

# 5

## Occupation Measures: Asymptotic Properties and Ramification

### 5.1 Introduction

Chapter 4 deals with the probability distribution of  $\alpha^\varepsilon(\cdot)$  through the corresponding forward equation and is mainly an analytical approach, whereas the current chapter is largely probabilistic in nature. The central theme of this chapter is limit results of unscaled as well as scaled sequences of occupation measures, which include the law of large numbers for an unscaled sequence, exponential upper bounds, and asymptotic distribution of a suitably scaled sequence of occupation times. It further exploits the deviation of the functional occupation times from its quasi-stationary distribution. We obtain estimates of centered deviations, prove the convergence of a properly scaled and centered sequence of occupation times, characterize the limit process by deriving the limit distribution and providing explicit formulas for the mean and covariance functions, and provide exponential bounds for the normalized process. It is worthwhile to note that the limit covariance function depends on the initial-layer terms in contrast with most of the existing results of central limit type.

The rest of the chapter is arranged as follows. We first study the asymptotic properties of irreducible Markov chains in Section 5.2. In view of the developments in Remarks 4.34 and 4.39, the Markov chain with recurrent states is the most illustrative and representative one. As a result, in the remaining chapters, we mainly treat problems associated with this model. Starting in Section 5.3.1, we consider Markov chains with weak and strong

interactions with generators consisting of multiple irreducible blocks. After treating aggregation of the Markov states, we study the corresponding exponential bounds. We deal with asymptotic distributions. Then in Section 5.4, we treat Markov chains with generators that are merely measurable. Next, remarks on inclusion of transient and absorbing states are provided in Section 5.5. Applications of the weak convergence results and a related stability problem are provided in Section 5.6. Finally, Section 5.7 concludes the chapter with notes and further remarks.

## 5.2 The Irreducible Case

The notion of occupation measure is set forth first. We consider a sequence of unscaled occupation measures and establish its convergence in probability to that of the accumulative quasi-stationary distribution. This is followed by exponential bounds of the function occupation time and moment estimates. In addition, asymptotic normality is derived. Although the prelimit process has nonzero mean and is nonstationary, using the results of Section 4.2, the quasi-stationary regime is established after a short period (of order  $O(\varepsilon)$ ). We also calculate explicitly the covariance representation of the limit process, and prove that the process  $\alpha^\varepsilon(\cdot)$  satisfies a mixing condition. The tightness of the sequence and the w.p.1 continuity of the sample paths of the limit process are proved by estimating the fourth moment. The limit of the finite-dimensional distributions is then calculated and shown to be Gaussian. By proving a series of lemmas, we derive the desired asymptotic normality.

As was mentioned in previous chapters, the process  $\alpha^\varepsilon(\cdot)$  arises from pervasive practical use that involves a rapidly fluctuating finite-state Markov chain. In these applications, the asymptotic behavior of the Markov chain  $\alpha^\varepsilon(\cdot)$  has a major influence. Further investigation and understanding of the asymptotic properties of  $\alpha^\varepsilon(\cdot)$ , in particular, the probabilistic structures, play an important role in the in-depth study.

In Section 4.2, using singular perturbation techniques, we examined the asymptotic properties of  $p_i^\varepsilon(t) = P(\alpha^\varepsilon(t) = i)$ . It has been proved that  $p^\varepsilon(t) = (p_1^\varepsilon(t), \dots, p_m^\varepsilon(t))$  converges to the quasi-stationary distribution  $\nu(t)$  as  $\varepsilon \rightarrow 0$  for each  $t > 0$  and  $p^\varepsilon(t)$  admits an asymptotic expansion in terms of  $\varepsilon$ . To gain further insight, we ask whether there is a limit result for the occupation measure  $\int_0^t I_{\{\alpha^\varepsilon(s)=i\}} ds$ . If a convergence is expected to take place, then what is the rate of convergence? Does one have a central limit theorem associated with the  $\alpha^\varepsilon(\cdot)$ -process? The answers to these questions are affirmative. We will prove a number of limit results related to an unscaled sequence, and a suitably scaled and normalized sequence. Owing to the asymptotic expansions, the scaling factor is  $\sqrt{\varepsilon}$ . The limit process is Gaussian with zero mean, and the covariance of the limit process depends on the asymptotic expansion in an essential way, which reflects

one of the distinct features of the central limit theorem. It appears that it is virtually impossible to calculate the asymptotic covariance of the Gaussian process without the help of the asymptotic expansion, which reveals a salient characteristic of the two-time-scale Markov chain.

A related problem is to examine the exponential bounds on the scaled occupation measure process. This is similar to the estimation of the moment generating function. Such estimates have been found useful in studying hierarchical controls of manufacturing systems. Using the asymptotic expansion and the martingale representation of finite-state Markov chains, we are able to establish such exponential bounds for the scaled occupation measures.

### 5.2.1 Occupation Measure

Let  $(\Omega, \mathcal{F}, P)$  denote the underlying probability space. As in Section 4.2,  $\alpha^\varepsilon(\cdot)$  is a nonstationary Markov chain on  $(\Omega, \mathcal{F}, P)$  with finite-state space  $\mathcal{M} = \{1, \dots, m\}$  and generator  $Q^\varepsilon(t) = Q(t)/\varepsilon$ .

For each  $i \in \mathcal{M}$ , let  $\beta_i(\cdot)$  denote a bounded Borel measurable deterministic function and define a sequence of centered (around the quasi-stationary distribution) occupation measures  $Z_i^\varepsilon(t)$  as

$$Z_i^\varepsilon(t) = \int_0^t (I_{\{\alpha^\varepsilon(s)=i\}} - \nu_i(s)) \beta_i(s) ds. \quad (5.1)$$

Set  $Z^\varepsilon(t) = (Z_1^\varepsilon(t), \dots, Z_m^\varepsilon(t))$ . It is a measure of the functional occupancy for the process  $\alpha^\varepsilon(\cdot)$ . Our interest lies in the asymptotic properties of the sequence defined in (5.1). To proceed, we first present some conditions and preliminary results needed in the sequel.

Note that a special choice of  $\beta_i(\cdot)$  is  $\beta_i(t) = 1$ , for  $t \in [0, T]$ . To insert  $\beta_i(\cdot)$  in sequence allows one to treat various situations in some applications. For example, in the manufacturing problem,  $\beta_i(t)$  is often given by a function of a control process; see Chapter 8 for further details.

### 5.2.2 Conditions and Preliminary Results

To proceed, we make the following assumptions.

- (A5.1) For each  $t \in [0, T]$ ,  $Q(t)$  is weakly irreducible.
- (A5.2)  $Q(\cdot)$  is continuously differentiable on  $[0, T]$ , and its derivative is Lipschitz.

Recall that  $p^\varepsilon(t) = (P(\alpha^\varepsilon(t) = 1), \dots, P(\alpha^\varepsilon(t) = m))$  and let

$$p_{ij}^\varepsilon(t, t_0) = P(\alpha^\varepsilon(t) = j | \alpha^\varepsilon(t_0) = i) \quad \text{for all } i, j \in \mathcal{M}.$$

Use  $P^\varepsilon(t, t_0)$  to denote the transition matrix  $(p_{ij}^\varepsilon(t, t_0))$ . The following lemma is on the asymptotic expansion of  $P^\varepsilon(t, t_0)$ .

**Lemma 5.1.** *Assume (A5.1) and (A5.2). Then there exists a positive constant  $\kappa_0$  such that for each fixed  $0 \leq T < \infty$ ,*

$$P^\varepsilon(t, t_0) = P_0(t) + O\left(\varepsilon + \exp\left(-\frac{\kappa_0(t - t_0)}{\varepsilon}\right)\right) \tag{5.2}$$

*uniformly in  $(t_0, t)$  where  $0 \leq t_0 \leq t \leq T$  and*

$$P_0(t) = \begin{pmatrix} \nu(t) \\ \vdots \\ \nu(t) \end{pmatrix}.$$

*In addition, assume  $Q(\cdot)$  to be twice continuously differentiable on  $[0, T]$  with the second derivative being Lipschitz. Then*

$$\begin{aligned} P^\varepsilon(t, t_0) &= P_0(t) + \varepsilon P_1(t) \\ &+ Q_0\left(\frac{t - t_0}{\varepsilon}, t_0\right) + \varepsilon Q_1\left(\frac{t - t_0}{\varepsilon}, t_0\right) + O(\varepsilon^2) \end{aligned} \tag{5.3}$$

*uniformly in  $(t_0, t)$ , where  $0 \leq t_0 \leq t \leq T$ ,*

$$P_1(t) = \begin{pmatrix} \varphi_1(t) \\ \vdots \\ \varphi_1(t) \end{pmatrix},$$

$$\frac{dQ_0(\tau, t_0)}{d\tau} = Q_0(\tau, t_0)Q(t_0), \quad \tau \geq 0,$$

$$Q_0(0, t_0) = I - P_0(t_0),$$

*and*

$$\frac{dQ_1(\tau, t_0)}{d\tau} = Q_1(\tau, t_0)Q(t_0) + \tau Q_0(\tau, t_0) \frac{dQ(t_0)}{dt}, \quad \tau \geq 0$$

$$Q_1(0, t_0) = -P_1(t_0),$$

*where  $\varphi_1(t)$  is given in (4.13) (with  $\tau := (t - t_0)/\varepsilon$ ). Furthermore, for  $i = 0, 1$ , the  $P_i(\cdot)$  are  $(2 - i)$  times continuously differentiable on  $[0, T]$  and there exist constants  $K > 0$  and  $\kappa_0 > 0$  such that*

$$|Q_i(\tau, t_0)| \leq K \exp(-\kappa_0\tau), \tag{5.4}$$

*uniformly for  $t_0 \in [0, T]$ .*

**Remark 5.2.** Recall that  $\nu(t)$  and  $\varphi_1(t)$  are row vectors. As a result,  $P_0(\cdot)$  and  $P_1(\cdot)$  have identical rows. This is a consequence of the convergence of  $p^\varepsilon(t)$  to the quasi-stationary distribution and the asymptotic expansions.

Proof of Lemma 5.1: It suffices to verify (5.3) because (5.2) can be derived similarly. The asymptotic expansion of  $P^\varepsilon(t, t_0)$  can be obtained as in Section 4.2. Thus only the exponential bound (5.4) needs to be proved. The main task is to verify the uniformity in  $t_0$ . To this end, it suffices to treat each row of  $Q_i(\tau, t_0)$  separately. For a fixed  $i = 0, 1$ , let

$$\eta(\tau, t_0) = (\eta_1(\tau, t_0), \dots, \eta_m(\tau, t_0))$$

denote any row of  $Q_i(\tau, t_0)$  and  $\eta_0(t_0)$  the corresponding row in  $Q_i(0, t_0)$  with

$$Q_0(0, t_0) = I - P_0(t_0) \quad \text{and}$$

$$Q_1(0, t_0) = -P_1(t_0).$$

Then  $\eta(\tau, t_0)$  satisfies the differential equation

$$\frac{d\eta(\tau, t_0)}{d\tau} = \eta(\tau, t_0)Q(t_0), \quad \tau \geq 0,$$

$$\eta(0, t_0) = \eta_0(t_0).$$

By virtue of the assumptions of Lemma 5.1 and the asymptotic expansion, it follows that  $\eta_0(t_0)$  is uniformly bounded and  $\eta_0(t_0)\mathbf{1} = 0$ .

Define

$$\widehat{\kappa} = -\max \left\{ \text{The real parts of eigenvalues of } Q(t), \quad t \in [0, T] \right\}.$$

Then Lemma A.6 implies that  $\widehat{\kappa} > 0$ . In view of Theorem 4.5, it suffices to show that for all  $\tau \geq 0$  and for some constant  $K > 0$  independent of  $t_0$ ,

$$|\eta(\tau, t_0)| \leq K \exp \left( -\frac{\widehat{\kappa}\tau}{2} \right). \tag{5.5}$$

To verify (5.5), note that for any  $\varsigma_0 \in [0, T]$ ,

$$\frac{d\eta(\tau, t_0)}{d\tau} = \eta(\tau, t_0)Q(\varsigma_0) + \eta(\tau, t_0)[Q(t_0) - Q(\varsigma_0)].$$

Solving this differential equation by treating  $\eta(\tau, t_0)[Q(t_0) - Q(\varsigma_0)]$  as the driving term, we have

$$\begin{aligned} \eta(\tau, t_0) &= \eta_0(t_0) \exp(Q(\varsigma_0)\tau) \\ &\quad + \int_0^\tau \eta(\varsigma, t_0)[Q(t_0) - Q(\varsigma_0)] \exp(Q(\varsigma_0)(\tau - \varsigma)) \, d\varsigma. \end{aligned} \tag{5.6}$$

In view of (A5.2), for some  $K_0 > 0$ ,

$$\left| Q(t_0) - Q(s_0) \right| \leq K_0 |t_0 - s_0|.$$

Noting that  $\eta_0(t_0)\mathbb{1} = 0$  and that  $P_0(t)$  has identical rows, we have

$$\eta_0(t_0)P_0(t) = 0, \quad \text{for } t \geq 0.$$

Thus the equation in (5.6) is equivalent to

$$\begin{aligned} \eta(\tau, t_0) &= \eta_0(t_0)(\exp(Q(s_0)s) - P_0(s_0)) \\ &\quad + \int_0^\tau \eta(s, t_0)[Q(t_0) - Q(s_0)](\exp(Q(s_0)(\tau - s)) - P_0(s_0))ds. \end{aligned}$$

From Lemma A.2, we have

$$|\eta(\tau, t_0)| \leq K_1 \exp(-\widehat{\kappa}\tau) + K_2 |t_0 - s_0| \int_0^\tau |\eta(s, t_0)| \exp(-\widehat{\kappa}(\tau - s)) ds,$$

for some constants  $K_1$  and  $K_2$  which may depend on  $s_0$  but are independent of  $t_0$ . By Gronwall's inequality (see Hale [79, p. 36]),

$$|\eta(\tau, t_0)| \leq K_1 \exp(-(\widehat{\kappa} - K_2 |t_0 - s_0|)\tau), \tag{5.7}$$

for all  $t_0 \in [0, T]$  and  $\tau > 0$ .

If (5.5) does not hold uniformly, then there exist  $\tau_n > 0$  and  $t_n \in [0, T]$  such that

$$|\eta(\tau_n, t_n)| \geq n \exp\left(-\frac{\widehat{\kappa}\tau_n}{2}\right).$$

Since  $T$  is finite, we may assume  $t_n \rightarrow s_0$ , as  $n \rightarrow \infty$ . This contradicts (5.7) for  $n$  large enough satisfying  $|t_n - s_0| < \widehat{\kappa}/(2K_2)$  and  $K_1 < n$ . Thus the proof is complete.  $\square$

### Unscaled Occupation Measure

To study the unscaled occupation measure  $Z_i^\varepsilon(t)$  in (5.1), we define a related sequence  $\{\widehat{Z}^\varepsilon(t)\}$  of  $\mathbb{R}^m$ -valued processes with its  $i$ th component  $\widehat{Z}_i^\varepsilon(t)$  given by

$$\widehat{Z}_i^\varepsilon(t) = \int_0^t (I_{\{\alpha^\varepsilon(s)=i\}} - P(\alpha^\varepsilon(s) = i)) \beta_i(s) ds.$$

Assume the conditions (A5.1) and (A5.2). We claim that for any  $\delta > 0$ ,

$$\lim_{\varepsilon \rightarrow 0} \left( \sup_{0 \leq t \leq T} P(|\widehat{Z}^\varepsilon(t)| \geq \delta) \right) = 0 \text{ and} \tag{5.8}$$

$$\lim_{\varepsilon \rightarrow 0} \left( \sup_{0 \leq t \leq T} E|\widehat{Z}^\varepsilon(t)|^2 \right) = 0. \tag{5.9}$$

Note that (5.8) follows from (5.9) using Tchebyshev's inequality. The verification of (5.9), which mainly depends on a mixing property of the underlying sequence, is almost the same as the moment estimates in the proof of asymptotic normality in Lemma 5.13. The details of the verifications of (5.8) and (5.9) are omitted here.

With (5.9) in hand for any  $\delta > 0$ , to study the asymptotic properties of  $Z^\varepsilon(\cdot)$ , it remains to show that

$$\lim_{\varepsilon \rightarrow 0} \left( \sup_{0 \leq t \leq T} P(|Z^\varepsilon(t)| \geq \delta) \right) = 0 \text{ and}$$

$$\lim_{\varepsilon \rightarrow 0} \left( \sup_{0 \leq t \leq T} E|Z^\varepsilon(t)|^2 \right) = 0.$$

In fact, it is enough to work with each component of  $Z^\varepsilon(t)$ . Note that both  $Z^\varepsilon(t)$  and  $\widehat{Z}^\varepsilon(t)$  are bounded. This together with the boundedness of  $\beta(t)$  and Lemma 5.1 implies that for each  $i \in \mathcal{M}$ ,

$$\begin{aligned} & \sup_{0 \leq t \leq T} E|Z_i^\varepsilon(t)|^2 \\ & \leq 2 \left( \sup_{0 \leq t \leq T} E|\widehat{Z}_i^\varepsilon(t)|^2 + \sup_{0 \leq t \leq T} E \left| \int_0^t (P(\alpha^\varepsilon(s) = i) - \nu_i(s)) \beta_i(s) ds \right|^2 \right) \\ & \leq 2 \left( \sup_{0 \leq t \leq T} E|\widehat{Z}_i^\varepsilon(t)|^2 + \int_0^T O(\varepsilon) ds \right) \rightarrow 0, \end{aligned}$$

as  $\varepsilon \rightarrow 0$ , which yields the desired results.

The limit result above is of the law-of-large-numbers type. What has been obtained is that as  $\varepsilon \rightarrow 0$ ,

$$\int_0^t I_{\{\alpha^\varepsilon(s)=i\}} ds \rightarrow \int_0^t \nu_i(s) ds \quad \text{in probability as } \varepsilon \rightarrow 0,$$

for  $0 < t \leq T$ . In fact, a somewhat stronger result on uniform convergence in terms of the second moment is established. To illustrate, suppose that  $\alpha^\varepsilon(t) = \alpha(t/\varepsilon)$  such that  $\alpha(\cdot)$  is a stationary process with stationary distribution  $\bar{\nu} = (\bar{\nu}_1, \dots, \bar{\nu}_m)$ . Then via a change of variable  $\varsigma = s/\varepsilon$ , we have

$$\begin{aligned} \frac{1}{t} \int_0^t I_{\{\alpha^\varepsilon(s)=i\}} ds &= \frac{\varepsilon}{t} \int_0^{t/\varepsilon} I_{\{\alpha(\varsigma)=i\}} d\varsigma \\ &= \frac{\varepsilon}{t} \int_0^{t/\varepsilon} I_{\{\alpha(\varsigma)=i\}} d\varsigma \rightarrow \bar{\nu}_i \quad \text{in probability as } \varepsilon \rightarrow 0, \end{aligned}$$

for  $0 < t \leq T$ . This is exactly the continuous-time version of the law of large numbers.

**Example 5.3.** Let us return to the singularly perturbed Cox process of Section 3.3. Recall that the compensator of the singularly perturbed Cox process is given by

$$G^\varepsilon(t) = G_0 + \sum_{i=1}^m \int_0^t a_i I_{\{\alpha^\varepsilon(s)=i\}} ds,$$

where  $a_i > 0$  for  $i = 1, \dots, m$ . Assume that all the conditions in Lemma 5.1 hold. Then Theorem 4.5 implies that  $P(\alpha^\varepsilon(t) = i) \rightarrow \nu_i(t)$  as  $\varepsilon \rightarrow 0$ . What we have discussed thus far implies that for each  $i \in \mathcal{M}$ ,

$$\int_0^t a_i I_{\{\alpha^\varepsilon(s)=i\}} ds \rightarrow \int_0^t a_i \nu_i(s) ds \quad \text{in probability as } \varepsilon \rightarrow 0 \text{ and}$$

$$G^\varepsilon(t) \rightarrow G(t) = G_0 + \sum_{i=1}^m \int_0^t a_i \nu_i(s) ds \quad \text{in probability.}$$

Moreover,

$$\lim_{\varepsilon \rightarrow 0} \left( \sup_{0 \leq t \leq T} E|G^\varepsilon(t) - G(t)|^2 \right) = 0.$$

In the rest of this chapter, we treat suitably scaled occupation measures; the corresponding results for the Cox process can also be derived.

