d\psi_0(\tau)}{d\tau} = \psi_0(\tau)Q(0),$$

$$\psi_0(0) = p^0 - \varphi_0(0),$$

with $\tau = t/\varepsilon$. We obtain

$$\psi_0(\tau) = (p^0 - \varphi_0(0)) \exp(Q(0)\tau),$$

where

$$\exp(Q(0)\tau) = \frac{1}{\lambda(0) + \mu(0)}$$

$$\times \begin{pmatrix} \mu(0) + \lambda(0)e^{-(\lambda(0)+\mu(0))\tau} & \lambda(0) - \lambda(0)e^{-(\lambda(0)+\mu(0))\tau} \\ \mu(0) - \mu(0)e^{-(\lambda(0)+\mu(0))\tau} & \lambda(0) + \mu(0)e^{-(\lambda(0)+\mu(0))\tau} \end{pmatrix}.$$

Similarly $\psi_1(\cdot)$ can be obtained from (4.21) with the exponential matrix given above.

It is interesting to note that either $\lambda(t)$ or $\mu(t)$ can be equal to 0 for some t as long as $\lambda(t) + \mu(t) > 0$. For example, if we take $\mu(\cdot)$ to be the repair rate of a machine in a manufacturing model, then $\mu(t) = 0$ corresponds to the repair workers taking breaks or waiting for parts on order to arrive. The minors of $Q(t)$ are $\lambda(t)$, $-\lambda(t)$, $\mu(t)$, and $-\mu(t)$. As long as not all of them are zero at the same time, the weak irreducibility condition will be met.

4.2.7   Two-Time-Scale Expansion

The asymptotic expansion derived in the preceding sections is separable in the sense that it is the sum of a regular part and initial-layer corrections. Naturally, one is interested in the relationship between such an expansion and the so-called two-time-scale expansion (see, for example, Smith [199]). To answer this question, we first obtain the two-time-scale asymptotic expansion for the forward equation (4.3), proceed with the exploration of the relationships between these two expansions, and conclude with a discussion of the connection between these two methods.

Two-Time-Scale Expansion. Following the literature on asymptotic expansion (e.g., Kevorkian and Cole [108, 109] and Smith [199] among others), consider two scales t and $\tau = t/\varepsilon$, both as "times." One of them is in a normal time scale and the other is a stretched one. We seek asymptotic expansions of the form

$$y^\varepsilon(t, \tau) = \sum_{i=0}^{n} \varepsilon^i y_i(t, \tau), \qquad (4.32)$$

where $\{y_i(t, \tau)\}_{i=0}^{n}$ is a sequence of two-time-scale functions. Treating t and τ as independent variables, one has

$$\frac{d}{dt} = \frac{\partial}{\partial t} + \frac{1}{\varepsilon} \frac{\partial}{\partial \tau}. \qquad (4.33)$$

Formally substituting (4.32) into (4.3) and equating coefficients of like powers of ε^i results in

$$\frac{\partial y_0(t, \tau)}{\partial \tau} = y_0(t, \tau) Q(t),$$

$$\frac{\partial y_1(t, \tau)}{\partial \tau} = y_1(t, \tau) Q(t) - \frac{\partial y_0(t, \tau)}{\partial t},$$

$$\cdots$$

$$\frac{\partial y_i(t, \tau)}{\partial \tau} = y_i(t, \tau) Q(t) - \frac{\partial y_{i-1}(t, \tau)}{\partial t}, \quad 1 \leq i \leq n.$$

$$(4.34)$$

The initial conditions are

$$y_0(t, 0) = p^0 \quad \text{and}$$

$$y_i(t, 0) = 0, \quad \text{for } 1 \leq i \leq n. \tag{4.35}$$

Holding t constant and solving the first equation in (4.34) (with the first equation in (4.35) as the initial condition) yields

$$y_0(t, \tau) = p^0 \exp(Q(t)\tau). \tag{4.36}$$

By virtue of (A4.4), $(\partial/\partial t)y_0(t, \tau)$ exists and

$$\frac{\partial y_0(t, \tau)}{\partial t} = p^0 \exp(Q(t)\tau)\left(\frac{dQ(t)}{dt}\right)\tau.$$

As a result, $(\partial/\partial t)y_0(t, \tau)$ is orthogonal to $\mathbb{1}$. We continue the procedure recursively. It can be verified that for $1 \leq i \leq n$,

$$y_i(t, \tau) = -\int_0^\tau \frac{\partial y_{i-1}(t, s)}{\partial t} \exp(Q(t)(\tau - s))ds. \tag{4.37}$$

Furthermore, for $i = 1, \ldots, n$, $(\partial/\partial t)y_i(t, \tau)$ exists and is continuous; it is also orthogonal to $\mathbb{1}$. It should be emphasized that in the equations above, t is viewed as being "frozen," and as a consequence, $Q(t)$ is a "constant" matrix.

Parallel to the previous development, one can show that for all $1 \leq i \leq n$,

$$|y_i(t, \tau)| \leq K(t) \exp(-\kappa_0(t)\tau).$$

Compared with the separable expansions presented before, note the t-dependence of $K(\cdot)$ and $\kappa_0(\cdot)$ above. Furthermore, the asymptotic series is valid. We summarize this as the following theorem.

Theorem 4.18. *Under the conditions of Theorem 4.5, a sequence of functions $\{y_i(t, \tau)\}_{i=0}^n$ can be constructed so that*

$$\sup_{t \in [0,T]}\left|p^\varepsilon(t) - \sum_{i=0}^n \varepsilon^i y_i(t, \tau)\right| = O(\varepsilon^{n+1}).$$

Example 4.19. We return to Example 4.16. It is readily verified that the zeroth-order two-time-scale expansion coincides with that of the analytic solution, in fact, with

$$y_0(t, \tau) = p^0 \exp\left(\frac{Qt}{\varepsilon}\right) \quad \text{and } y_i(t, \tau) \equiv 0 \text{ for all } i \geq 1.$$

Relationship between the Two Methods. Now we have two different asymptotic expansions. Do they in some sense produce similar asymptotic results? Note that each term in $y_i(t, \tau)$ contains the regular part $\varphi_i(t)$ as well as the initial-layer corrections. Examining the zeroth-order approximation leads to

$$\exp(Q(t)\tau) \to \overline{P}(t) \text{ as } \tau \to \infty$$

via the same argument employed in the proof of Lemma 4.4. The matrix has identical rows, and is given by $\overline{P}(t) = \mathbb{1}\nu(t)$. In fact, owing to $p^0 \mathbb{1} = \sum_{i=1}^m p_i^0 = 1$, we have

$$y_0(t, \tau) = \nu(t) + p^0 \left(\exp(Q(t)\tau) - \overline{P}(t)\right) = \nu(t) + \widetilde{y}_0(t, \tau), \qquad (4.38)$$

where $\widetilde{y}_0(t, \tau)$ decays exponentially fast as $\tau \to \infty$ for $t < \tau$.

In view of (4.38), the two methods produce the same limit as $\tau \to \infty$, namely, the quasi-stationary distribution. To explore further, we study a special case (a two-state Markov chain) so as to keep the notation simple. Consider the two-state Markov chain model Example 4.17. In view of (4.38) and the formulas in Example 4.17, we have

$$y_0(t, \tau) = \nu(t) + \widetilde{y}_0(t, \tau) = \varphi_0(t) + \widetilde{y}_0(t, \tau).$$

Owing to (4.37), direct calculation yields that

$$y_1(t, \tau) = -\int_0^\tau \frac{d\varphi_0(t)}{dt} \exp(Q(t)(\tau - s)) ds$$

$$- \int_0^\tau \frac{\partial \widetilde{y}_0(t, \tau)}{\partial t} \exp(Q(t)(\tau - s)) ds.$$

It can be verified that the second term on the right-hand side of the equal sign above decays exponentially fast, while the first term yields $\varphi_1(t)$ plus another term tending to 0 exponentially fast as $\tau \to \infty$. Using the result of Example 4.17 yields

$$-\int_0^\tau \frac{d\varphi_0(t)}{dt} \exp(Q(t)(\tau - s)) ds$$

$$= \frac{d\varphi_0(t)}{dt} \left(\frac{1 - \exp(-(\lambda(t) + \mu(t))\tau)}{\lambda(t) + \mu(t)}\right) \begin{pmatrix} \lambda(t) & -\lambda(t) \\ -\mu(t) & \mu(t) \end{pmatrix}$$

$$= \varphi_1(t) - \frac{d\varphi_0(t)}{dt} \left(\frac{\exp(-(\lambda(t) + \mu(t))\tau)}{\lambda(t) + \mu(t)}\right) \begin{pmatrix} \lambda(t) & -\lambda(t) \\ -\mu(t) & \mu(t) \end{pmatrix}.$$

Thus, it follows that

$$y_1(t, \tau) = \varphi_1(t) + \widetilde{y}_1(t, \tau),$$

where

$$\widetilde{y}_1(t,\tau) = -\int_0^\tau \frac{\partial \widetilde{y}_0(t,\tau)}{\partial t} \exp(Q(t)(\tau - s))ds$$

$$-\frac{d\varphi_0(t)}{dt}\left(\frac{\exp(-(\lambda(t)+\mu(t))\tau)}{\lambda(t)+\mu(t)}\right)\begin{pmatrix}\lambda(t) & -\lambda(t) \\ -\mu(t) & \mu(t)\end{pmatrix}.$$

Similarly, we can obtain

$$y_i(t,\tau) = \varphi_i(t) + \widetilde{y}_i(t,\tau), \text{ for } 1 \le i \le n,$$

where $\widetilde{y}_i(t,\tau)$ decay exponentially fast as $\tau \to \infty$ for all $t < \tau$. This establishes the connection between these two different expansions.

Comparison and Additional Remark. A moment of reflection reveals that:

- The conditions required to obtain the asymptotic expansions are the same.
- Except for the actual forms, there is no significant difference between these two methods.
- No matter which method is employed, in one way or another the results for stationary Markov chains are used. In the separable expansion, this is accomplished by using $Q(0)$, and in the two-time-scale expansion, this is carried out by holding t constant and hence treating $Q(t)$ as a constant matrix.
- Although the two-time-scale expansion admits a seemingly more general form, the separable expansion is more transparent as far as the quasi-stationary distribution is concerned.
- When a more complex problem, for example the case of weak and strong interactions, is encountered, the separable expansion becomes more advantageous.
- To study asymptotic normality, etc., in the sequel, the separable expansion will prove to be more convenient than the two-time-scale expansion.

In view of the items mentioned above, we choose to use the separable form of the expansion throughout.

4.3 Markov Chains with Multiple Weakly Irreducible Classes

This section presents the asymptotic expansions of two-time-scale Markov chains with slow and fast components subject to weak and strong interactions. We assume that all the states of the Markov chain are recurrent. In contrast to Section 4.2, the states belong to multiple weakly

irreducible classes. As was mentioned in the introductory chapter, such time-scale separation stems from various applications in production planning, queueing networks, random fatigue, system reliability, competing risk theory, control and optimization of large-scale dynamical systems, and related fields. The sunderlying models in which some components change very rapidly whereas others vary relatively slowly, are more complex than those of Section 4.2. The weak and strong interactions of the systems are modeled by assuming the generator of the underlying Markov chain to be of the form

$$Q^\varepsilon(t) = \frac{1}{\varepsilon}\widetilde{Q}(t) + \widehat{Q}(t), \tag{4.39}$$

where $\widetilde{Q}(t)$ governs the rapidly changing part and $\widehat{Q}(t)$ describes the slowly changing components. They have the appropriate forms to be mentioned in the sequel.

This section extends the results in Section 4.2 to incorporate the cases in which the generator $\widetilde{Q}(t)$ is not irreducible. Our study focuses on the forward equation, similar to (4.3); now the forward equation takes the form

$$\frac{dp^\varepsilon(t)}{dt} = p^\varepsilon(t)\left(\frac{1}{\varepsilon}\widetilde{Q}(t) + \widehat{Q}(t)\right), \quad p^\varepsilon(0) = p^0 \tag{4.40}$$

such that

$$p_i^0 \geq 0 \text{ for each } i \text{ and } \sum_{i=1}^m p_i^0 = 1.$$

To illustrate, we present a simple example below.

Example 4.20. Consider a two-machine flowshop with machines that are subject to breakdown and repair. The production capacity of the machines is described by a finite-state Markov chain. If the machine is up, then it can produce parts with production rate $u(t)$; its production rate is zero if the machine is under repair. For simplicity, suppose each of the machines is either in operating condition (denoted by 1) or under repair (denoted by 0). Then the capacity of the workshop becomes a four-state Markov chain with state space $\{(1,1),(0,1),(1,0),(0,0)\}$. Suppose that the first machine breaks down much more often than the second one. To reflect this situation, consider a Markov chain $\alpha^\varepsilon(\cdot)$ generated by $Q^\varepsilon(t)$ in (4.39), with $\widetilde{Q}(\cdot)$ and $\widehat{Q}(\cdot)$ given by

$$\widetilde{Q}(t) = \begin{pmatrix} -\lambda_1(t) & \lambda_1(t) & 0 & 0 \\ \mu_1(t) & -\mu_1(t) & 0 & 0 \\ 0 & 0 & -\lambda_1(t) & \lambda_1(t) \\ 0 & 0 & \mu_1(t) & -\mu_1(t) \end{pmatrix}$$

and

$$\widehat{Q}(t) = \begin{pmatrix} -\lambda_2(t) & 0 & \lambda_2(t) & 0 \\ 0 & -\lambda_2(t) & 0 & \lambda_2(t) \\ \mu_2(t) & 0 & -\mu_2(t) & 0 \\ 0 & \mu_2(t) & 0 & -\mu_2(t) \end{pmatrix},$$

where $\lambda_i(\cdot)$ and $\mu_i(\cdot)$ are the rates of repair and breakdown, respectively. The matrices $\widetilde{Q}(t)$ and $\widehat{Q}(t)$ are themselves generators of Markov chains. Note that

$$\widetilde{Q}(t) = \text{diag}\left(\begin{pmatrix} -\lambda_1(t) & \lambda_1(t) \\ \mu_1(t) & -\mu_1(t) \end{pmatrix}, \begin{pmatrix} -\lambda_1(t) & \lambda_1(t) \\ \mu_1(t) & -\mu_1(t) \end{pmatrix} \right)$$

is a block-diagonal matrix, representing the fast motion, and $\widehat{Q}(t)$ governs the slow components. In order to obtain any meaningful results for controlling and optimizing the performance of the underlying systems, the foremost task is to determine the asymptotic behavior (as $\varepsilon \to 0$) of the probability distribution of the underlying chain.

In this example, a first glance reveals that $\widetilde{Q}(t)$ is reducible, hence the results in Section 4.2 are not applicable. However, closer scrutiny indicates that $\widetilde{Q}(t)$ consists of two irreducible submatrices. One expects that the asymptotic expansions may still be established. Our main objective is to develop asymptotic expansions of such systems and their variants. The corresponding procedure is, however, much more involved compared with the irreducible cases.

Examining (4.39), it is seen that the asymptotic properties of the underlying Markov chains largely depend on the structure of the matrix $\widetilde{Q}(t)$. In accordance with the classification of states, we may consider three different cases: the chains with recurrent states only, the inclusion of absorbing states, and the inclusion of transient states. We treat the recurrent-state cases in this section, and then extend the results to notationally more involved cases including absorbing states and transient states in the following two sections.

Suppose $\alpha^\varepsilon(\cdot)$ is a finite-state Markov chain with generator given by (4.39), where both $\widetilde{Q}(t)$ and $\widehat{Q}(t)$ are generators of appropriate Markov chains. In view of the results in Section 4.2, it is intuitively clear that the structure of the generator $\widetilde{Q}(t)$ governs the fast-changing part of the Markov chain. As mentioned in the previous section, our study of the finite-state-space cases is naturally divided into the recurrent cases, the inclusion of absorbing states, and the inclusion of transient states of the generator $\widetilde{Q}(t)$. In accordance with classical results (see Chung [31] and Karlin and Taylor [105, 106]), one can always decompose the states of a finite-state Markov chain into recurrent (including absorbing) and transient classes. Inspired by Seneta's approach to nonnegative matrices (see Seneta [189]), we aim to put the matrix $\widetilde{Q}(t)$ into some sort of "canonical" form so that a

systematic study can be carried out. In a finite-state Markov chain, not all states are transient, and it has at least one recurrent state. Similar to the argument of Iosifescu [95, p. 94] (see also Goodman [75], Karlin and Mc-Gregor [104], Keilson [107] among others), if there are no transient states, then after suitable permutations and rearrangements (i.e., by appropriately relabeling the states), $\widetilde{Q}(t)$ can be put into the block-diagonal form

$$\widetilde{Q}(t) = \begin{pmatrix} \widetilde{Q}^1(t) & & & \\ & \widetilde{Q}^2(t) & & \\ & & \ddots & \\ & & & \widetilde{Q}^l(t) \end{pmatrix} \qquad (4.41)$$

$$= \mathrm{diag}\left(\widetilde{Q}^1(t), \ldots, \widetilde{Q}^l(t)\right),$$

where $\widetilde{Q}^k(t) \in \mathbb{R}^{m_k \times m_k}$ are weakly irreducible, for $k = 1, 2, \ldots, l$, and $\sum_{k=1}^l m_k = m$. Here and hereinafter, $\widetilde{Q}^k(t)$, (a superscript without parentheses) denotes the kth block matrix in $\widetilde{Q}(t)$. The rest of this section deals with the generator $Q^\varepsilon(t)$ given by (4.39) with $\widetilde{Q}(t)$ taking the form (4.41). Note that an example of the recurrent case is that of the irreducible (or weakly irreducible) generators treated in Section 4.2.

Let $\mathcal{M}_k = \{s_{k1}, \ldots, s_{km_k}\}$ for $k = 1, \ldots, l$ denote the states corresponding to $\widetilde{Q}^k(t)$ and let $\mathcal{M}$ denote the state space of the underlying chains given by

$$\mathcal{M} = \mathcal{M}_1 \cup \cdots \cup \mathcal{M}_l$$

$$= \left\{s_{11}, \ldots, s_{1m_1}, \ldots, s_{l1}, \ldots, s_{lm_l}\right\}.$$

Since $\widetilde{Q}^k(t) = (\tilde{q}_{ij}^k(t))_{m_k \times m_k}$ and $\widehat{Q}(t) = (\hat{q}_{ij}(t))_{m \times m}$ are generators, for $k = 1, 2, \ldots, l$, we have

$$\sum_{j=1}^{m_k} \tilde{q}_{ij}^k(t) = 0, \quad \text{for } i = 1, \ldots, m_k, \quad \text{and}$$

$$\sum_{j=1}^{m} \hat{q}_{ij}(t) = 0, \quad \text{for } i = 1, \ldots, m.$$

The slow and fast components are coupled through weak and strong interactions in the sense that the underlying Markov chain fluctuates rapidly within a single group $\mathcal{M}_k$ and jumps less frequently between groups $\mathcal{M}_k$ and $\mathcal{M}_j$ for $k \neq j$. The states in $\mathcal{M}_k$, $k = 1, \ldots, l$, are not isolated or independent of each other. More precisely, if we consider the states in $\mathcal{M}_k$ as a single "state," then these "states" are coupled through the matrix $\widehat{Q}(t)$, and transitions from $\mathcal{M}_k$ to $\mathcal{M}_j$, $k \neq j$ are possible.

In fact, $\widehat{Q}(\cdot)$, together with the quasi-stationary distributions of $\widetilde{Q}^k(t)$, determines the transition rates among states in $\mathcal{M}_k$, for $k = 1, \ldots, l$.

Consider the forward equation (4.40). Our goal here is to develop an asymptotic series for the solution $p^\varepsilon(\cdot)$ of (4.40). Working with the interval $[0, T]$ for some $T < \infty$, we will need the following conditions:

(**A4.3**) For each $t \in [0, T]$ and $k = 1, 2, \ldots, l$, $\widetilde{Q}^k(t)$ is weakly irreducible.

(**A4.4**) For some positive integer n, $\widetilde{Q}(\cdot)$ and $\widehat{Q}(\cdot)$ are $(n + 1)$-times and n-times continuously differentiable on $[0, T]$, respectively. Moreover, $(d^{n+1}/dt^{n+1})\widetilde{Q}(\cdot)$ and $(d^n/dt^n)\widehat{Q}(\cdot)$ are Lipschitz on $[0, T]$.

Compared with the irreducible models in Section 4.2, the main difficulty in this chapter lies in the interactions among different blocks. In constructing the expansion in Section 4.2, for $i = 1, \ldots, n$, the two sets of functions $\{\varphi_i(\cdot)\}$ and $\{\psi_i(\cdot)\}$ are obtained independently except the initial conditions $\psi_i(0) = -\varphi_i(0)$. For Markov chains with weak and strong interactions, $\varphi_i(\cdot)$ and $\psi_i(\cdot)$ are highly intertwined. The essence is to find $\varphi_i(\cdot)$ and $\psi_i(\cdot)$ jointly and recursively. In the process of construction, one of the crucial and delicate points is to select the "right" initial conditions. This is done by demanding that $\psi_i(\tau)$ decay to 0 as $\tau \to \infty$. For future use, we define a differential operator $\mathcal{L}^\varepsilon$ on $\mathbb{R}^{1 \times m}$-valued functions by

$$\mathcal{L}^\varepsilon f = \varepsilon \frac{df}{dt} - f(\widetilde{Q} + \varepsilon \widehat{Q}). \tag{4.42}$$

Then it follows that $\mathcal{L}^\varepsilon f = 0$ iff f is a solution to the differential equation in (4.40). We are now in a position to derive the asymptotic expansion.