With the limit results in hand, the next question is this: How fast does the convergence take place? The rate of convergence issue together with more detailed asymptotic properties is examined fully in the following sections.

### 5.2.3 Exponential Bounds

This section is devoted to the derivation of exponential bounds for the normalized occupation measure (or occupation time)  $n^\varepsilon(\cdot)$ . Given a deterministic process  $\beta(\cdot)$ , we consider the “centered” and “scaled” functional occupation-time process  $n^\varepsilon(t, i)$  defined by

$$n^\varepsilon(t, i) = \frac{1}{\sqrt{\varepsilon}} \int_0^t (I_{\{\alpha^\varepsilon(s)=i\}} - \nu_i(s)) \beta_i(s) ds \quad \text{and} \tag{5.10}$$

$$n^\varepsilon(t) = (n^\varepsilon(t, 1), \dots, n^\varepsilon(t, m)) \in \mathbb{R}^{1 \times m}.$$

In view of Lemma 5.1, we have, for  $0 \leq s \leq t \leq T$ ,

$$P^\varepsilon(t, s) - P_0(t) = O \left( \varepsilon + \exp \left( -\frac{\kappa_0(t-s)}{\varepsilon} \right) \right),$$

for some  $\kappa_0 > 0$ . Note that the big  $O(\cdot)$  usually depends on  $T$ . Let  $K_T$  denote an upper bound of

$$\frac{P^\varepsilon(t, s) - P_0(t)}{\varepsilon + \exp(-\kappa_0(t-s)/\varepsilon)}$$



for  $0 \leq s \leq t \leq T$ . For convenience, we use the notation  $O_1(y)$  to denote a function of  $y$  such that  $|O_1(y)|/|y| \leq 1$ . The rationale is that  $K_T$  represents the magnitude of the bounding constant and the rest of the bound is in terms of a function with norm bounded by 1. Using this notation and  $K_T$ , we write

$$P^\varepsilon(t, s) - P_0(t) = K_T O_1 \left( \varepsilon + \exp \left( -\frac{\kappa_0(t-s)}{\varepsilon} \right) \right). \tag{5.11}$$

Let  $y(t) = (y_{ij}(t))$  and  $z(t) = (z_i(t))$  denote a matrix-valued function and a vector-valued function defined on  $[0, T]$ , respectively. Their norms are defined by

$$\begin{aligned} |y|_T &= \max_{i,j} \sup_{0 \leq t \leq T} |y_{ij}(t)|, \\ |z|_T &= \max_i \sup_{0 \leq t \leq T} |z_i(t)|. \end{aligned} \tag{5.12}$$

For future use, define  $\beta(t) = \text{diag}(\beta_1(t), \dots, \beta_m(t))$ . The following theorem is concerned with the exponential bound of  $n^\varepsilon(t)$  for  $\varepsilon$  sufficiently small.

**Theorem 5.4.** *Assume that (A5.1) and (A5.2) are satisfied. Then there exist  $\varepsilon_0 > 0$  and  $K > 0$  such that for all  $0 < \varepsilon \leq \varepsilon_0$ ,  $T \geq 0$ , and any bounded and measurable deterministic function  $\beta(\cdot) = \text{diag}(\beta_1(\cdot), \dots, \beta_m(\cdot))$ , the following exponential bound holds:*

$$E \exp \left\{ \frac{\theta_T}{(T+1)^{\frac{3}{2}}} \sup_{0 \leq t \leq T} |n^\varepsilon(t)| \right\} \leq K, \tag{5.13}$$

where  $\theta_T$  is a constant satisfying

$$0 \leq \theta_T \leq \frac{\min\{1, \kappa_0\}}{K_T |\beta|_T} \tag{5.14}$$

with  $\kappa_0$  being the exponential constant in Lemma 5.1.

**Remark 5.5.** Note that the constants  $\varepsilon_0$  and  $K$  are independent of  $T$ . This is a convenient feature of the estimate in certain applications. The result is in terms of a fixed but otherwise arbitrary  $T$ , which is particularly useful for systems in an infinite horizon.

Proof: The proof is divided into several steps.

*Step 1.* In the first step, we show that (5.13) holds when the ‘sup’ is absent. Let  $\chi^\varepsilon(\cdot)$  denote the indicator vector of  $\alpha^\varepsilon(\cdot)$ , that is,

$$\chi^\varepsilon(t) = (I_{\{\alpha^\varepsilon(t)=1\}}, \dots, I_{\{\alpha^\varepsilon(t)=m\}}) \quad \text{and}$$

$$w^\varepsilon(t) = \chi^\varepsilon(t) - \chi^\varepsilon(0) - \frac{1}{\varepsilon} \int_0^t \chi^\varepsilon(s) Q(s) ds.$$

It is well known (see Elliott [56]) that  $w^\varepsilon(t) = (w_1^\varepsilon(t), \dots, w_m^\varepsilon(t))$ , for  $t \geq 0$ , is a  $\sigma\{\alpha^\varepsilon(s) : s \leq t\}$ -martingale. In view of a result of Kunita and Watanabe [134] (see also Ikeda and Watanabe [91, p. 55]), one can define a stochastic integral with respect to  $w^\varepsilon(t)$ . Moreover, the solution of the linear stochastic differential equation

$$d\chi^\varepsilon(t) = \chi^\varepsilon(t)Q^\varepsilon(t)dt + dw^\varepsilon(t)$$

is given by

$$\chi^\varepsilon(t) = \chi^\varepsilon(0)P^\varepsilon(t, 0) + \int_0^t (dw^\varepsilon(s))P^\varepsilon(t, s),$$

where  $P^\varepsilon(t, s)$  is the principal matrix solution to the equation

$$\frac{dP^\varepsilon(t, s)}{dt} = \frac{1}{\varepsilon}P^\varepsilon(t, s)Q(t), \quad \text{with } P^\varepsilon(s, s) = I$$

representing the transition probability matrix.

Note that for  $t \geq s \geq 0$ ,

$$\chi^\varepsilon(s)P_0(t) = (\chi^\varepsilon(s)\mathbb{1})\nu(t) = \nu(t).$$

Using this and (5.11), we have

$$\begin{aligned} & \chi^\varepsilon(t) - \nu(t) \\ &= \chi^\varepsilon(0)(P^\varepsilon(t, 0) - P_0(t)) + \int_0^t (dw^\varepsilon(s))[(P^\varepsilon(t, s) - P_0(t)) + P_0(t)] \\ &= K_T O_1 \left( \varepsilon + \exp\left(-\frac{\kappa_0 t}{\varepsilon}\right) \right) \\ & \quad + K_T \int_0^t (dw^\varepsilon(s)) O_1 \left( \varepsilon + \exp\left(-\frac{\kappa_0(t-s)}{\varepsilon}\right) \right) + w^\varepsilon(t)P_0(t) \\ &= K_T O_1 \left( \varepsilon + \exp\left(-\frac{\kappa_0 t}{\varepsilon}\right) \right) \\ & \quad + K_T \int_0^t (dw^\varepsilon(s)) O_1 \left( \varepsilon + \exp\left(-\frac{\kappa_0(t-s)}{\varepsilon}\right) \right). \end{aligned}$$

The last equality above follows from the observation that

$$Q(s)P_0(t) = 0 \text{ for all } t \geq s \geq 0,$$

and

$$\begin{aligned} w^\varepsilon(t)P_0(t) &= \left( \chi^\varepsilon(t) - \chi^\varepsilon(0) - \frac{1}{\varepsilon} \int_0^t \chi^\varepsilon(r)Q(r)dr \right) P_0(t) \\ &= \nu(t) - \nu(t) - \frac{1}{\varepsilon} \int_0^t \chi^\varepsilon(r)Q(r)P_0(t)dr = 0. \end{aligned}$$

Recall that  $\beta(t) = \text{diag}(\beta_1(t), \dots, \beta_m(t))$ . Then it follows that

$$\begin{aligned}
& \int_0^t (\chi^\varepsilon(s) - \nu(s))\beta(s)ds \\
&= K_T O_1(\varepsilon(t+1)) \\
&\quad + K_T \int_0^t \int_0^s (dw^\varepsilon(r)) O_1\left(\varepsilon + \exp\left(-\frac{\kappa_0(s-r)}{\varepsilon}\right)\right) \beta(s)ds \\
&= K_T O_1(\varepsilon(t+1)) \\
&\quad + K_T \int_0^t (dw^\varepsilon(r)) \left( \int_r^t O_1\left(\varepsilon + \exp\left(-\frac{\kappa_0(s-r)}{\varepsilon}\right)\right) \beta(s)ds \right) \\
&= K_T O_1(\varepsilon(t+1)) \\
&\quad + \varepsilon K_T \int_0^t (dw^\varepsilon(r)) O_1\left((t-r) + \frac{1}{\kappa_0}\left(1 - \exp\left(-\frac{\kappa_0(t-r)}{\varepsilon}\right)\right)\right) |\beta|_T \\
&= K_T O_1(\varepsilon(t+1)) + \varepsilon K_T |\beta|_T \left(T + \frac{1}{\kappa_0}\right) \int_0^t (dw^\varepsilon(r)) b(r, t),
\end{aligned}$$

where  $b(s, t)$  is a measurable function and  $|b(s, t)| \leq 1$  for all  $s$  and  $t$ . Dividing both sides by  $(T+1)$ , we obtain

$$\begin{aligned}
& \frac{1}{T+1} \left| \int_0^t (\chi^\varepsilon(s) - \nu(s))\beta(s)ds \right| \\
&= \varepsilon K_T O_1(1) + \varepsilon K_T |\beta|_T \left( \frac{T + (1/\kappa_0)}{T+1} \right) \left| \int_0^t (dw^\varepsilon(s)) b(s, t) \right|.
\end{aligned} \tag{5.15}$$

Therefore, we have

$$\begin{aligned}
& E \exp \left\{ \frac{\theta_T}{\sqrt{\varepsilon}(T+1)^{\frac{3}{2}}} \left| \int_0^t (\chi^\varepsilon(s) - \nu(s))\beta(s)ds \right| \right\} \\
&\leq E \exp \left\{ \frac{1}{\sqrt{\varepsilon}\sqrt{T+1}} \left[ \varepsilon O_1(1) + \varepsilon \left| \int_0^t (dw^\varepsilon(s)) b(s, t) \right| \right] \right\}.
\end{aligned}$$

In view of the choice of  $\theta_T$ , it follows that

$$\begin{aligned}
& E \exp \left\{ \frac{\theta_T}{\sqrt{\varepsilon}(T+1)^{\frac{3}{2}}} \left| \int_0^t (\chi^\varepsilon(s) - \nu(s))\beta(s)ds \right| \right\} \\
&\leq \exp \left( \frac{\sqrt{\varepsilon}}{\sqrt{T+1}} \right) E \exp \left\{ \frac{\sqrt{\varepsilon}}{\sqrt{T+1}} \left| \int_0^t (dw^\varepsilon(s)) b(s, t) \right| \right\} \\
&\leq e E \exp \left\{ \frac{\sqrt{\varepsilon}}{\sqrt{T+1}} \left| \int_0^t (dw^\varepsilon(s)) b(s, t) \right| \right\}.
\end{aligned} \tag{5.16}$$

Recall that

$$w^\varepsilon(t) = (w_1^\varepsilon(t), \dots, w_m^\varepsilon(t)).$$

It suffices to work out the estimate for each component  $w_i^\varepsilon(t)$ . That is, it is enough to show that for each  $i = 1, \dots, m$ ,

$$E \exp \left\{ \frac{\sqrt{\varepsilon}}{\sqrt{T+1}} \left| \int_0^t b(s, t) dw_i^\varepsilon(s) \right| \right\} \leq K, \tag{5.17}$$

for all measurable functions  $b(\cdot, \cdot)$  with  $|b(s, t)| \leq 1$  and  $0 \leq t \leq T$ . For each  $t_0 \geq 0$ , let  $b_0(s) = b(s, t_0)$ .

For any nonnegative random variable  $\xi$ ,

$$\begin{aligned} E e^\xi &= \sum_{j=0}^\infty \int_{\{j \leq \xi < j+1\}} e^\xi dP \\ &\leq \sum_{j=0}^\infty \int_{\{j \leq \xi < j+1\}} e^{j+1} dP \\ &= \sum_{j=0}^\infty e^{j+1} P(j \leq \xi < j+1) \\ &= \sum_{j=0}^\infty e^{j+1} [P(\xi \geq j) - P(\xi \geq j+1)] \\ &\leq e + (e-1) \sum_{j=1}^\infty e^j P(\xi \geq j). \end{aligned}$$

By virtue of the inequality above, we have

$$\begin{aligned} E \exp \left\{ \frac{\sqrt{\varepsilon}}{\sqrt{T+1}} \left| \int_0^t b_0(s) dw_i^\varepsilon(s) \right| \right\} \\ \leq e + (e-1) \sum_{j=1}^\infty e^j P \left( \frac{\sqrt{\varepsilon}}{\sqrt{T+1}} \left| \int_0^t b_0(s) dw_i^\varepsilon(s) \right| \geq j \right). \end{aligned} \tag{5.18}$$

To proceed, let us concentrate on the estimate of

$$P \left( \frac{\sqrt{\varepsilon}}{\sqrt{T+1}} \left| \int_0^t b_0(s) dw_i^\varepsilon(s) \right| \geq j \right).$$

For each  $i = 1, \dots, m$ , let

$$\tilde{p}_i(t) = \int_0^t b_0(s) dw_i^\varepsilon(s)$$

and let  $\tilde{q}_i(\cdot)$  denote the only solution to the following equation (see Elliott [55, Chapter 13])

$$\tilde{q}_i(t) = 1 + \zeta \int_0^t \tilde{q}_i(s^-) d\tilde{p}_i(s),$$

where  $\tilde{q}_i(s^-)$  is the left-hand limit of  $\tilde{q}_i$  at  $s$  and  $\zeta$  is a positive constant to be determined later. In what follows, we suppress the  $i$ -dependence and write  $\tilde{p}_i(\cdot)$  and  $\tilde{q}_i(\cdot)$  as  $\tilde{p}(\cdot)$  and  $\tilde{q}(\cdot)$  whenever there is no confusion.

Note that  $\tilde{p}(t)$ , for  $t \geq 0$ , is a local martingale. Since

$$\zeta \int_0^t \tilde{q}(s^-) d\tilde{p}(s), \quad t \geq 0,$$

is a local martingale, we have  $E\tilde{q}(t) \leq 1$  for all  $t \geq 0$ . Moreover,  $\tilde{q}(t)$  can be written as follows (see Elliott [55, Chapter 13]):

$$\tilde{q}(t) = \exp(\zeta\tilde{p}(t)) \prod_{s \leq t} (1 + \zeta\Delta\tilde{p}(s)) \exp(-\zeta\Delta\tilde{p}(s)), \quad (5.19)$$

where  $\Delta\tilde{p}(s) := \tilde{p}(s) - \tilde{p}(s^-)$ , with  $|\Delta\tilde{p}(s)| \leq 1$ .

Now observe that there exist positive constants  $\zeta_0$  and  $\kappa_1$  such that for  $0 < \zeta \leq \zeta_0$  and for all  $s > 0$ ,

$$(1 + \zeta\Delta\tilde{p}(s)) \exp(-\zeta\Delta\tilde{p}(s)) \geq \exp(-\kappa_1\zeta^2). \quad (5.20)$$

Combining (5.19) and (5.20), we obtain

$$\tilde{q}(t) \geq \exp\{\zeta\tilde{p}(t) - \kappa_1\zeta^2N(t)\}, \quad \text{for } 0 < \zeta \leq \zeta_0, \quad t > 0,$$

where  $N(t)$  is the number of jumps of  $\tilde{p}(s)$  in  $s \in [0, t]$ . Since  $N(t)$  is a monotonically increasing process, we have

$$\tilde{q}(t) \geq \exp\{\zeta\tilde{p}(t) - \kappa_1\zeta^2N(T)\}, \quad \text{for } 0 < \zeta \leq \zeta_0.$$

Note also that for each  $i = 1, \dots, m$ ,

$$\begin{aligned} & P\left(\left|\frac{\sqrt{\varepsilon}}{\sqrt{T+1}} \int_0^t b_0(s) dw_i^\varepsilon(s)\right| \geq j\right) \\ &= P\left(|\tilde{p}(t)| \geq \frac{j\sqrt{T+1}}{\sqrt{\varepsilon}}\right) \\ &\leq P\left(\tilde{p}(t) \geq \frac{j\sqrt{T+1}}{\sqrt{\varepsilon}}\right) + P\left(-\tilde{p}(t) \geq \frac{j\sqrt{T+1}}{\sqrt{\varepsilon}}\right). \end{aligned}$$

Consider the first term on the right-hand side of the inequality above. Let  $a_j = j(T+1)/(8\kappa_1\varepsilon)$ . Then

$$\begin{aligned}
 & P\left(\tilde{p}(t) \geq \frac{j\sqrt{T+1}}{\sqrt{\varepsilon}}\right) \\
 & \leq P\left(\tilde{q}(t) \geq \exp\left\{\frac{j\zeta\sqrt{T+1}}{\sqrt{\varepsilon}} - \kappa_1\zeta^2N(T)\right\}\right) \\
 & \leq P\left(\tilde{q}(t) \geq \exp\left\{\frac{j\zeta\sqrt{T+1}}{\sqrt{\varepsilon}} - \kappa_1\zeta^2N(T)\right\}, N(T) \leq a_j\right) \\
 & \quad + P(N(T) \geq a_j) \\
 & \leq P\left(\tilde{q}(t) \geq \exp\left(\frac{j\zeta\sqrt{T+1}}{\sqrt{\varepsilon}} - \kappa_1\zeta^2a_j\right)\right) + P(N(T) \geq a_j) \\
 & \leq 2 \exp\left(-\frac{j\zeta\sqrt{T+1}}{\sqrt{\varepsilon}} + \kappa_1\zeta^2a_j\right) + P(N(T) \geq a_j).
 \end{aligned}$$

The last inequality follows from the local martingale property (see Elliott [55, Theorem 4.2]).