4.3.1   Asymptotic Expansions

As in Section 4.2, we seek expansions of the form

$$y_n^\varepsilon(t) = \Phi_n^\varepsilon(t) + \Psi_n^\varepsilon(t) = \sum_{i=0}^n \varepsilon^i \varphi_i(t) + \sum_{i=0}^n \varepsilon^i \psi_i\left(\frac{t}{\varepsilon}\right). \tag{4.43}$$

For the purpose of estimating the remainder (or error), the terms $\varphi_{n+1}(\cdot)$ and $\psi_{n+1}(\cdot)$ are needed. Set $\mathcal{L}^\varepsilon y_{n+1}^\varepsilon(t) = 0$. Parallel to the approach in

Section 4.2, equating like powers of ε^i (for $i = 0, 1, \ldots, n+1$) leads to the equations for the regular part:

$$\varepsilon^0 : \; \varphi_0(t)\widetilde{Q}(t) = 0,$$

$$\varepsilon^1 : \; \varphi_1(t)\widetilde{Q}(t) = \frac{d\varphi_0(t)}{dt} - \varphi_0(t)\widehat{Q}(t),$$

$$\cdots$$

$$\varepsilon^i : \; \varphi_i(t)\widetilde{Q}(t) = \frac{d\varphi_{i-1}(t)}{dt} - \varphi_{i-1}(t)\widehat{Q}(t).$$

(4.44)

As discussed in Section 4.2, the approximation above is good for t away from 0. When t is sufficiently close to 0, an initial layer of thickness ε develops. Thus for the singular part of the expansion we enlarge the picture near 0 using the stretched variable τ defined by $\tau = t/\varepsilon$. Identifying the initial-layer terms in $\mathcal{L}^\varepsilon y_{n+1}^\varepsilon = 0$, we obtain

$$\frac{d}{d\tau} \left(\psi_0(\tau) + \varepsilon\psi_1(\tau) + \cdots + \varepsilon^{n+1}\psi_{n+1}(\tau) \right)$$

$$= \left(\psi_0(\tau) + \varepsilon\psi_1(\tau) + \cdots + \varepsilon^{n+1}\psi_{n+1}(\tau) \right) \left(\widetilde{Q}(\varepsilon\tau) + \varepsilon\widehat{Q}(\varepsilon\tau) \right).$$

By means of the Taylor expansion, we have

$$\widetilde{Q}(\varepsilon\tau) = \widetilde{Q}(0) + \varepsilon\tau\frac{d\widetilde{Q}(0)}{dt} + \cdots$$

$$+ \frac{(\varepsilon\tau)^{n+1}}{(n+1)!}\frac{d^{n+1}\widetilde{Q}(0)}{dt^{n+1}} + \widetilde{R}_{n+1}(\varepsilon\tau),$$

$$\varepsilon\widehat{Q}(\varepsilon\tau) = \varepsilon\widehat{Q}(0) + \varepsilon^2\tau\frac{d\widehat{Q}(0)}{dt} + \cdots$$

$$+ \frac{\varepsilon(\varepsilon\tau)^n}{n!}\frac{d^n\widehat{Q}(0)}{dt^n} + \widehat{R}_n(\varepsilon\tau),$$

where

$$\widetilde{R}_{n+1}(t) = \frac{t^{n+1}}{(n+1)!}\left(\frac{d^{n+1}\widetilde{Q}(\xi)}{dt^{n+1}} - \frac{d^{n+1}\widetilde{Q}(0)}{dt^{n+1}} \right),$$

$$\widehat{R}_n(t) = \frac{\varepsilon t^n}{n!}\left(\frac{d^n\widehat{Q}(\zeta)}{dt^n} - \frac{d^n\widehat{Q}(0)}{dt^n} \right),$$

for some $0 \le \xi \le t$ and $0 \le \zeta \le t$. Note that in view of (A4.4),

$$\widetilde{R}_{n+1}(t) = O(t^{n+2}) \text{ and } \widehat{R}_n(t) = O(\varepsilon t^{n+1}).$$

Equating coefficients of like powers of ε^i, for $i = 0, 1, \ldots, n + 1$, and using the Taylor expansion above, we obtain

$$
\varepsilon^0 : \quad \frac{d\psi_0(\tau)}{d\tau} = \psi_0(\tau)\widetilde{Q}(0),
$$

$$
\varepsilon^1 : \quad \frac{d\psi_1(\tau)}{d\tau} = \psi_1(\tau)\widetilde{Q}(0)
$$
$$
+ \psi_0(\tau)\left(\widehat{Q}(0) + \tau\frac{d\widetilde{Q}(0)}{dt}\right),
$$

$$
\cdots
$$

$$
\varepsilon^i : \quad \frac{d\psi_i(\tau)}{d\tau} = \psi_i(\tau)\widetilde{Q}(0) + \sum_{j=0}^{i-1}\psi_{i-j-1}(\tau)
$$
$$
\times\left(\frac{\tau^j}{j!}\frac{d^j\widehat{Q}(0)}{dt^j} + \frac{\tau^{j+1}}{(j+1)!}\frac{d^{j+1}\widetilde{Q}(0)}{dt^{j+1}}\right).
$$

(4.45)

In view of the essence of matched asymptotic expansion, we have necessarily at $t = 0$ that

$$
\sum_{i=0}^{n+1}\varepsilon^i\left(\varphi_i(0) + \psi_i(0)\right) = p^0.
$$

(4.46)

This equation implies

$$
p^0 = \varphi_0(0) + \psi_0(0) \text{ and } \varphi_i(0) + \psi_i(0) = 0,
$$

for $i \geq 1$. Moreover, note that $p^\varepsilon(t)\mathbb{1} = 1$ for all $t \in [0, T]$. Sending $\varepsilon \to 0$ in the asymptotic expansion, one necessarily has to have the following conditions: For all $t \in [0, T]$,

$$
\varphi_0(t)\mathbb{1} = 1 \text{ and } \varphi_i(t)\mathbb{1} = 0, \ i \geq 1.
$$

(4.47)

Our task now is to determine the functions $\varphi_i(\cdot)$ and $\psi_i(\cdot)$.

Determining $\varphi_0(\cdot)$ and $\psi_0(\cdot)$. Write $v = (v^1, \ldots, v^l)$ for a vector $v \in \mathbb{R}^{1\times m}$, where v^k denotes the subvector corresponding to the kth block of the partition. Meanwhile, a superscript with parentheses denotes a sequence. Thus v_n^k denotes the kth subblock of the corresponding partitioned vector of the sequence v_n.

Let us start with the first equation in (4.44). In view of (4.47), we have

$$
\varphi_0(t)\widetilde{Q}(t) = 0,
$$
$$
\sum_{i=1}^m\varphi_0^i(t) = 1.
$$

(4.48)

Note that the system above depends only on the generator $\widetilde{Q}(t)$. However, by itself, the system is not uniquely solvable. Since for each $t \in [0, T]$ and $k = 1, \ldots, l$, $\widetilde{Q}^k(t)$ is weakly irreducible, it follows that $\text{rank}(\widetilde{Q}^k(t)) = m_k - 1$ and $\text{rank}(\widetilde{Q}(t)) = m - l$. Therefore, to get a unique solution, we need to supply l auxiliary equations. Where can we find these equations? Upon dividing the system (4.48) into l subsystems, one can apply the Fredholm alternative (see Lemma A.37 and Corollary A.38) and use the orthogonality condition to choose l additional equations to replace l equations in the system represented by the first equation in (4.48).

Since for each k, $\widetilde{Q}^k(t)$ is weakly irreducible, there exists a unique quasi-stationary distribution $\nu^k(t)$. Therefore any solution to the equation

$$\varphi_0^k(t)\widetilde{Q}^k(t) = 0$$

can be written as the product of $\nu^k(t)$ and a scalar "multiplier," say $\vartheta_0^k(t)$. It follows from the second equation in (4.48) that $\sum_{k=1}^{l} \vartheta_0^k(t) = 1$. These $\vartheta_0^k(t)$'s can be interpreted as the probabilities of the "grouped states" (or "aggregated states") $\mathcal{M}_k$.

As will be seen in the sequel, $\vartheta_0^k(t)$ becomes an important spinoff in the process of construction. Effort will subsequently be devoted to finding the unique solution $(\vartheta_0^1(t), \ldots, \vartheta_0^l(t))$. Let $\mathbb{1}_{m_k} = (1, \ldots, 1)' \in \mathbb{R}^{m_k \times 1}$.

Lemma 4.21. *Under* (A4.3) *and* (A4.4), *for each* $k = 1, \ldots, l$, *the solution of the equation*

$$
\begin{aligned}
\varphi_0^k(t)\widetilde{Q}^k(t) &= 0, \\[2mm]
\varphi_0^k(t)\mathbb{1}_{m_k} &= \vartheta_0^k(t),
\end{aligned}
\tag{4.49}
$$

can be uniquely expressed as $\varphi_0^k(t) = \nu^k(t)\vartheta_0^k(t)$, *where* $\nu^k(t)$ *is the quasi-stationary distribution corresponding to* $\widetilde{Q}^k(t)$. *Moreover,* $\varphi_0^k(t)$ *is* $(n+1)$-*times continuously differentiable on* $[0, T]$, *provided that* $\vartheta_0^k(\cdot)$ *is* $(n+1)$-*times continuously differentiable.*

Proof: For each k, let us regard $\vartheta_0^k(\cdot)$ as a known function temporarily. For $t \in [0, T]$, let $\widetilde{Q}_c^k(t) = (\mathbb{1}_{m_k} \vdots \widetilde{Q}^k(t))$. Then the solution can be written as

$$\varphi_0^k(t) = (\vartheta_0^k(0) \vdots 0'_{m_k})\widetilde{Q}_c^{k,\prime}(t) \left(\widetilde{Q}_c^k(t)\widetilde{Q}_c^{k,\prime}(t) \right)^{-1},$$

where $0_{m_k} = (0, \ldots, 0)' \in \mathbb{R}^{m_k \times 1}$. Moreover, $\varphi_0(\cdot)$ is $(n+1)$-times continuously differentiable. The lemma is thus concluded. $\qquad\square$

Remark 4.22. This lemma indicates that for each k, the subvector $\varphi_0^k(\cdot)$ is a multiple of the quasi-stationary distribution $\nu^k(\cdot)$ for each $k = 1, \ldots, l$. The multipliers $\vartheta_0^k(\cdot)$ are to be determined. Owing to the interactions

among different "aggregated states" corresponding to the block matrices, piecing together quasi-stationary distributions does not produce a quasi-stationary distribution for the entire system (i.e., $(\nu^1(t), \ldots, \nu^k(t))$ is not a quasi-stationary distribution for the entire system). Therefore, the leading term in the asymptotic expansion is proportional to (or a "multiple" of) the quasi-stationary distributions of the Markov chains generated by $\widetilde{Q}^k(t)$, for $k = 1, \ldots, l$. The multiplier $\vartheta_0^k(t)$ reflects the interactions of the Markov chain among the "aggregated states." The probabilistic meaning of the leading term $\varphi_0(\cdot)$ is in the sense of total probability. Intuitively, $\vartheta_0^k(t)$ is the corresponding probability of the chain belonging to $\mathcal{M}_k$, and $\varphi_0^k(t)$ is the probability distribution of the chain belonging to $\mathcal{M}_k$ and the transitions taking place within this group of states.

We proceed to determining $\vartheta_0^k(\cdot)$ for $k = 1, \ldots, l$. Define an $m \times l$ matrix

$$\widetilde{\mathbb{1}} = \begin{pmatrix} \mathbb{1}_{m_1} & & & \\ & \mathbb{1}_{m_2} & & \\ & & \ddots & \\ & & & \mathbb{1}_{m_l} \end{pmatrix} = \mathrm{diag}(\mathbb{1}_{m_1}, \ldots, \mathbb{1}_{m_l}).$$

A crucial observation is that $\widetilde{Q}(t)\widetilde{\mathbb{1}} = 0$, that is, $\widetilde{Q}(t)$ and $\widetilde{\mathbb{1}}$ are orthogonal. Thus postmultiplying by $\widetilde{\mathbb{1}}$ leads to

$$\mathcal{L}^\varepsilon \left(\sum_{i=0}^{n+1} \varepsilon^i \varphi_i(t) \widetilde{\mathbb{1}} \right) = 0.$$

Recall that

$$\varphi_0^k(t) = \vartheta_0^k(t)\nu^k(t) \text{ and } \varphi_0^k(t)\mathbb{1} = \vartheta_0^k(t).$$

Equating the coefficients of ε in the above equation yields

$$\frac{d}{dt}(\vartheta_0^1(t), \ldots, \vartheta_0^l(t)) = (\vartheta_0^1(t), \ldots, \vartheta_0^l(t))\overline{Q}(t), \qquad (4.50)$$

where

$$\overline{Q}(t) = \begin{pmatrix} \nu^1(t) & & & \\ & \nu^2(t) & & \\ & & \ddots & \\ & & & \nu^l(t) \end{pmatrix} \widehat{Q}(t)\widetilde{\mathbb{1}} \qquad (4.51)$$

$$= \mathrm{diag}(\nu^1(t), \ldots, \nu^l(t))\widehat{Q}(t)\widetilde{\mathbb{1}}.$$

Remark 4.23. Intuitively, $\overline{Q}(t)$ is the "average" of $\widehat{Q}(t)$ weighted by the collection of quasi-stationary distributions $(\nu^1(t), \ldots, \nu^l(t))$. In fact, (4.50)

is merely a requirement that the equations in (4.44) be consistent in the sense of Fredholm. This can be seen as follows. Denote by $N(\widetilde{Q}(t))$ the null space of the matrix $Q(t)$. Since $\text{rank}(\widetilde{Q}(t)) = m - l$, the dimension of $N(\widetilde{Q}(t))$ is l. Observe that $\widetilde{\mathbb{1}} = \text{diag}(\widetilde{\mathbb{1}}_{m_1}, \ldots, \widetilde{\mathbb{1}}_{m_l})$ where

$$
\begin{aligned}
\widetilde{\mathbb{1}}_{m_1} &= (\underbrace{1, \ldots, 1}_{m_1}, \underbrace{0, \ldots, 0}_{m_2 + \cdots + m_l})', \\
\widetilde{\mathbb{1}}_{m_2} &= (\underbrace{0, \ldots, 0}_{m_1}, \underbrace{1, \ldots, 1}_{m_2}, \underbrace{0, \ldots, 0}_{m_3 + \cdots + m_l})', \\
&\cdots \\
\widetilde{\mathbb{1}}_{m_l} &= (\underbrace{0, \ldots, 0}_{m_1 + \cdots + m_{l-1}}, \underbrace{1, \ldots, 1}_{m_l})'
\end{aligned}
\tag{4.52}
$$

are linearly independent and span the null space of $\widetilde{Q}(t)$. The equations in (4.44) have solutions only if the right-hand side of each equation is orthogonal to $\widetilde{\mathbb{1}}$. Hence, (4.50) must hold.

Next we determine the initial value $\vartheta_0(0)$. Assuming that the asymptotic expansion of $p^\varepsilon(\cdot)$ is given by $y_n^\varepsilon(\cdot)$ (see (4.43)), then it is necessary that

$$
\varphi_0(0)\widetilde{\mathbb{1}} = \lim_{\delta \to 0} \lim_{\varepsilon \to 0} p^\varepsilon(\delta)\widetilde{\mathbb{1}}.
\tag{4.53}
$$

We will refer to such a condition as an initial-value *consistency* condition. Moreover, in view of (4.40) and $\widetilde{Q}(t)\widetilde{\mathbb{1}} = 0$,

$$
p^\varepsilon(t)\widetilde{\mathbb{1}} = p^0\widetilde{\mathbb{1}} + \int_0^\delta p^\varepsilon(s)\widehat{Q}(s)ds\widetilde{\mathbb{1}}.
$$

Since $p^\varepsilon(\cdot)$ and $\widehat{Q}(\cdot)$ are both bounded, it follows that

$$
\lim_{\delta \to 0}\left(\limsup_{\varepsilon \to 0} \int_0^\delta p^\varepsilon(s)\widehat{Q}(s)ds\widetilde{\mathbb{1}}\right) = 0.
$$

Therefore, the initial-value consistency condition (4.53) yields

$$
\varphi_0(0)\widetilde{\mathbb{1}} = \lim_{\delta \to 0}\left(\lim_{\varepsilon \to 0} p^\varepsilon(\delta)\widetilde{\mathbb{1}}\right) = p^0\widetilde{\mathbb{1}}.
$$

Note that $(\vartheta_0^1(0), \ldots, \vartheta_0^l(0)) = \varphi_0(0)\widetilde{\mathbb{1}}$. So the initial value for $\vartheta_0(t)$ should be

$$
(\vartheta_0^1(0), \ldots, \vartheta_0^l(0)) = p^0\widetilde{\mathbb{1}}.
$$

Using this initial condition and solving (4.50) yields that

$$
(\vartheta_0^1(t), \ldots, \vartheta_0^l(t)) = p^0\widetilde{\mathbb{1}}X(t, 0),
$$

where $X(t, s)$ is the principal matrix solution of (4.50) (see Hale [79]). Since the smoothness of $X(\cdot, \cdot)$ depends solely on the smoothness properties of $\widetilde{Q}(t)$ and $\widehat{Q}(t)$, $(\vartheta_0^1(\cdot), \ldots, \vartheta_0^l(\cdot))$ is $(n+1)$-times continuously differentiable on $[0, T]$. Up to now, we have shown that $\varphi_0(\cdot)$ can be constructed that is $(n+1)$-times continuously differentiable on $[0, T]$. Set $\vartheta_0(t) = (\vartheta_0^1(t), \ldots, \vartheta_0^l(t))$. We now summarize the discussion above as follows:

Proposition 4.24. *Assume conditions* (A4.3) *and* (A4.4). *Then for* $t \in [0, T]$, $\varphi_0(t)$ *can be obtained uniquely by solving the following system of equations:*

$$\varphi_0^k(t)\widetilde{Q}^k(t) = 0,$$

$$\varphi_0^k(t)\mathbb{1}_{m_k} = \vartheta_0^k(t),$$

$$\frac{d\vartheta_0(t)}{dt} = \vartheta_0(t)\overline{Q}(t), \tag{4.54}$$

$$\text{with } \vartheta_0(0) = p^0\widetilde{\mathbb{1}},$$

such that $\varphi_0(\cdot)$ *is* $(n+1)$-*times continuously differentiable.* $\qquad\square$

We next consider the initial-layer term $\psi_0(\cdot)$. First note that solving (4.45) for each $i = 0, 1 \ldots, n+1$ leads to

$$\psi_0(\tau) = \psi_0(0) \exp(\widetilde{Q}(0)\tau),$$

$$\cdots$$

$$\psi_i(\tau) = \psi_i(0) \exp(\widetilde{Q}(0)\tau)$$
$$+ \sum_{j=0}^{i-1} \int_0^\tau \psi_{i-j-1}(s) \left(\frac{s^j}{j!} \frac{d^j \widehat{Q}(0)}{dt} + \frac{s^{j+1}}{(j+1)!} \frac{d^{j+1}\widetilde{Q}(0)}{dt^{j+1}} \right) \tag{4.55}$$

$$\times \exp(\widetilde{Q}(0)(\tau - s))ds.$$

Once again, to match the asymptotic expansion requires that (4.46) hold and hence

$$p^0 = p^\varepsilon(0) = \varphi_0(0) + \psi_0(0).$$

Solving the first equation in (4.45) together with the above initial condition, one obtains

$$\psi_0(\tau) = (p^0 - \varphi_0(0)) \exp(\widetilde{Q}(0)\tau). \tag{4.56}$$

Note that in Proposition 4.25 to follow, we still use $\kappa_{0,0}$ as a positive constant, which is generally a different constant from that in Section 4.2.