Now if we choose  $\zeta = 4\sqrt{\varepsilon}/\sqrt{T+1}$ , then

$$\exp\left(-\frac{j\zeta\sqrt{T+1}}{\sqrt{\varepsilon}} + \kappa_1\zeta^2a_j\right) = e^{-2j}.$$

In view of the construction of Markov chains in Section 2.4, there exists a Poisson process  $N_0(\cdot)$  with parameter (i.e., mean)  $a/\varepsilon$  for some  $a > 0$ , such that  $N(t) \leq N_0(t)$ . Assume  $a = 1$  without loss of generality (otherwise one may replace  $\varepsilon$  by  $\varepsilon a^{-1}$ ). Using the Poisson distribution of  $N_0(t)$ , we have

$$P(N_0(T) \geq k) \leq \frac{(T/\varepsilon)^k}{k!} \text{ for } k \geq 0.$$

In view of Stirling's formula (see Chow and Teicher [30] or Feller [60]), for  $\varepsilon$  small enough,

$$P(N(T) \geq a_j) \leq \frac{(T/\varepsilon)^{\lfloor a_j \rfloor}}{\lfloor a_j \rfloor!} \leq 2 \left(\frac{8\kappa_1}{j}\right)^{a_j-1} := 2\gamma_0^{a_j-1},$$

where  $\lfloor a_j \rfloor$  is the integer part of  $a_j$  and

$$\gamma_0 = \frac{8e\kappa_1}{j_0} \in (0, 1) \quad \text{for } j_0 > \max\{1, 8e\kappa_1\}.$$

Thus, for  $j \geq j_0$ ,

$$P\left(\frac{\sqrt{\varepsilon}}{\sqrt{T+1}} \int_0^t b_0(s)dw_i^\varepsilon(s) \geq j\right) \leq 2e^{-2j} + 2\gamma_0^{a_j-1}.$$

Repeating the same argument for the martingale  $(-\tilde{p}(\cdot))$ , we get for  $j \geq j_0$ ,

$$P\left(-\frac{\sqrt{\varepsilon}}{\sqrt{T+1}} \int_0^t b_0(s)dw_i^\varepsilon(s) \geq j\right) \leq 2e^{-2j} + 2\gamma_0^{a_j-1}.$$

Combining the above two inequalities yields that for  $j \geq j_0$ ,

$$P \left( \frac{\sqrt{\varepsilon}}{\sqrt{T+1}} \left| \int_0^t b_0(s) dw_i^\varepsilon(s) \right| \geq j \right) \leq 4(e^{-2j} + \gamma_0^{a_j-1}).$$

Then by (5.18),

$$\begin{aligned} E \exp \left\{ \frac{\sqrt{\varepsilon}}{\sqrt{T+1}} \left| \int_0^t b_0(s) dw_i^\varepsilon(s) \right| \right\} \\ \leq K_0 + 4(e-1) \sum_{j=1}^{\infty} e^j (e^{-2j} + \gamma_0^{a_j-1}), \end{aligned}$$

where  $K_0$  is the sum corresponding to those terms with  $j \leq j_0$ . Now choose  $\varepsilon$  small enough that  $e\gamma_0^{1/(8\kappa_1\varepsilon)} \leq 1/2$ . Then

$$E \exp \left\{ \frac{\sqrt{\varepsilon}}{\sqrt{T+1}} \left| \int_0^t b_0(s) dw_i^\varepsilon(s) \right| \right\} \leq K_0 + 4e\gamma_0^{-1}.$$

Since  $t_0$  is arbitrary, we may take  $t_0 = t$  in the above inequality. Then

$$E \exp \left\{ \frac{\sqrt{\varepsilon}}{\sqrt{T+1}} \left| \int_0^t b(s, t) dw_i^\varepsilon(s) \right| \right\} \leq K_0 + 4e\gamma_0^{-1}.$$

Combining this inequality with (5.16) leads to

$$\begin{aligned} E \exp \left\{ \frac{\theta_T}{\sqrt{\varepsilon}(T+1)^{\frac{3}{2}}} \left| \int_0^t (\chi^\varepsilon(s) - \nu(s)) \beta(s) ds \right| \right\} \\ \leq e(K_0 + 4e\gamma_0^{-1}) := K. \end{aligned}$$

*Step 2.* Recall that

$$n^\varepsilon(t, i) = \frac{1}{\sqrt{\varepsilon}} \int_0^t (I_{\{\alpha(\varepsilon, s)=i\}} - \nu_i(s)) \beta_i(s) ds.$$

Then, for each  $i \in \mathcal{M}$ ,  $n^\varepsilon(t, i)$  is nearly a martingale, i.e., for  $\varepsilon$  small enough,

$$|E[n^\varepsilon(t, i) | \mathcal{F}_s] - n^\varepsilon(s, i)| \leq O(\sqrt{\varepsilon}), \quad \text{for all } \omega \in \Omega \text{ and } 0 \leq s \leq t \leq T. \tag{5.21}$$

Here  $O(\sqrt{\varepsilon})$  is deterministic, i.e., it does not depend on the sample point  $\omega$ . The reason is that it is obtained from the asymptotic expansions. In fact, for all  $i_0 \in \mathcal{M}$ ,

$$\begin{aligned}
 E \left[ \int_s^t (I_{\{\alpha(\varepsilon,r)=i\}} - \nu_i(r))\beta_i(r) dr \mid \alpha(\varepsilon, s) = i_0 \right] \\
 &= \int_s^t (E[I_{\{\alpha(\varepsilon,r)=i\}} \mid \alpha(\varepsilon, s) = i_0] - \nu_i(r))\beta_i(r) dr \\
 &= \int_s^t [P(\alpha(\varepsilon, r) = i \mid \alpha(\varepsilon, s) = i_0) - \nu_i(r)]\beta_i(r) dr \\
 &= \int_s^t O(\varepsilon + \exp(-\kappa_0(r - s)/\varepsilon)) dr = O(\varepsilon).
 \end{aligned}$$

So, (5.21) follows.

Step 3. We show that for each  $a > 0$ ,

$$E[\exp\{a|n^\varepsilon(t, i)|\} \mid \mathcal{F}_s] \geq \exp\{a|n^\varepsilon(s, i)|\}(1 + O(\sqrt{\varepsilon})).$$

First of all, note that  $\phi(x) = |x|$  is a convex function. There exists a vector function  $\phi_0(x)$  bounded by 1 such that

$$\phi(x) \geq \phi(a) + \phi_0(a) \cdot (x - a),$$

for all  $x$  and  $a$ . Noting that  $O(\sqrt{\varepsilon}) = -O(\sqrt{\varepsilon})$ , we have

$$\begin{aligned}
 E[|n^\varepsilon(t, i)| \mid \mathcal{F}_s] &\geq |n^\varepsilon(s, i)| + \phi_0(n^\varepsilon(s, i)) \cdot E[n^\varepsilon(t, i) - n^\varepsilon(s, i) \mid \mathcal{F}_s] \\
 &\geq |n^\varepsilon(s, i)| + O(\sqrt{\varepsilon}).
 \end{aligned}$$

Moreover, note that  $e^{ax}$  is also convex. It follows that

$$\begin{aligned}
 E[\exp(a|n^\varepsilon(t, i)|) \mid \mathcal{F}_s] \\
 &\geq \exp(a|n^\varepsilon(s, i)|) + a \exp\{a|n^\varepsilon(s, i)|\} E[|n^\varepsilon(t, i)| - |n^\varepsilon(s, i)| \mid \mathcal{F}_s] \\
 &\geq \exp(a|n^\varepsilon(s, i)|)(1 + O(\sqrt{\varepsilon})).
 \end{aligned}$$

Step 4. Let  $x^\varepsilon(t) = \exp(a|n^\varepsilon(t, i)|)$  for  $a > 0$ . Then, for any  $\mathcal{F}_t$  stopping time  $\tau \leq T$ ,

$$E[x^\varepsilon(T) \mid \mathcal{F}_\tau] \geq x^\varepsilon(\tau)(1 + O(\sqrt{\varepsilon})). \tag{5.22}$$

Note that  $x^\varepsilon(t)$  is continuous. Therefore, it suffices to show the above inequality when  $\tau$  takes values in a countable set  $\{t_1, t_2, \dots\}$ . To this end, note that, for each  $t_i$ ,

$$E[x^\varepsilon(T) \mid \mathcal{F}_{t_i}] \geq x^\varepsilon(t_i)(1 + O(\sqrt{\varepsilon})).$$

For all  $A \in \mathcal{F}_\tau$ , we have  $A \cap \{\tau = t_i\} \in \mathcal{F}_{t_i}$ . Therefore,

$$\int_{A \cap \{\tau=t_i\}} x^\varepsilon(T) dP \geq \left( \int_{A \cap \{\tau=t_i\}} x^\varepsilon(\tau) dP \right) (1 + O(\sqrt{\varepsilon})).$$



Thus

$$\int_A x^\varepsilon(T) dP \geq \left( \int_A x^\varepsilon(\tau) dP \right) (1 + O(\sqrt{\varepsilon})),$$

and (5.22) follows.

*Step 5.* Let  $a = \theta/\sqrt{(T+1)^3}$  in Step 3. Then, for  $\varepsilon$  small enough, there exists  $K$  such that

$$P\left(\sup_{t \leq T} x^\varepsilon(t) \geq x\right) \leq \frac{K}{x}, \quad (5.23)$$

for all  $x > 0$ .

In fact, let  $\tau = \inf\{t > 0 : x^\varepsilon(t) \geq x\}$ . We adopt the convention that  $\tau = \infty$  if  $\{t > 0 : x^\varepsilon(t) \geq x\} = \emptyset$ . Then we have

$$E[x^\varepsilon(T)] \geq (E[x^\varepsilon(T \wedge \tau)])(1 + O(\sqrt{\varepsilon})),$$

and write

$$E[x^\varepsilon(T \wedge \tau)] = E[x^\varepsilon(\tau)I_{\{\tau < T\}}] + E[x^\varepsilon(T)I_{\{\tau \geq T\}}] \geq E[x^\varepsilon(\tau)I_{\{\tau < T\}}].$$

Moreover, in view of the definition of  $\tau$ , we have

$$E[x^\varepsilon(\tau)I_{\{\tau < T\}}] \geq xP(\tau < T) \geq xP\left(\sup_{t \leq T} x^\varepsilon(t) \geq x\right).$$

It follows that

$$P\left(\sup_{t \leq T} x^\varepsilon(t) \geq x\right) \leq \frac{E[x^\varepsilon(T)]}{(1 + O(\sqrt{\varepsilon}))x} \leq \frac{K}{x}.$$

Thus, (5.23) follows.

Finally, to complete the proof of (5.13), note that, for  $0 < \kappa < 1$ ,

$$E \exp\left(\frac{\kappa\theta}{\sqrt{(1+T)^3}} \sup_{t \leq T} |n^\varepsilon(t, i)|\right) = E \left[ \sup_{t \leq T} (x^\varepsilon(t))^\kappa \right].$$

It follows that

$$\begin{aligned} E \left[ \sup_{t \leq T} (x^\varepsilon(t))^\kappa \right] &= \int_0^\infty P\left(\sup_{t \leq T} (x^\varepsilon(t))^\kappa \geq x\right) dx \\ &\leq 1 + \int_1^\infty P\left(\sup_{t \leq T} (x^\varepsilon(t))^\kappa \geq x\right) dx \\ &\leq 1 + \int_1^\infty P\left(\sup_{t \leq T} x^\varepsilon(t) \geq x^{1/\kappa}\right) dx \\ &\leq 1 + \int_1^\infty Kx^{-1/\kappa} dx < \infty. \end{aligned}$$

This completes the proof.  $\square$

Next we give several corollaries to the theorem. Such estimates are useful for establishing exponential bounds of asymptotic optimal hierarchical controls in manufacturing models (see Sethi and Zhang [192]).

**Corollary 5.6.** *In Theorem 5.4, if  $Q(t) = Q$ , a constant matrix, then the following stronger estimate holds:*

$$E \exp \left\{ \frac{\theta_T}{\sqrt{1+T}} \sup_{0 \leq t \leq T} |n^\varepsilon(t)| \right\} \leq K. \tag{5.24}$$

Moreover, the constant  $\theta_T = \theta$  is independent of  $T$  for  $T > 0$ .

Proof: If  $Q(t) = Q$ , then  $\varphi_1(t)$  in Lemma 5.1 is identically 0. Therefore, the estimate (5.11) can be replaced by

$$P^\varepsilon(t, s) - P_0(t) = K_T O_1 \left( \exp \left( -\frac{\kappa_0(t-s)}{\varepsilon} \right) \right).$$

As a result, the estimate in (5.15) can be replaced by

$$\sup_{0 \leq t \leq T} \left| \int_0^t (\chi^\varepsilon(s) - \nu(s)) \beta(s) ds \right| = \varepsilon K_T O_1(1) + \varepsilon K_T \sup_{0 \leq t \leq T} \left| \int_0^t O_1(1) dw^\varepsilon(s) \right|.$$

The proof of (5.24) follows in essentially the same way as that of Theorem 5.4 (from equation (5.15) on).

To see that  $\theta_T$  in (5.24) is independent of  $T$ , it suffices to note that in (5.11) the constant  $K_T$  is independent of  $T$ , which can be seen by examining closely Example 4.16. □

**Corollary 5.7.** *Under the conditions of Theorem 5.4, there exist constants  $K_j$ , such that for  $j = 1, 2, \dots$ ,*

$$E \sup_{0 \leq t \leq T} |n^\varepsilon(t)|^{2j} \leq K_j (1+T)^{3j}. \tag{5.25}$$

Moreover, if  $Q(t) = Q$ , then for some  $K_j$  independent of  $T$  and

$$E \sup_{0 \leq t \leq T} |n^\varepsilon(t)|^{2j} \leq K_j (1+T)^j. \tag{5.26}$$

Proof: Since (5.26) follows from a similar argument to that of Corollary 5.6, it suffices to verify (5.25) using Theorem 5.4. Note that for each  $j = 1, 2, \dots$ , there exists  $K_j^0$  such that for all  $x$ , we have  $x^{2j} \leq K_j^0 e^x$ . Thus,

$$\left( \frac{\theta_T}{(T+1)^{\frac{3}{2}}} \sup_{0 \leq t \leq T} |n^\varepsilon(t)| \right)^{2j} \leq K_j^0 \exp \left\{ \frac{\theta_T}{(T+1)^{\frac{3}{2}}} \sup_{0 \leq t \leq T} |n^\varepsilon(t)| \right\}.$$

Taking expectations on both sides of the above inequality yields the desired estimate. □

**Corollary 5.8.** *Under the conditions of Theorem 5.4, for each  $0 < \delta < 1/2$ , we have*

$$\begin{aligned}
 P \left( \sup_{0 \leq t \leq T} \left| \int_0^t (I_{\{\alpha^\varepsilon(s)=i\}} - \nu_i(s)) \beta_i(s) ds \right| \geq \varepsilon^{\frac{1}{2}-\delta} \right) \\
 \leq K \exp \left\{ -\frac{\theta_T}{\varepsilon^\delta (T+1)^{\frac{3}{2}}} \right\}.
 \end{aligned}
 \tag{5.27}$$

Moreover, if  $Q(t) = Q$ , then  $\theta_T = \theta$  is independent of  $T$  and

$$\begin{aligned}
 P \left( \sup_{0 \leq t \leq T} \left| \int_0^t (I_{\{\alpha^\varepsilon(s)=i\}} - \nu_i(s)) \beta_i(s) ds \right| \geq \varepsilon^{\frac{1}{2}-\delta} \right) \\
 \leq K \exp \left\{ -\frac{\theta}{\varepsilon^\delta \sqrt{1+T}} \right\}.
 \end{aligned}
 \tag{5.28}$$

Proof: Using Theorem 5.4, we obtain

$$\begin{aligned}
 P \left( \sup_{0 \leq t \leq T} \left| \int_0^t (I_{\{\alpha^\varepsilon(s)=i\}} - \nu_i(s)) \beta_i(s) ds \right| \geq \varepsilon^{\frac{1}{2}-\delta} \right) \\
 = P \left( \exp \left\{ \frac{\theta_T}{(T+1)^{\frac{3}{2}}} \sup_{0 \leq t \leq T} |n^\varepsilon(t)| \right\} \geq \exp \left\{ \frac{\theta_T \varepsilon^{\frac{1}{2}-\delta}}{\sqrt{\varepsilon} (T+1)^{\frac{3}{2}}} \right\} \right) \\
 \leq K \exp \left\{ -\frac{\theta_T}{\varepsilon^\delta (T+1)^{\frac{3}{2}}} \right\}.
 \end{aligned}$$

This proves (5.27). Similarly, (5.28) follows from Corollary 5.6. □

### 5.2.4 Asymptotic Normality

Recall that the  $i$ th component of  $n^\varepsilon(\cdot)$  is given by

$$n^\varepsilon(t, i) = \frac{1}{\sqrt{\varepsilon}} \int_0^t (I_{\{\alpha^\varepsilon(s)=i\}} - \nu_i(s)) \beta_i(s) ds.$$

It is expected that the sequence of centered and scaled occupation measures will display certain “central limit type” phenomena. The goal here is to study the asymptotic properties of  $n^\varepsilon(\cdot)$  as  $\varepsilon \rightarrow 0$ . To be more specific, we show that  $n^\varepsilon(\cdot)$  converges to a Gaussian process as  $\varepsilon$  goes to 0. The following theorem is the main result of this section.

**Theorem 5.9.** *Suppose that (A5.1) is satisfied and  $Q(\cdot)$  is twice continuously differentiable in  $[0, T]$  with the second derivative being Lipschitz. Then for  $t \in [0, T]$ , the process  $n^\varepsilon(\cdot)$  converges weakly to a Gaussian process  $n(\cdot)$  with independent increments such that*

$$E n(t) = 0 \text{ and } E[n'(t)n(t)] = \int_0^t A(s) ds,
 \tag{5.29}$$

where  $A(t) = (A_{ij}(t))$  with

$$A_{ij}(t) = \beta_i(t)\beta_j(t) \left[ \nu_i(t) \int_0^\infty q_{0,ij}(r, t) dr + \nu_j(t) \int_0^\infty q_{0,ji}(r, t) dr \right], \quad (5.30)$$

and  $Q_0(r, t) = (q_{0,ij}(r, t))$ .

**Remark 5.10.** In view of (5.29) and the independent increment property of  $n(t)$ , it follows that

$$E[n'(t_1)n(t_2)] = \int_0^{\min\{t_1, t_2\}} A(s) ds. \quad (5.31)$$

The form of the covariance matrix (between  $t_1$  and  $t_2$ ) reveals the nonstationarity of the limit process  $n(\cdot)$ . Note that the limit covariance of  $n(t)$  given in (5.31) is an integral of the function  $A(s)$  defined in (5.30). For simplicity, with a slight abuse of notation, we shall also call  $A(t)$  as the covariance. This convention will be used throughout the chapter.

**Remark 5.11.** The additional assumptions on the second derivative of  $Q(\cdot)$  in Theorem 5.9 are required for computing or characterizing the function  $A(\cdot)$ . It is not crucial for the convergence of  $n^\varepsilon(\cdot)$ ; see Remark 5.44 in Section 5.3.3 for details.

Proof of Theorem 5.9: We divide the proof into several steps, which are presented by a number of lemmas.

*Step 1.* Show that the limit of the mean of  $n^\varepsilon(\cdot)$  is 0.

**Lemma 5.12.** For each  $t \in [0, T]$ ,

$$\lim_{\varepsilon \rightarrow 0} En^\varepsilon(t) = 0.$$

Proof: Using Theorem 4.5 and the boundedness of  $\beta_i(\cdot)$ , for  $t \in [0, T]$ ,

$$\begin{aligned} En^\varepsilon(t, i) &= \frac{1}{\sqrt{\varepsilon}} \int_0^t (EI_{\{\alpha^\varepsilon(s)=i\}} - \nu_i(s))\beta_i(s) ds \\ &= \frac{1}{\sqrt{\varepsilon}} \int_0^t (P(\alpha^\varepsilon(s) = i) - \nu_i(s))\beta_i(s) ds \\ &= \frac{1}{\sqrt{\varepsilon}} \int_0^t \left[ O(\varepsilon) + O\left(\exp\left(-\frac{\kappa_0 s}{\varepsilon}\right)\right) \right] \beta_i(s) ds \\ &= O(\sqrt{\varepsilon}) + \frac{1}{\sqrt{\varepsilon}} \int_0^t O\left(\exp\left(-\frac{\kappa_0 s}{\varepsilon}\right)\right) ds = O(\sqrt{\varepsilon}) \rightarrow 0, \end{aligned}$$

for each  $i \in \mathcal{M}$ . □

*Step 2.* Calculate the limit covariance function.

**Lemma 5.13.** For each  $t \in [0, T]$ ,

$$\lim_{\varepsilon \rightarrow 0} E(n^{\varepsilon, \prime}(t)n^{\varepsilon}(t)) = \int_0^t A(s)ds, \quad (5.32)$$

where  $A(t)$  is given by (5.30).

Proof: For each  $i, j \in \mathcal{M}$ ,

$$\begin{aligned} E[n^{\varepsilon}(t, i)n^{\varepsilon}(t, j)] &= \frac{1}{\varepsilon} E \left[ \left( \int_0^t (I_{\{\alpha^{\varepsilon}(\varsigma)=i\}} - \nu_i(\varsigma))\beta_i(\varsigma)d\varsigma \right) \right. \\ &\quad \left. \times \left( \int_0^t (I_{\{\alpha^{\varepsilon}(r)=j\}} - \nu_j(r))\beta_j(r)dr \right) \right] \\ &= \frac{1}{\varepsilon} E \left[ \int_0^t \int_0^t \left( I_{\{\alpha^{\varepsilon}(\varsigma)=i, \alpha^{\varepsilon}(r)=j\}} - \nu_i(\varsigma)I_{\{\alpha^{\varepsilon}(r)=j\}} \right. \right. \\ &\quad \left. \left. - \nu_j(r)I_{\{\alpha^{\varepsilon}(\varsigma)=i\}} + \nu_i(\varsigma)\nu_j(r) \right) \beta_i(\varsigma)\beta_j(r)d\varsigma dr \right]. \end{aligned}$$

Let

$$D_1 = \{(\varsigma, r) : 0 \leq r \leq \varsigma \leq t\},$$

$$D_2 = \{(\varsigma, r) : 0 \leq \varsigma \leq r \leq t\},$$

and let

$$\begin{aligned} \Phi^{\varepsilon}(\varsigma, r) &= P(\alpha^{\varepsilon}(\varsigma) = i, \alpha^{\varepsilon}(r) = j) - \nu_i(\varsigma)P(\alpha^{\varepsilon}(r) = j) \\ &\quad - \nu_j(r)P(\alpha^{\varepsilon}(\varsigma) = i) + \nu_i(\varsigma)\nu_j(r). \end{aligned}$$

Then it follows that

$$\begin{aligned} E[n^{\varepsilon}(t, i)n^{\varepsilon}(t, j)] &= \frac{1}{\varepsilon} \left[ \int_0^t \int_0^t \Phi^{\varepsilon}(\varsigma, r)\beta_i(\varsigma)\beta_j(r)d\varsigma dr \right] \\ &= \frac{1}{\varepsilon} \left( \int_{D_1} + \int_{D_2} \right) \Phi^{\varepsilon}(\varsigma, r)\beta_i(\varsigma)\beta_j(r)d\varsigma dr. \end{aligned}$$

Note that if  $(\varsigma, r) \in D_1$ , then  $\varsigma \geq r$  and

$$\begin{aligned} P(\alpha^{\varepsilon}(\varsigma) = i, \alpha^{\varepsilon}(r) = j) \\ = P(\alpha^{\varepsilon}(\varsigma) = i | \alpha^{\varepsilon}(r) = j)P(\alpha^{\varepsilon}(r) = j). \end{aligned}$$

Hence, for  $(\varsigma, r) \in D_1$  we have

$$\begin{aligned} \Phi^\varepsilon(\varsigma, r) &= [P(\alpha^\varepsilon(\varsigma) = i | \alpha^\varepsilon(r) = j) - \nu_i(\varsigma)]P(\alpha^\varepsilon(r) = j) \\ &\quad + \nu_j(r)[\nu_i(\varsigma) - P(\alpha^\varepsilon(\varsigma) = i)]. \end{aligned}$$

Using Theorem 4.5 and Lemma 5.1, for  $(\varsigma, r) \in D_1$ ,

$$\begin{aligned} \Phi^\varepsilon(\varsigma, r) &= \left( \varepsilon \varphi_1^i(\varsigma) + q_{0,ji} \left( \frac{\varsigma - r}{\varepsilon}, r \right) + \varepsilon q_{1,ji} \left( \frac{\varsigma - r}{\varepsilon}, r \right) + O(\varepsilon^2) \right) \\ &\quad \times \left( \nu_j(r) + \varepsilon \varphi_1^j(r) + \psi_0^j \left( \frac{r}{\varepsilon} \right) + \varepsilon \psi_1^j \left( \frac{r}{\varepsilon} \right) + O(\varepsilon^2) \right) \\ &\quad - \nu_j(r) \left( \varepsilon \varphi_1^i(\varsigma) + \psi_0^i \left( \frac{\varsigma}{\varepsilon} \right) + \varepsilon \psi_1^i \left( \frac{\varsigma}{\varepsilon} \right) + O(\varepsilon^2) \right) \\ &= \nu_j(r) q_{0,ji} \left( \frac{\varsigma - r}{\varepsilon}, r \right) + \left[ O \left( \varepsilon \exp \left( -\frac{\kappa_0 r}{\varepsilon} \right) \right) \right. \\ &\quad \left. + O \left( \varepsilon \exp \left( -\frac{\kappa_0(\varsigma - r)}{\varepsilon} \right) \right) + O \left( \exp \left( -\frac{\kappa_0 \varsigma}{\varepsilon} \right) \right) + O(\varepsilon^2) \right]. \end{aligned}$$

In the above,  $\varphi_\ell^i$  and  $\psi_\ell^i$  denote the  $i$ th components of the vectors  $\varphi_\ell$  and  $\psi_\ell$ , respectively. By elementary integration, we have

$$\begin{aligned} \int_0^t \left( \int_0^\varsigma \exp \left( -\frac{\kappa_0 \varsigma}{\varepsilon} \right) dr \right) d\varsigma &= \int_0^t \varsigma \exp \left( -\frac{\kappa_0 \varsigma}{\varepsilon} \right) d\varsigma = O(\varepsilon^2), \\ \varepsilon \int_0^t \left( \int_0^\varsigma \exp \left( -\frac{\kappa_0 r}{\varepsilon} \right) dr \right) d\varsigma &= \frac{\varepsilon^2}{\kappa_0} \int_0^t \left( 1 - \exp \left( -\frac{\kappa_0 \varsigma}{\varepsilon} \right) \right) d\varsigma = O(\varepsilon^2), \end{aligned}$$

and

$$\varepsilon \int_0^t \left( \int_0^\varsigma \exp \left( -\frac{\kappa_0(\varsigma - r)}{\varepsilon} \right) dr \right) d\varsigma = \varepsilon \int_0^t \left( \int_0^\varsigma \exp \left( -\frac{\kappa_0 r}{\varepsilon} \right) dr \right) d\varsigma = O(\varepsilon^2).$$

Thus, it follows that

$$\begin{aligned} &\int_{D_1} \Phi^\varepsilon(\varsigma, r) \beta_i(\varsigma) \beta_j(r) d\varsigma dr \\ &= \int_0^t \left( \int_0^\varsigma q_{0,ji} \left( \frac{\varsigma - r}{\varepsilon}, r \right) \nu_j(r) \beta_i(\varsigma) \beta_j(r) dr \right) d\varsigma + O(\varepsilon^2). \end{aligned}$$

Exchanging the order of integration leads to

$$\begin{aligned} & \int_0^t \left( \int_0^\varsigma q_{0,ji} \left( \frac{\varsigma-r}{\varepsilon}, r \right) \nu_j(r) \beta_i(\varsigma) \beta_j(r) dr \right) d\varsigma \\ &= \int_0^t \left( \int_r^t q_{0,ji} \left( \frac{\varsigma-r}{\varepsilon}, r \right) \nu_j(r) \beta_i(\varsigma) \beta_j(r) d\varsigma \right) dr \\ &= \int_0^t \beta_j(r) \nu_j(r) \left( \int_r^t q_{0,ji} \left( \frac{\varsigma-r}{\varepsilon}, r \right) \beta_i(\varsigma) d\varsigma \right) dr. \end{aligned}$$

Making a change of variables (via  $\varsigma - r = \varepsilon s$ ) yields

$$\int_r^t q_{0,ji} \left( \frac{\varsigma-r}{\varepsilon}, r \right) \beta_i(\varsigma) d\varsigma = \varepsilon \int_0^{(t-r)/\varepsilon} q_{0,ji}(s, r) \beta_i(r + \varepsilon s) ds.$$

We note that  $\beta_i(\cdot)$  is bounded and  $\beta_i(r + \varepsilon s) \rightarrow \beta_i(r)$  in  $L^1$  for each  $r \in [0, T]$ , as  $\varepsilon \rightarrow 0$ . Since  $q_{0,ji}(\cdot)$  decays exponentially fast, as in Lemma 5.1, we have

$$\int_0^{(t-r)/\varepsilon} q_{0,ji}(s, r) \beta_i(r + \varepsilon s) ds \rightarrow \beta_i(r) \int_0^\infty q_{0,ji}(s, r) ds.$$

Therefore, we obtain

$$\begin{aligned} & \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} \int_{D_1} \Phi^\varepsilon(\varsigma, r) \beta_i(\varsigma) \beta_j(r) d\varsigma dr \\ &= \int_0^t \beta_i(r) \beta_j(r) \nu_j(r) \left( \int_0^\infty q_{0,ji}(s, r) ds \right) dr. \end{aligned} \tag{5.33}$$

Similarly, we can show that

$$\begin{aligned} & \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} \int_{D_2} \Phi^\varepsilon(\varsigma, r) \beta_i(\varsigma) \beta_j(r) d\varsigma dr \\ &= \int_0^t \beta_i(r) \beta_j(r) \nu_i(r) \left( \int_0^\infty q_{0,ij}(s, r) ds \right) dr. \end{aligned} \tag{5.34}$$

Combining (5.33) and (5.34), we obtain

$$\lim_{\varepsilon \rightarrow 0} E[n^\varepsilon(t, i) n^\varepsilon(t, j)] = \int_0^t A_{ij}(s) ds,$$

with  $A(t) = (A_{ij}(t))$  given by (5.30). □

*Step 3.* Establish a mixing condition for the sequence  $\{n^\varepsilon(\cdot)\}$ .

**Lemma 5.14.** *For any  $\varsigma \geq 0$  and  $\sigma\{\alpha^\varepsilon(s) : s \geq t + \varsigma\}$ -measurable  $\eta$  with  $|\eta| \leq 1$ ,*

$$\left| E(\eta | \alpha^\varepsilon(s) : s \leq t) - E\eta \right| \leq K \exp\left(-\frac{\kappa\varsigma}{\varepsilon}\right) \quad w.p.1. \tag{5.35}$$

**Remark 5.15.** It follows from (5.35) that for any  $\sigma\{\alpha^\varepsilon(s) : 0 \leq s \leq t\}$ -measurable  $\xi$  with  $|\xi| \leq 1$  and  $\eta$  given in Lemma 5.14,

$$\left| E\xi\eta - E\xi E\eta \right| \leq K \exp\left(-\frac{\kappa\zeta}{\varepsilon}\right). \tag{5.36}$$

We will make crucial use of (5.35) and (5.36) in what follows.

Proof of Lemma 5.14: For any

$$0 \leq s_1 \leq s_2 \leq \dots \leq s_n = t \leq t + \zeta = t_0 \leq t_1 \leq \dots \leq t_l < \infty,$$

let

$$E_1 = \{\alpha^\varepsilon(t) = i, \alpha^\varepsilon(s_{n-1}) = i_{n-1}, \dots, \alpha^\varepsilon(s_1) = i_1\} \text{ and}$$

$$E_2 = \{\alpha^\varepsilon(t + \zeta) = j, \alpha^\varepsilon(t_1) = j_1, \dots, \alpha^\varepsilon(t_l) = j_l\}.$$

Then in view of the Markovian property of  $\alpha^\varepsilon(\cdot)$ ,

$$\begin{aligned} P(E_2|E_1) &= P(E_2|\alpha^\varepsilon(t) = i) \\ &= P(\alpha^\varepsilon(t + \zeta) = j|\alpha^\varepsilon(t) = i)[p_{j,j_1}^\varepsilon(t_1, t + \zeta) \cdots p_{j_{l-1},j_l}^\varepsilon(t_l, t_{l-1})]. \end{aligned}$$

Similarly, we have

$$P(E_2) = P(\alpha^\varepsilon(t + \zeta) = j)[p_{j,j_1}^\varepsilon(t_1, t + \zeta) \cdots p_{j_{l-1},j_l}^\varepsilon(t_l, t_{l-1})].$$

We first show that

$$\left| P(E_2|E_1) - P(E_2) \right| \leq K \exp\left(-\frac{\kappa\zeta}{\varepsilon}\right), \tag{5.37}$$

for some positive constants  $K$  and  $\kappa$  that are independent of  $i, j \in \mathcal{M}$  and  $t \in [0, T]$ .

To verify (5.37), it suffices to show that for any  $k \in \mathcal{M}$ ,

$$\left| p_{ij}^\varepsilon(t + \zeta, t) - p_{kj}^\varepsilon(t + \zeta, t) \right| \leq K \exp\left(-\frac{2\kappa\zeta}{\varepsilon}\right). \tag{5.38}$$

Since  $P_0(\cdot)$  and  $P_1(\cdot)$  have identical rows, the asymptotic expansion in (5.3) implies that  $p_{ij}^\varepsilon(t + \zeta, t) - p_{kj}^\varepsilon(t + \zeta, t)$  is determined by  $Q_0(\zeta/\varepsilon, t)$ . By virtue of the asymptotic expansion (see Theorem 4.5 and Lemma 5.1), there exist a  $K_1 > 0$  and a  $\kappa_0 > 0$  such that

$$\left| Q_0\left(\frac{(t + \tilde{t}) - t}{\varepsilon}, t\right) \right| \leq K_1 \exp\left(-\frac{\kappa_0 \tilde{t}}{\varepsilon}\right), \text{ for all } \tilde{t} \geq 0.$$



Choose  $N > 0$  sufficiently large that  $K_1 \exp(-\kappa_0 N) < 1$ . Then for  $\varepsilon > 0$  sufficiently small, there is a  $0 < \rho < 1$  such that

$$\left| p_{ij}^\varepsilon(t + N\varepsilon, t) - p_{kj}^\varepsilon(t + N\varepsilon, t) \right| \leq \rho.$$

To proceed, subdivide  $[t + N\varepsilon, t + \varsigma]$  into intervals of length  $N\varepsilon$ .

In view of the Chapman–Kolmogorov equation,

$$\begin{aligned} & |p_{ij}^\varepsilon(t + 2N\varepsilon, t) - p_{kj}^\varepsilon(t + 2N\varepsilon, t)| \\ &= \left| \sum_{l_0=1}^m [p_{il_0}^\varepsilon(t + N\varepsilon, t) - p_{kl_0}^\varepsilon(t + N\varepsilon, t)] p_{l_0j}^\varepsilon(t + 2N\varepsilon, t + N\varepsilon) \right| \\ &= \left| \sum_{l_0=1}^m [p_{il_0}^\varepsilon(t + N\varepsilon, t) - p_{kl_0}^\varepsilon(t + N\varepsilon, t)] \right. \\ &\quad \left. \times [p_{l_0j}^\varepsilon(t + 2N\varepsilon, t + N\varepsilon) - p_{l_1j}^\varepsilon(t + 2N\varepsilon, t + N\varepsilon)] \right| \leq K\rho^2, \end{aligned}$$

for any  $l_1 \in \mathcal{M}$ . Iterating on the inequality above, we arrive at

$$\left| p_{ij}^\varepsilon(t + k_0N\varepsilon, t) - p_{kj}^\varepsilon(t + k_0N\varepsilon, t) \right| \leq K\rho^{k_0}, \quad \text{for } k_0 \geq 1.$$

Choose  $\kappa = -1/(2N) \log \rho$ , and note that  $\kappa > 0$ . Then for any  $\varsigma$  satisfying  $k_0N\varepsilon \leq \varsigma < (k_0 + 1)N\varepsilon$ ,

$$\left| p_{ij}^\varepsilon(t + \varsigma, t) - p_{kj}^\varepsilon(t + \varsigma, t) \right| \leq K \exp\left(-\frac{2\kappa\varsigma}{\varepsilon}\right).$$

Thus (5.37) holds. This implies that  $\alpha^\varepsilon(\cdot)$  is a mixing process with exponential mixing rate. By virtue of Lemma A.16, (5.35) holds.  $\square$

*Step 4.* Prove that the sequence  $n^\varepsilon(\cdot)$  is tight, and any weakly convergent subsequence of  $\{n^\varepsilon(\cdot)\}$  has continuous paths with probability 1.

**Lemma 5.16.** *The following assertions hold:*

- (a)  $\{n^\varepsilon(t); t \in [0, T]\}$  is tight in  $D([0, T]; \mathbb{R}^m)$ , where  $D([0, T]; \mathbb{R}^m)$  denotes the space of functions that are defined on  $[0, T]$  and that are right continuous with left limits.
- (b) The limit  $n(\cdot)$  of any weakly convergent subsequence of  $n^\varepsilon(\cdot)$  has continuous sample paths with probability 1.

*Proof:* For  $i \in \mathcal{M}$ , define

$$\tilde{n}^\varepsilon(t, i) = \frac{1}{\sqrt{\varepsilon}} \int_0^t (I_{\{\alpha^\varepsilon(s)=i\}} - P(\alpha^\varepsilon(s) = i)) \beta_i(s) ds.$$

By virtue of Theorem 4.5,

$$\frac{1}{\sqrt{\varepsilon}} \int_0^t (P(\alpha^\varepsilon(s) = i) - \nu_i(s)) \beta_i(s) ds = O(\sqrt{\varepsilon}).$$

Thus  $n^\varepsilon(t, i) = \tilde{n}^\varepsilon(t, i) + O(\sqrt{\varepsilon})$ , and as a result the tightness of  $\{n^\varepsilon(\cdot)\}$  will follow from the tightness of  $\{\tilde{n}^\varepsilon(\cdot)\}$  (see Kushner [139, Lemma 5, p. 50]).

For the tightness of  $\{\tilde{n}^\varepsilon(\cdot)\}$ , in view of Kushner [139, Theorem 5, p. 32], it suffices to show that

$$E|\tilde{n}^\varepsilon(t + \varsigma) - \tilde{n}^\varepsilon(t)|^4 \leq K\varsigma^2. \tag{5.39}$$

To verify this assertion, it is enough to prove that for each  $i \in \mathcal{M}$ ,  $\tilde{n}^\varepsilon(\cdot, i)$  satisfies the condition.

Fix  $i \in \mathcal{M}$  and for any  $0 \leq t \leq T$ , let

$$\theta(t) = (I_{\{\alpha^\varepsilon(t)=i\}} - P(\alpha^\varepsilon(t) = i)) \beta_i(t).$$

We have suppressed the  $i$  and  $\varepsilon$  dependence in  $\theta(t)$  for ease of presentation. Let  $D = \{(s_1, s_2, s_3, s_4) : t \leq s_i \leq t + \varsigma, i = 1, 2, 3, 4\}$ . It follows that

$$\begin{aligned} E|\tilde{n}^\varepsilon(t + \varsigma, i) - \tilde{n}^\varepsilon(t, i)|^4 & \\ & \leq \frac{1}{\varepsilon^2} \int_D |E\theta(s_1)\theta(s_2)\theta(s_3)\theta(s_4)| ds_1 ds_2 ds_3 ds_4. \end{aligned} \tag{5.40}$$

Let  $(i_1, i_2, i_3, i_4)$  denote a permutation of  $(1, 2, 3, 4)$  and

$$D_{i_1 i_2 i_3 i_4} = \{(s_1, s_2, s_3, s_4) : t \leq s_{i_1} \leq s_{i_2} \leq s_{i_3} \leq s_{i_4} \leq t + \varsigma\}.$$

Then it is easy to see that  $D = \cup D_{i_1 i_2 i_3 i_4}$ . This and (5.40) leads to

$$\begin{aligned} E|\tilde{n}^\varepsilon(t + \varsigma, i) - \tilde{n}^\varepsilon(t, i)|^4 & \\ & \leq \frac{K}{\varepsilon^2} \int_{D_0} |E\theta(s_1)\theta(s_2)\theta(s_3)\theta(s_4)| ds_1 ds_2 ds_3 ds_4, \end{aligned}$$

where  $D_0 = D_{1234}$ .