Proposition 4.25. *Assume conditions* (A4.3) *and* (A4.4). *Then* $\psi_0(\cdot)$ *can be obtained uniquely by* (4.56). *In addition, there is a positive number* $\kappa_{0,0}$ *such that*

$$|\psi_0(\tau)| \leq K \exp(-\kappa_{0,0}\tau), \ \tau \geq 0.$$

Proof: We prove only the exponential decay property, since the rest is obvious. Let $\nu^k(0)$ be the stationary distribution corresponding to the generator $\widetilde{Q}^k(0)$. Define

$$
\begin{aligned}
\pi &= \widetilde{\mathbb{1}}
\begin{pmatrix}
\nu^1(0) & & & \\
 & \nu^2(0) & & \\
 & & \ddots & \\
 & & & \nu^l(0)
\end{pmatrix} \\
&=
\begin{pmatrix}
\mathbb{1}_{m_1}\nu^1(0) & & & \\
 & \mathbb{1}_{m_2}\nu^2(0) & & \\
 & & \ddots & \\
 & & & \mathbb{1}_{m_l}\nu^l(0)
\end{pmatrix},
\end{aligned}
\tag{4.57}
$$

where

$$
\mathbb{1}_{m_k}\nu^k(0) =
\begin{pmatrix}
\nu_1^k(0) & \cdots & \nu_{m_k}^k(0) \\
 & \vdots & \\
\nu_1^k(0) & \cdots & \nu_{m_k}^k(0)
\end{pmatrix}.
$$

Noting the block-diagonal structure of $\widetilde{Q}(0)$, we have

$$
\exp(\widetilde{Q}(0)\tau) =
\begin{pmatrix}
\exp(\widetilde{Q}^1(0)\tau) & & & \\
 & \exp(\widetilde{Q}^2(0)\tau) & & \\
 & & \ddots & \\
 & & & \exp(\widetilde{Q}^l(0)\tau)
\end{pmatrix}.
$$

It is easy to see that

$$(p^0 - \varphi_0(0))\widetilde{\mathbb{1}} = p^0\widetilde{\mathbb{1}} - \varphi_0(0))\widetilde{\mathbb{1}} = p^0\widetilde{\mathbb{1}} - \vartheta_0(0) = 0.$$

Owing to the choice of initial condition, $(p^0 - \varphi_0(0))$ is orthogonal to π, and by virtue of Lemma 4.4, for each $k = 1, \ldots, l$ and some $\kappa_{0,k} > 0$,

$$\left| \exp(\widetilde{Q}^k(0)\tau) - \mathbb{1}_{m_k}\nu^k(0) \right| \leq K \exp(-\kappa_{0,k}\tau),$$

we have

$$|\psi_0(\tau)| = \left|(p^0 - \varphi_0(0))[\exp(\widetilde{Q}(0)\tau) - \pi]\right|$$

$$\le K \sup_{k \le l} \left|\exp(\widetilde{Q}^k(0)\tau) - \mathbb{1}_{m_k}\nu^k(0)\right|$$

$$\le K \exp(-\kappa_{0,0}\tau),$$

where $\kappa_{0,0} = \min_{k \le l} \kappa_{0,k} > 0$.   □

Determining $\varphi_i(\cdot)$ and $\psi_i(\cdot)$ for $i \ge 1$. In contrast to the situation encountered in Section 4.2, the sequence $\{\varphi_i(\cdot)\}$ cannot be obtained without the involvement of $\{\psi_i(\cdot)\}$. We thus obtain the sequences pairwise. While the determination of $\varphi_0(\cdot)$ and $\psi_0(\cdot)$ is similar to that of Section 4.2, the solutions for the rest of the functions show distinct features resulting from the underlying weak and strong interactions. With known

$$b_0(t) = \frac{d\varphi_0(t)}{dt} - \varphi_0(t)\widehat{Q}(t),$$

we proceed to solve the second equation in (4.44) together with the constraint $\sum_{i=1}^m \varphi_1^i(t) = 0$ due to (4.47). Partition the vectors $\varphi_1(t)$ and $b_0(t)$ as

$$\varphi_1(t) = (\varphi_1^1(t), \dots, \varphi_1^l(t)),$$

$$b_0(t) = (b_0^1(t), \dots, b_0^l(t)).$$

In view of the definition of $\overline{Q}(t)$ in (4.51) and $\varphi_0^k(t) = \nu^k(t)\vartheta_0^k(t)$, it follows that $b_0(t)\widetilde{\mathbb{1}} = 0$, thus,

$$b_0^k(t)\mathbb{1}_{m_k} = 0, \ k = 1, \dots, l.$$

Let $\vartheta_1^k(t)$ denote the function such that $\sum_{k=1}^l \vartheta_1^k(t) = 0$ because $\varphi_1(t)\mathbb{1} = 0$. Then for each $k = 1, \dots, l$, the solution to

$$\varphi_1^k(t)\widetilde{Q}^k(t) = b_0^k(t),$$

$$\varphi_1^k(t)\mathbb{1}_{m_k} = \vartheta_1^k(t),$$

(4.58)

can be expressed as

$$\varphi_1^k(t) = \widetilde{b}_0^k(t) + \vartheta_1^k(t)\nu^k(t),$$ (4.59)

where $\widetilde{b}_0^k(t)$ is a solution to the following equation:

$$\widetilde{b}_0^k(t)\widetilde{Q}^k(t) = b_0^k(t),$$

$$\widetilde{b}_0^k(t)\mathbb{1}_{m_k} = 0,$$

or equivalently,

$$\widetilde{b}_0^k(t)(\mathbb{1}_{m_k} \vdots \widetilde{Q}^k(t)) = (0 \vdots b_0^k(t)).$$

The procedure for solving this equation is similar to that for $\varphi_0(\cdot)$.

Analogously to the previous treatment, we proceed to determine $\vartheta_1^k(t)$ by solving the system of equations

$$\mathcal{L}^\varepsilon \left(\sum_{i=0}^{n+1} \varepsilon^i \varphi_i(t)\widetilde{\mathbb{1}} \right) = 0. \tag{4.60}$$

Using the conditions

$$\widetilde{b}_0^k(t)\mathbb{1}_{m_k} = 0 \text{ and } \nu^k(t)\mathbb{1}_{m_k} = 1,$$

we have

$$\varphi_1(t)\widetilde{\mathbb{1}} = (\vartheta_1^1(t), \ldots, \vartheta_1^l(t))$$

and

$$\varphi_1(t)\widehat{Q}(t)\widetilde{\mathbb{1}} = (\vartheta_1^1(t), \ldots, \vartheta_1^l(t))\overline{Q}(t) + (\widetilde{b}_0^1(t), \ldots, \widetilde{b}_0^l(t))\widehat{Q}(t)\widetilde{\mathbb{1}},$$

where $\overline{Q}(t)$ was defined in (4.51).

By equating the coefficients of ε^2 in (4.60), we obtain a system of linear inhomogeneous equations

$$\frac{d}{dt}(\vartheta_1^1(t), \ldots, \vartheta_1^l(t)) = (\vartheta_1^1(t), \ldots, \vartheta_1^l(t))\overline{Q}(t) \tag{4.61}$$
$$+ (\widetilde{b}_0^1(t), \ldots, \widetilde{b}_0^l(t))\widehat{Q}(t)\widetilde{\mathbb{1}},$$

with initial conditions

$$\vartheta_1^k(0), \text{ for } k = 1, 2, \ldots, l \text{ such that } \sum_{k=1}^{l} \vartheta_1^k(0) = 0.$$

Again, as observed in Remark 4.23, equation (4.61) comes from the consideration in the sense of Fredholm since the functions on the right-hand sides in (4.44) must be orthogonal to $\widetilde{\mathbb{1}}$.

The initial conditions $\vartheta_1^k(0)$ for $k = 1, \ldots, l$ have not been completely specified yet. We do this later to ensure the matched asymptotic expansion. Once the $\vartheta_1^k(0)$'s are given, the solution of the above equation is

$$(\vartheta_1^1(t), \ldots, \vartheta_1^l(t)) = (\vartheta_1^1(0), \ldots, \vartheta_1^l(0))X(t, 0)$$
$$+ \int_0^t (\widetilde{b}_0^1(s), \ldots, \widetilde{b}_0^l(s))\widehat{Q}(s)\widetilde{\mathbb{1}}X(t, s)ds.$$

Thus if the initial value $\vartheta_1^k(0)$ is given, then $\vartheta_1^k(\cdot)$, $k = 1, \ldots, l$ can be found, and so can $\varphi_1(\cdot)$. Moreover, $\varphi_1(\cdot)$ is n-times continuously differentiable on $[0, T]$. The problem boils down to finding the initial condition of $\vartheta_1(0)$.

So far, with the proviso of specified initial conditions $\vartheta_1^k(0)$, for $k = 1, \ldots, l$, the construction of $\varphi_1(\cdot)$ has been completed, and its smoothness has been established. Compared with the determination of $\varphi_0(\cdot)$, the multipliers $\vartheta_1^k(\cdot)$ can no longer be determined using the information about the regular part alone because its initial values have to be determined in conjunction with that of the singular part. This will be seen as follows.

In view of (4.55),

$$\psi_1(\tau) = \psi_1(0) \exp(\widetilde{Q}(0)\tau)$$
$$+ \int_0^\tau \psi_0(0) \exp(\widetilde{Q}(0)s)\widehat{Q}(0) \exp(\widetilde{Q}(0)(\tau - s))ds \qquad (4.62)$$
$$+ \int_0^\tau s\psi_0(0) \exp(\widetilde{Q}(0)s)\frac{d\widetilde{Q}(0)}{dt} \exp(\widetilde{Q}(0)(\tau - s))ds.$$

Recall that $\psi_1(0)$ has not been specified yet.

Similar to Section 4.2, for each $t \in [0, T]$, $\widetilde{Q}(t)\widetilde{\mathbb{1}} = 0$. Therefore,

$$\left(\frac{d^i\widetilde{Q}(t)}{dt^i}\right)\widetilde{\mathbb{1}} = 0 \quad \text{and} \quad \left(\frac{d^i\widetilde{Q}(0)}{dt^i}\right)\pi = 0,$$

for $i = 1, \ldots, n+1$, where π is defined in (4.57). This together with $\psi_0(0)\pi = 0$ yields

$$\left|\int_0^\tau s\psi_0(0) \exp(\widetilde{Q}(0)s)\frac{d\widetilde{Q}(0)}{dt} \exp(\widetilde{Q}(0)(\tau - s))ds\right|$$
$$\leq \int_0^\tau s\left|\psi_0(0)[\exp(\widetilde{Q}(0)s) - \pi]\right|$$
$$\times \left|\frac{d\widetilde{Q}(0)}{dt}[\exp(\widetilde{Q}(0)(\tau - s)) - \pi]\right|ds \qquad (4.63)$$
$$\leq K\tau^2 \exp(-\kappa_{0,0}\tau).$$

To obtain the desired property, we need only work with the first two terms on the right side of the equal sign of (4.62). Noting the exponential decay property of $\psi_0(\tau) = \psi_0(0)\exp(\widetilde{Q}(0)\tau)$, we have

$$\int_0^\infty \left|\psi_0(0)\exp(\widetilde{Q}(0)s)\right|ds < \infty,$$

that is, the improper integral converges absolutely. Set

$$\overline{\psi}_0 = \left(\int_0^\infty \psi_0(0)\exp(\widetilde{Q}(0)s)ds\right)\widehat{Q}(0) \in \mathbb{R}^{1\times m}. \tag{4.64}$$

Consequently,

$$\lim_{\tau\to\infty} \psi_1(0)\exp(\widetilde{Q}(0)\tau) = \psi_1(0)\pi \quad \text{and}$$

$$\begin{aligned}
\lim_{\tau\to\infty} &\int_0^\tau \psi_0(0)\exp(\widetilde{Q}(0)s)\widehat{Q}(0)\exp(\widetilde{Q}(0)(\tau-s))ds \\
&= \left(\int_0^\infty \psi_0(0)\exp(\widetilde{Q}(0)s)ds\right)\widehat{Q}(0)\pi \\
&:= \overline{\psi}_0\pi.
\end{aligned} \tag{4.65}$$

Recall that $\pi = \mathrm{diag}(\mathbb{1}_{m_1}\nu^1(0),\ldots,\mathbb{1}_{m_l}\nu^l(0))$. Partitioning the vector $\overline{\psi}_0$ as $(\overline{\psi}_0^1,\ldots,\overline{\psi}_0^l)$ for $k = 1,\ldots,l$, we have

$$\begin{aligned}
\psi_1(0)\pi &= \left(\left(\psi_1^1(0)\mathbb{1}_{m_1}\right)\nu^1(0),\ldots,\left(\psi_1^l(0)\mathbb{1}_{m_l}\right)\nu^l(0)\right) \\
\overline{\psi}_0\pi &= \left(\left(\overline{\psi}_0^1\mathbb{1}_{m_1}\right)\nu^1(0),\ldots,\left(\overline{\psi}_0^l\mathbb{1}_{m_l}\right)\nu^l(0)\right).
\end{aligned} \tag{4.66}$$

Our expansion requires that $\lim_{\tau\to\infty}\psi_1(\tau) = 0$. As a result,

$$\psi_1(0)\pi + \overline{\psi}_0\pi = 0, \tag{4.67}$$

which implies, by virtue of (4.66),

$$\psi_1^k(0)\mathbb{1}_{m_k} = -\overline{\psi}_0^k\mathbb{1}_{m_k},$$

for $k = 1,\ldots,l$. Solving these equations and in view of

$$\vartheta_1^k(0) = \varphi_1^k(0)\mathbb{1}_{m_k},$$

we choose

$$\vartheta_1^k(0) = -\psi_1^k(0)\mathbb{1}_{m_k} = \overline{\psi}_0^k\mathbb{1}_{m_k} \text{ for } k = 1,\ldots,l.$$

Substituting these into (4.59), we obtain $\varphi_1(\cdot)$. Finally, we use $\psi_1(0) = -\varphi_1(0)$. The process of choosing initial conditions for $\varphi_1(\cdot)$ and $\psi_1(\cdot)$ is complete. Furthermore,

$$|\psi_1(\tau)| \leq K \exp(-\kappa_{1,0}\tau) \quad \text{for some } 0 < \kappa_{1,0} < \kappa_{0,0}.$$

This procedure can be applied to $\varphi_i(\cdot)$ and $\psi_i(\cdot)$ for $i = 2, \ldots, n+1$. We proceed recursively to solve for $\varphi_i(\cdot)$ and $\psi_i(\cdot)$ jointly. Using exactly the same methods as the solution for $\varphi_1(\cdot)$, we define

$$\vartheta_i^k(t) = \varphi_i^k(t)\mathbb{1}_{m_k},$$

for each $k = 1, \ldots, l$ and $i = 2, \ldots, n+1$. Similar to $\widetilde{b}_0^k(\cdot)$, we define $\widetilde{b}_i^k(\cdot)$. and write

$$\widetilde{b}_i(t) = (\widetilde{b}_i^1(t), \ldots, \widetilde{b}_i^l(t)).$$

Proceeding inductively, suppose that $\vartheta_i^k(0)$ is selected and in view of (4.55), it has been shown that

$$|\psi_i(\tau)| \leq K \exp(-\kappa_{i,0}\tau), \quad i \leq n \tag{4.68}$$

for some $0 < \kappa_{i,0} < \kappa_{i-1,0}$. Solve

$$\psi_{i+1}(0)\pi = -\left(\sum_{j=0}^{i} \int_0^\infty \frac{s^j}{j!} \psi_{i-j}(s)ds \frac{d^j \widehat{Q}(0)}{dt^j} \right)\pi := -\overline{\psi}_i \pi$$

to obtain $\psi_{i+1}^k(0)\mathbb{1}_{m_k} = -\overline{\psi}_i^k \mathbb{1}_{m_k}$. Set

$$\vartheta_{i+1}^k(0) = -\psi_{i+1}^k(0)\mathbb{1}_{m_k} = \overline{\psi}_i^k \mathbb{1}_{m_k}, \text{ for } k = 1, \ldots, l.$$

Finally choose $\psi_{i+1}(0) = -\varphi_{i+1}(0)$. We thus have determined the initial conditions for $\varphi_i(\cdot)$. Exactly the same arguments as in Proposition 4.25 lead to

$$|\psi_{i+1}(\tau)| \leq K \exp(-\kappa_{i+1,0}\tau) \text{ for some } 0 < \kappa_{i+1,0} < \kappa_{i,0}.$$

Proposition 4.26. *Assume* (A4.3) *and* (A4.4). *Then the following assertions hold:*

(a) *The sequences of row-vector-valued functions $\varphi_i(\cdot)$ and $\vartheta_i(\cdot)$ for $i = 1, 2, \ldots, n$ can be obtained by solving the system of algebraic differential equations*

$$\varphi_i(t)\widetilde{Q}(t) = \frac{d\varphi_{i-1}(t)}{dt} - \varphi_{i-1}(t)\widehat{Q}(t),$$

$$\vartheta_i^k(t) = \varphi_i^k(t)\mathbb{1}_{m_k}, \tag{4.69}$$

$$\frac{d\vartheta_i(t)}{dt} = \vartheta_i(t)\overline{Q}(t) + \widetilde{b}_{i-1}(t)\widehat{Q}(t)\widetilde{\mathbb{1}}.$$

(b) *For $i = 1, \ldots, n$, the initial conditions are selected as follows:*

    – *For $k = 1, 2, \ldots, l$, find $\psi_i^k(0)\mathbb{1}_{m_k}$ from the equation*

$$\psi_i(0)\overline{\pi} = -\left(\sum_{j=0}^{i-1} \int_0^\infty \frac{s^j}{j!}\psi_{i-j-1}(s)ds\frac{d^j\widehat{Q}(0)}{dt^j} \right)\overline{\pi} := -\overline{\psi}_{i-1}\overline{\pi}.$$

    – *Choose*

$$\vartheta_i^k(0) = -\psi_i^k(0)\mathbb{1}_{m_k} = \overline{\psi}_{i-1}^k\mathbb{1}_{m_k}, \; for \; k = 1, \ldots, l.$$

    – *Choose $\psi_i(0) = -\varphi_i(0)$.*

(c) *There is a positive real number $0 < \kappa_0 < \kappa_{i,0}$ (given in (4.68)) for $i = 0, 1, \ldots, n+1$ such that*

$$|\psi_i(\tau)| \leq K\exp(-\kappa_0\tau).$$

(d) *The choice of initial conditions yields that $\vartheta_i^k(\cdot)$ is $(n+1-i)$-times continuously differentiable on $[0, T]$ and hence $\varphi_i(\cdot)$ is $(n+1-i)$-times continuously differentiable on $[0, T]$.*   □

4.3.2   Analysis of Remainder

The objective here is to carry out the error analysis and validate the asymptotic expansion. Since the details are quite similar to those of Section 4.2, we make no attempt to spell them out. Only the following lemma and proposition are presented.

Lemma 4.27. *Suppose that (A4.3) and (A4.4) are satisfied. Let $\eta^\varepsilon(\cdot)$ be a function such that*

$$\sup_{t\in[0,T]} |\eta^\varepsilon(t)| = O(\varepsilon^{k+1}) \; for \; k \leq n$$

and let $\mathcal{L}^\varepsilon$ be an operator defined in (4.42). If $f^\varepsilon(\cdot)$ is a solution to the equation

$$\mathcal{L}^\varepsilon f^\varepsilon(t) = \eta^\varepsilon(t) \; with \; f^\varepsilon(0) = 0,$$

then $f^\varepsilon(\cdot)$ satisfies

$$\sup_{t\in[0,T]} |f^\varepsilon(t)| = O(\varepsilon^k).$$

Proof: Note that using $Q^\varepsilon(t) = \widetilde{Q}(t)/\varepsilon + \widehat{Q}(t)$, the differential equation can be written as

$$\frac{df^\varepsilon(t)}{dt} = f^\varepsilon(t)Q^\varepsilon(t) + \frac{\eta^\varepsilon(t)}{\varepsilon}.$$

We can then proceed as in the proof of Lemma 4.13.   □

Lemma 4.27 together with detailed computation similar to that of Section 4.2 yields the following proposition.

Proposition 4.28. *For each* $i = 0, 1, \ldots, n$, *define*

$$e_i^\varepsilon(t) = p^\varepsilon(t) - y_i^\varepsilon(t). \tag{4.70}$$

Under conditions (A4.3) *and* (A4.4),

$$\sup_{0 \le t \le T} |e_i^\varepsilon(t)| = O(\varepsilon^{i+1}).$$

4.3.3 Computational Procedure: User's Guide

Since the constructions of $\varphi_i(\cdot)$ and $\psi_i(\cdot)$ are rather involved, and the choice of initial conditions is tricky, we summarize the procedure below. This procedure, which can be used as a user's guide for developing the asymptotic expansion, comprises two main stages.

Step 1: Initialization: finding $\varphi_0(\cdot)$ and $\psi_0(\cdot)$.

1. Obtain the unique solution $\varphi_0(\cdot)$ via (4.54).
2. Obtain the unique solution $\psi_0(\cdot)$ via (4.55) and the initial condition $\psi_0(0) = p^0 - \varphi_0(0)$.

Step 2. Iteration: finding $\varphi_i(\cdot)$ and $\psi_i(\cdot)$ for $1 \le i \le n$.
   While $i \le n$, do the following:

1. Find $\varphi_i(\cdot)$ the solution of (4.69) with temporarily unspecified $\vartheta_i^k(0)$ for $k = 1, \ldots, l$.
2. Obtain $\psi_i(\cdot)$ from (4.55) with temporarily unspecified $\psi_i(0)$.
3. Use the equation

$$\psi_i(0)\pi = -\left(\sum_{j=0}^{i-1} \int_0^\infty \frac{s^j}{j!} \psi_{i-j-1}(s) ds \frac{d^j \widehat{Q}(0)}{dt^j}\right)\pi := -\overline{\psi}_{i-1}\pi$$

   to obtain $\psi_i^k(0)\mathbb{1}_{m_k} = -\overline{\psi}_{i-1}^k \mathbb{1}_{m_k}$.
4. Set $\vartheta_i^k(0) = -\psi_i^k(0)\mathbb{1}_{m_k} = \overline{\psi}_{i-1}^k \mathbb{1}_{m_k}$. By now, $\varphi_i(\cdot)$ has been determined uniquely.
5. Choose $\psi_i(0) = -\varphi_i(0)$. By now, $\psi_i(\cdot)$ has also been determined uniquely.
6. Set $i = i + 1$.
7. If $i > n$, stop.

4.3.4 Summary of Results

While the previous subsection gives the computational procedure, this subsection presents the main theorem. It establishes the validity of the asymptotic expansion.

Theorem 4.29. *Suppose conditions* (A4.3) *and* (A4.4) *are satisfied. Then the asymptotic expansion*

$$y_n^\varepsilon(t) = \sum_{i=0}^{n} \left(\varepsilon^i \varphi_i(t) + \varepsilon^i \psi_i\left(\frac{t}{\varepsilon}\right) \right)$$

can be constructed as in the computational procedure such that

(a) $\varphi_i(\cdot)$ *is* $(n+1-i)$-*times continuously differentiable on* $[0, T]$;

(b) $|\psi_i(t)| \leq K \exp(-\kappa_0 t)$ *for some* $\kappa_0 > 0$;

(c) $|p^\varepsilon(t) - y_n^\varepsilon(t)| = O(\varepsilon^{n+1})$ *uniformly in* $t \in [0, T]$.

Remark 4.30 In general, in view of Proposition 4.11, the error bound is of the form $c_{2n}(t) \exp(-\kappa_0 t)$, where $c_{2n}(t)$ is a polynomial of degree $2n$. The exponential constant κ_0 typically depends on n. The larger n is, the smaller κ_0 will be to account for the polynomial $c_{2n}(t)$.