Note that

$$\begin{aligned} & |E\theta(s_1)\theta(s_2)\theta(s_3)\theta(s_4)| \\ & \leq |E\theta(s_1)\theta(s_2)\theta(s_3)\theta(s_4) - E\theta(s_1)\theta(s_2)E\theta(s_3)\theta(s_4)| \\ & \quad + |E\theta(s_1)\theta(s_2)||E\theta(s_3)\theta(s_4)|. \end{aligned} \tag{5.41}$$

By virtue of (5.36) and  $E\theta(t) = 0, t \geq 0$ ,

$$\begin{aligned} |E\theta(s_1)\theta(s_2)| &= |E\theta(s_1)\theta(s_2) - E\theta(s_1)E\theta(s_2)| \\ &\leq K \exp\left(-\frac{\kappa(s_2 - s_1)}{\varepsilon}\right). \end{aligned}$$

Similarly, we have

$$\begin{aligned} |E\theta(s_3)\theta(s_4)| &= |E\theta(s_3)\theta(s_4) - E\theta(s_3)E\theta(s_4)| \\ &\leq K \exp\left(-\frac{\kappa(s_4 - s_3)}{\varepsilon}\right). \end{aligned}$$

Therefore, it follows that

$$\frac{K}{\varepsilon^2} \int_{D_0} |E\theta(s_1)\theta(s_2)| \cdot |E\theta(s_3)\theta(s_4)| ds_1 ds_2 ds_3 ds_4 \leq K\zeta^2. \quad (5.42)$$

The elementary inequality  $(a+b)^{1/2} \leq a^{1/2} + b^{1/2}$  for nonnegative numbers  $a$  and  $b$  yields that

$$\begin{aligned} &|E\theta(s_1)\theta(s_2)\theta(s_3)\theta(s_4) - E\theta(s_1)\theta(s_2)E\theta(s_3)\theta(s_4)| \\ &= \left(|E\theta(s_1)\theta(s_2)\theta(s_3)\theta(s_4) - E\theta(s_1)\theta(s_2)E\theta(s_3)\theta(s_4)|^{\frac{1}{2}}\right)^2 \\ &\leq |E\theta(s_1)\theta(s_2)\theta(s_3)\theta(s_4) - E\theta(s_1)\theta(s_2)E\theta(s_3)\theta(s_4)|^{\frac{1}{2}} \\ &\quad \times \left(|E\theta(s_1)\theta(s_2)\theta(s_3)\theta(s_4)|^{\frac{1}{2}} + |E\theta(s_1)\theta(s_2)E\theta(s_3)\theta(s_4)|^{\frac{1}{2}}\right). \end{aligned}$$

In view of (5.36), we obtain

$$\begin{aligned} &|E\theta(s_1)\theta(s_2)\theta(s_3)\theta(s_4) - E\theta(s_1)\theta(s_2)E\theta(s_3)\theta(s_4)|^{\frac{1}{2}} \\ &\leq K \exp\left(-\frac{\kappa(s_3 - s_2)}{2\varepsilon}\right). \end{aligned}$$

Similarly, by virtue of (5.35) and the boundedness of  $\theta(s)$ ,

$$\begin{aligned} &|E\theta(s_1)\theta(s_2)\theta(s_3)\theta(s_4)|^{\frac{1}{2}} \\ &= |E\theta(s_1)\theta(s_2)\theta(s_3)(E(\theta(s_4)|\alpha^\varepsilon(s) : s \leq s_3) - E\theta(s_4))|^{\frac{1}{2}} \\ &\leq K \exp\left(-\frac{\kappa(s_4 - s_3)}{2\varepsilon}\right), \end{aligned}$$

and

$$\begin{aligned}
 & |E\theta(s_1)\theta(s_2)E\theta(s_3)\theta(s_4)|^{\frac{1}{2}} \\
 &= |(E\theta(s_1)\theta(s_2) - E\theta(s_1)E\theta(s_2))(E\theta(s_3)\theta(s_4) - E\theta(s_3)E\theta(s_4))|^{\frac{1}{2}} \\
 &\leq K \exp\left(-\frac{\kappa(s_2 - s_1)}{2\varepsilon}\right) \exp\left(-\frac{\kappa(s_4 - s_3)}{2\varepsilon}\right).
 \end{aligned}$$

By virtue of the estimates above, we arrive at

$$\begin{aligned}
 & \frac{K}{\varepsilon^2} \int_{D_0} |E\theta(s_1)\theta(s_2)\theta(s_3)\theta(s_4) \\
 & \quad - E\theta(s_1)\theta(s_2)E\theta(s_3)\theta(s_4)| ds_1 ds_2 ds_3 ds_4 \leq K\zeta^2.
 \end{aligned} \tag{5.43}$$

The estimate (5.39) then follows from (5.42) and (5.43), and so does the desired tightness of  $\{n^\varepsilon(\cdot)\}$ .

Since  $\{n^\varepsilon(\cdot)\}$  is tight, by Prohorov’s theorem, we extract a convergent subsequence, and for notational simplicity, we still denote the sequence by  $\{n^\varepsilon(\cdot)\}$  whose limit is  $n(\cdot)$ . By virtue of Kushner [139, Theorem 5, p. 32] or Ethier and Kurtz [59, Proposition 10.3, p. 149],  $n(\cdot)$  has continuous paths with probability 1. □

**Remark 5.17.** Step 4 implies that both  $n^\varepsilon(\cdot)$  and  $n(\cdot)$  have continuous sample paths with probability 1. It follows, in view of Prohorov’s theorem (see Billingsley [13]), that  $n^\varepsilon(\cdot)$  is tight in  $C([0, T]; \mathbb{R}^m)$ .

*Step 5.* Show that the finite-dimensional distributions of  $n^\varepsilon(\cdot)$  converge to that of a Gaussian process with independent increments.

This part of the proof is similar to Khasminskii [112] (see also Friedlin and Wentzel [67, pp. 224]). Use  $\iota$  to denote the imaginary number  $\iota^2 = -1$ . To prove the convergence of the finite-dimensional distributions, we use the characteristic function  $E \exp(\iota \langle z, n^\varepsilon(t) \rangle)$ , where  $z \in \mathbb{R}^m$  and  $\langle \cdot, \cdot \rangle$  denotes the usual inner product in  $\mathbb{R}^m$ . Owing to the mixing property and repeated applications of Remark 5.15, for arbitrary positive real numbers  $s_l$  and  $t_l$  satisfying

$$0 \leq s_0 \leq t_0 \leq s_1 \leq t_1 \leq s_2 \leq \dots \leq s_n \leq t_n,$$

we have

$$\begin{aligned}
 & \left| E \exp\left(\iota \sum_{l=0}^n \langle z_l, (n^\varepsilon(t_l) - n^\varepsilon(s_l)) \rangle\right) \right. \\
 & \quad \left. - \prod_{l=0}^n E \exp\left(\iota \langle z_l, (n^\varepsilon(t_l) - n^\varepsilon(s_l)) \rangle\right) \right| \rightarrow 0
 \end{aligned}$$

as  $\varepsilon \rightarrow 0$ , for  $z_l \in \mathbb{R}^m$ . This, in turn, implies that the limit process  $n(\cdot)$  has independent increments. Moreover, in view of Lemma 5.16, the limit process has continuous path with probability 1. In accordance with a result in Skorohod [197, p. 7], if a process with independent increments has continuous paths w.p.1, then it must necessarily be a Gaussian process. This implies that the limits of the finite-dimensional distribution of  $n(\cdot)$  are Gaussian.

Consequently,  $n(\cdot)$  is a process having Gaussian finite-dimensional distributions, with mean zero and covariance  $\int_0^t A(s)ds$  given by Lemma 5.13. Moreover, the limit does not depend on the chosen subsequence. Thus  $n^\varepsilon(\cdot)$  converges weakly to the Gaussian process  $n(\cdot)$ . This completes the proof of the theorem.  $\square$

To illustrate, we give an example in which the covariance function of the limit process can be calculated explicitly.

**Example 5.18.** Let  $\alpha^\varepsilon(t) \in \mathcal{M} = \{1, 2\}$  be a two-state Markov chain with a generator

$$Q(t) = \begin{pmatrix} -\mu_1(t) & \mu_1(t) \\ \mu_2(t) & -\mu_2(t) \end{pmatrix}$$

where  $\mu_1(t) \geq 0$ ,  $\mu_2(t) \geq 0$ , and  $\mu_1(t) + \mu_2(t) > 0$  for each  $t \in [0, T]$ . Moreover,  $\mu_1(\cdot)$  and  $\mu_2(\cdot)$  are twice continuously differentiable with Lipschitz continuous second derivatives. It is easy to see that assumptions (A5.1) and (A5.2) are satisfied. Therefore the desired asymptotic normality follows.

In this example,

$$\nu(t) = (\nu_1(t), \nu_2(t)) = \left( \frac{\mu_2(t)}{\mu_1(t) + \mu_2(t)}, \frac{\mu_1(t)}{\mu_1(t) + \mu_2(t)} \right).$$

Moreover,

$$Q_0(s, t_0) = -\frac{\exp(-(\mu_1(t_0) + \mu_2(t_0))s)}{\mu_1(t_0) + \mu_2(t_0)} Q(t_0).$$

Thus,

$$A(t) = \frac{2\mu_1(t)\mu_2(t)}{(\mu_1(t) + \mu_2(t))^3} \begin{pmatrix} (\beta_1(t))^2 & -\beta_1(t)\beta_2(t) \\ -\beta_1(t)\beta_2(t) & (\beta_2(t))^2 \end{pmatrix}.$$

### 5.2.5 Extensions

In this section, we generalize our results in the previous sections including asymptotic expansions, asymptotic normality, and exponential bounds, to the Markov chain  $\alpha^\varepsilon(\cdot)$  with generator given by  $Q^\varepsilon(t) = Q(t)/\varepsilon + \widehat{Q}(t)$  with

weakly irreducible generator  $Q(t)$ . Recall that the vector of probabilities  $p^\varepsilon(t) = (P(\alpha^\varepsilon(t) = 1), \dots, P(\alpha^\varepsilon(t) = m))$  satisfies the differential equation

$$\begin{aligned} \frac{dp^\varepsilon(t)}{dt} &= p^\varepsilon(t)Q^\varepsilon(t), \quad p^\varepsilon(t) \in \mathbb{R}^m, \\ p^\varepsilon(0) &= p^0 \text{ with } p_i^0 \geq 0 \text{ for } i \in \mathcal{M} \text{ and } \sum_{i=1}^m p_i^0 = 1, \end{aligned}$$

To proceed, the following conditions are needed.

**(A5.3)** Both  $Q(t)$  and  $\widehat{Q}(t)$  are generators. For each  $t \in [0, T]$ ,  $Q(t)$  is weakly irreducible.

**(A5.4)** For some positive integer  $n_0$ ,  $Q(\cdot)$  is  $(n_0 + 1)$ -times continuously differentiable on  $[0, T]$  and  $(d^{n_0+1}/dt^{n_0+1})Q(\cdot)$  is Lipschitz. Moreover,  $\widehat{Q}(\cdot)$  is  $n_0$ -times continuously differentiable on  $[0, T]$  and  $(d^{n_0}/dt^{n_0})\widehat{Q}(\cdot)$  is Lipschitz.

Similarly to Section 4.2 for  $k = 1, \dots, n_0 + 1$ , the outer expansions lead to equations

$$\begin{aligned} \varepsilon^0 : \quad & \varphi_0(t)Q(t) = 0, \\ \varepsilon^1 : \quad & \varphi_1(t)Q(t) + \varphi_0(t)\widehat{Q}(t) = \frac{d\varphi_0(t)}{dt}, \\ & \dots \\ \varepsilon^k : \quad & \varphi_k(t)Q(t) + \varphi_{k-1}(t)\widehat{Q}(t) = \frac{d\varphi_{k-1}(t)}{dt}, \end{aligned} \tag{5.44}$$

with constraints

$$\sum_{i=1}^m \varphi_{0,i}(t) = 1$$

and

$$\sum_{i=1}^m \varphi_{k,i}(t) = 0, \text{ for } k \geq 1.$$

The initial-layer correction terms are

$$\begin{aligned}
 \varepsilon^0 : \quad & \frac{d\psi_0(\tau)}{d\tau} = \psi_0(\tau)Q(0), \\
 \varepsilon^1 : \quad & \frac{d\psi_1(\tau)}{d\tau} = \psi_1(\tau)Q(0) + \psi_0(\tau) \left( \tau \frac{dQ(0)}{dt} + \widehat{Q}(0) \right), \\
 & \dots \\
 \varepsilon^k : \quad & \frac{d\psi_k(\tau)}{d\tau} = \psi_k(\tau)Q(0) + r_k(\tau),
 \end{aligned} \tag{5.45}$$

where

$$r_k(\tau) = \sum_{i=1}^k \psi_{k-i}(\tau) \left( \frac{\tau^i}{i!} \frac{d^i Q(0)}{dt^i} + \frac{\tau^{i-1}}{(i-1)!} \frac{d^{i-1} \widehat{Q}(0)}{dt^{i-1}} \right),$$

with initial conditions

$$\begin{aligned}
 \psi_0(0) &= p^0 - \varphi_0(0), \text{ and} \\
 \psi_k(0) &= -\varphi_k(0) \text{ for } k \geq 1.
 \end{aligned}$$

**Theorem 5.19.** *Suppose that (A5.3) and (A5.4) are satisfied. Then*

- (a)  $\varphi_i(\cdot)$  is  $(n_0 + 1 - i)$ -times continuously differentiable on  $[0, T]$ ,
- (b) for each  $i$ , there is a  $\widehat{\kappa} > 0$  such that

$$\left| \psi_i \left( \frac{t}{\varepsilon} \right) \right| \leq K \exp \left( -\frac{\widehat{\kappa}t}{\varepsilon} \right), \text{ and}$$

- (c) the approximation error satisfies

$$\sup_{t \in [0, T]} \left| p^\varepsilon(t) - \sum_{i=0}^{n_0} \varepsilon^i \varphi_i(t) - \sum_{i=0}^{n_0} \varepsilon^i \psi_i \left( \frac{t}{\varepsilon} \right) \right| \leq K \varepsilon^{n_0+1}. \tag{5.46}$$

The proof of this theorem is similar to those of Theorem 4.5, and is thus omitted. We also omit the proofs of the following two theorems because they are similar to that of Theorem 5.4 and Theorem 5.9, respectively.

**Theorem 5.20.** *Suppose (A5.3) and (A5.4) are satisfied with  $n_0 = 0$ . Then there exist positive constants  $\varepsilon_0$  and  $K$  such that for  $0 < \varepsilon \leq \varepsilon_0$ ,  $i \in \mathcal{M}$ ,*

and for any deterministic process  $\beta_i(\cdot)$  satisfying  $|\beta_i(t)| \leq 1$  for all  $t \geq 0$ , we have

$$E \exp \left\{ \frac{\theta_T}{(T+1)^{\frac{3}{2}}} \sup_{0 \leq t \leq T} |n^\varepsilon(t)| \right\} \leq K,$$

where  $\theta_T$  and  $n^\varepsilon(\cdot)$  are as defined previously.

**Corollary 5.21.** Consider  $Q^\varepsilon = Q/\varepsilon + \widehat{Q}$  with constant generators  $Q$  and  $\widehat{Q}$  such that  $Q$  is weakly irreducible. Then (5.25) and (5.27) hold with constants  $K$  and  $K_j$  independent of  $T$ .

**Theorem 5.22.** Suppose (A5.3) and (A5.4) are satisfied with  $n_0 = 1$ . Then for  $t \in [0, T]$ , the process  $n^\varepsilon(\cdot)$  converges weakly to a Gaussian process  $n(\cdot)$  such that

$$En(t) = 0 \text{ and } E[n'(t)n(t)] = \int_0^t A(s)ds,$$

where  $A(t) = (A_{ij}(t))$  with

$$A_{ij}(t) = \beta_i(t)\beta_j(t) \left[ \nu_i(t) \int_0^\infty q_{0,ij}(r, t)dr + \nu_j(t) \int_0^\infty q_{0,ji}(r, t)dr \right],$$

and  $Q_0(r, t) = (q_{0,ij}(r, t))$  satisfying

$$\frac{dQ_0(r, t)}{dr} = Q_0(r, t)Q(t), \quad r \geq 0,$$

$$Q_0(0, t) = I - P_0(t),$$

with  $P_0(t) = (\nu'(t), \dots, \nu'(t))'$ .

**Remark 5.23.** In view of Theorem 5.22, the asymptotic covariance is determined by the quasi-stationary distribution  $\nu(t)$  and  $Q_0(r, t)$ . Both  $\nu(t)$  and  $Q_0(r, t)$  are determined by  $Q(t)$ , the dominating term in  $Q^\varepsilon(t)$ . In the asymptotic normality analysis, it is essential to have the irreducibility condition of  $Q(t)$ , whereas the role of  $\widehat{Q}(t)$  is not as important. If  $Q(t)$  is weakly irreducible, then there exists an  $\varepsilon_0 > 0$  such that  $Q^\varepsilon(t) = Q(t)/\varepsilon + \widehat{Q}(t)$  is weakly irreducible for  $0 < \varepsilon \leq \varepsilon_0$ , as shown in Sethi and Zhang [192, Lemma J.10].

By introducing another generator  $\widehat{Q}(t)$ , we are dealing with a singularly perturbed Markovian system with fast and slow motions. Nevertheless, the entire system under consideration is still weakly irreducible. This irreducibility allows us to extend our previous results with minor modifications.



Although most of the results in this section can be extended to the case with  $Q^\varepsilon(t) = Q(t)/\varepsilon + \widehat{Q}(t)$ , there are some exceptions. For example, Corollary 5.6 would not go through because even with constant matrix  $\widehat{Q}(t) = \widehat{Q}$ ,  $\varphi_1(t)$  in Lemma 5.1 does not equal 0 when  $\widehat{Q} \neq 0$ .

One may wonder what happens if  $Q(t)$  in  $Q^\varepsilon(t)$  is *not* weakly irreducible. In particular, one can consider the case in which  $Q(t)$  consists of several blocks of irreducible submatrices. Related results of asymptotic normality and the exponential bounds are treated in subsequent sections.

### 5.3 Markov Chains with Weak and Strong Interactions

For brevity, unless otherwise noted, in the rest of the book, whenever the phrase “weak and strong interaction” is used, it refers to the case of two-time-scale Markov chains with all states being recurrent. Similar approaches can be used for the other cases as well. The remainder of the chapter concentrates on exploiting detailed structures of the weak and strong interactions. In addition, it deals with convergence of the probability distribution with merely measurable generators.

We continue our investigation of asymptotic properties of the Markov chain  $\alpha^\varepsilon(\cdot)$  generated by  $Q^\varepsilon(\cdot)$ , with

$$Q^\varepsilon(t) = \frac{1}{\varepsilon} \widetilde{Q}(t) + \widehat{Q}(t), \text{ for } t \geq 0, \tag{5.47}$$

where  $\widetilde{Q}(t) = \text{diag}(\widetilde{Q}^1(t), \dots, \widetilde{Q}^l(t))$  is a block-diagonal matrix such that  $\widehat{Q}(t)$  and  $\widetilde{Q}^k(t)$ , for  $k = 1, \dots, l$ , are themselves generators. The state space of  $\alpha^\varepsilon(\cdot)$  is given by

$$\mathcal{M} = \left\{ s_{11}, \dots, s_{1m_1}, \dots, s_{l1}, \dots, s_{lm_l} \right\}.$$

For each  $k = 1, \dots, l$ , let  $\mathcal{M}_k = \{s_{k1}, \dots, s_{km_k}\}$ , representing the group of states corresponding to  $\widetilde{Q}^k(t)$ .