The following result is a corollary to Theorem 4.29 and will be used in Chapters 5 and 7. Denote the jth component of $\nu^k(t)$ by $\nu_j^k(t)$.

Corollary 4.31. *Assume, in addition to the conditions in Theorem 4.29 with* $n = 0$, *that* $\widetilde{Q}(t) = \widetilde{Q}$ *and* $\widehat{Q}(t) = \widehat{Q}$ *are time independent. Then there exist positive constants* K *and* κ_0 (*both independent of* ε *and* t) *such that*

$$\left| P(\alpha^\varepsilon(t) = s_{kj}) - \nu_j^k(t) \vartheta^k(t) \right| \leq K \left(\varepsilon(t+1) + \exp\left(-\frac{\kappa_0 t}{\varepsilon}\right) \right), \quad (4.71)$$

where $\vartheta^k(t)$ *satisfies*

$$\frac{d}{dt}(\vartheta^1(t), \ldots, \vartheta^l(t)) = (\vartheta^1(t), \ldots, \vartheta^l(t))\overline{Q},$$

with $(\vartheta^1(0), \ldots, \vartheta^l(0)) = (P(\alpha^\varepsilon(0) \in \mathcal{M}_1), \ldots, P(\alpha^\varepsilon(0) \in \mathcal{M}_l))$.

Proof: By a slight modification of the analysis of remainder in Section 4.3, we can obtain (4.71) with a constant K independent of ε and t. The second part of the lemma follows from the uniqueness of the solution to the ordinary differential equation (4.71). □

Remark 4.32. We mention an alternative approach to establishing the asymptotic expansion. In lieu of the constructive procedure presented previously, one may wish to write $\varphi_i(t)$ as a sum of solutions of the homogeneous part and the inhomogeneous part. For instance, one may set

$$\varphi_i(t) = v_i(t)\text{diag}(\nu^1(t), \ldots, \nu^l(t)) + U_i(t), \quad (4.72)$$

where $v_i(t) \in \mathbb{R}^l$ and $U_i(t)$ is a particular solution of the inhomogeneous equation. For $i \geq 0$, the equation

$$\varphi^{(i+1)}(t)\widetilde{Q}(t) = \frac{d\varphi_i(t)}{dt} - \varphi_i(t)\widehat{Q}(t)$$

and $\widetilde{Q}(t)\widetilde{\mathbb{1}} = 0$ lead to

$$0 = \left(\frac{d\varphi_i(t)}{dt} - \varphi_i(t)\widehat{Q}(t)\right)\widetilde{\mathbb{1}}.$$

Substituting (4.72) into the equation above, and noting that $\nu^k(t)\mathbb{1}_{m_k} = 1$ for $k = 1, \ldots, l$, and that $\operatorname{diag}(\nu^1(t), \ldots, \nu^l(t))\widetilde{\mathbb{1}} = I_l$, the $l \times l$ identity matrix, one obtains

$$\frac{dv_i(t)}{dt} = v_i(t)\overline{Q}(t) + U_i(t)\widehat{Q}(t)\widetilde{\mathbb{1}} - \left(\frac{dU_i(t)}{dt}\right)\widetilde{\mathbb{1}}.$$

One then proceeds to determine $v_i(0)$ via the matching condition. The main ideas are similar, and the details are slightly different.

4.3.5 An Example

Consider Example 4.20 again. Note that the conditions in (A4.3) and (A4.4) require that

$$\lambda_1(t) + \mu_1(t) > 0 \text{ for all } t \in [0, T],$$

and the jump rates $\lambda(t)$ and $\mu(t)$ be smooth enough.

The probability distribution of the state process is given by $p^\varepsilon(t)$ satisfying

$$\frac{dp^\varepsilon(t)}{dt} = p^\varepsilon(t)Q^\varepsilon(t),$$

$$p^\varepsilon(0) = p^0 \text{ such that}$$

$$p_i^0 \geq 0 \text{ and } \sum_{i=1}^{4} p_i^0 = 1.$$

To solve this set of equations, note that

$$\frac{d}{dt}(p_1^\varepsilon(t) + p_2^\varepsilon(t)) = -\lambda_2(t)(p_1^\varepsilon(t) + p_2^\varepsilon(t)) + \mu_2(t)(p_3^\varepsilon(t) + p_4^\varepsilon(t)),$$

$$\frac{d}{dt}(p_1^\varepsilon(t) + p_3^\varepsilon(t)) = -\frac{\lambda_1(t)}{\varepsilon}(p_1^\varepsilon(t) + p_3^\varepsilon(t)) + \frac{\mu_1(t)}{\varepsilon}(p_2^\varepsilon(t) + p_4^\varepsilon(t)),$$

$$\frac{d}{dt}(p_2^\varepsilon(t) + p_4^\varepsilon(t)) = \frac{\lambda_1(t)}{\varepsilon}(p_1^\varepsilon(t) + p_3^\varepsilon(t)) - \frac{\mu_1(t)}{\varepsilon}(p_2^\varepsilon(t) + p_4^\varepsilon(t)),$$

$$\frac{d}{dt}(p_3^\varepsilon(t) + p_4^\varepsilon(t)) = \lambda_2(t)(p_1^\varepsilon(t) + p_2^\varepsilon(t)) - \mu_2(t)(p_3^\varepsilon(t) + p_4^\varepsilon(t)).$$

To proceed, define functions $a_{12}(t)$, $a_{13}(t)$, $a_{24}(t)$, and $a_{34}(t)$ as follows:

$$a_{12}(t) = (p_1^0 + p_2^0) \exp\left(-\int_0^t (\lambda_2(s) + \mu_2(s))ds\right)$$

$$+ \int_0^t \mu_2(u) \exp\left(-\int_u^t (\lambda_2(s) + \mu_2(s))ds\right) du,$$

$$a_{13}(t) = (p_1^0 + p_3^0) \exp\left(-\frac{1}{\varepsilon} \int_0^t (\lambda_1(s) + \mu_1(s))ds\right)$$

$$+ \int_0^t \frac{\mu_1(u)}{\varepsilon} \exp\left(-\frac{1}{\varepsilon} \int_u^t (\lambda_1(s) + \mu_1(s))ds\right) du,$$

$$a_{24}(t) = (p_2^0 + p_4^0) \exp\left(-\frac{1}{\varepsilon} \int_0^t (\lambda_1(s) + \mu_1(s))ds\right)$$

$$+ \int_0^t \frac{\lambda_1(u)}{\varepsilon} \exp\left(-\frac{1}{\varepsilon} \int_u^t (\lambda_1(s) + \mu_1(s))ds\right) du,$$

$$a_{34}(t) = (p_3^0 + p_4^0) \exp\left(-\int_0^t (\lambda_2(s) + \mu_2(s))ds\right)$$

$$+ \int_0^t \lambda_2(u) \exp\left(-\int_u^t (\lambda_2(s) + \mu_2(s))ds\right) du.$$

Then using the fact that $p_1^\varepsilon(t) + p_2^\varepsilon(t) + p_3^\varepsilon(t) + p_4^\varepsilon(t) = 1$, we have

$$\begin{aligned}
p_1^\varepsilon(t) + p_2^\varepsilon(t) &= a_{12}(t), \\
p_1^\varepsilon(t) + p_3^\varepsilon(t) &= a_{13}(t), \\
p_2^\varepsilon(t) + p_4^\varepsilon(t) &= a_{24}(t), \\
p_3^\varepsilon(t) + p_4^\varepsilon(t) &= a_{34}(t).
\end{aligned} \tag{4.73}$$

Note also that

$$\frac{dp_1^\varepsilon(t)}{dt} = -\left(\frac{\lambda_1(t)}{\varepsilon} + \frac{\mu_1(t)}{\varepsilon} + \lambda_2(t) + \mu_2(t)\right) p_1^\varepsilon(t)$$

$$+ \frac{\mu_1(t)}{\varepsilon} a_{12}(t) + \mu_2(t) a_{13}(t).$$

The solution to this equation is

$$p_1^\varepsilon(t) = p_1^0 \exp\left(-\int_0^t \left(\frac{\lambda_1(s) + \mu_1(s)}{\varepsilon} + \lambda_2(s) + \mu_2(s)\right) ds\right)$$

$$+ \int_0^t \left(\frac{\mu_1(u)}{\varepsilon} a_{12}(u) + \mu_2(u) a_{13}(u)\right)$$

$$\times \exp\left(-\int_u^t \left(\frac{\lambda_1(s) + \mu_1(s)}{\varepsilon} + \lambda_2(s) + \mu_2(s)\right) ds\right) du.$$

Consequently, in view of (4.73), it follows that

$$p_2^\varepsilon(t) = a_{12}(t) - p_1^\varepsilon(t),$$
$$p_3^\varepsilon(t) = a_{13}(t) - p_1^\varepsilon(t),$$
$$p_4^\varepsilon(t) = a_{24}(t) - p_2^\varepsilon(t).$$

In this example, the zeroth-order term is given by

$$\varphi_0(t) = (\nu^1(t)\vartheta_0^1(t), \nu^2(t)\vartheta_0^2(t)),$$

where the quasi-stationary distributions are given by

$$\nu^1(t) = \nu^2(t) = \left(\frac{\mu_1(t)}{\lambda_1(t) + \mu_1(t)}, \frac{\lambda_1(t)}{\lambda_1(t) + \mu_1(t)} \right),$$

and the multipliers $(\vartheta_0^1(t), \vartheta_0^2(t))$ are determined by the differential equation

$$\frac{d}{dt}(\vartheta_0^1(t), \vartheta_0^2(t)) = (\vartheta_0^1(t), \vartheta_0^2(t)) \begin{pmatrix} -\lambda_2(t) & \lambda_2(t) \\ \mu_2(t) & -\mu_2(t) \end{pmatrix},$$

with initial value $(\vartheta_0^1(0), \vartheta_0^2(0)) = (p_1^0 + p_2^0, p_3^0 + p_4^0)$.
  The inner expansion term $\psi_0(\tau)$ is given by

$$\frac{d\psi_0(\tau)}{d\tau} = \psi_0(\tau)\widetilde{Q}(0), \ \psi_0(0) = p^0 - \varphi_0(0).$$

By virtue of Theorem 4.29,

$$p^\varepsilon(t) - \varphi_0(t) - \psi_0\left(\frac{t}{\varepsilon}\right) = O(\varepsilon),$$

provided that $Q^\varepsilon(t)$ is continuously differentiable on $[0, T]$. Noting the exponential decay of $\psi_0(t/\varepsilon)$, we further have

$$p^\varepsilon(t) = \varphi_0(t) + O\left(\varepsilon + \exp\left(-\frac{\kappa_0 t}{\varepsilon}\right)\right).$$

In particular, for any $t > 0$,

$$\lim_{\varepsilon \to 0} p^\varepsilon(t) = \varphi_0(t).$$

Namely, $\varphi_0(t)$ is the limit distribution of the Markov chain generated by $Q^\varepsilon(t)$.

4.4   Inclusion of Absorbing States

While the case of recurrent states was considered in the previous section, this section concerns the asymptotic expansion in which the Markov chain generated by $Q^\varepsilon(t)$ in which $\widetilde{Q}(t)$ includes components corresponding to absorbing states. By rearrangement, the matrix $\widetilde{Q}(t)$ takes the form

$$
\widetilde{Q}(t) = \begin{pmatrix} \widetilde{Q}^1(t) & & & & \\ & \widetilde{Q}^2(t) & & & \\ & & \ddots & & \\ & & & \widetilde{Q}^l(t) & \\ & & & & 0_{m_a \times m_a} \end{pmatrix}, \tag{4.74}
$$

where $\widetilde{Q}^k(t) \in \mathbb{R}^{m_k \times m_k}$ for $k = 1, 2, \ldots, l$, $0_{m_a \times m_a}$ is an $m_a \times m_a$ zero matrix, and

$$
m_1 + m_2 + \cdots + m_l + m_a = m.
$$

Let $\mathcal{M}_a = \{s_{a1}, \ldots, s_{am_a}\}$ denote the set of absorbing states. We may, as in Section 4.3, represent the state space as

$$
\mathcal{M} = \mathcal{M}_1 \cup \cdots \cup \mathcal{M}_l \cup \mathcal{M}_a
$$

$$
= \Big\{ s_{11}, \ldots, s_{1m_1}, \ldots, s_{l1}, \ldots, s_{lm_l}, s_{a1}, \ldots, s_{am_a} \Big\}.
$$

Following the development of Section 4.3, suppose that $\alpha^\varepsilon(\cdot)$ is a Markov chain generated by $Q^\varepsilon(\cdot) = \widetilde{Q}(\cdot)/\varepsilon + \widehat{Q}(\cdot)$. Compared with Section 4.3, the difference is that now the dominant part in the generator includes absorbing states corresponding to the $m_a \times m_a$ matrix $0_{m_a \times m_a}$. As in the previous case, our interest is to obtain an asymptotic expansion of the probability distribution.

Remark 4.33. The motivation of the current study stems from the formulation of competitive risk theory discussed in Section 3.3. The idea is that within the m states, there are several groups. Some of them are much riskier than the others (in the sense of frequency of the occurrence of the corresponding risks). The different rates (sensitivity) of risks are modeled by the use of a small parameter $\varepsilon > 0$.

Denote by $p^\varepsilon(\cdot)$ the solution of (4.40). The objective here is to obtain an asymptotic expansion

$$
y_n^\varepsilon = \sum_{i=0}^n \varepsilon^i \varphi_i(t) + \sum_{i=0}^n \varepsilon^i \psi_i\left(\frac{t}{\varepsilon}\right).
$$

Since the techniques employed are essentially the same as in the previous section, it will be most instructive here to highlight the main ideas. Thus, we only note the main steps and omit most of the details.

Assume conditions (A4.3) and (A4.4) for the current matrices $\widetilde{Q}^k(t)$, $\widetilde{Q}(t)$, and $\widehat{Q}(t)$. For $t \in [0, T]$, substituting the expansion above into (4.40) and equating coefficients of ε^i, for $i = 1, \ldots, n+1$, yields

$$\varphi_0(t)\widetilde{Q}(t) = 0,$$

$$\varphi_i(t)\widetilde{Q}(t) = \frac{d\varphi_{i-1}(t)}{dt} - \varphi_{i-1}(t)\widehat{Q}(t),$$

(4.75)

and (with the use of the stretched variable $\tau = t/\varepsilon$)

$$\frac{d\psi_0(\tau)}{d\tau} = \psi_0(\tau)\widetilde{Q}(0),$$

$$\frac{d\psi_i(\tau)}{d\tau} = \psi_i(\tau)\widetilde{Q}(0) + \sum_{j=0}^{i-1} \psi_{i-j-1}(\tau)$$

$$\times \left(\frac{\tau^j}{j!} \frac{d^j \widehat{Q}(0)}{dt^j} + \frac{\tau^{j+1}}{(j+1)!} \frac{d^{j+1}\widetilde{Q}(0)}{dt^{j+1}} \right).$$

(4.76)

For each $i \geq 0$, we use the following notation for the partitioned vectors:

$$\varphi_i(t) = (\varphi_i^1(t), \ldots, \varphi_i^l(t), \varphi_i^a(t)),$$

$$\psi_i(\tau) = (\psi_i^1(\tau), \ldots, \psi_i^l(\tau), \psi_i^a(\tau)).$$

In the above $\varphi_i^a(t)$ and $\psi_i^a(\tau)$ are vectors in $\mathbb{R}^{1 \times m_a}$.

To determine the outer- and the initial-layer expansions, let us start with $i = 0$. For each $t \in [0, T]$, the use of the partitioned vector $\varphi_0(t)$ leads to

$$\varphi_0^k(t)\widetilde{Q}^k(t) = 0, \text{ for } k = 1, \ldots, l.$$

Note that $\varphi_0^a(t)$ does not show up in any of these equations owing to the $0_{m_a \times m_a}$ matrix in $\widetilde{Q}(t)$. It will have to be obtained from the equation in (4.75) corresponding to $i = 1$. Put another way, $\varphi_0^a(t)$ is determined mainly by the matrix $\widehat{Q}(t)$.

Similar to Section 4.3, $\varphi_0^k(t) = \vartheta_0^k(t)\nu^k(t)$, where $\nu^k(t)$ are the quasi-stationary distributions corresponding to the generators $\widetilde{Q}^k(t)$ for $k = 1, \ldots, l$ and $\vartheta_0^k(t)$ are the corresponding multipliers. Define

$$\widetilde{\mathbb{1}}_a = \begin{pmatrix} \mathbb{1}_{m_1} & & & \\ & \ddots & & \\ & & \mathbb{1}_{m_l} & \\ & & & I_{m_a} \end{pmatrix},$$

where I_{m_a} is an $m_a \times m_a$ identity matrix. Clearly, $\widetilde{\mathbb{1}}_a$ is orthogonal to $\widetilde{Q}(t)$ for each $t \in [0, T]$. As a result, multiplying (4.75) by $\widetilde{\mathbb{1}}_a$ from the right with $i = 1$ leads to

$$\frac{d\varphi_0(t)}{dt}\widetilde{\mathbb{1}}_a = \varphi_0(t)\widehat{Q}(t)\widetilde{\mathbb{1}}_a,$$

$$(\vartheta_0(0), \varphi_0^a(0)) = p^0\widetilde{\mathbb{1}}_a, \tag{4.77}$$

where $\vartheta_0(0) = (\vartheta_0^1(0), \ldots, \vartheta_0^l(0))$.

The above initial condition is a consequence of the initial-value consistency condition in (4.53). It is readily seen that

$$\sum_{k=1}^{l} \vartheta_0^k(0) = 1 - \varphi_0^a(0)\mathbb{1}_{m_a} = 1 - p^{0,a}\mathbb{1}_{m_a},$$

where $p^0 = (p^{0,1}, \ldots, p^{0,l}, p^{0,a})$.

We write

$$\varphi_0(t) = (\vartheta_0^1(t), \ldots, \vartheta_0^l(t), \varphi_0^a(t))\mathrm{diag}(\nu^1(t), \ldots, \nu^l(t), I_{m_a}).$$

Define

$$\overline{Q}(t) = \mathrm{diag}(\nu^1(t), \ldots, \nu^l(t), I_{m_a})\widehat{Q}(t)\widetilde{\mathbb{1}}_a. \tag{4.78}$$

Then (4.77) is equivalent to

$$\frac{d}{dt}(\vartheta_0(t), \varphi_0^a(t)) = (\vartheta_0(t), \varphi_0^a(t))\overline{Q}(t),$$

$$(\vartheta_0(0), \varphi_0^a(0)) = p^0\widetilde{\mathbb{1}}_a.$$

This is a linear system of differential equations. Therefore it has a unique solution given by

$$(\vartheta_0(t), \varphi_0^a(t)) = p^0\widetilde{\mathbb{1}}_a X(t, 0),$$

where $X(t, 0)$ is the principal matrix solution of the homogeneous equation. Thus $\varphi_0(t)$ has been found and is $(n+1)$-times continuously differentiable.

Remark 4.34. Note that in $\varphi_0(t)$, the term $\varphi_0^a(t)$ corresponds to the set of absorbing states $\mathcal{M}_a$. Clearly, these states cannot be aggregated to a single state as in the case of recurrent states. Nevertheless, the function $\varphi_0^a(t)$ tends to be stabilized in a neighborhood of a constant for t large enough. To illustrate, let us consider a stationary case, that is, both $\widetilde{Q}(t) = \widetilde{Q}$ and $\widehat{Q}(t) = \widehat{Q}$ are independent of t. Partition $\widehat{Q}$ as blocks of submatrices

$$\widehat{Q} = \begin{pmatrix} \widehat{Q}^{11} & \widehat{Q}^{12} \\ \widehat{Q}^{21} & \widehat{Q}^{22} \end{pmatrix},$$

where $\widehat{Q}^{22}$ is an $m_a \times m_a$ matrix. Assume that the eigenvalues of $\widehat{Q}^{22}$ have negative real parts. Then, in view of the definition of $\overline{Q}(t) = \overline{Q}$ in (4.78), it follows that

$$\varphi_0^a(t) \to \text{ a constant as } t \to \infty.$$

Using the partition $\psi_0(\tau) = (\psi_0^1(\tau), \ldots, \psi_0^l(\tau), \psi_0^a(\tau))$, consider the zeroth-order initial-layer term given by

$$\frac{d\psi_0(\tau)}{d\tau} = \frac{d}{d\tau}(\psi_0^1(\tau), \ldots, \psi_0^l(\tau), \psi_0^a(\tau))$$

$$= \psi_0(\tau)\widetilde{Q}(0) = (\psi_0^1(\tau)\widetilde{Q}^1(\tau), \ldots, \psi_0^l(\tau)\widetilde{Q}^l(0), 0_{m_a}).$$

We obtain

$$\psi_0^k(\tau) = \psi_0^k(0)\exp(\widetilde{Q}^k(0)\tau), \text{ for } k = 1, \ldots, l, \text{ and}$$

$$\psi_0^a(\tau) = \text{constant}.$$

Noting that $p^{0,a} = \varphi_0^a(0)$ and choosing $\psi_0(0) = p^0 - \varphi_0(0)$ lead to $\psi_0^a(\tau) = 0_{m_a}$. Thus

$$\psi_0(\tau) = (\psi_0^1(0)\exp(\widetilde{Q}^1(0)\tau), \ldots, \psi_0^l(0)\exp(\widetilde{Q}^l(0)\tau), 0_{m_a}).$$

Similar to the result in Section 4.3, the following lemma holds. The proof is analogous to that of Proposition 4.25.

Lemma 4.35. *Define*

$$\pi_a = \text{diag}(\mathbb{1}_{m_1}\nu^1(0), \ldots, \mathbb{1}_{m_l}\nu^l(0), I_{m_a}).$$

Then there exist positive constants K and $\kappa_{0,0}$ such that

$$|\exp(\widetilde{Q}(0)\tau) - \pi_a| \leq K\exp(-\kappa_{0,0}\tau).$$

By virtue of the lemma above and the orthogonality $(p^0 - \varphi_0(0))\pi_a = 0$, we have

$$|\psi_0(\tau)| = |(p^0 - \varphi_0(0))(\exp(\widetilde{Q}(0)\tau) - \pi_a)|$$

$$\leq K\exp(-\kappa_{0,0}\tau)$$

for some $K > 0$ and $\kappa_{0,0} > 0$ given in Lemma 4.35; that is, $\psi_0(\tau)$ decays exponentially fast. Therefore, $\psi_0(\tau)$ has the desired property.

We continue in this fashion and proceed to determine the next term $\varphi_1(t)$ as well as $\psi_1(t/\varepsilon)$. Let

$$b_0(t) = \frac{d\varphi_0(t)}{dt} - \varphi_0(t)\widehat{Q}(t) \quad \text{with}$$

$$b_0(t) = (b_0^1(t), \dots, b_0^l(t), b_0^a(t)).$$

It is easy to check that $b_0^a(t) = 0_{m_a}$. The equation $\varphi_1(t)\widetilde{Q}(t) = b_0(t)$ then leads to

$$\varphi_1^k(t)\widetilde{Q}^k(t) = b_0^k(t), \text{ for } k = 1, \dots, l,$$

$$b_0^a(t) = 0_{m_a}.$$

(4.79)

The solutions of the l inhomogeneous equations in (4.79) above are of the form

$$\varphi_1^k(t) = \vartheta_1^k(t)\nu^k(t) + \widetilde{b}_0^k(t), \ k = 1, \dots, l,$$

where $\vartheta_1^k(t)$ for $k = 1, \dots, l$ are scalar multipliers. Again, $\varphi_1^a(t)$ cannot be obtained from the equation above, it must come from the contribution of the matrix-valued function $\widehat{Q}(t)$.

Note that

$$\widetilde{b}_0^k(t)\widetilde{Q}^k(t) = b_0^k(t) \quad \text{and} \quad \widetilde{b}_0^k(t)\mathbb{1}_{m_k} = 0.$$

Using the equation

$$\varphi_2(t)\widetilde{Q}(t) = \frac{d\varphi_1(t)}{dt} - \varphi_1(t)\widehat{Q}(t),$$

one obtains

$$0 = \varphi_2(t)\widetilde{Q}(t)\widetilde{\mathbb{1}}_a = \frac{d\varphi_1(t)}{dt}\widetilde{\mathbb{1}}_a - \varphi_1(t)\widehat{Q}(t)\widetilde{\mathbb{1}}_a,$$

which in turn implies that

$$\frac{d}{dt}(\vartheta_1(t), \varphi_1^a(t)) = (\vartheta_1(t), \varphi_1^a(t))\overline{Q}(t)$$

$$+ (\widetilde{b}_0(t), 0_{m_a})\widehat{Q}(t)\widetilde{\mathbb{1}}_a,$$

(4.80)

where

$$\vartheta_1(t) = (\vartheta_1^1(t), \dots, \vartheta_1^l(t)) \text{ and } \widetilde{b}_0(t) = (\widetilde{b}_0^1(t), \dots, \widetilde{b}_0^l(t)).$$

Let $X(t, s)$ denote the principal matrix solution to the homogeneous differential equation

$$\frac{dy(t)}{dt} = y(t)\overline{Q}(t).$$

Then the solution to (4.80) can be represented by $X(t, s)$ as follows:

$$(\vartheta_1(t), \varphi_1^a(t)) = (\vartheta_1(0), \varphi_1^a(0))X(t, 0)$$
$$+ \int_0^t (\widetilde{b}_0(s), 0_{m_a})\widehat{Q}(s)\widetilde{\mathbb{1}}_a X(t, s)ds.$$

Note that the initial conditions $\varphi_1^a(0)$ and $\vartheta_1^k(0)$ for $k = 1, \ldots, l$ need to be determined using the initial-layer terms just as in Section 4.3.

Using (4.76) with $i = 1$, one obtains an equation that has the same form as that of (4.62). That is,

$$\psi_1(\tau) = \psi_1(0)\exp(\widetilde{Q}(0)\tau)$$
$$+ \int_0^\tau \psi_0(0)\exp(\widetilde{Q}(0)s)\widehat{Q}(0)\exp(\widetilde{Q}(0)(\tau - s))ds$$
$$+ \int_0^\tau s\psi_0(0)\exp(\widetilde{Q}(0)s)\frac{d\widetilde{Q}(0)}{dt}\exp(\widetilde{Q}(0)(\tau - s))ds.$$

As in Section 4.3, with the use of π_a, it can be shown that $|\psi_1(\tau)| \leq K\exp(-\kappa_{1,0}\tau)$ for some $K > 0$ and $0 < \kappa_{1,0} < \kappa_{0,0}$. By requiring that $\psi_1(\tau)$ decay to 0 as $\tau \to \infty$, we obtain the equation

$$\psi_1(0)\pi_a = -\overline{\psi}_0\pi_a, \tag{4.81}$$

where

$$\overline{\psi}_0 = \int_0^\infty \psi_0(0)\exp(\widetilde{Q}(0)s)ds\widehat{Q}(0).$$

Owing to (4.81) and the known form of $\psi_0(\tau)$,

$$\overline{\psi}_0 = (\overline{\psi}_0^1, \ldots, \overline{\psi}_0^l, \overline{\psi}_0^a)$$
$$= (p^{0,1} - \varphi_0^1(0), \ldots, p^{0,l} - \varphi_0^l, 0_{m_a})\left(\int_0^\infty \exp(\widetilde{Q}(0)s)ds\right)\widehat{Q}(0),$$

which is a completely known vector. Thus the solution to (4.81) is

$$\psi_1^k(0)\mathbb{1}_{m_k} = -\overline{\psi}_0^k\mathbb{1}_{m_k} \text{ for } k = 1, \ldots, l, \text{ and } \psi_1^a(0) = -\overline{\psi}_0^a.$$

To obtain the desired matching property for the inner-outer expansions, choose

$$\vartheta_1^k(0) = -\psi_1^k(0)\mathbb{1}_{m_k} = \overline{\psi}_0^k \mathbb{1}_{m_k} \text{ for } k = 1, \ldots, l,$$

$$\varphi_1^a(0) = -\psi_1^a(0) = \overline{\psi}_0^a.$$

In general, for $i = 2, \ldots, n$, the initial conditions are selected as follows: For $k = 1, 2, \ldots, l$, find $\psi_i^k(0)\mathbb{1}_{m_k}$ from the equation

$$\psi_i(0)\pi_a = -\left(\sum_{j=0}^{i-1} \int_0^\infty \frac{s^j}{j!}\psi_{i-j-1}(s)ds\frac{d^j\widehat{Q}(0)}{dt^j}\right)\pi_a := -\overline{\psi}_{i-1}\pi_a.$$

Choose

$$\vartheta_i^k(0) = -\psi_i^k(0)\mathbb{1}_{m_k} = \overline{\psi}_{i-1}^k \mathbb{1}_{m_k},$$

for $k = 1, \ldots, l,$

$$\phi_i^a(0) = -\overline{\psi}_{i-1}^a, \quad \text{and } \psi_i(0) = -\varphi_i(0).$$

Proceeding inductively, we then construct all $\varphi_i(t)$ and $\psi_i(\tau)$. Moreover, we can verify that there exists $0 < \kappa_{i,0} < \kappa_{i-1,0} < \kappa_{0,0}$ such that $|\psi_i(\tau)| \leq K \exp(-\kappa_{i,0}\tau)$. This indicates that the inclusion of absorbing states is very similar to the case of all recurrent states. In the zeroth-order outer expansion, there is a component $\varphi_0^a(t)$ that "takes care of" the absorbing states. Note, however, that starting from the leading term (zeroth-order approximation), the matching will be determined not only by the multipliers $\vartheta_i(0)$ but also by the vector $\psi_i(0)$ associated with the absorbing states. We summarize the results in the following theorem.

Theorem 4.36. *Consider $\widetilde{Q}(t)$ given by (4.74), and suppose conditions (A4.3) and (A4.4) are satisfied for the matrix-valued functions $\widetilde{Q}^k(\cdot)$ for $k = 1, \ldots, l$ and $\widehat{Q}(\cdot)$. An asymptotic expansion*

$$y_n^\varepsilon(t) = \sum_{i=0}^n \left(\varepsilon^i\varphi_i(t) + \varepsilon^i\psi_i\left(\frac{t}{\varepsilon}\right)\right)$$

exists such that

(a) *$\varphi_i(\cdot)$ is $(n+1-i)$-times continuously differentiable on $[0,T]$;*

(b) *$|\psi_i(t)| \leq K \exp(-\kappa_0 t)$ for some $0 < \kappa_0 < \kappa_{i,0}$;*

(c) *$|p^\varepsilon(t) - y_n^\varepsilon(t)| = O(\varepsilon^{n+1})$ uniformly in $t \in [0,T]$.*

Finally, at the end of this section, we give a simple example to illustrate the result.

Example 4.37. Let us consider a Markov chain generated by

$$Q^\varepsilon = \frac{1}{\varepsilon}\widetilde{Q} + \widehat{Q},$$

where

$$\widetilde{Q} = \begin{pmatrix} -1 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \text{ and } \widehat{Q} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & -1 \end{pmatrix}.$$

Not being irreducible, the chain generated by $\widetilde{Q}$ includes an absorbing state. In this example, $\overline{Q} = \begin{pmatrix} 0 & 0 \\ 1 & -1 \end{pmatrix}$. Let $p^0 = (p_1^0, p_2^0, p^{0,a})$ denote the initial distribution of $\alpha^\varepsilon(\cdot)$. Then solving the forward equation (4.40) gives us

$$p^\varepsilon(t) = (p_1^\varepsilon(t), p_2^\varepsilon(t), p_3^\varepsilon(t)),$$

where

$$p_1^\varepsilon(t) = \frac{p_1^0 + p_2^0 + p^{0,a}}{2}$$
$$- \left(\frac{-p_1^0 + p_2^0 - p^{0,a}}{2} + \frac{p^{0,a}}{2-\varepsilon}\right)\exp\left(-\frac{2t}{\varepsilon}\right) - \left(\frac{(1-\varepsilon)p^{0,a}}{2-\varepsilon}\right)\exp(-t),$$

$$p_2^\varepsilon(t) = \frac{p_1^0 + p_2^0 + p^{0,a}}{2}$$
$$+ \left(\frac{-p_1^0 + p_2^0 - p^{0,a}}{2} + \frac{p^{0,a}}{2-\varepsilon}\right)\exp\left(-\frac{2t}{\varepsilon}\right) - \left(\frac{p^{0,a}}{2-\varepsilon}\right)\exp(-t),$$

$$p_3^\varepsilon(t) = p^{0,a}\exp(-t).$$

Computing $\varphi_0(t)$ yields

$$\varphi_0(t) = \left(\frac{p_1^0 + p_2^0 + p^{0,a}}{2}, \frac{p_1^0 + p_2^0 + p^{0,a}}{2}, 0\right)$$
$$+ \left(-\frac{p^{0,a}}{2}, -\frac{p^{0,a}}{2}, p^{0,a}\right)\exp(-t).$$

It is easy to see that for $t > 0$,

$$\lim_{\varepsilon \to 0} |p^\varepsilon(t) - \varphi_0(t)| = 0.$$

The limit behavior of the underlying Markov chain as $\varepsilon \to 0$ is determined by $\varphi_0(t)$ (for $t > 0$). Moreover, when t is large, the influence from $\widehat{Q}$ corresponding to the absorbing state (the vector multiplied by $\exp(-t)$) can be ignored because $\exp(-t)$ goes to 0 exponentially fast as $t \to \infty$.

4.5   Inclusion of Transient States

If a Markov chain has transient states, then, relabeling the states through suitable permutations, one can decompose the states into several groups of recurrent states, each of which is weakly irreducible, and a group of transient states. Naturally, we consider the generator $\widetilde{Q}(t)$ in $Q^\varepsilon(t)$ having the form

$$\widetilde{Q}(t) = \begin{pmatrix} \widetilde{Q}^1(t) & & & \\ & \ddots & & \\ & & \widetilde{Q}^l(t) & \\ \widetilde{Q}^1_*(t) & \cdots & \widetilde{Q}^l_*(t) & \widetilde{Q}_*(t) \end{pmatrix} \qquad (4.82)$$

such that for each $t \in [0, T]$, and each $k = 1, \ldots, l$, $\widetilde{Q}^k(t)$ is a generator with dimension $m_k \times m_k$, $\widetilde{Q}_*(t)$ is an $m_* \times m_*$ matrix, $\widetilde{Q}^k_*(t) \in \mathbb{R}^{m_* \times m_k}$, and

$$m_1 + m_2 + \cdots + m_l + m_* = m.$$

We continue our study of singularly perturbed chains with weak and strong interactions by incorporating the transient states into the model. Let $\alpha^\varepsilon(\cdot)$ be a Markov chain generated by $Q^\varepsilon(\cdot)$, with $Q^\varepsilon(t) \in \mathbb{R}^{m \times m}$ given by (4.39) with $\widetilde{Q}(t)$ given by (4.82). The state space of the underlying Markov chain is given by

$$\mathcal{M} = \mathcal{M}_1 \cup \cdots \cup \mathcal{M}_l \cup \mathcal{M}_*$$

where $\mathcal{M}_k = \{s_{k1}, \ldots, s_{km_k}\}$ are the states corresponding to the recurrent states and $\mathcal{M}_* = \{s_{*1}, \ldots, s_{*m_*}\}$ are those corresponding to the transient states.

Since $\widetilde{Q}(t)$ is a generator, for each $k = 1, \ldots, l$, $\widetilde{Q}^k(t)$ is a generator. Thus the matrix $\widetilde{Q}^k_*(t) = (\widetilde{q}^k_{*,ij})$ satisfies $\widetilde{q}^k_{*,ij} \geq 0$ for each $i = 1, \ldots, m_*$ and $j = 1, \ldots, m_k$, and $\widetilde{Q}_*(t) = (\widetilde{q}_{*,ij})$ satisfies

$$\widetilde{q}_{*,ij}(t) \geq 0 \text{ for } i \neq j, \widetilde{q}_{*,ii}(t) < 0, \text{ and } \widetilde{q}_{*,ii}(t) \leq -\sum_{j \neq i} \widetilde{q}_{*,ij}(t).$$

Roughly, the block matrix $(\widetilde{Q}^1_*(t), \ldots, \widetilde{Q}^l_*(t), \widetilde{Q}_*(t))$ is "negatively dominated" by the matrix $\widetilde{Q}_*(t)$. Thus it is natural to assume that $\widetilde{Q}_*(t)$ is a stable matrix (or Hurwitz, i.e., all its eigenvalues have negative real parts). Comparing with the setups of Sections 4.3 and 4.4, the difference in $\widetilde{Q}(t)$ is the additional matrices $\widetilde{Q}^k_*(t)$ for $k = 1, \ldots, l$ and $\widetilde{Q}_*(t)$. Note that $\widetilde{Q}^k_*(t)$ are nonsquare matrices, and $\widetilde{Q}(t)$ no longer has block-diagonal form.

The formulation here is inspired by the work of Phillips and Kokotovic [175] and Delebecque and Quadrat [44]; see also the recent work of Pan and Başar [164], in which the authors treated time-invariant $\widetilde{Q}$ matrix of a similar form. Sections 4.3 and 4.4 together with this section essentially include

generators of finite-state Markov chains of the most practical concerns. It ought to be pointed out that just as one cannot in general simultaneously diagonalize two matrices, for Markov chains with weak and strong interactions, one cannot put both $\widetilde{Q}(t)$ and $\widehat{Q}(t)$ into the forms mentioned above simultaneously. Although the model to be studied in this section is slightly more complex compared with the block-diagonal $\widetilde{Q}(t)$ in (4.41), we demonstrate that an asymptotic expansion of the probability distribution can still be obtained by using the same techniques of the previous sections. Moreover, it can be seen from the expansion that the underlying Markov chain stays in the transient states only with very small probability. In some cases, for example $\widehat{Q}(t) = 0$, these transient states can be ignored; see Remark 4.40 for more details.

To incorporate the transient states, we need the following conditions. The main addition is the assumption that $\widetilde{Q}_*(t)$ is stable.

(**A4.5**) For each $t \in [0, T]$ and $k = 1, \ldots, l$, $\widetilde{Q}(t)$, $\widehat{Q}(t)$, and $\widetilde{Q}^k(t)$ satisfy (A4.3) and (A4.4).

(**A4.6**) For each $t \in [0, T]$, $\widetilde{Q}_*(t)$ is Hurwitz (i.e., all of its eigenvalues have negative real parts).

Remark 4.38. Condition (A4.6) indicates the inclusion of transient states. Since $\widetilde{Q}_*(t)$ is Hurwitz, it is nonsingular. Thus the inverse matrix $\widetilde{Q}_*^{-1}(t)$ exists for each $t \in [0, T]$.

Let $p^\varepsilon(\cdot)$ denote the solution to (4.40) with $\widetilde{Q}(t)$ specified in (4.82). We seek asymptotic expansions of $p^\varepsilon(\cdot)$ having the form

$$y_n^\varepsilon(t) = \sum_{i=0}^{n} \varepsilon^i \varphi_i(t) + \sum_{i=0}^{n} \varepsilon^i \psi_i \left(\frac{t}{\varepsilon} \right).$$

The development is very similar to that of Section 4.3, so no attempt is made to give verbatim details. Instead, only the salient features will be brought out.

Substituting $y_n^\varepsilon(t)$ into the forward equation and equating coefficients of ε^i for $i = 1, \ldots, n$ lead to the equations

$$
\begin{aligned}
\varphi_0(t)\widetilde{Q}(t) &= 0, \\
\varphi_i(t)\widetilde{Q}(t) &= \frac{d\varphi_{i-1}(t)}{dt} - \varphi_{i-1}(t)\widehat{Q}(t),
\end{aligned}
\tag{4.83}
$$

and with the change of time scale $\tau = t/\varepsilon$,

$$\frac{d\psi_0(\tau)}{d\tau} = \psi_0(\tau)\widetilde{Q}(0),$$

$$\frac{d\psi_i(\tau)}{d\tau} = \psi_i(\tau)\widetilde{Q}(0) + \sum_{j=0}^{i-1}\psi_{i-j-1}(\tau) \qquad (4.84)$$

$$\times\left(\frac{\tau^j}{j!}\frac{d^j\widehat{Q}(0)}{dt^j} + \frac{\tau^{j+1}}{(j+1)!}\frac{d^{j+1}\widetilde{Q}(0)}{dt^{j+1}}\right).$$

As far as the expansions are concerned, the equations have exactly the same form as that of Section 4.3. Note, however, that the partitioned vector $\varphi_i(t)$ has the form

$$\varphi_i(t) = (\varphi_i^1(t), \dots, \varphi_i^l(t), \varphi_i^*(t)), \ i = 0, 1, \dots, n,$$

where $\varphi_i^k(t)$, $k = 1, \dots, l$, is an m_k row vector and $\varphi_i^*(t)$ is an m_* row vector. A similar partition holds for the vector $\psi_i(t)$. To construct these functions, we begin with $i = 0$. Writing $\varphi_0(t)\widetilde{Q}(t) = 0$ in terms of the corresponding partition, we have

$$\varphi_0^k(t)\widetilde{Q}^k(t) + \varphi_0^*(t)\widetilde{Q}_*^k(t) = 0, \ \text{for } k = 1, \dots, l, \ \text{and}$$

$$\varphi_0^*(t)\widetilde{Q}_*(t) = 0.$$

Since $\widetilde{Q}_*(t)$ is stable, it is nonsingular. The last equation above implies $\varphi_0^*(t) = 0_{m_*} = (0, \dots, 0) \in \mathbb{R}^{1 \times m_*}$. Consequently, as in the previous section, for each $k = 1, \dots, l$, the weak irreducibility of $\widetilde{Q}^k(t)$ implies that $\varphi_0^k(t) = \vartheta_0^k(t)\nu^k(t)$, for some scalar function $\vartheta_0^k(t)$. Equivalently,

$$\varphi_0(t) = (\vartheta_0^1(t)\nu^1(t), \dots, \vartheta_0^l(t)\nu^l(t), 0_{m_*}).$$

Comparing the equation above with the corresponding expression of $\varphi_0(t)$ in Section 4.3, the only difference is the addition of the m_*-dimensional row vector 0_{m_*}.

Remark 4.39. Note that the dominant term in the asymptotic expansion is $\varphi_0(t)$, in which the probabilities corresponding to the transient states are 0. Thus, the probability corresponding to $\alpha^\varepsilon(t) \in \{$ transient states $\}$ is negligibly small.

Define

$$\widetilde{\mathbb{1}}_*(t) = \begin{pmatrix} \mathbb{1}_{m_1} & & & \\ & \ddots & & \\ & & \mathbb{1}_{m_l} & \\ a_{m_1}(t) & \cdots & a_{m_l}(t) & 0_{m_* \times m_*} \end{pmatrix} \qquad (4.85)$$

where $a_{m_k}(t) = -\widetilde{Q}_*^{-1}(t)\widetilde{Q}_*^k(t)\mathbb{1}_{m_k}$ for $k = 1, \ldots, l$, and $0_{m_* \times m_*}$ is the zero matrix in $\mathbb{R}^{m_* \times m_*}$.