The results in Section 5.3.1 reveal the structures of the Markov chains with weak and strong interactions based on the following observations. Intuitively, for small  $\varepsilon$ , the Markov chain  $\alpha^\varepsilon(\cdot)$  jumps more frequently within the states in  $\mathcal{M}_k$  and less frequently from  $\mathcal{M}_k$  to  $\mathcal{M}_j$  for  $j \neq k$ . Therefore, the states in  $\mathcal{M}_k$  can be aggregated and represented by a single state  $k$  (one may view the state  $k$  as a super state). That is, one can approximate  $\alpha^\varepsilon(\cdot)$  by an aggregated process, say,  $\bar{\alpha}^\varepsilon(\cdot)$ . Furthermore, by examining the tightness and finite-dimensional distribution of  $\bar{\alpha}^\varepsilon(\cdot)$ , it will be shown that  $\bar{\alpha}^\varepsilon(\cdot)$  converges weakly to a Markov chain  $\bar{\alpha}(\cdot)$  generated by

$$\overline{Q}(t) = \text{diag}(\nu^1(t), \dots, \nu^l(t)) \widehat{Q}(t) \text{diag}(\mathbf{1}_{m_1}, \dots, \mathbf{1}_{m_l}). \tag{5.48}$$

Section 5.3.2 continues the investigation along the line of estimating the error bounds of the approximation. Our interest lies in finding how closely one can approximate an unscaled sequence of occupation measures. The study is through the examination of appropriate exponential-type bounds. To take a suitable scaled sequence, one first centers the sequence around the “mean,” and then compares the actual sequence of occupation measures with this “mean.” In contrast to the results of Section 5.2, in lieu of taking the difference of the occupation measure with that of a deterministic function, it is compared with a random process. One of the key points here is the utilization of solutions of linear time-varying stochastic differential equations, in which the stochastic integration is with respect to a square-integrable martingale.

In comparison with the central limit theorem obtained in Section 5.2, it is interesting to know whether these results still hold under the structure of weak and strong interactions. The answer to this question is in Section 5.3.3, which also contains further study on related scaled sequences of occupation measures. The approach is quite different from that of Section 5.2. We use the martingale formulation and apply the techniques of perturbed test functions. It is interesting to note that the limit process is a switching diffusion process, which does not have independent increments. When the generator is weakly irreducible as in Section 5.2, the motion of jumping around the grouped states disappears and the diffusion becomes the dominant force.

We have considered only Markov chains with smooth generators up to now. However, there are cases in certain applications in which the generators may be merely measurable. Section 5.4 takes care of the scenario in which the Markov chains are governed by generators that are only measurable. Formulation via weak derivatives is also discussed briefly. Finally the chapter is concluded with a few more remarks. Among other things, additional references are given.

### 5.3.1 Aggregation of Markov Chains

This section deals with an aggregation of  $\alpha^\varepsilon(\cdot)$ . The following assumptions will be needed:

(A5.5) For each  $k = 1, \dots, l$  and  $t \in [0, T]$ ,  $\tilde{Q}^k(t)$  is weakly irreducible.

(A5.6)  $\tilde{Q}(\cdot)$  is differentiable on  $[0, T]$  and its derivative is Lipschitz. Moreover,  $\hat{Q}(\cdot)$  is also Lipschitz.

The assumptions above guarantee the existence of an asymptotic expansion up to zeroth order. To prepare for the subsequent study, we first provide the following error estimate. Since only the zeroth-order expansion is needed here, the estimate is confined to such an approximation. Higher-order terms can be obtained in a similar way.

**Lemma 5.24.** *Assume (A5.5) and (A5.6). Let  $P^\varepsilon(t, t_0)$  denote the transition probability of  $\alpha^\varepsilon(\cdot)$ . Then for some  $\kappa_0 > 0$ ,*

$$P^\varepsilon(t, t_0) = P_0(t, t_0) + O\left(\varepsilon + \exp\left(-\frac{\kappa_0(t - t_0)}{\varepsilon}\right)\right),$$

where

$$\begin{aligned} P_0(t, t_0) &= \tilde{\mathbf{1}}\Theta(t, t_0)\text{diag}(\nu^1(t), \dots, \nu^l(t)) \\ &= \begin{pmatrix} \mathbb{1}_{m_1}\nu^1(t)\vartheta_{11}(t, t_0), & \dots, & \mathbb{1}_{m_1}\nu^l(t)\vartheta_{1l}(t, t_0) \\ \vdots & \dots & \vdots \\ \mathbb{1}_{m_l}\nu^1(t)\vartheta_{l1}(t, t_0), & \dots, & \mathbb{1}_{m_l}\nu^l(t)\vartheta_{ll}(t, t_0) \end{pmatrix}, \end{aligned} \tag{5.49}$$

where  $\nu^k(t)$  is the quasi-stationary distribution of  $\tilde{Q}^k(t)$ , and  $\Theta(t, t_0) = (\vartheta_{ij}(t, t_0)) \in \mathbb{R}^{l \times l}$  is the solution to the following initial value problem:

$$\begin{aligned} \frac{d\Theta(t, t_0)}{dt} &= \Theta(t, t_0)\bar{Q}(t), \\ \Theta(t_0, t_0) &= I. \end{aligned} \tag{5.50}$$

Proof: The proof is similar to those of Lemma 5.1 and Theorem 4.29, except that the notation is more involved. □

Define an aggregated process of  $\alpha^\varepsilon(\cdot)$  on  $[0, T]$  by

$$\bar{\alpha}^\varepsilon(t) = k \text{ if } \alpha^\varepsilon(t) \in \mathcal{M}_k. \tag{5.51}$$

The idea to follow is to treat a related Markov chain having only  $l$  states. The transitions among its states correspond to the jumps from one group  $\mathcal{M}_k$  to another  $\mathcal{M}_j$ ,  $j \neq k$ , in the original Markov chain.

**Theorem 5.25.** *Assume (A5.5) and (A5.6). Then, for any  $i = 1, \dots, l$ ,  $j = 1, \dots, m_i$ , and bounded and measurable deterministic function  $\beta_{ij}(\cdot)$ ,*

$$E\left(\int_0^T \left(I_{\{\alpha^\varepsilon(t)=s_{ij}\}} - \nu_j^i(t)I_{\{\bar{\alpha}^\varepsilon(t)=i\}}\right) \beta_{ij}(t) dt\right)^2 = O(\varepsilon).$$

Proof: For any  $i, j$  and  $0 \leq t \leq T$ , let

$$\eta^\varepsilon(t) = E\left(\int_0^t \left(I_{\{\alpha^\varepsilon(r)=s_{ij}\}} - \nu_j^i(r)I_{\{\bar{\alpha}^\varepsilon(r)=i\}}\right) \beta_{ij}(r) dr\right)^2. \tag{5.52}$$

We have suppressed the  $i, j$  dependence of  $\eta^\varepsilon(\cdot)$  for notational simplicity. Loosely speaking, the argument used below is a Liapunov stability one, and  $\eta^\varepsilon(\cdot)$  can be viewed as a Liapunov function. By differentiating  $\eta^\varepsilon(\cdot)$ , we have

$$\begin{aligned} \frac{d\eta^\varepsilon(t)}{dt} = 2E & \left[ \left( \int_0^t (I_{\{\alpha^\varepsilon(r)=s_{ij}\}} - \nu_j^i(r)I_{\{\bar{\alpha}^\varepsilon(r)=i\}}) \beta_{ij}(r)dr \right) \right. \\ & \left. \times (I_{\{\alpha^\varepsilon(t)=s_{ij}\}} - \nu_j^i(t)I_{\{\bar{\alpha}^\varepsilon(t)=i\}}) \beta_{ij}(t) \right]. \end{aligned}$$

The definition of  $\bar{\alpha}^\varepsilon(\cdot)$  yields that  $\{\bar{\alpha}^\varepsilon(t) = i\} = \{\alpha^\varepsilon(t) \in \mathcal{M}_i\}$ . Thus,

$$\frac{d\eta^\varepsilon(t)}{dt} = 2 \int_0^t \Phi^\varepsilon(t, r) \beta_{ij}(t) \beta_{ij}(r) dr,$$

where  $\Phi^\varepsilon(t, r) = \Phi_1^\varepsilon(t, r) + \Phi_2^\varepsilon(t, r)$  with

$$\begin{aligned} \Phi_1^\varepsilon(t, r) &= P(\alpha^\varepsilon(t) = s_{ij}, \alpha^\varepsilon(r) = s_{ij}) \\ &\quad - \nu_j^i(t) P(\alpha^\varepsilon(t) \in \mathcal{M}_i, \alpha^\varepsilon(r) = s_{ij}), \end{aligned} \tag{5.53}$$

and

$$\begin{aligned} \Phi_2^\varepsilon(t, r) &= -\nu_j^i(r) P(\alpha^\varepsilon(t) = s_{ij}, \alpha^\varepsilon(r) \in \mathcal{M}_i) \\ &\quad + \nu_j^i(r) \nu_j^i(t) P(\alpha^\varepsilon(t) \in \mathcal{M}_i, \alpha^\varepsilon(r) \in \mathcal{M}_i). \end{aligned} \tag{5.54}$$

Note that the Markov property of  $\alpha^\varepsilon(\cdot)$  implies that for  $0 \leq r \leq t$ ,

$$\begin{aligned} P(\alpha^\varepsilon(t) = s_{ij}, \alpha^\varepsilon(r) = s_{ij}) \\ = P(\alpha^\varepsilon(t) = s_{ij} | \alpha^\varepsilon(r) = s_{ij}) P(\alpha^\varepsilon(r) = s_{ij}). \end{aligned}$$

In view of the asymptotic expansion, we have

$$\begin{aligned} P(\alpha^\varepsilon(t) = s_{ij} | \alpha^\varepsilon(r) = s_{ij}) \\ = \nu_j^i(t) \vartheta_{ii}(t, r) + O\left(\varepsilon + \exp\left(-\frac{\kappa_0(t-r)}{\varepsilon}\right)\right). \end{aligned} \tag{5.55}$$

It follows that

$$\begin{aligned}
 & P(\alpha^\varepsilon(t) \in \mathcal{M}_i | \alpha^\varepsilon(r) = s_{ij}) \\
 &= \sum_{k=1}^{m_i} \nu_k^i(t) \vartheta_{ii}(t, r) + O\left(\varepsilon + \exp\left(-\frac{\kappa_0(t-r)}{\varepsilon}\right)\right) \\
 &= \vartheta_{ii}(t, r) + O\left(\varepsilon + \exp\left(-\frac{\kappa_0(t-r)}{\varepsilon}\right)\right).
 \end{aligned} \tag{5.56}$$

Combining (5.55) and (5.56) leads to

$$\Phi_1^\varepsilon(t, r) = O\left(\varepsilon + \exp\left(-\frac{\kappa_0(t-r)}{\varepsilon}\right)\right).$$

Similarly, we can show that

$$\Phi_2^\varepsilon(t, r) = O\left(\varepsilon + \exp\left(-\frac{\kappa_0(t-r)}{\varepsilon}\right)\right),$$

by noting that

$$\begin{aligned}
 \Phi_2^\varepsilon(t, r) &= -\nu_j^i(r) \sum_{k=1}^{m_i} P(\alpha^\varepsilon(t) = s_{ij}, \alpha^\varepsilon(r) = s_{ik}) \\
 &\quad + \nu_j^i(r) \sum_{k=1}^{m_i} \nu_j^i(t) P(\alpha^\varepsilon(t) \in \mathcal{M}_i, \alpha^\varepsilon(r) = s_{ik})
 \end{aligned}$$

and

$$\begin{aligned}
 & P(\alpha^\varepsilon(t) = s_{ij} | \alpha^\varepsilon(r) = s_{ik}) \\
 &= \nu_j^i(t) \vartheta_{ii}(t, r) + O\left(\varepsilon + \exp\left(-\frac{\kappa_0(t-r)}{\varepsilon}\right)\right),
 \end{aligned}$$

for any  $k = 1, \dots, m_i$ . Therefore,

$$\frac{d\eta^\varepsilon(t)}{dt} = 2 \int_0^t O\left(\varepsilon + \exp\left(-\frac{\kappa_0(t-r)}{\varepsilon}\right)\right) dr = O(\varepsilon). \tag{5.57}$$

This together with  $\eta^\varepsilon(0) = 0$  implies that  $\eta^\varepsilon(t) = O(\varepsilon)$ .  $\square$

Theorem 5.25 indicates that  $\nu^k(t)$  together with  $\bar{\alpha}^\varepsilon(\cdot)$  approximates well the Markov chain  $\alpha^\varepsilon(\cdot)$  in an appropriate sense. Nevertheless, in general,  $\{\alpha^\varepsilon(\cdot)\}$  is not tight. The following example provides a simple illustration.

**Example 5.26.** Let  $\alpha^\varepsilon(\cdot) \in \{1, 2\}$  denote a Markov chain generated by

$$\frac{1}{\varepsilon} \begin{pmatrix} -\lambda & \lambda \\ \mu & -\mu \end{pmatrix},$$

for some  $\lambda, \mu > 0$ . Then  $\alpha^\varepsilon(\cdot)$  is *not* tight.

Proof: If  $\alpha^\varepsilon(\cdot)$  is tight, then there exists a sequence  $\varepsilon_k \rightarrow 0$  such that  $\alpha^{\varepsilon_k}(\cdot)$  converges weakly to a stochastic process  $\alpha(\cdot) \in D([0, T]; \mathcal{M})$ . In view of the Skorohod representation (without changing notation for simplicity), Theorem A.11, we may assume  $\alpha^{\varepsilon_k}(\cdot) \rightarrow \alpha(\cdot)$  w.p.1. It follows from Lemma A.41 that

$$E \left| \int_0^t \alpha^{\varepsilon_k}(s) ds - \int_0^t \alpha(s) ds \right|^2 \rightarrow 0,$$

for all  $t \in [0, T]$ . Moreover, similarly as in Theorem 5.25, we obtain

$$E \left| \int_0^t \alpha^{\varepsilon_k}(s) ds - \int_0^t (\nu_1 + 2\nu_2) ds \right|^2 \rightarrow 0,$$

where  $(\nu_1, \nu_2)$  is the stationary distribution of  $\alpha^\varepsilon(\cdot)$  and  $\nu_1 + 2\nu_2$  is the mean with respect to the stationary distribution. As a consequence, it follows that  $\alpha(t) = \nu_1 + 2\nu_2$  for all  $t \in [0, T]$  w.p.1. Let

$$\delta_0 = \min\{|1 - (\nu_1 + 2\nu_2)|, |2 - (\nu_1 + 2\nu_2)|\} > 0.$$

Then for  $t \in [0, T]$ ,

$$|\alpha^\varepsilon(t) - (\nu_1 + 2\nu_2)| \geq \delta_0.$$

Hence, under the Skorohod topology

$$d(\alpha^{\varepsilon_k}(\cdot), \nu_1 + 2\nu_2) \geq \delta_0.$$

This contradicts the fact that  $\alpha^{\varepsilon_k}(\cdot) \rightarrow \alpha(\cdot) = \nu_1 + 2\nu_2$  w.p.1. Therefore,  $\alpha^\varepsilon(\cdot)$  cannot be tight.  $\square$

Although  $\alpha^\varepsilon(\cdot)$  is not tight because it fluctuates in  $\mathcal{M}_k$  very rapidly for small  $\varepsilon$ , its aggregation  $\bar{\alpha}^\varepsilon(\cdot)$  is tight, and converges weakly to  $\bar{\alpha}(t)$ ,  $t \geq 0$ , a Markov chain generated by  $\bar{Q}(t)$ ,  $t \geq 0$ , where  $\bar{Q}(t)$  is defined in (5.48). The next theorem shows that  $\bar{\alpha}^\varepsilon(\cdot)$  can be further approximated by  $\bar{\alpha}(\cdot)$ .

**Theorem 5.27.** *Assume (A5.5) and (A5.6). Then  $\bar{\alpha}^\varepsilon(\cdot)$  converges weakly to  $\bar{\alpha}(\cdot)$  in  $D([0, T]; \bar{\mathcal{M}})$ , as  $\varepsilon \rightarrow 0$ .*

Proof: The proof is divided into two steps. First, we show that  $\bar{\alpha}^\varepsilon(\cdot)$  defined in (5.51) is tight in  $D([0, T]; \bar{\mathcal{M}})$ . The definition of  $\bar{\alpha}^\varepsilon(\cdot)$  implies that

$$\{\bar{\alpha}^\varepsilon(t) = i\} = \{\alpha^\varepsilon(t) \in \mathcal{M}_i\} = \{\alpha^\varepsilon(t) = s_{ij} \text{ for some } j = 1, \dots, m_i\}.$$

Consider the conditional expectation

$$\begin{aligned}
 & E \left[ (\bar{\alpha}^\varepsilon(t+s) - \bar{\alpha}^\varepsilon(s))^2 \mid \alpha^\varepsilon(s) = s_{ij} \right] \\
 &= E \left[ (\bar{\alpha}^\varepsilon(t+s) - i)^2 \mid \alpha^\varepsilon(s) = s_{ij} \right] \\
 &= \sum_{k=1}^l E \left[ (\bar{\alpha}^\varepsilon(t+s) - i)^2 I_{\{\bar{\alpha}^\varepsilon(t+s)=k\}} \mid \alpha^\varepsilon(s) = s_{ij} \right] \\
 &= \sum_{k=1}^l (k-i)^2 P(\bar{\alpha}^\varepsilon(t+s) = k \mid \alpha^\varepsilon(s) = s_{ij}) \\
 &\leq l^2 \sum_{k \neq i} P(\bar{\alpha}^\varepsilon(t+s) = k \mid \alpha^\varepsilon(s) = s_{ij}).
 \end{aligned}$$

Since  $\{\bar{\alpha}^\varepsilon(t+s) = k\} = \{\alpha^\varepsilon(t+s) \in \mathcal{M}_k\}$ , it follows that

$$\begin{aligned}
 & P(\bar{\alpha}^\varepsilon(t+s) = k \mid \alpha^\varepsilon(s) = s_{ij}) \\
 &= \sum_{k_1=1}^{m_k} P(\alpha^\varepsilon(t+s) = s_{kk_1} \mid \alpha^\varepsilon(s) = s_{ij}) \\
 &= \sum_{k_1=1}^{m_k} \nu_{k_1}^k(t+s) \vartheta_{ik}(t+s, s) + O\left(\varepsilon + \exp\left(-\frac{\kappa_0 t}{\varepsilon}\right)\right) \\
 &= \vartheta_{ik}(t+s, s) + O\left(\varepsilon + \exp\left(-\frac{\kappa_0 t}{\varepsilon}\right)\right).
 \end{aligned}$$

Therefore, we obtain

$$\begin{aligned}
 & E \left[ (\bar{\alpha}^\varepsilon(t+s) - \bar{\alpha}^\varepsilon(s))^2 \mid \alpha^\varepsilon(s) = s_{ij} \right] \\
 &\leq l^2 \sum_{k \neq i} \vartheta_{ik}(t+s, s) + O\left(\varepsilon + \exp\left(-\frac{\kappa_0 t}{\varepsilon}\right)\right).
 \end{aligned}$$

Note that  $\lim_{t \rightarrow 0} \vartheta_{ik}(t+s, s) = 0$  for  $i \neq k$ .

$$\lim_{t \rightarrow 0} \left( \lim_{\varepsilon \rightarrow 0} E \left( (\bar{\alpha}^\varepsilon(t+s) - \bar{\alpha}^\varepsilon(s))^2 \mid \alpha^\varepsilon(s) = s_{ij} \right) \right) = 0.$$

Thus, the Markov property of  $\alpha^\varepsilon(\cdot)$  implies

$$\lim_{t \rightarrow 0} \left( \lim_{\varepsilon \rightarrow 0} E \left( (\bar{\alpha}^\varepsilon(t+s) - \bar{\alpha}^\varepsilon(s))^2 \mid \alpha^\varepsilon(r) : r \leq s \right) \right) = 0. \tag{5.58}$$

Recall that  $\bar{\alpha}^\varepsilon(\cdot)$  is bounded. The tightness of  $\bar{\alpha}^\varepsilon(\cdot)$  follows from Kurtz' tightness criterion (see Lemma A.17).