It is readily seen that

$$\widetilde{Q}(t)\widetilde{\mathbb{1}}_*(t) = 0 \text{ for each } t \in [0, T].$$

In view of (4.83), it follows that

$$\frac{d}{dt}(\vartheta_0^1(t), \ldots, \vartheta_0^l(t), 0_{m_*})$$

$$= (\vartheta_0^1(t), \ldots, \vartheta_0^l(t), 0_{m_*})\overline{Q}(t),$$

(4.86)

where

$$\overline{Q}(t) = \operatorname{diag}(\nu^1(t), \ldots, \nu^l(t), 0_{m_* \times m_*})\widehat{Q}(t)\widetilde{\mathbb{1}}_*(t).$$

We write $\widehat{Q}(t)$ as follows:

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

where for each $t \in [0, T]$,

$$\widehat{Q}^{11}(t) \in \mathbb{R}^{(m-m_*) \times (m-m_*)}, \ \widehat{Q}^{12}(t) \in \mathbb{R}^{(m-m_*) \times m_*},$$

$$\widehat{Q}^{21}(t) \in \mathbb{R}^{m_* \times (m-m_*)}, \text{ and } \widehat{Q}^{22}(t) \in \mathbb{R}^{m_* \times m_*}.$$

Let

$$\overline{Q}_*(t) = \operatorname{diag}(\nu^1(t), \ldots, \nu^l(t)) \left(\widehat{Q}^{11}(t)\widetilde{\mathbb{1}} + \widehat{Q}^{12}(t)(a_{m_1}(t), \ldots, a_{m_l}(t)) \right).$$

Then $\overline{Q}(t) = \operatorname{diag}(\overline{Q}_*(t), 0_{m_* \times m_*})$. Moreover, the differential equation (4.86) becomes

$$\frac{d}{dt}(\vartheta_0^1(t), \ldots, \vartheta_0^l(t)) = (\vartheta_0^1(t), \ldots, \vartheta_0^l(t))\overline{Q}_*(t).$$

Remark 4.40. Note that the submatrix $\widehat{Q}^{12}(t)$ in $\widehat{Q}(t)$ determines the jump rates of the underlying Markov chain from a recurrent state in $\mathcal{M}_1 \cup \cdots \cup \mathcal{M}_l$ to a transient state in $\mathcal{M}_*$. If the magnitude of the entries of $\widehat{Q}^{12}(t)$ is small, then the transient state can be safely ignored because the contribution of $\widehat{Q}^{12}(t)$ to $\overline{Q}(t)$ is small. On the other hand, if $\widehat{Q}^{12}(t)$ is not negligible, then one has to be careful to include the corresponding terms in $\overline{Q}(t)$.

We now determine the initial value $\vartheta_0^k(0)$. In view of the asymptotic expansions $y_n^\varepsilon(t)$ and the initial-value consistency condition in (4.53), it is necessary that for $k = 1, \ldots, l$,

$$\vartheta_0^k(0) = \varphi_0^k(0)\mathbb{1}_{m_k} = \lim_{\delta \to 0} \lim_{\varepsilon \to 0} p^{\varepsilon,k}(\delta)\mathbb{1}_{m_k}, \tag{4.87}$$

where $p^\varepsilon(t) = (p^{\varepsilon,1}(t), \ldots, p^{\varepsilon,l}(t), p^{\varepsilon,*}(t))$ is a solution to (4.40). Here $p^{\varepsilon,k}(t)$ has dimensions compatible with $\varphi_0^k(0)$ and $\psi_0^k(0)$. Similarly, we write the partition of the initial vector as $p^0 = (p^{0,1}, \ldots, p^{0,l}, p^{0,*})$. The next theorem establishes the desired *consistency* of the initial values. Its proof is placed in Appendix A.4.

Theorem 4.41. *Assume* (A4.5) *and* (A4.6). *Then for* $k = 1, \ldots, l$,

$$\lim_{\delta \to 0} \left(\limsup_{\varepsilon \to 0} \left| p^{\varepsilon,k}(\delta)\mathbb{1}_{m_k} - \left(p^{0,k}\mathbb{1}_{m_k} - p^{0,*}\widetilde{Q}_*^{-1}(0)\widetilde{Q}_*^k(0)\mathbb{1}_{m_k} \right) \right| \right) = 0.$$

Remark 4.42. In view of this theorem, the initial value should be given as

$$\vartheta_0^k(0) = p^{0,k}\mathbb{1}_{m_k} - p^{0,*}\widetilde{Q}_*^{-1}(0)\widetilde{Q}_*^k(0)\mathbb{1}_{m_k}. \tag{4.88}$$

Therefore, in view of (4.88), to make sure that the initial condition satisfies the probabilistic interpretation, it is necessary that

$$\vartheta_0^k(t) \geq 0 \text{ for } t \in [0, T] \text{ and } k = 1, \ldots, l \text{ and } \sum_{k=1}^{l} \vartheta_0^k(0) = 1.$$

In view of the structure of the $\widetilde{Q}(0)$ matrix, for each $k = 1, \ldots, l$, all components of the vector $\widetilde{Q}_*^k(0)\mathbb{1}_{m_k}$ are nonnegative. Note that the solution of the differential equation

$$\frac{dy(t)}{dt} = y(t)\widetilde{Q}(0),$$

$$y(0) = p^0$$

is $p^0 \exp(\widetilde{Q}(0)t)$. This implies that all components of $p^{0,*} \exp(\widetilde{Q}_*(0)t)$ are nonnegative. By virtue of the stability of $\widetilde{Q}_*(0)$,

$$-\widetilde{Q}_*^{-1}(0) = \int_0^\infty \exp(\widetilde{Q}_*(0)t)dt.$$

Thus all components of $-p^{0,*}\widetilde{Q}_*^{-1}(0)$ are nonnegative, and as a result, the inner product

$$-p^{0,*}\widetilde{Q}_*^{-1}(0)\widetilde{Q}_*^k(0)\mathbb{1}_{m_k}$$

is nonnegative. It follows that for each $k = 1, \ldots, l$, $\vartheta_0^k(0) \geq p^{0,k} \mathbb{1}_{m_k} \geq 0$. Moreover,

$$\sum_{k=1}^{l} \vartheta_0^k(0) = \sum_{k=1}^{l} p^{0,k} \mathbb{1}_{m_k} - p^{0,*} \widetilde{Q}_*^{-1}(0) \left(\sum_{k=1}^{l} \widetilde{Q}_*^k(0) \mathbb{1}_{m_k} \right)$$

$$= (1 - p^{0,*} \mathbb{1}_{m_*}) - p^{0,*} \widetilde{Q}_*^{-1}(0)(-\widetilde{Q}_*(0) \mathbb{1}_{m_*}) = 1. \tag{4.89}$$

Before treating the terms in $\psi_0(\cdot)$, let us give an estimate on $\exp(\widetilde{Q}(0)t)$.

Lemma 4.43. *Set*

$$\pi_* = \begin{pmatrix} \mathbb{1}_{m_1} \nu^1(0) & & & \\ & \ddots & & \\ & & \mathbb{1}_{m_l} \nu^l(0) & \\ a_{m_1}(0)\nu^1(0) & \cdots & a_{m_l}(0)\nu^l(0) & \mathbb{1}_{m_*} 0_{m_*} \end{pmatrix}.$$

Then there exist positive constants K and $\kappa_{0,0}$ such that

$$\left| \exp(\widetilde{Q}(0)\tau) - \pi_* \right| \leq K \exp(-\kappa_{0,0}\tau), \tag{4.90}$$

for $\tau \geq 0$.

Proof: To prove (4.90), it suffices to show for any m-row vector y^0,

$$\left| y^0(\exp(\widetilde{Q}(0)\tau) - \pi_*) \right| \leq K|y^0| \exp(-\kappa_0\tau).$$

Given $y^0 = (y^{0,1}, \ldots, y^{0,l}, y^{0,*}) \in \mathbb{R}^{1 \times m}$, let

$$y(\tau) = (y^1(\tau), \ldots, y^l(\tau), y^*(\tau)) = y^0 \exp(\widetilde{Q}(0)\tau).$$

Then, $y(\tau)$ is a solution to

$$\frac{dy(\tau)}{d\tau} = y(\tau)\widetilde{Q}(0), \quad y(0) = y^0.$$

It follows that

$$y^*(\tau) = y^{0,*} \exp(\widetilde{Q}_*(0)\tau)$$

and for $k = 1, \ldots, l$,

$$y^k(\tau) = y^{0,k} \exp(\widetilde{Q}^k(0)\tau) + \int_0^\tau y^*(s)\widetilde{Q}_*^k(0) \exp(\widetilde{Q}^k(0)(\tau - s))ds.$$

For each $k = 1, \ldots, l$, we have

$$y^k(\tau) - \left(y^{0,k} \mathbb{1}_{m_k} \nu^k(0) + y^{0,*} \int_0^\infty \exp(\widetilde{Q}_*(0)s) ds \widetilde{Q}_*^k(0) \mathbb{1}_{m_k} \nu^k(0) \right)$$

$$= y^{0,k} \left(\exp(\widetilde{Q}^k(0)\tau) - \mathbb{1}_{m_k} \nu^k(0) \right)$$

$$+ y^{0,*} \int_0^\tau \exp(\widetilde{Q}_*(0)s) \widetilde{Q}_*^k(0) \left(\exp(\widetilde{Q}^k(0)(\tau - s)) - \mathbb{1}_{m_k} \nu^k(0) \right) ds$$

$$- y^{0,*} \int_\tau^\infty \exp(\widetilde{Q}_*(0)s) \widetilde{Q}_*^k(0) \mathbb{1}_{m_k} \nu^k(0) ds.$$

By virtue of the stability of $\widetilde{Q}_*(0)$, the last term above is bounded above by $K |y^{0,*}| \exp(-\kappa_* \tau)$. Recall that by virtue of Lemma 4.4, for some $\kappa_{0,k} > 0$,

$$\left| \exp(\widetilde{Q}^k(0)\tau) - \mathbb{1}_{m_k} \nu^k(0) \right| \leq K \exp(-\kappa_{0,k} \tau).$$

Choose $\kappa_{0,0} = \min(\kappa_*, \min_k \{\kappa_{0,k}\})$. The terms in the second and the third lines above are bounded by $K |y^0| \exp(-\kappa_{0,0} \tau)$. The desired estimate thus follows. $\qquad\square$

Next consider the first equation in the initial-layer expansions:

$$\frac{d\psi_0(\tau)}{d\tau} = \psi_0(\tau) \widetilde{Q}(0).$$

The solution to this equation can be written as

$$\psi_0(\tau) = \psi_0(0) \exp(\widetilde{Q}(0)\tau).$$

To be able to match the asymptotic expansion, choose

$$\psi_0(0) = p^0 - \varphi_0(0).$$

Thus,

$$\psi_0(\tau) = (p^0 - \varphi_0(0)) \exp(\widetilde{Q}(0)\tau)$$

$$= (p^0 - \varphi_0(0)) \left(\exp(\widetilde{Q}(0)\tau) - \pi_* \right) + (p^0 - \varphi_0(0)) \pi_*.$$

By virtue of the choice of $\varphi_0(0)$, it is easy to show that

$$(p^0 - \varphi_0(0)) \pi_* = 0.$$

Therefore, in view of Lemma 4.43, $\psi_0(\cdot)$ decays exponentially fast in that for some constants K and $\kappa_{0,0} > 0$ given in Lemma 4.43,

$$|\psi_0(\tau)| \leq K \exp(-\kappa_{0,0} \tau), \ \tau \geq 0.$$

We have obtained $\varphi_0(\cdot)$ and $\psi_0(\cdot)$. To proceed, set

$$b_0(t) = \frac{d\varphi_0(t)}{dt} - \varphi_0(t)\widehat{Q}(t)$$

and

$$b_0(t) = (b_0^1(t), \ldots, b_0^l(t), b_0^*(t)).$$

Note that $b_0(t)$ is a completely known function.

In view of the second equation in (4.83),

$$\varphi_1^k(t)\widetilde{Q}^k(t) + \varphi_1^*(t)\widetilde{Q}_*^k(t) = b_0^k(t) \text{ for } k = 1, \ldots, l,$$

$$\varphi_1^*(t)\widetilde{Q}_*(t) = b_0^*(t).$$

(4.91)

Solving the last equation in (4.91) yields

$$\varphi_1^*(t) = b_0^*(t)\widetilde{Q}_*^{-1}(t).$$

Putting this back into the first l equations of (4.91) leads to

$$\varphi_1^k(t)\widetilde{Q}^k(t) = b_0^k(t) - b_0^*(t)\widetilde{Q}_*^{-1}(t)\widetilde{Q}_*^k(t). \qquad (4.92)$$

Again, the right side is a known function. In view of the choice of $\varphi_0(\cdot)$ and (4.86), we have $b_0(t)\widetilde{\mathbb{1}}_*(t) = 0$. This implies

$$b_0^k(t)\mathbb{1}_{m_k} - b_0^*(t)\widetilde{Q}_*^{-1}(t)\widetilde{Q}_*^i(t)\mathbb{1}_{m_k}$$

$$= b_0^k(t)\mathbb{1}_{m_k} + b_0^*(t)a_{m_k}(t) = 0.$$

Therefore, (4.92) has a particular solution $\widetilde{b}_0^k(t)$ with

$$\widetilde{b}_0^k(t)\mathbb{1}_{m_k} = 0, \text{ for } k = 1, \ldots, l.$$

As in the previous section, we write the solution of $\varphi_1^k(t)$ as a sum of the homogeneous solution and a solution of the inhomogeneous equation $\widetilde{b}_0^k(t)$, that is,

$$\varphi_1^k(t) = \vartheta_1^k(t)\nu^k(t) + \widetilde{b}_0^k(t) \text{ for } k = 1, \ldots, l.$$

In view of

$$\widetilde{Q}(t)\widetilde{\mathbb{1}}_*(t) = 0 \text{ and}$$

$$\widetilde{b}_0^k(t)\mathbb{1}_{m_k} = 0,$$

using the equation

$$\varphi_2(t)\widetilde{Q}(t) = \frac{d\varphi_1(t)}{dt} - \varphi_1(t)\widehat{Q}(t),$$

we obtain that

$$\frac{d}{dt}(\vartheta_1^1(t), \ldots, \vartheta_1^l(t), 0)$$

$$= (\vartheta_1^1(t), \ldots, \vartheta_1^l(t), 0)\overline{Q}(t) + \widetilde{b}_0(t)\widehat{Q}(t)\widetilde{1}_*(t) \qquad (4.93)$$

$$- \left(\frac{d\widetilde{b}_0^*(t)}{dt}\right)\left(a_{m_1}(t), \ldots, a_{m_l}(t), 0'_{m_*}\right).$$

The initial value $\vartheta_1(0)$ will be determined in conjunction with the initial value of $\psi_1(\cdot)$ next.

Note that in comparison with the differential equation governing $\vartheta_1(t)$ in Section 4.3, the equation (4.93) has an extra term involving the derivative of $\widetilde{b}_0^*(t)$.

To determine $\psi_1(\cdot)$, solving the equation in (4.84) with $i = 1$, we have

$$\psi_1(\tau) = \psi_1(0)\exp(\widetilde{Q}(0)\tau)$$

$$+ \int_0^\tau \psi_0(0)\exp(\widetilde{Q}(0)s)\widehat{Q}(0)\exp(\widetilde{Q}(0)(\tau - s))ds$$

$$+ \int_0^\tau s\psi_0(0)\exp(\widetilde{Q}(0)s)\left(\frac{d\widetilde{Q}(0)}{dt}\right)\exp(\widetilde{Q}(0)(\tau - s))ds.$$

Choose the initial values of $\psi_1(0)$ and $\vartheta_1^k(0)$ as follows:

$$\psi_1(0) = -\varphi_1(0),$$

$$\vartheta_1^k(0) = -\psi_1^k(0)\mathbb{1}_{m_k},$$

$$\psi_1(0)\pi_* = -\left(\int_0^\infty \psi_0(0)\exp(\widetilde{Q}(0)s)ds\right)\widehat{Q}(0)\pi_* \qquad (4.94)$$

$$- \left(\int_0^\infty s\psi_0(0)\exp(\widetilde{Q}(0)s)ds\right)\frac{d\widetilde{Q}(0)}{dt}\pi_*$$

$$:= -\overline{\psi}_0\pi_*.$$

Write $\overline{\psi}_0 = (\overline{\psi}_0^1, \ldots, \overline{\psi}_0^l, \overline{\psi}_0^*)$. Then the definition of π_* implies that

$$\psi_1^k(0)\mathbb{1}_{m_k} + \psi_1^*(0)a_{m_k}(0) = -(\overline{\psi}_0^k\mathbb{1}_{m_k} + \overline{\psi}_0^*a_{m_k}(0)).$$

Recall that

$$\varphi_1^*(0) + \psi_1^*(0) = 0$$

and

$$\varphi_1^*(t) = b_0^*(t)\widetilde{Q}_*^{-1}(t).$$

It follows that

$$\psi_1^k(0)\mathbb{1}_{m_k} = -(\overline{\psi}_0^k \mathbb{1}_{m_k} + \overline{\psi}_0^* a_{m_k}(0)) + b_0^*(0)\widetilde{Q}_*^{-1}(0)a_{m_k}(0).$$

Moreover, it can be verified that $|\psi_1(\tau)| \leq K \exp(-\kappa_{1,0}\tau)$ for some $0 < \kappa_{1,0} < \kappa_{0,0}$.

Remark 4.44. Note that there is an extra term

$$\left(\int_0^\infty s\psi_0(0)\exp(\widetilde{Q}(0)s)ds\right)\frac{d\widetilde{Q}(0)}{dt}\pi_*$$

involved in the equation determining $\vartheta_1(0)$ in (4.94). This term does not vanish as in Section 4.3 because generally $((d/dt)\widetilde{Q}(0))\pi_* \neq 0$.

To obtain the desired asymptotic expansion, continue inductively. For each $i = 2, \ldots, n$, we first obtain the solution of $\varphi_i(t)$ with the "multiplier" given by the solution of the differential equation but with unspecified condition $\vartheta_i(0)$; solve $\psi_i(t)$ with the as yet unavailable initial condition $\psi_i(0) = -\varphi_i(0)$. Next jointly prove the exponential decay properties of $\psi_i(\tau)$ and obtain the solution $\vartheta_i(0)$. The equation to determine $\vartheta_i(0)$ with transient states becomes

$$\psi_i(0)\pi_*$$
$$= -\left(\sum_{j=0}^{i-1}\int_0^\infty \psi_{i-j-1}(s)\left(\frac{s^j}{j!}\frac{d^j\widehat{Q}(0)}{dt^j} + \frac{s^{j+1}}{(j+1)!}\frac{d^{j+1}\widetilde{Q}(0)}{dt^{j+1}}\right)ds\right)\pi_*.$$

In this way, we have constructed the asymptotic expansion with transient states. In addition, we can show that $\varphi_i(\cdot)$ are smooth and $\psi_i(\cdot)$ satisfies $|\psi_i(\tau)| \leq K \exp(-\kappa_{i,0}\tau)$ for some $0 < \kappa_{i,0} < \kappa_{i-1,0} < \kappa_{0,0}$. Similarly as in the case with all recurrent states, we establish the following theorem.

Theorem 4.45. *Suppose* (A4.5) *and* (A4.6) *hold. Then an asymptotic expansion*

$$y_n^\varepsilon(t) = \sum_{i=0}^n \left(\varepsilon^i\varphi_i(t) + \varepsilon^i\psi_i\left(\frac{t}{\varepsilon}\right)\right)$$

can be constructed such that for $i = 0, \ldots, n$,

(a) $\varphi_i(\cdot)$ is $(n + 1 - i)$-times continuously differentiable on $[0, T]$;

(b) $|\psi_i(t)| \leq K \exp(-\kappa_0 t)$ for some $K > 0$ and $0 < \kappa_0 < \kappa_{i,0}$;

(c) $|p^\varepsilon(t) - y_n^\varepsilon(t)| = O(\varepsilon^{n+1})$ uniformly in $t \in [0, T]$.

Example 4.46. Let $\widetilde{Q}(t) = \widetilde{Q}$, a constant matrix such that

$$\widetilde{Q} = \begin{pmatrix} -1 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 1 & 0 & -2 & 1 \\ 0 & 1 & 1 & -2 \end{pmatrix} \quad \text{and} \quad \widehat{Q} = 0.$$

In this example,

$$\widetilde{Q}^1 = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}, \quad \widetilde{Q}_* = \begin{pmatrix} -2 & 1 \\ 1 & -2 \end{pmatrix}, \quad \text{and} \quad \widetilde{Q}_*^1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

The last two rows in $\widetilde{Q}$ represent the jump rates corresponding to the transient states. The matrix $\widetilde{Q}^1$ is weakly irreducible and $\widetilde{Q}_*$ is stable. Solving the forward equation gives us

$$p^\varepsilon(t) = (p_1^\varepsilon(t), p_2^\varepsilon(t), p_3^\varepsilon(t), p_4^\varepsilon(t)),$$

where

$$p_1^\varepsilon(t) = \frac{1}{2} + \frac{1}{2}\left[(-p_3^0 - p_4^0)\exp\left(-\frac{t}{\varepsilon}\right)\right.$$

$$+(p_1^0 - p_2^0 + p_3^0 - p_4^0)\exp\left(-\frac{2t}{\varepsilon}\right)$$

$$\left.+(-p_3^0 + p_4^0)\exp\left(-\frac{3t}{\varepsilon}\right)\right],$$

$$p_2^\varepsilon(t) = \frac{1}{2} + \frac{1}{2}\left[(-p_3^0 - p_4^0)\exp\left(-\frac{t}{\varepsilon}\right)\right.$$

$$+(-p_1^0 + p_2^0 - p_3^0 + p_4^0)\exp\left(-\frac{2t}{\varepsilon}\right)$$

$$\left.+(p_3^0 - p_4^0)\exp\left(-\frac{3t}{\varepsilon}\right)\right],$$

$$p_3^\varepsilon(t) = \frac{1}{2}\left[(p_3^0 + p_4^0)\exp\left(-\frac{t}{\varepsilon}\right) + (p_3^0 - p_4^0)\exp\left(-\frac{3t}{\varepsilon}\right)\right],$$

$$p_4^\varepsilon(t) = \frac{1}{2}\left[(p_3^0 + p_4^0)\exp\left(-\frac{t}{\varepsilon}\right) + (-p_3^0 + p_4^0)\exp\left(-\frac{3t}{\varepsilon}\right)\right].$$

It is easy to see that $\varphi_0(t) = (1/2, 1/2, 0, 0)$ and

$$|p^\varepsilon(t) - \varphi_0(t)| \le K \exp\left(-\frac{t}{\varepsilon}\right).$$

The limit behavior of the underlying Markov chain as $\varepsilon \to 0$ is determined by $\varphi_0(t)$ for $t > 0$. It is clear that the probability of the Markov chain staying at the transient states is very small for small ε.