To complete the proof, it remains to show that the finite-dimensional distributions of  $\bar{\alpha}^\varepsilon(\cdot)$  converge to that of  $\bar{\alpha}(\cdot)$ . In fact, for any

$$0 \leq t_1 < t_2 < \dots < t_n \leq T \text{ and } i_1, i_2, \dots, i_n \in \bar{\mathcal{M}} = \{1, \dots, l\},$$

we have

$$\begin{aligned} P(\bar{\alpha}^\varepsilon(t_n) = i_n, \dots, \bar{\alpha}^\varepsilon(t_1) = i_1) &= P(\alpha^\varepsilon(t_n) \in \mathcal{M}_{i_n}, \dots, \alpha^\varepsilon(t_1) \in \mathcal{M}_{i_1}) \\ &= \sum_{j_1, \dots, j_n} P(\alpha^\varepsilon(t_n) = s_{i_n j_n}, \dots, \alpha^\varepsilon(t_1) = s_{i_1 j_1}) \\ &= \sum_{j_1, \dots, j_n} P(\alpha^\varepsilon(t_n) = s_{i_n j_n} | \alpha^\varepsilon(t_{n-1}) = s_{i_{n-1} j_{n-1}}) \\ &\quad \times \dots \times P(\alpha^\varepsilon(t_2) = s_{i_2 j_2} | \alpha^\varepsilon(t_1) = s_{i_1 j_1}) P(\alpha^\varepsilon(t_1) = s_{i_1 j_1}). \end{aligned}$$

In view of Lemma 5.24, for each  $k$ , we have

$$P(\alpha^\varepsilon(t_k) = s_{i_k j_k} | \alpha^\varepsilon(t_{k-1}) = s_{i_{k-1} j_{k-1}}) \rightarrow \nu_{j_k}^{i_k}(t_k) \vartheta_{i_{k-1} i_k}(t_k, t_{k-1}).$$

Moreover, note that

$$\sum_{j_k=1}^{m_{i_k}} \nu_{j_k}^{i_k}(t_k) = 1.$$

It follows that

$$\begin{aligned} &\sum_{j_1, \dots, j_n} P(\alpha^\varepsilon(t_n) = s_{i_n j_n} | \alpha^\varepsilon(t_{n-1}) = s_{i_{n-1} j_{n-1}}) \\ &\quad \times \dots \times P(\alpha^\varepsilon(t_2) = s_{i_2 j_2} | \alpha^\varepsilon(t_1) = s_{i_1 j_1}) P(\alpha^\varepsilon(t_1) = s_{i_1 j_1}) \\ &\rightarrow \sum_{j_1, \dots, j_n} \nu_{j_n}^{i_n}(t_n) \vartheta_{i_{n-1} i_n}(t_n, t_{n-1}) \dots \nu_{j_2}^{i_2}(t_2) \vartheta_{i_1 i_2}(t_2, t_1) \nu_{j_1}^{i_1}(t_1) \tilde{\vartheta}_{i_1}(t_1) \\ &= \vartheta_{i_{n-1} i_n}(t_n, t_{n-1}) \dots \vartheta_{i_1 i_2}(t_2, t_1) \tilde{\vartheta}_{i_1}(t_1) \\ &= P(\bar{\alpha}(t_n) = i_n, \dots, \bar{\alpha}(t_1) = i_1), \end{aligned}$$

where  $\sum_{j_1, \dots, j_n} = \sum_{j_1=1}^{m_{i_1}} \dots \sum_{j_n=1}^{m_{i_n}}$  and  $\tilde{\vartheta}_{i_1}(t_1)$  denotes the initial distribution (also known as absolute probability in the literature of Markov chains). Thus,  $\bar{\alpha}^\varepsilon(\cdot) \rightarrow \bar{\alpha}(\cdot)$  in distribution.  $\square$

This theorem implies that  $\bar{\alpha}^\varepsilon(\cdot)$  converges to a Markov chain, although  $\bar{\alpha}^\varepsilon(\cdot)$  itself is not a Markov chain in general. If, however, the generator  $Q^\varepsilon(t)$



has some specific structure, then  $\bar{\alpha}^\varepsilon(\cdot)$  is a Markov chain. The following example demonstrates this point.

**Example 5.28.** Let  $\tilde{Q}(t) = (\tilde{q}_{ij}(t))$  and  $\bar{Q}(t) = (\bar{q}_{ij}(t))$  denote generators with the corresponding state spaces  $\{a_1, \dots, a_{m_0}\}$  and  $\{1, \dots, l\}$ , respectively. Consider

$$Q^\varepsilon(t) = \frac{1}{\varepsilon} \begin{pmatrix} \tilde{Q}(t) & & \\ & \ddots & \\ & & \tilde{Q}(t) \end{pmatrix} + \begin{pmatrix} \bar{q}_{11}(t)I_{m_0} & \cdots & \bar{q}_{1l}(t)I_{m_0} \\ \vdots & \vdots & \vdots \\ \bar{q}_{l1}(t)I_{m_0} & \cdots & \bar{q}_{ll}(t)I_{m_0} \end{pmatrix}, \quad (5.59)$$

where  $I_{m_0}$  is the  $m_0 \times m_0$  identity matrix. In this case

$$m_1 = m_2 = \cdots = m_l = m_0.$$

Then  $\bar{\alpha}^\varepsilon(\cdot)$  is a Markov chain generated by  $\bar{Q}(t)$ . In fact, let

$$\chi^\varepsilon(t) = \left( I_{\{\alpha^\varepsilon(t)=s_{11}\}}, \dots, I_{\{\alpha^\varepsilon(t)=s_{1m_0}\}}, \dots, I_{\{\alpha^\varepsilon(t)=s_{l1}\}}, \dots, I_{\{\alpha^\varepsilon(t)=s_{lm_0}\}} \right).$$

Note that  $s_{ij} = (i, a_j)$  for  $j = 1, \dots, m_0$  and  $i = 1, \dots, l$ . In view of Lemma 2.4, we obtain that

$$\chi^\varepsilon(t) - \int_0^t \chi^\varepsilon(s) Q^\varepsilon(s) ds \quad (5.60)$$

is a martingale. Postmultiplying (multiplying from the right) (5.60) by

$$\tilde{\mathbb{1}} = \text{diag}(\mathbb{1}_{m_0}, \dots, \mathbb{1}_{m_0})$$

and noting that  $\{\bar{\alpha}^\varepsilon(t) = i\} = \{\alpha^\varepsilon(t) \in \mathcal{M}_i\}$  and

$$\chi^\varepsilon(t) \tilde{\mathbb{1}} = (I_{\{\bar{\alpha}^\varepsilon(t)=1\}}, \dots, I_{\{\bar{\alpha}^\varepsilon(t)=l\}}),$$

we obtain that

$$(I_{\{\bar{\alpha}^\varepsilon(t)=1\}}, \dots, I_{\{\bar{\alpha}^\varepsilon(t)=l\}}) - \int_0^t \chi^\varepsilon(s) Q^\varepsilon(s) ds \tilde{\mathbb{1}}$$

is still a martingale. In view of the special structure of  $Q^\varepsilon(t)$  in (5.59),

$$\tilde{Q}(t) \mathbb{1}_{m_0} = 0, \quad Q^\varepsilon(t) \tilde{\mathbb{1}} = \hat{Q}(t) \tilde{\mathbb{1}},$$

and

$$\chi^\varepsilon(s) \hat{Q}(s) \tilde{\mathbb{1}} = (I_{\{\bar{\alpha}^\varepsilon(s)=1\}}, \dots, I_{\{\bar{\alpha}^\varepsilon(s)=l\}}) \bar{Q}(s).$$

Therefore, (5.60) implies that

$$(I_{\{\bar{\alpha}^\varepsilon(t)=1\}}, \dots, I_{\{\bar{\alpha}^\varepsilon(t)=l\}}) - \int_0^t (I_{\{\bar{\alpha}^\varepsilon(s)=1\}}, \dots, I_{\{\bar{\alpha}^\varepsilon(s)=l\}}) \bar{Q}(s) ds$$

is a martingale. This implies, in view of Lemma 2.4, that  $\bar{\alpha}^\varepsilon(\cdot)$  is a Markov chain generated by  $\bar{Q}(t)$ ,  $t \geq 0$ .

### 5.3.2 Exponential Bounds

For each  $i = 1, \dots, l, j = 1, \dots, m_i, \alpha \in \mathcal{M}$ , and  $t \geq 0$ , let  $\beta_{ij}(t)$  be a bounded, Borel measurable, deterministic function and let

$$W_{ij}(t, \alpha) = \left( I_{\{\alpha = s_{ij}\}} - \nu_j^i(t) I_{\{\alpha \in \mathcal{M}_i\}} \right) \beta_{ij}(t). \tag{5.61}$$

Consider normalized occupation measures

$$n^\varepsilon(t) = \left( n_{11}^\varepsilon(t), \dots, n_{1m_1}^\varepsilon(t), \dots, n_{l1}^\varepsilon(t), \dots, n_{lm_l}^\varepsilon(t) \right),$$

where

$$n_{ij}^\varepsilon(t) = \frac{1}{\sqrt{\varepsilon}} \int_0^t W_{ij}(s, \alpha^\varepsilon(s)) ds.$$

In this section, we establish the exponential error bound for  $n^\varepsilon(\cdot)$ , a sequence of suitably scaled occupation measures for the singularly perturbed Markov chains with weak and strong interactions.

In view of Theorem 4.29, there exists  $\kappa_0 > 0$  such that

$$\left| P^\varepsilon(t, s) - P_0(t, s) \right| = O \left( \varepsilon + \exp \left( -\frac{\kappa_0(t-s)}{\varepsilon} \right) \right). \tag{5.62}$$

Similar to Section 5.3.2, for fixed but otherwise arbitrary  $T > 0$ , let

$$K_T = \max \left\{ 1, \sup_{0 \leq s \leq t \leq T} \left( \frac{|P^\varepsilon(t, s) - P_0(t, s)|}{\varepsilon + \exp(-\kappa_0(t-s)/\varepsilon)} \right) \right\}. \tag{5.63}$$

We may write (5.62) in terms of  $K_T$  and  $O_1(\cdot)$  as follows:

$$\left| P^\varepsilon(t, s) - P_0(t, s) \right| = K_T O_1 \left( \varepsilon + \exp \left( -\frac{\kappa_0(t-s)}{\varepsilon} \right) \right), \tag{5.64}$$

where  $|O_1(y)|/|y| \leq 1$ . The notation of  $K_T$  and  $O_1(\cdot)$  above emphasizes the separation of the dependence of the constant and a “norm 1” function. Essentially,  $K_T$  serves as a magnitude of the bound indicating the size of the bounding region, and the rest is absorbed into the function  $O_1(\cdot)$ .

**Theorem 5.29.** *Assume (A5.5) and (A5.6). Then there exist  $\varepsilon_0 > 0$  and  $K > 0$  such that for  $0 < \varepsilon \leq \varepsilon_0, T \geq 0$ , and for any bounded, Borel measurable, and deterministic process  $\beta_{ij}(\cdot)$ ,*

$$E \exp \left( \frac{\theta_T}{(T+1)^3} \sup_{0 \leq t \leq T} |n^\varepsilon(t)| \right) \leq K, \tag{5.65}$$

where  $\theta_T$  is any constant satisfying

$$0 \leq \theta_T \leq \frac{\min\{1, \kappa_0\}}{K_T |\beta|_T (1 + |\widehat{Q}|_T)}, \tag{5.66}$$

and where  $|\cdot|_T$  denotes the matrix norm as defined in (5.12), that is,

$$|\beta|_T = \max_{i,j} \sup_{0 \leq t \leq T} |\beta_{ij}(t)|,$$

similarly for  $|\widehat{Q}|_T$ .

**Remark 5.30.** This theorem is a natural extension to Theorem 5.4. Owing to the existence of the weak and strong interactions, slightly stronger conditions on  $K_T$  and  $\theta_T$  are made in (5.63) and (5.66). Also the exponential constant in (5.65) is changed to  $(T + 1)^3$ .

Proof of Theorem 5.29: Here the proof is again along the lines of Theorem 5.4. Since Steps 2-5 in the proof are similar to those of Theorem 5.4, we will only give the proof for Step 1.

Let  $\chi^\varepsilon(\cdot)$  denote the vector of indicators corresponding to  $\alpha^\varepsilon(\cdot)$ , that is,

$$\chi^\varepsilon(t) = \left( I_{\{\alpha^\varepsilon(t)=s_{11}\}}, \dots, I_{\{\alpha^\varepsilon(t)=s_{1m_1}\}}, \dots, I_{\{\alpha^\varepsilon(t)=s_{l1}\}}, \dots, I_{\{\alpha^\varepsilon(t)=s_{lm_l}\}} \right).$$

Then  $w^\varepsilon(\cdot)$  defined by

$$w^\varepsilon(t) = \chi^\varepsilon(t) - \chi^\varepsilon(0) - \int_0^t \chi^\varepsilon(s)Q^\varepsilon(s)ds \tag{5.67}$$

is an  $\mathbb{R}^m$ -valued martingale. In fact,  $w^\varepsilon(\cdot)$  is square integrable on  $[0, T]$ . It then follows from a well-known result (see Elliott [55] or Kunita and Watanabe [134]) that a stochastic integral with respect to  $w^\varepsilon(t)$  can be defined. In view of the defining equation (5.67), the linear stochastic differential equation

$$d\chi^\varepsilon(t) = \chi^\varepsilon(t)Q^\varepsilon(t)dt + dw^\varepsilon(t) \tag{5.68}$$

makes sense. Recall that  $P^\varepsilon(t, s)$  is the principal matrix solution of the matrix differential equation

$$\frac{dy(t)}{dt} = y(t)Q^\varepsilon(t). \tag{5.69}$$

The solution of this stochastic differential equation is

$$\begin{aligned} \chi^\varepsilon(t) &= \chi^\varepsilon(0)P^\varepsilon(t, 0) + \int_0^t (dw^\varepsilon(s))P^\varepsilon(t, s) \\ &= \chi^\varepsilon(0)(P^\varepsilon(t, 0) - P_0(t, 0)) \\ &\quad + \int_0^t (dw^\varepsilon(s))(P^\varepsilon(t, s) - P_0(t, s)) \\ &\quad + \chi^\varepsilon(0)P_0(t, 0) + \int_0^t (dw^\varepsilon(s))P_0(t, s). \end{aligned} \tag{5.70}$$

Use  $\vartheta_{ij}(t, s)$  defined in Lemma 5.24 and write  $\Theta(t, s) = (\vartheta_{ij}(t, s))$ . Then it is easy to check that

$$P_0(t, s) = \tilde{\mathbb{1}}\Theta(t, s)\text{diag}(\nu^1(t), \dots, \nu^l(t)). \tag{5.71}$$

Set

$$\bar{\chi}^\varepsilon(t) = (\nu^1(t)I_{\{\bar{\alpha}^\varepsilon(t)=1\}}, \dots, \nu^l(t)I_{\{\bar{\alpha}^\varepsilon(t)=l\}}) \in \mathbb{R}^m$$

and

$$\tilde{\chi}^\varepsilon(t) = (I_{\{\bar{\alpha}^\varepsilon(t)=1\}}, \dots, I_{\{\bar{\alpha}^\varepsilon(t)=l\}}) \in \mathbb{R}^l.$$

Then it follows that

$$\tilde{\chi}^\varepsilon(t) = \chi^\varepsilon(t)\tilde{\mathbb{1}} \quad \text{and} \tag{5.72}$$

$$\bar{\chi}^\varepsilon(t) = \tilde{\chi}^\varepsilon(t)\text{diag}(\nu^1(t), \dots, \nu^l(t)).$$

Moreover, postmultiplying both sides of (5.67) by  $\tilde{\mathbb{1}}$  yields that

$$\chi^\varepsilon(t)\tilde{\mathbb{1}} - \chi^\varepsilon(0)\tilde{\mathbb{1}} - \int_0^t \chi^\varepsilon(s)Q^\varepsilon(s)\tilde{\mathbb{1}}ds = w^\varepsilon(t)\tilde{\mathbb{1}}. \tag{5.73}$$

Here  $w^\varepsilon(\cdot)\tilde{\mathbb{1}}$  is also a square-integrable martingale. Note that  $\tilde{Q}(s)\tilde{\mathbb{1}} = 0$  and hence

$$Q^\varepsilon(s)\tilde{\mathbb{1}} = \hat{Q}(s)\tilde{\mathbb{1}} \quad \text{and}$$

$$\bar{\chi}^\varepsilon(s)\hat{Q}(s)\tilde{\mathbb{1}} = \tilde{\chi}^\varepsilon(s)\text{diag}(\nu^1(s), \dots, \nu^l(s))\hat{Q}(s)\tilde{\mathbb{1}} = \tilde{\chi}^\varepsilon(s)\bar{Q}(s).$$

We obtain from (5.73) that

$$\tilde{\chi}^\varepsilon(t) - \tilde{\chi}^\varepsilon(0) - \int_0^t \left( (\chi^\varepsilon(s) - \bar{\chi}^\varepsilon(s))\hat{Q}(s)\tilde{\mathbb{1}} + \tilde{\chi}^\varepsilon(s)\bar{Q}(s) \right) ds = w^\varepsilon(t)\tilde{\mathbb{1}}.$$

Since  $\Theta(t, s)$  is the principal matrix solution to

$$\frac{d\Theta(t, s)}{dt} = \Theta(t, s)\bar{Q}(t), \quad \text{with } \Theta(s, s) = I,$$

similar to (5.68), solving the stochastic differential equation for  $\tilde{\chi}^\varepsilon(\cdot)$  leads to the equation:

$$\begin{aligned} \tilde{\chi}^\varepsilon(t) &= \tilde{\chi}^\varepsilon(0)\Theta(t, 0) + \int_0^t (dw^\varepsilon(s)\tilde{\mathbb{1}})\Theta(t, s) \\ &\quad + \int_0^t (\chi^\varepsilon(s) - \bar{\chi}^\varepsilon(s))\hat{Q}(s)\tilde{\mathbb{1}}\Theta(t, s)ds. \end{aligned} \tag{5.74}$$

Let us now return to the last two terms in (5.70) and use (5.71), (5.72), and (5.74) to obtain

$$\begin{aligned}
 & \chi^\varepsilon(0)P_0(t, 0) + \int_0^t (dw^\varepsilon(s))P_0(t, s) \\
 &= \left( \chi^\varepsilon(0)\tilde{\mathbb{I}}\Theta(t, 0) + \int_0^t (dw^\varepsilon(s))\tilde{\mathbb{I}}\Theta(t, s) \right) \text{diag}(\nu^1(t), \dots, \nu^l(t)) \\
 &= \left( \tilde{\chi}^\varepsilon(0)\Theta(t, 0) + \int_0^t (dw^\varepsilon(s))\tilde{\mathbb{I}}\Theta(t, s) \right) \text{diag}(\nu^1(t), \dots, \nu^l(t)) \\
 &= \left( \tilde{\chi}^\varepsilon(t) - \int_0^t (\chi^\varepsilon(s) - \bar{\chi}^\varepsilon(s))\widehat{Q}(s)\tilde{\mathbb{I}}\Theta(t, s)ds \right) \text{diag}(\nu^1(t), \dots, \nu^l(t)) \\
 &= \bar{\chi}^\varepsilon(t) - \int_0^t (\chi^\varepsilon(s) - \bar{\chi}^\varepsilon(s))\widehat{Q}(s)\tilde{\mathbb{I}}\Theta(t, s)\text{diag}(\nu^1(t), \dots, \nu^l(t))ds \\
 &= \bar{\chi}^\varepsilon(t) - \int_0^t (\chi^\varepsilon(s) - \bar{\chi}^\varepsilon(s))\widehat{Q}(s)P_0(t, s)ds.
 \end{aligned}$$