Remark 4.47. The model discussed in this section has the extra ingredient of including transient states as compared with that of Section 4.3. The main feature is embedded in the last few rows of the $\widetilde{Q}(t)$ matrix. One of the crucial points here is that the matrix $\widetilde{Q}_*(t)$ in the right corner is Hurwitzian. This stability condition guarantees the exponential decay properties of the boundary layers. As far as the regular part (or the outer) expansion is concerned, we have that the last subvector $\varphi_0^*(t) = 0$. The determination of the initial conditions $\vartheta_i(0)$ uses the same technique as before, namely, matching the outer terms and inner layers. The procedure involves recursively solving a sequence of algebraic and differential equations. Although the model is seemingly more general, the methods and techniques involved in obtaining the asymptotic expansion and proof of the results are essentially the same as in the previous section. The notation is slightly more complex, nevertheless.

4.6   Remarks on Countable-State-Space Cases

4.6.1   Countable-State Spaces: Part I

This section presents an extension of the singularly perturbed Markov chains with fast and slow components and finite-state spaces. In this section, the generator $\widetilde{Q}(\cdot)$ is a block-diagonal matrix consisting of infinitely many blocks each of which is of finite dimension. The generator $Q^\varepsilon(t)$ still has the form (4.39). However,

$$\widetilde{Q}(t) = \begin{pmatrix} \widetilde{Q}^1(t) & & & & \\ & \widetilde{Q}^2(t) & & & \\ & & \ddots & & \\ & & & \widetilde{Q}^k(t) & \\ & & & & \ddots \end{pmatrix}, \qquad (4.95)$$

where $\widetilde{Q}^k(t) \in \mathbb{R}^{m_k \times m_k}$ is a generator of an appropriate Markov chain with finite-state space, and $\widehat{Q}(t)$ is an infinite-dimensional matrix and is a generator of a Markov chain having a countable-state space, that is,

$\widehat{Q}(t) = (\widehat{q}_{ij}(t))$ such that

$$\widehat{q}_{ij}(t) \geq 0 \text{ for } i \neq j, \text{ and } \sum_j \widehat{q}_{ij}(t) = 0.$$

We aim at deriving asymptotic results under the current setting. To do so, assume that the following condition holds:

(**A4.7**) For $t \in [0, T]$, $\widetilde{Q}^k(t)$, for $k = 1, 2, \ldots$, are weakly irreducible.

Parallel to the development of Section 4.3, the solution of $\varphi_i(\cdot)$ can be constructed similar to that of Theorem 4.29 as in (4.44) and (4.45). In fact, we obtain $\varphi_0(\cdot)$ from (4.49) and (4.50) with $l = \infty$; the difference is that now we have an infinite number of equations. Similarly, for all $k = 1, 2, \ldots$ and $i = 0, 1, \ldots, n+1$, $\varphi_i(\cdot)$ can be obtained from

$$
\begin{aligned}
&\varphi_0(t)\widetilde{Q}(t) = 0, \text{ if } i = 0 \\
&\varphi_i(t)\widetilde{Q}(t) = \frac{d\varphi_{i-1}(t)}{dt} - \varphi_{i-1}(t)\widehat{Q}(t), \text{ if } i \geq 1 \\
&\varphi_i^k(t)\mathbb{1}_{m_k} = \vartheta_i^k(t), \\
&\frac{d\vartheta_i(t)}{dt} = \vartheta_i(t)\overline{Q}(t) + \widetilde{b}_{i-1}(t)\widehat{Q}(t)\widetilde{\mathbb{1}}.
\end{aligned}
\tag{4.96}
$$

The problem is converted to one that involves infinitely many algebraic differential equations. The same technique as presented before still works.

Nevertheless, the boundary layer corrections deserve more attention. Let us start with $\psi_0(\cdot)$, which is the solution of the abstract Cauchy problem

$$
\begin{aligned}
&\frac{d\psi_0(\tau)}{d\tau} = \psi_0(\tau)\widetilde{Q}(0), \\
&\psi_0(0) = p^0 - \varphi_0(0).
\end{aligned}
\tag{4.97}
$$

To continue our study, one needs the notion of semigroup (see Dunford and Schwartz [52], and Pazy [172]). Recall that for a Banach space $\mathbb{B}$, a one-parameter family $T(t)$, $0 \leq t < \infty$, of bounded linear operators from $\mathbb{B}$ into $\mathbb{B}$ is a semigroup of bounded linear operators on $\mathbb{B}$ if (i) $T(0) = I$ and (ii) $T(t + s) = T(t)T(s)$ for every $t, s \geq 0$.

Let $\mathbb{R}^\infty$ be the sequence space with a canonical element $x = (x_1, x_2, \ldots) \in \mathbb{R}^\infty$. Let $A = (a_{ij})$ satisfying $A : \mathbb{R}^\infty \mapsto \mathbb{R}^\infty$, equipped with the l_1-norm

$$|A|_1 = \sup_j \sum_i |a_{ij}|;$$

(see Hutson and Pym [90, p. 74]) Using the definition of semigroup above, the solution of (4.97) is

$$\psi_0(\tau) = T(\tau)\psi_0(0),$$

where $T(\tau)$ is a one-parameter family of semigroups generated by $\widetilde{Q}(0)$. Moreover, since $\widetilde{Q}(0)$ is a bounded linear operator, $\exp(\widetilde{Q}(0)\tau)$ still makes sense. Thus $T(\tau)\psi_0(0) = \psi_0(0)\exp(\widetilde{Q}(0)\tau)$, where

$$T(\tau) = \exp(\widetilde{Q}(0)\tau) = \sum_{j=0}^{\infty} \frac{\left(\widetilde{Q}(0)\tau\right)^j}{j!}$$

$$= \mathrm{diag}\left(\exp\left(\widetilde{Q}^1(0)\tau\right), \ldots, \exp\left(\widetilde{Q}^k(0)\tau\right), \ldots\right).$$

Therefore, the solution has the same form as in the previous section. Under (A4.7), exactly the same argument as in the proof of Lemma 4.4 yields that for each $k = 1, 2, \ldots,$

$$\exp(\widetilde{Q}^k(0)\tau) \to \mathbb{1}_{m_k}\nu^k(0) \text{ as } \tau \to \infty$$

and the convergence takes place at an exponential rate, that is,

$$\left|\exp(\widetilde{Q}^k(0)\tau) - \mathbb{1}_{m_k}\nu^k(0)\right| \leq K\exp(-\kappa_k\tau),$$

for some $\kappa_k > 0$. In order to obtain a valid asymptotic expansion, another piece of assumption is needed. That is, these κ_k, for all $k = 1, 2, \ldots,$ are uniformly bounded below by a positive constant κ_0.

(**A4.8**) There exists a positive number $\kappa_0 = \min_k\{\kappa_k\} > 0$.

Set

$$\widetilde{\mathbb{1}} = \mathrm{diag}\left(\mathbb{1}_{m_1}, \ldots, \mathbb{1}_{m_k}, \ldots\right) \text{ and } \nu(0) = \mathrm{diag}\left(\nu^1(0), \ldots, \nu^k(0), \ldots\right).$$

In view of (A4.8)

$$\left|\exp(\widetilde{Q}(0)\tau) - \widetilde{\mathbb{1}}\nu(0)\right|_1 \leq \sup_k \left|\exp(\widetilde{Q}^k(0)\tau) - \mathbb{1}_{m_k}\nu^k(0)\right|$$

$$\leq K\exp(-\kappa_0\tau). \tag{4.98}$$

The exponential decay property of $\psi_0(\cdot)$ is thus established. Likewise, it can be proved that all $\psi_i(\cdot)$ for $i = 1, \ldots, n+1$, satisfy the exponential decay property. From here on, we can proceed as in the previous section to get the error estimate and verify the validity of the asymptotic expansion. In short the following theorem is obtained.

Theorem 4.48. *Suppose conditions* (A4.7) *and* (A4.8) *are satisfied. Then the results in Theorem 4.29 hold for the countable-state-space model with* $\widetilde{Q}(\cdot)$ *given by* (4.95).

4.6.2   Countable-State Spaces: Part II

The aim of this section is to develop further results on singularly perturbed Markov chains with fast and slow components whose generators are infinite-dimensional matrices but in different form from that described in Section 4.6.1. The complexity as well as difficulty increase. A number of technical issues also arise. One idea arises almost immediately: to approximating the underlying system via a Galerkin-kind procedure, that is, to approximate an infinite-dimensional system by finite-dimensional truncations. Unfortunately, this does not work in the setting of this section. We will return to this question at the end of this section.

To proceed, as in the previous sections, the first step invariably involves the solution of algebraic differential equations in the constructions of the approximating functions. One of the main ideas used is the Fredholm alternative. There are analogues to the general setting in Banach spaces for compact operators. Nevertheless, the infinite-dimensional matrices are in fact more difficult to handle.

Throughout this section, we treat the class of generators with $|Q(t)|_1 < \infty$ only. We use $\mathbb{1}$ to denote the column vector with all components equal to

1. Consider $(\mathbb{1} \vdots Q(t))$ as an operator for a generator $Q(t)$ of a Markov chain with state space $\mathcal{M} = \{1, 2, \ldots\}$. To proceed, we first give the definitions

of irreducibility and quasi-stationary distribution. Set $Q_c(t) := (\mathbb{1} \vdots Q(t))$.

Definition 4.49. The generator $Q(t)$ is said to be weakly irreducible at $t_0 \in [0, T]$, for $w \in \mathbb{R}^\infty$, if the equation $wQ_c(t_0) = 0$ has only the zero solution. If $Q(t)$ is weakly irreducible for each $t \in [0, T]$, then it is said to be weakly irreducible on $[0, T]$.

Definition 4.50. A quasi-stationary distribution $\nu(t)$ (with respect to $Q(t)$) is a solution to (2.8) with the finite summation replaced by $\sum_{i=1}^\infty \nu_i(t) = 1$ that satisfies $\nu(t) \geq 0$.

As was mentioned before, the Fredholm alternative plays an important role in our study. For infinite-dimensional systems, we state another definition to take this into account.

Definition 4.51. A generator $Q(t)$ satisfies the F-Property if $wQ_c(t) = b$ has a unique solution for each $b \in \mathbb{R}^\infty$.

Note that for all weakly irreducible generators of finite dimension (i.e., generators for Markov chains with finite-state space), the F-Property above is automatically satisfied.

Since $\mathbb{1} \in l_\infty$ (l_∞ denotes the sequence space equipped with the l_∞ norm) for each $t \in [0, T]$, $Q(t) \in \mathbb{R}^\infty \times \mathbb{R}^\infty$. Naturally, we use the norm

$$|(z \vdots A)|_{\infty,1} = \max\left\{\sup_{z_j} |z_j|, \sup_j \sum_{i=1}^\infty |a_{ij}(t)|\right\}.$$

It is easily seen that

$$|Q_c(t)|_{\infty,1} \le \max\Big\{1, \sup_j \sum_i |q_{ij}(t)|\Big\}.$$

If a generator $Q(t)$ satisfies the F-Property, then it is weakly irreducible. In fact if $Q(t)$ satisfies the F-Property on $t \in [0, T]$, then $yQ_c(t) = 0$ has a unique solution $y = 0$.

By the definition of the generator, in particular the q-Property, $Q_c(t)$ is a bounded linear operator for each $t \in [0, T]$. If $Q_c(t)$ is bijective (i.e., one-to-one and onto), then it has a bounded inverse. This, in turn, implies that $Q_c(t)$ exhibits the F-Property. Roughly, the F-Property is a generalization of the conditions in dealing with finite-dimensional spaces. Recall from Section 4.2 that although $fQ(t) = b$ is not solvable uniquely, by adding an equation $f\mathbb{1} = c$, the system has a unique solution.

Owing to the inherited difficulty caused by the infinite dimensionality, the irreducibility and smoothness of $Q(\cdot)$ are not sufficient to guarantee the existence of asymptotic expansions. Stronger conditions are needed. In the sequel, for ease of presentation, we consider the model with $\widetilde{Q}(\cdot)$ irreducible and both $\widetilde{Q}(\cdot)$ and $\widehat{Q}(\cdot)$ infinite-dimensional.

For each t, we denote the spectrum of $Q(t)$ by $\sigma(Q(t))$. In view of Pazy [172] and Hutson and Pym [90], we have

$$\sigma(Q(t)) = \sigma_d(Q(t)) \cap \sigma_c(Q(t)) \cap \sigma_r(Q(t)),$$

where $\sigma_d(Q(t))$, $\sigma_c(Q(t))$, and $\sigma_r(Q(t))$ denote the discrete, continuous, and residue spectrum of $Q(t)$, respectively. The well-known linear operator theory implies that for a compact operator A, $\sigma_r(A) = \emptyset$, and the only possible candidate for $\sigma_c(A)$ is 0. Keeping this in mind, we assume that the following condition holds.

(**A4.9**) The following condition holds.
  (a) The smoothness condition (A4.4) is satisfied.
  (b) The generator $\widetilde{Q}(t)$ exhibits the F-Property.
  (c) $\sup_{t \in [0,T]} |\widetilde{Q}(t)|_1 < \infty$ and $\sup_{t \in [0,T]} |\widehat{Q}(t)| < \infty$.
  (d) The eigenvalue 0 of $\widetilde{Q}(t)$ has multiplicity 1 and 0 is not an accumulation point of the eigenvalues.
  (e) $\sigma_r(\widetilde{Q}(t)) = \emptyset$.

Remark 4.52. Item (a) above requires that the smoothness condition be satisfied and Item (b) requires the operator $(\mathbb{1}\vdots\widetilde{Q}(t))$ satisfy a Fredholm-alternative-like condition. Finally, (d) indicates the spectrum of $(\mathbb{1}\vdots\widetilde{Q}(t))$ is like a compact operator. Recall that for a compact linear operator, 0 is in its spectrum, and the only possible accumulation point is 0. Our conditions

mimic such a condition. It will be used when we prove the exponential decay property of the initial-layer terms.

Theorem 4.53. *Under condition* (A4.9), *the results in Theorem 4.29 hold for Markov chains with countable-state space.*

Proof: The proof is very similar to its finite-dimensional counterpart. We only point out the difference here.

As far as the regular part is concerned, we get the same equation (4.44). One thing to note is that we can no longer use Cramer's rule to solve the systems of equations. Without such an explicit representation of the solution, the smoothness of $\varphi_i(\cdot)$ needs to be proved by examining (4.44) directly. For example,

$$\sum_{i=1}^{\infty} \varphi_{0,i}(t) = 1,$$

$$\varphi_0(t)\widetilde{Q}(t) = 0,$$

can be rewritten as

$$\varphi_0(t)\left(\mathbb{1} \vdots \widetilde{Q}(t) \right) = (1, 0, \ldots). \tag{4.99}$$

Since $\widetilde{Q}(t)$ satisfies the F-Property, this equation has a unique solution.

To verify the differentiability, consider also

$$\varphi_0(t+\delta)\left(\mathbb{1} \vdots \widetilde{Q}(t+\delta) \right) = (1, 0, \ldots).$$

Examining the difference quotient leads to

$$0 = \frac{\varphi_0(t+\delta)\left(\mathbb{1} \vdots \widetilde{Q}(t+\delta) \right) - \varphi_0(t)\left(\mathbb{1} \vdots \widetilde{Q}(t) \right)}{\delta}$$

$$= \frac{[\varphi_0(t+\delta) - \varphi_0(t)]\left(\mathbb{1} \vdots \widetilde{Q}(t+\delta) \right)}{\delta}$$

$$+ \frac{\varphi_0(t)\left((\mathbb{1} \vdots \widetilde{Q}(t+\delta)) - (\mathbb{1} \vdots \widetilde{Q}(t)) \right)}{\delta}.$$

Taking the limit as $\delta \to 0$ and by virtue of the smoothness of $\widetilde{Q}(\cdot)$, we have

$$\lim_{\delta \to 0} \frac{[\varphi_0(t+\delta) - \varphi_0(t)]\left(\mathbb{1} \vdots \widetilde{Q}(t+\delta) \right)}{\delta} = -\varphi_0(t)\left(0 \vdots \frac{d\widetilde{Q}(t)}{dt} \right).$$

That is $(d/dt)\varphi_0(t)$ exists and is given by the solution of

$$\frac{d\varphi_0(t)}{dt}\left(\mathbb{1}\vdots\widetilde{Q}(t)\right) = -\varphi_0(t)\left(0\vdots\frac{d\widetilde{Q}(t)}{dt}\right).$$

Again by the F-Property, there is a unique solution for this equation. Higher-order derivatives of $\varphi_0(\cdot)$ and smoothness of $\varphi_i(\cdot)$ can be proved in a similar way.

As far as the initial-layer terms are concerned, since $\widetilde{Q}(0)$ is a bounded linear operator, the semigroup interpretation $\exp(\widetilde{Q}(0)\tau)$ makes sense. It follows from Theorem 1.4 of Pazy [172, p. 104] that the equation

$$\frac{d\psi_0(\tau)}{d\tau} = \psi_0(\tau)\widetilde{Q}(0), \quad \psi_0(0) = p_0 - \varphi_0(0)$$

has a unique solution.

To show that $\psi_0(\cdot)$ decays exponentially fast, we use an argument that is analogous to the finite-dimensional counterpart. Roughly, since the multiplicity of the eigenvalue 0 is 1, the subspace generated by the corresponding eigenvector v_0 is one-dimensional. Similar to the situation of Section 4.2, $\lim_{\tau\to\infty}\exp(\widetilde{Q}(0)\tau)$ exists and the limit must have identical rows. Denote the limit by $\overline{P}$. It then follows that

$$\left|\exp(\widetilde{Q}(0)\tau) - \overline{P}\right| \le K\exp(-\kappa_0\tau).$$

The meaning should be very clear. Upon "subtracting" the subspace generated by v_0, it ought to behave like $\exp(-\kappa_0\tau)$. A similar argument works for $i = 1, \ldots, n+1$, so the $\psi_i(\cdot)$ decay exponentially fast.   □

4.6.3   A Remark on Finite-Dimensional Approximation

Concerning the cases in Section 4.6.2, a typical way of dealing with infinite-dimensional Markov chains is to make a finite-dimensional approximation. Let $Q(t) = (q_{ij}(t))$, $t \ge 0$, denote a generator of a Markov chain with countable-state space. We consider an $N \times N$, $N = 1, 2, \ldots$, truncation matrix $Q_N(t) = (q_{ij}(t))_{i,j=1}^{N}$. Then $Q_N(t)$ is a subgenerator in the sense that $\sum_{j=1}^{N} q_{ij}(t) \le 0$, $i = 1, 2, \ldots, N$.

A first glance seems to indicate that the idea of subgenerator provides a way to treat the problem of approximating an infinite-dimensional generator by finite-dimensional matrices. In fact, Reuter and Ledermann used such an idea to derive the existence and uniqueness of the solution to the forward equation (see Bharucha-Reid [10]). Dealing with singularly perturbed chains with countable-state space, one would be interested in knowing whether a Galerkin-like approximation would work in the sense that an asymptotic expansion of a finite-dimensional system would provide

an approximation to the probability distribution. To be more precise, let $\alpha^\varepsilon(\cdot)$ denote the Markov chain generated by $Q(t)/\varepsilon$ and let

$$p^\varepsilon(t) = (P(\alpha^\varepsilon(t) = 1), \ldots, P(\alpha^\varepsilon(t) = k), \ldots).$$

Consider the following approximation via N-dimensional systems

$$\frac{dp^{\varepsilon,N}(t)}{dt} = \frac{1}{\varepsilon}p^{\varepsilon,N}(t)Q_N(t), \ p^{\varepsilon,N}(0) = p^0. \tag{4.100}$$

Using the techniques presented in the previous sections, we can find outer and inner expansions to approximate $p^{\varepsilon,N}(t)$. The questions are these: For small ε and large N, can we approximate $p^\varepsilon(t)$ by $p^{\varepsilon,N}(t)$? Can we approximate $p^{\varepsilon,N}(t)$ by $y_n^{\varepsilon,N}(t)$, where $y_n^{\varepsilon,N}(t)$ is an expansion of the form (4.43) when subgenerators are used? More importantly, can we use $y_n^{\varepsilon,N}(t)$ to approximate $p^\varepsilon(t)$?

Although $p_i^\varepsilon(t)$ can be approximated by its truncation $p_i^{\varepsilon,N}(t)$ for large N and $p^{\varepsilon,N}(t)$ can be expanded as $y_n^{\varepsilon,N}(t)$ for small ε, the approximation of $y_n^{\varepsilon,N}(t)$ to $p^\varepsilon(t)$ does not work in general because the limits as $\varepsilon \to 0$ and $N \to \infty$ are not interchangeable. This can be seen by considering the following example.

Let

$$Q(t) = Q = \begin{pmatrix} -1 & \dfrac{1}{2} & \dfrac{1}{2^2} & \cdots \\[2mm] \dfrac{1}{2} & -1 & \dfrac{1}{2^2} & \cdots \\[2mm] \dfrac{1}{2^2} & \dfrac{1}{2} & -1 & \cdots \\[2mm] \vdots & \vdots & \vdots & \vdots \end{pmatrix}.$$

Then for any N, the truncation matrix Q_N has only negative eigenvalues. It follows that the solution $p^{\varepsilon,N}(t)$ decays exponentially fast, i.e.,

$$\left|p^{\varepsilon,N}(t)\right| \leq C \exp\left(-\frac{\kappa_0 t}{\varepsilon}\right).$$

Thus, all terms in the regular part of $y_n^{\varepsilon,N}$ vanish. It is clear from this example that $y_n^{\varepsilon,N}(t)$ cannot be used to approximate $p^\varepsilon(t)$.

4.7   Remarks on Singularly Perturbed Diffusions

In this section, we present some related results on singular perturbations of diffusions. If in lieu of a discrete state space, one considers a continuous-state space, then naturally the singularly perturbed Markov chains become

singularly perturbed Markov processes. We illustrate the idea of matched asymptotic expansions for singularly perturbed diffusions. In this section, we only summarize the results and refer the reader to Khasminskii and Yin [116] for details of proofs. To proceed, consider the following example.

Example 4.54. This example discusses a model arising from stochastic control, namely, a controlled singularly perturbed system. As pointed out in Kushner [140] and Kokotovic, Bensoussan, and Blankenship [127], many control problems can be modeled by systems of differential equations, where the state variables can be divided into two coupled groups, consisting of "fast" and "slow" variables. A typical system takes the form

$$dx_1^\varepsilon = f_1(x_1^\varepsilon, x_2^\varepsilon, u)dt + \sigma_1(x_1^\varepsilon, x_2^\varepsilon)dw_1, \ x_1^\varepsilon(0) = x_1,$$

$$dx_2^\varepsilon = \frac{1}{\varepsilon}f_2(x_1^\varepsilon, x_2^\varepsilon, u)dt + \frac{1}{\sqrt{\varepsilon}}\sigma_2(x_1^\varepsilon, x_2^\varepsilon)dw_2, \ x_2^\varepsilon(0) = x_2,$$

where $w_1(\cdot)$ and $w_2(\cdot)$ are independent Brownian motions, $f_i(\cdot)$ and $\sigma_i(\cdot)$ for $i = 1, 2$ are suitable functions, u is the control variable, and $\varepsilon > 0$ is a small parameter. The underlying control problem is to minimize the cost function

$$J^\varepsilon(x_1, x_2, u) = E \int_0^T R(x_1^\varepsilon(t), x_2^\varepsilon(t), u)dt,$$

where $R(\cdot)$ is the running cost function. The small parameter $\varepsilon > 0$ signifies the relative rates of x_1^ε and x_2^ε. Such singularly perturbed systems have drawn much attention (see Bensoussan [8], Kushner [140], and the references therein). The system is very difficult to analyze directly; the approach of Kushner [140] is to use weak convergence methods to approximate the total system by the reduced system that is obtained using the differential equation for the slow variable, where the fast variable is fixed at its steady-state value as a function of the slow variable. In order to gain further insight, it is crucial to understand the asymptotic behavior of the rapidly changing process x_2^ε through the transition density given by the solution of the corresponding Kolmogorov-Fokker-Planck equations.

As demonstrated in the example above, a challenge common to many applications is to study the asymptotic behavior of the following problem. Let $\varepsilon > 0$ be a small parameter, and let $X_1^\varepsilon(\cdot)$ and $X_2^\varepsilon(\cdot)$ be real-valued diffusion processes satisfying

$$\begin{cases} dX_1^\varepsilon = a_1(t, X_1^\varepsilon, X_2^\varepsilon)dt + \sigma_1(t, X_1^\varepsilon, X_2^\varepsilon)dw_1, \\ dX_2^\varepsilon = \frac{1}{\varepsilon}a_2(t, X_1^\varepsilon, X_2^\varepsilon)dt + \frac{1}{\sqrt{\varepsilon}}\sigma_2(t, X_1^\varepsilon, X_2^\varepsilon)dw_2, \end{cases}$$

where the real-valued functions $a_1(t, x_1, x_2)$, $a_2(t, x_1, x_2)$, $\sigma_1(t, x_1, x_2)$, and $\sigma_2(t, x_1, x_2)$ represent the drift and diffusion coefficients, respectively, and $w_1(\cdot)$ and $w_2(\cdot)$ are independent and standard Brownian motions. Define a vector X as $X = (X_1, X_2)'$. Then $X^\varepsilon(\cdot) = (X_1^\varepsilon(\cdot), X_2^\varepsilon(\cdot))'$ is a diffusion process. This is a model treated in Khasminskii [113], in which a probabilistic approach was employed. It was shown that as $\varepsilon \to 0$, the fast component is averaged out and the slow component $X_1^\varepsilon(\cdot)$ has a limit $X_1^0(\cdot)$ such that

$$dX_1^0(t) = \bar{a}_1(X_1^0(t))dt + \bar{\sigma}_1(X_1^0(t))dw_1,$$

where

$$\bar{a}_1(t, x_1) = \int a_1(t, x_1, x_2)\mu(t, x_1, x_2)dx_2,$$

$$\bar{\sigma}_1(t, x_1) = \int \sigma_1(t, x_1, x_2)\mu(t, x_1, x_2)dx_2,$$

and $\mu(\cdot)$ is a limit density of the fast process $X_2^\varepsilon(\cdot)$.

To proceed further, it is necessary to investigate the limit properties of the rapidly changing process $X_2^\varepsilon(\cdot)$. To do so, consider the transition density of the underlying diffusion process. It is known that it satisfies the forward equation

$$\frac{\partial p^\varepsilon}{\partial t} = \frac{1}{\varepsilon}\mathcal{L}_2^* p^\varepsilon + \mathcal{L}_1^* p^\varepsilon,$$

$$p^\varepsilon(0, x_1, x_2) = p_0(x_1, x_2) \text{ with } p_0(x_1, x_2) \geq 0 \text{ and} \qquad (4.101)$$

$$\int\int p_0(x_1, x_2)dx_1 dx_2 = 1,$$

where

$$\mathcal{L}_1^*(t, x_1, x_2) \cdot = \frac{1}{2}\frac{\partial^2}{\partial x_1^2}(\sigma_1^2(t, x_1, x_2) \cdot) - \frac{\partial}{\partial x_1}(a_1(t, x_1, x_2) \cdot),$$

$$\mathcal{L}_2^*(t, x_1, x_2) \cdot = \frac{1}{2}\frac{\partial^2}{\partial x_2^2}(\sigma_2^2(t, x_1, x_2) \cdot) - \frac{\partial}{\partial x_2}(a_2(t, x_1, x_2) \cdot).$$

Similar to the discrete-state-space cases, the basic problems to be addressed are these: As $\varepsilon \to 0$, does the system display certain asymptotic properties? Is there any equilibrium distribution? If $p^\varepsilon(t, x_1, x_2) \to p(t, x_1, x_2)$ for some function $p(\cdot)$, can one get a handle on the error bound (i.e., a bound on $|p^\varepsilon(t, x_1, x_2) - p(t, x_1, x_2)|$)?

To obtain the desired asymptotic expansion in this case, one needs to make sure the quasi-stationary density exists. Note that for diffusions in unbounded domains, the quasi-stationary density may not exist. Loosely for the existence of the quasi-stationary distribution, it is necessary that

the Markov processes corresponding to $\mathcal{L}_2^*$ be positive recurrent for each fixed t. Certain sufficient conditions for the existence of the quasi-stationary density are provided in Il'in and Khasminskii [93]. An alternative way of handling the problem is to concentrate on a compact manifold. In doing so we are able to establish the existence of the quasi-stationary density. To illustrate, we choose the second alternative and suppose the following conditions are satisfied.

For each $t \in [0, T]$, $i, j = 1, 2$, and

- for each $x_2 \in \mathbb{R}$, $a_1(t, \cdot, x_2)$, $\sigma_1^2(t, \cdot, x_2)$ and $p_0(\cdot, x_2)$ are periodic with period 1;
- for each $x_1 \in \mathbb{R}$, $a_2(t, x_1, \cdot)$, $\sigma_2^2(t, x_1, \cdot)$ and $p_0(x_1, \cdot)$ are periodic with period 1.

There is an $n \in \mathbb{Z}_+$ such that for each $i = 1, 2$,

$$a_i(\cdot), \ \sigma_i^2(\cdot) \in C^{n+1,2(n+1),2(n+1)}, \ \text{for all } t \in [0, T], \ x_1, x_2 \in [0, 1], \quad (4.102)$$

the $(n + 1)$st partial with respect to t of $a_i(\cdot, x_1, x_2)$, and $\sigma_i^2(\cdot, x_1, x_2)$ are Lipschitz continuous uniformly in $x_1, x_2 \in [0, 1]$. In addition, for each $t \in [0, T]$ and each $x_1, x_2 \in [0, 1]$, $\sigma_i^2(t, x_1, x_2) > 0$.

Definition 4.55. A function $\mu(\cdot)$ is said to be a quasi-stationary density for the periodic diffusion corresponding to the Kolmogorov-Fokker-Planck operator $\mathcal{L}_2^*$ if it is periodic in x_1 and x_2 with period 1,

$$0 \le \mu(t, x_1, x_2) \text{ for each } (t, x_1, x_2) \in [0, T] \times [0, 1] \times [0, 1],$$

and for each fixed t and x_1,

$$\int_0^1 \mu(t, x_1, x_2) dx_2 = 1 \ \text{ and } \ \mathcal{L}_2^* \mu(t, x_1, x_2) = 0.$$

To proceed, let $\mathcal{H}$ be the space of functions that are bounded and continuous and are Hölder continuous in $(x_1, x_2) \in [0, 1] \times [0, 1]$ (with Hölder exponent Δ for some $0 < \Delta < 1$), uniformly with respect to t. For each $h_1, h_2 \in \mathcal{H}$ define $\langle h_1, h_2 \rangle_{\mathcal{H}}$ as

$$\langle h_1, h_2 \rangle_{\mathcal{H}} = \int_0^T \int_0^1 \int_0^1 h_1(t, x_1, x_2) h_2(t, x_1, x_2) dx_1 dx_2 dt.$$

Under the assumptions mentioned above, two sequences of functions $\varphi_i(\cdot)$ (periodic in x_1 and x_2) and $\psi_i(\cdot)$ for $i = 0, \ldots, n$ can be found such that

(a) $\varphi_i(\cdot, \cdot, \cdot) \in C^{n+1-i,2(n+1-i),2(n+1-i)}$;

(b) $\psi_i(t/\varepsilon, x_1, x_2)$ decay exponentially fast in that for some $c_1 > 0$ and $c_2 > 0$,

$$\sup_{x_1, x_2 \in [0,1]} \left| \psi_i \left(\frac{t}{\varepsilon}, x_1, x_2 \right) \right| \le c_1 \exp \left(-\frac{c_2 t}{\varepsilon} \right);$$

(c) define $\widetilde{s}_n^{\varepsilon}$ by

$$\widetilde{s}_n^{\varepsilon}(t, x_1, x_2) = \sum_{i=0}^{n} \left(\varepsilon^i \varphi_i(t, x_1, x_2) + \varepsilon^i \psi_i \left(\frac{t}{\varepsilon}, x_1, x_2 \right) \right);$$

for each $h \in \mathcal{H}$, the following error bound holds:

$$\left| \langle p^{\varepsilon} - \widetilde{s}_n^{\varepsilon}, h \rangle_{\mathcal{H}} \right| = O(\varepsilon^{n+1}). \tag{4.103}$$

It is interesting to note that the leading term of the approximation $\varphi_0(\cdot)$ is approximately the probability density of X_1, namely, $v_0(t, x_1)$ multiplied by the conditional density of X_2 given $X_1 = x_1$ (i.e., holding x_1 as a parameter), the quasi-stationary density $\mu(t, x_1, x_2)$. The rest of the terms in the regular part of the expansion assume the form

$$\mu(t, x_1, x_2) v_i(t, x_1) + U_i(t, x_1, x_2),$$

where $U_i(\cdot)$ is a particular solution of an inhomogeneous equation. Note the resemblance of the form to that of the Markov-chain cases studied in this chapter. A detailed proof of the assertion is in Khasminskii and Yin [116]. In fact, more complex systems (allowing interaction of X_1^{ε} and X_2^{ε}, the mixed partial derivatives of x_1 and x_2 as well as extension to multidimensional systems) are treated in [116]. In addition, in lieu of $\langle \cdot, \cdot \rangle_{\mathcal{H}}$, convergence under the uniform topology can be considered via the use of stochastic representation of solutions of partial differential equations or energy integration methods (see, for example, the related treatment of singularly perturbed switching diffusion systems in Il'in, Khasminskii, and Yin [94]).

4.8   Notes

Two-time-scale Markov chains are dealt with in this chapter using purely analytic methods, which are closely connected with the singular perturbation methods. The literature of singular perturbation for ordinary differential equations is rather rich. For an extensive list of references in singular perturbation methods for ordinary differential equations and various techniques such as initial-layer etc., we refer to Vasi'leva and Butuzov [209], Wasow [215, 216], O'Malley [163], and the references therein. The development of singular perturbation methods has been intertwined with advances in technology and progress in various applications. It can be traced back to the beginning of the twentieth century when Prandtl dealt with fluid motion with small friction (see Prandtl [178]). Nowadays, the averaging principle developed by Krylov, Bogoliubov, and Mitropolskii (see Bogoliubov and

Mitropolskii [18]) has become a popular technique, taught in standard graduate applied mathematics courses and employed widely. General results on singular perturbations can be found in Bensoussan, Lion, and Papanicolaou [7], Bogoliubov and Mitropolskii [18], Eckhaus [54], Erdélyi [58], Il'in [92], Kevorkian and Cole [108, 109], Krylov and Bogoliubov [133], O'Malley [163], Smith [199], Vasil'eava and Butuzov [209, 210], Wasow [215, 216]; applications to control theory and related fields are in Bensoussan [8], Bielecki and Filar [11], Delebecque and Quadrat [44], Delebecque, Quadrat, and Kokotovic [45], Kokotovic [126], Kokotovic, Bensoussan, and Blankenship [127], Kokotovic and Khalil [128], Kokotovic, Khalil, and O'Reilly [129], Kushner [140], Pan and Başar [164, 165, 166], Pervozvanskii and Gaitsgori [174], Phillips and Kokotovic [175], Yin and Zhang [233], among others; the vast literature on applications to different branches of physics are in Risken [182], van Kampen [208]; the survey by Hänggi, Talkner, and Borkovec [80] contains hundreds of references concerning applications in physics; related problems via large deviations theory are in Lerman and Schuss [151]; some recent work of singular perturbations to queueing networks, and heavy traffic, etc., is in Harrison and Reiman [81], Knessel and Morrison [125], and the references therein; applications to manufacturing systems are in Sethi and Zhang [192], Soner [202], Zhang [248], and the references cited there; related problems for stochastic differential equations and diffusion approximations, etc., can be found in Day [42], Friedlin and Wentzell [67], Il'in and Khasminskii [93], Khaminskii [111, 112], Kushner [139], Ludwig [152], Matkowsky and Schuss [158], Naeh, Klosek, Matkowski, and Schuss [160], Papanicolaou [169, 170], Schuss [187, 188], Skorohod [198], Yin [222], Yin and Ramachandran [227], and Zhang [247], among others. Singularly perturbed Markov processes also appear in the context of random evolution, a generalization of the motion of a particle on a fixed line with a random velocity or a random diffusivity; see, for example, Griego and Hersh [76, 77] and Pinsky [177]; an extensive survey can be found in Hersh [85]. A first-order approximation of the distribution of the Cox process with rapid switching is in Di Masi and Kabanov [48]. Recently, modeling communication systems via two-time-scale Markov chains has gained renewed interest; see Tse, Gallager, and Tsitsiklis [206], and the references therein.

It should be pointed out that there is a distinct feature in the problem we are studying compared with the traditional study of singularly perturbed systems. In contrast to many singularly perturbed ordinary differential equations, the matrix $Q(t)$ in (4.3) is singular, and has an eigenvalue 0. Thus the usual stability condition does not hold. To circumvent this difficulty, we utilize the q-Property of the matrix $Q(t)$, which leads to a probabilistic interpretation. The main emphasis in this chapter is on developing approximations to the solutions of the forward equations. The underlying systems arise from a wide range of applications where a finite-state Markov chain is involved and a fast time scale t/ε is used. Asymptotic series of the probability distribution of the Markov chain have been developed by em-

ploying the techniques of matched expansions. An attempt to obtain the asymptotic expansion of (4.3) is initiated in Khasminskii, Yin, and Zhang [119] for time-inhomogeneous Markov chains. The result presented here is a refinement of the aforementioned reference.

Extending the results for irreducible generators, this chapter further discusses two-time-scale Markov chains with weak and strong interactions. The formulations substantially generalize the work of Khasminskii, Yin, and Zhang [120]. Section 4.3 discusses Markovian models with recurrent states belonging to several ergodic classes is a refinement of [120].

Previous work on singularly perturbed Markov chains with weak and strong interactions can be found in Delebecque, Quadrat, and Kokotovic [45], Gaitsgori and Pervozvanskii [69], Pervozvanskii and Gaitsgori [174], and Phillips and Kokotovic [175]. The essence is a decomposition and aggregation point of view. Their models are similar to that considered in this chapter. For example, translating the setup into our setting, the authors of [175] assumed that the Markov chain generated by $\widetilde{Q}/\varepsilon + \widehat{Q}$ has a single ergodic class for ε sufficiently small. Moreover, for each $j = 1, 2, \ldots, l$, the subchain has a single ergodic class. Their formulation requires that $\widetilde{Q}(t) = \widetilde{Q}$ and $\widehat{Q}(t) = \widehat{Q}$, and it requires essentially the irreducibility of $\widetilde{Q}/\varepsilon + \widehat{Q}$ for all $\varepsilon \leq \varepsilon_0$ for some $\varepsilon_0 > 0$ small enough in addition to the irreducibility of $\widetilde{Q}^j$ for $j = 1, 2, \ldots, l$. The problem considered in this chapter is nonstationary; the generators are time-varying. The irreducibility is in the weak sense, and only weak irreducibility of each subgenerator (or block matrix) $\widetilde{Q}^j(t)$ for $j = 1, 2, \ldots, l$ is needed. Thus our results generalize the existing theorems to nonstationary cases under weaker assumptions. The condition on $\widetilde{Q}(t)$ exploits the intrinsic properties of the underlying chains. Furthermore, our results also include Markov chains with countable-state spaces. The formulation and development of Section 4.5 are inspired by that of [175] (see also Pan and Başar [164]). This together with the consideration of chains with recurrent states and the inclusion of absorbing states includes most of practical concerns for the rapidly varying part of the generator. Although the forms of the generators with absorbing states and with transient states have more complex structures, the asymptotic expansion of the probability distributions can still be obtained via a similar approach to that of the case of block-diagonal $\widetilde{Q}(\cdot)$. Applications to manufacturing systems are discussed, for example, in Jiang and Sethi [99] and Sethi and Zhang [192] among others. As a complement of the development in this chapter, the work of Il'in, Khasminskii, and Yin [94] deals with the cases that the underlying Markov processes involve both diffusion and pure jump processes; see also Yin and Yang [229]. Previous work of singular perturbation of stochastic systems can be found in Day [42], Friedlin and Wentzel [67], Khasminskii [111, 112, 113], Kushner [139], Ludwig [152], Matkowsky and Schuss [158], Naeh, Klosek, Matkowski, and Schuss [160], Papanicolaou [169, 170], Schuss [187], Yin and Ramachandran [227], and

the references therein. Singular perturbation in connection with optimal control problems are contained in Bensoussan [8], Bielecki and Filar [11], Delebecque and Quadrat [44], Kokotovic [126], Kokotovic, Bensoussan, and Blankenship [127], Kushner [140], Lehoczky, Sethi, Soner, and Taksar [150], Martins and Kushner [156], Pan and Başar [164], Pervozvanskii and Gaitsgori [174], Sethi and Zhang [192], Soner [202], and Yin and Zhang [233] among others. For discrete-time two-time-scale Markov chains, we refer the reader to Yin and Zhang [238] Yin, Zhang, and Badowski [242] among others.

We note that one of the key points that enables us to solve these problems is the Fredholm alternative. This is even more crucial compared with the situation in Section 4.2 for irreducible generators. In Section 4.2, the consistency conditions are readily verified, whereas in the formulation under weak and strong interactions, the verification needs more work and we have to utilize the consistency to obtain the desired solution.

The discussions on Markov chains with countable-state spaces in this chapter focused on simple situations. For more general cases, see Yin and Zhang [230, 231], in which applications to quasi-birth-death queues were considered; see also Altman, Avrachenkov, and Nunez-Queija [4] for a different approach. The discussions on singularly perturbed diffusion processes dealt with mainly forward equations. For related work on singularly perturbed diffusions, see the papers of Khasminskii and Yin [115, 116] and the references therein; one of the motivations for studying singularly perturbed diffusion comes from wear process modeling (see Rishel [181]). For treatments of averaging principles and related backward equations, we refer the reader to Khasminskii and Yin [117, 118]. For a number of applications on queueing systems, financial engineering, and insurance risk, we refer the reader to Yin, Zhang, and Zhang [232] and references therein.