Combining this with (5.70), we have

$$(\chi^\varepsilon(t) - \bar{\chi}^\varepsilon(t)) + \int_0^t (\chi^\varepsilon(s) - \bar{\chi}^\varepsilon(s))\widehat{Q}(s)P_0(t, s)ds = \eta^\varepsilon(t), \quad (5.75)$$

where

$$\eta^\varepsilon(t) = \chi^\varepsilon(0)(P^\varepsilon(t, 0) - P_0(t, 0)) + \int_0^t (dw^\varepsilon(s))(P^\varepsilon(t, s) - P_0(t, s)).$$

Note that the matrix  $P^\varepsilon(t, s)$  is invertible but  $P_0(t, s)$  is not. The idea is to approximate the noninvertible matrix  $P_0(t, s)$  by the invertible  $P^\varepsilon(t, s)$ . Let

$$\eta_1^\varepsilon(t) = \int_0^t (\chi^\varepsilon(s) - \bar{\chi}^\varepsilon(s))\widehat{Q}(s)(P_0(t, s) - P^\varepsilon(t, s))ds \quad (5.76)$$

and

$$\phi^\varepsilon(t) = (\chi^\varepsilon(t) - \bar{\chi}^\varepsilon(t)) - (\eta^\varepsilon(t) - \eta_1^\varepsilon(t)).$$

Then  $\phi^\varepsilon(0) = 0$  and  $\phi^\varepsilon(t)$  satisfies the following equation:

$$\phi^\varepsilon(t) + \int_0^t \phi^\varepsilon(s)\widehat{Q}(s)P^\varepsilon(t, s)ds + \int_0^t (\eta^\varepsilon(s) - \eta_1^\varepsilon(s))\widehat{Q}(s)P^\varepsilon(t, s)ds = 0.$$

The properties of the principal matrix solution imply that

$$P^\varepsilon(t, s) = P^\varepsilon(0, s)P^\varepsilon(t, 0).$$

Set

$$\begin{aligned}\check{Q}^\varepsilon(t) &= P^\varepsilon(t, 0)\widehat{Q}(t)P^\varepsilon(0, t), \\ \psi^\varepsilon(t) &= \phi^\varepsilon(t)P^\varepsilon(0, t), \quad \text{and} \\ \eta_2^\varepsilon(t) &= (\eta^\varepsilon(t) - \eta_1^\varepsilon(t))\widehat{Q}(t)P^\varepsilon(0, t).\end{aligned}$$

Owing to the properties of the principal matrix solution, for any  $t \in [0, T]$ , we have

$$P^\varepsilon(0, t)P^\varepsilon(t, 0) = P^\varepsilon(t, t) = I, \tag{5.77}$$

$\psi^\varepsilon(0) = 0$  and  $\psi^\varepsilon(t)$  satisfies the equation

$$\psi^\varepsilon(t) + \int_0^t \psi^\varepsilon(s)\check{Q}^\varepsilon(s)ds + \int_0^t \eta_2^\varepsilon(s)ds = 0.$$

The solution to this equation is given by

$$\psi^\varepsilon(t) = - \int_0^t \eta_2^\varepsilon(s)\check{\Phi}^\varepsilon(t, s)ds, \tag{5.78}$$

where  $\check{\Phi}^\varepsilon(t, s)$  is the principal matrix solution to

$$\frac{d\check{\Phi}^\varepsilon(t, s)}{dt} = -\check{\Phi}^\varepsilon(t, s)\check{Q}^\varepsilon(t), \quad \text{with } \check{\Phi}^\varepsilon(s, s) = I.$$

Postmultiplying both sides of (5.78) by  $P^\varepsilon(t, 0)$  yields

$$\begin{aligned}\phi^\varepsilon(t) &= \psi^\varepsilon(t)P^\varepsilon(t, 0) \\ &= - \int_0^t \eta_2^\varepsilon(s)\check{\Phi}^\varepsilon(t, s)P^\varepsilon(t, 0)ds \\ &= - \int_0^t (\eta^\varepsilon(s) - \eta_1^\varepsilon(s))\widehat{Q}(s)\check{\Psi}^\varepsilon(t, s)ds,\end{aligned}$$

where

$$\check{\Psi}^\varepsilon(t, s) = P^\varepsilon(0, s)\check{\Phi}^\varepsilon(t, s)P^\varepsilon(t, 0).$$

Thus it follows that

$$\chi^\varepsilon(t) - \bar{\chi}^\varepsilon(t) = \eta^\varepsilon(t) - \eta_1^\varepsilon(t) - \int_0^t (\eta^\varepsilon(s) - \eta_1^\varepsilon(s))\widehat{Q}(s)\check{\Psi}^\varepsilon(t, s)ds. \tag{5.79}$$

Again using (5.77), we have

$$\begin{aligned}
 & \frac{d}{dt} (\check{\Phi}^\varepsilon(t, 0)P^\varepsilon(t, 0)) \\
 &= \left( \frac{d\check{\Phi}^\varepsilon(t, 0)}{dt} \right) P^\varepsilon(t, 0) + \check{\Phi}^\varepsilon(t, 0) \left( \frac{dP^\varepsilon(t, 0)}{dt} \right) \\
 &= -\check{\Phi}^\varepsilon(t, 0)\check{Q}^\varepsilon(t)P^\varepsilon(t, 0) + \check{\Phi}^\varepsilon(t, 0)P^\varepsilon(t, 0)Q^\varepsilon(t) \\
 &= -\check{\Phi}^\varepsilon(t, 0)P^\varepsilon(t, 0)\widehat{Q}(t)P^\varepsilon(0, t)P^\varepsilon(t, 0) + \check{\Phi}^\varepsilon(t, 0)P^\varepsilon(t, 0)Q^\varepsilon(t) \\
 &= -\check{\Phi}^\varepsilon(t, 0)P^\varepsilon(t, 0)\widehat{Q}(t) + \check{\Phi}^\varepsilon(t, 0)P^\varepsilon(t, 0)Q^\varepsilon(t) \\
 &= \check{\Phi}^\varepsilon(t, 0)P^\varepsilon(t, 0) \left( -\widehat{Q}(t) + Q^\varepsilon(t) \right) \\
 &= \check{\Phi}^\varepsilon(t, 0)P^\varepsilon(t, 0) \left( \frac{1}{\varepsilon}\widetilde{Q}(t) \right).
 \end{aligned}$$

This implies that  $\check{\Psi}^\varepsilon(t, s)$  is the principal matrix solution to the differential equation

$$\frac{d\check{\Psi}^\varepsilon(t, s)}{dt} = \check{\Psi}^\varepsilon(t, s) \left( \frac{1}{\varepsilon}\widetilde{Q}(t) \right), \quad \text{with } \check{\Psi}^\varepsilon(s, s) = I. \quad (5.80)$$

Therefore, all entries of  $\check{\Psi}^\varepsilon(t, s)$  are bounded below from 0 and bounded above by 1, and these bounds are uniform in  $0 \leq s \leq t \leq T$ . Thus,  $|\check{\Psi}^\varepsilon(t, s)|_T \leq 1$ .

Multiplying both sides of (5.79) by the  $m \times m$  matrix

$$\beta(t) := \text{diag}(\beta_{11}(t), \dots, \beta_{1m_1}(t), \dots, \beta_{l1}(t), \dots, \beta_{lm_l}(t))$$

from the right and integrating over the interval  $[0, \varsigma]$ , for each  $\varsigma \in [0, T]$ , we have

$$\begin{aligned}
 \int_0^\varsigma (\chi^\varepsilon(t) - \bar{\chi}^\varepsilon(t))\beta(t)dt &= \int_0^\varsigma \eta^\varepsilon(t)\beta(t)dt - \int_0^\varsigma \eta_1^\varepsilon(t)\beta(t)dt \\
 &\quad - \int_0^\varsigma \int_0^t (\eta^\varepsilon(s) - \eta_1^\varepsilon(s))\widehat{Q}(s)\check{\Psi}^\varepsilon(t, s)ds\beta(t)dt.
 \end{aligned}$$

By changing the order of integration, we write the last term in the above expression as

$$\begin{aligned}
 & \int_0^\varsigma \int_0^t (\eta^\varepsilon(s) - \eta_1^\varepsilon(s))\widehat{Q}(s)\check{\Psi}^\varepsilon(t, s)ds\beta(t)dt \\
 &= \int_0^\varsigma (\eta^\varepsilon(s) - \eta_1^\varepsilon(s)) \left( \int_s^\varsigma \widehat{Q}(s)\check{\Psi}^\varepsilon(t, s)\beta(t)dt \right) ds.
 \end{aligned}$$

Therefore, it follows that

$$\int_0^\varsigma (\chi^\varepsilon(t) - \bar{\chi}^\varepsilon(t))\beta(t)dt = \int_0^\varsigma \eta^\varepsilon(t)\tilde{\beta}(t)dt - \int_0^\varsigma \eta_1^\varepsilon(t)\tilde{\beta}(t)dt, \quad (5.81)$$

where

$$\tilde{\beta}(t) = \beta(t) + \int_t^\varsigma \widehat{Q}(t)\check{\Psi}^\varepsilon(r,t)\beta(r)dr.$$

Moreover, in view of the fact that  $|\check{\Psi}^\varepsilon(t,s)|_T \leq 1$ , it is easy to see that

$$|\tilde{\beta}|_T \leq (1+T)|\beta|_T(1+|\widehat{Q}|_T). \quad (5.82)$$

Note that  $n^\varepsilon(\cdot)$  can be written in terms of  $\chi^\varepsilon(\cdot)$  and  $\bar{\chi}^\varepsilon(\cdot)$  as

$$n^\varepsilon(\varsigma) = \frac{1}{\sqrt{\varepsilon}} \int_0^\varsigma (\chi^\varepsilon(t) - \bar{\chi}^\varepsilon(t))\beta(t)dt.$$

By virtue of (5.81), it follows that

$$|n^\varepsilon(\varsigma)| \leq \frac{1}{\sqrt{\varepsilon}} \left| \int_0^\varsigma \eta^\varepsilon(t)\tilde{\beta}(t)dt \right| + \frac{1}{\sqrt{\varepsilon}} \left| \int_0^\varsigma \eta_1^\varepsilon(t)\tilde{\beta}(t)dt \right|.$$

Note that in view of the definition of  $\eta_1^\varepsilon(\cdot)$  in (5.76),

$$|\eta_1^\varepsilon(t)| = \int_0^t O\left(\varepsilon + \exp\left(-\frac{\kappa_0(t-s)}{\varepsilon}\right)\right)ds = O(\varepsilon(t+1)).$$

Thus, in view of (5.82),

$$\begin{aligned} \sup_{0 \leq \varsigma \leq T} \left| \int_0^\varsigma \eta_1^\varepsilon(t)\tilde{\beta}(t)dt \right| &= |\tilde{\beta}|_T \sup_{0 \leq \varsigma \leq T} \int_0^\varsigma O(\varepsilon(t+1))dt \\ &= |\tilde{\beta}|_T \sup_{0 \leq \varsigma \leq T} O(\varepsilon(\varsigma^2 + \varsigma)) \\ &= |\tilde{\beta}|_T(T^2 + T)O(\varepsilon) \\ &\leq (1+T)^3|\beta|_T(1+|\widehat{Q}|_T)O(\varepsilon). \end{aligned} \quad (5.83)$$

Thus, in view of (5.63) and (5.66), for some  $\varepsilon_0 > 0$ , and all  $0 < \varepsilon \leq \varepsilon_0$ ,

$$\begin{aligned} &\exp\left(\frac{\theta_T}{(T+1)^3} \sup_{0 \leq \varsigma \leq T} \left| \int_0^\varsigma \frac{\eta_1^\varepsilon(t)\tilde{\beta}(t)}{\sqrt{\varepsilon}}dt \right| \right) \\ &\leq \exp\left(\frac{O(\sqrt{\varepsilon}) \min\{1, \kappa_0\}}{K_T}\right) \\ &\leq \exp\left(O(\sqrt{\varepsilon_0}) \min\{1, \kappa_0\}\right) \leq K. \end{aligned} \quad (5.84)$$



Moreover, using (5.64), as in the proof of Theorem 5.4, we obtain that

$$E \exp \left( \frac{\theta_T}{(T+1)^{\frac{3}{2}}} \sup_{0 \leq t \leq T} \left| \int_0^t \left( \frac{\eta^\varepsilon(s) \tilde{\beta}(s)}{\sqrt{\varepsilon}} \right) ds \right| \right) \leq K, \tag{5.85}$$

for

$$0 \leq \theta_T \leq \frac{\min\{1, \kappa_0\}}{K_T |\beta|_T (1 + |\widehat{Q}|_T)}.$$

Finally, combine (5.81), (5.83), and (5.85) to obtain

$$E \exp \left( \frac{\theta_T}{(T+1)^3} \sup_{0 \leq t \leq T} |n^\varepsilon(t)| \right) \leq K.$$

This completes the proof. □

**Remark 5.31.** It is easily seen that the error bound so obtained has a form similar to that of the martingale inequality. If  $n^\varepsilon(\cdot)$  were a martingale, the inequality would be obtained much more easily since  $\exp(\cdot)$  is a convex function. As in Section 5.2, the error bound is still a measure of “goodness” of approximation. However, one cannot compare the unscaled occupation measures with a deterministic function. A sensible alternative is to use an approximation by the aggregated process that is no longer deterministic. The exponential bounds obtained tell us exactly how closely one can carry out the approximation. It should be particularly useful for many applications in stochastic control problems with Markovian jump disturbances under discounted cost criteria.

The next two corollaries show that the error bound can be improved under additional conditions by having smaller exponential constants, e.g.,  $(T+1)^{3/2}$  or  $(T+1)^{5/2}$  instead of  $(T+1)^3$ .

**Corollary 5.32.** *Assume that the conditions of Theorem 5.29 hold. Let  $\tilde{Q}(t) = (\tilde{q}_{ij}(t))$  and  $\bar{Q}(t) = (\bar{q}_{ij}(t))$  denote generators with the corresponding state spaces  $\{a_1, \dots, a_{m_0}\}$  and  $\{1, \dots, l\}$ , respectively. Consider*

$$Q^\varepsilon(t) = \frac{1}{\varepsilon} \begin{pmatrix} \tilde{Q}(t) & & \\ & \ddots & \\ & & \tilde{Q}(t) \end{pmatrix} + \begin{pmatrix} \bar{q}_{11}(t)I_{m_0} & \cdots & \bar{q}_{1l}(t)I_{m_0} \\ \vdots & \cdots & \vdots \\ \bar{q}_{l1}(t)I_{m_0} & \cdots & \bar{q}_{ll}(t)I_{m_0} \end{pmatrix},$$

where  $I_{m_0}$  is the  $m_0 \times m_0$  identity matrix. Then there exist positive constants  $\varepsilon_0$  and  $K$  such that for  $0 < \varepsilon \leq \varepsilon_0$ , and  $T \geq 0$ ,

$$E \exp \left( \frac{\theta_T}{(T+1)^{\frac{3}{2}}} \sup_{0 \leq t \leq T} |n^\varepsilon(t)| \right) \leq K.$$

Proof: Under the special structure of the generator  $Q^\varepsilon$ , it is easy to see that

$$\widehat{Q}(s)\widetilde{\mathbb{1}} = \widetilde{\mathbb{1}}\overline{Q}(s),$$

where  $\widetilde{\mathbb{1}}$  now takes the form

$$\widetilde{\mathbb{1}} = \text{diag}(\mathbb{1}_{m_0}, \dots, \mathbb{1}_{m_0}).$$

Note that under current conditions on the fast-changing part of the generator  $\widehat{Q}(t)$ ,

$$\nu^1(t) = \nu^2(t) = \dots = \nu^l(t) \text{ and } \text{diag}(\nu^1(t), \dots, \nu^l(t))\widetilde{\mathbb{1}} = I_l,$$

where  $I_l$  denotes the  $l$ -dimensional identity matrix. This together with (5.72) implies that

$$(\chi^\varepsilon(s) - \overline{\chi}^\varepsilon(s))\widehat{Q}(s)\widetilde{\mathbb{1}} = 0.$$

It follows from (5.71) that

$$\int_0^t (\chi^\varepsilon(s) - \overline{\chi}^\varepsilon(s))\widehat{Q}(s)P_0(t, s)ds = 0.$$

Then (5.75) becomes

$$\chi^\varepsilon(t) - \overline{\chi}^\varepsilon(t) = \eta^\varepsilon(t).$$

The rest of the proof follows exactly that of Theorem 5.29. □

**Corollary 5.33.** *Assume the conditions of Theorem 5.29. Suppose  $\widetilde{Q}(t) = \widetilde{Q}$  and  $\widehat{Q}(t) = \widehat{Q}$  for some constant matrices  $\widetilde{Q}$  and  $\widehat{Q}$ . Then there exist positive constants  $\varepsilon_0$  and  $K$  such that for  $0 < \varepsilon \leq \varepsilon_0$ , and  $T \geq 0$ ,*

$$E \exp \left( \frac{\theta_T}{(T+1)^{\frac{5}{2}}} \sup_{0 \leq t \leq T} |n^\varepsilon(t)| \right) \leq K.$$

**Remark 5.34.** Note that in view of Corollary 4.31, one can show under the condition  $\widetilde{Q}(t) = \widetilde{Q}$  and  $\widehat{Q}(t) = \widehat{Q}$  that there exists a constant  $K$  such that

$$P^\varepsilon(t, s) - P_0(t, s) = K(T+1)O_1 \left( \varepsilon + \exp \left( -\frac{\kappa_0(t-s)}{\varepsilon} \right) \right).$$

In this case,  $\theta_T$  can be taken as

$$0 \leq \theta_T \leq \frac{\min\{1, \kappa_0\}}{K(T+1)|\beta|_T(1+|\widehat{Q}|_T)}.$$

That is, compared with the general result, the constant  $K_T$  can be further specified as  $K_T = K(T+1)$ .

Proof of Corollary 5.33: Note that when the generators are time independent, the quasi-stationary distribution  $\nu^i(t)$  is also independent of time and is denoted by  $\nu^i$ . In this case, the argument from (5.75) to (5.80) can be replaced by the following. Let

$$\check{Q}_0 = \widehat{Q}\widetilde{\mathbb{1}}\text{diag}(\nu^1, \dots, \nu^l).$$

Then it can be shown that

$$\widehat{Q}\widetilde{\mathbb{1}}(\overline{Q})^k \text{diag}(\nu^1, \dots, \nu^l) = (\check{Q}_0)^{k+1}, \text{ for } k \geq 0.$$

This implies that

$$\begin{aligned} \widehat{Q}P_0(t, s) &= \widehat{Q}\widetilde{\mathbb{1}}\exp(\overline{Q}(t-s)) \text{diag}(\nu^1, \dots, \nu^l) \\ &= \check{Q}_0 \exp(\check{Q}_0(t-s)). \end{aligned}$$

Let  $\phi^\varepsilon(t) = (\chi^\varepsilon(t) - \overline{\chi}^\varepsilon(t)) - \eta^\varepsilon(t)$ . Then  $\phi^\varepsilon(\cdot)$  satisfies the equation

$$\phi^\varepsilon(t) + \int_0^t (\phi^\varepsilon(s) + \eta^\varepsilon(s))\check{Q}_0 \exp(\check{Q}_0(t-s))ds = 0.$$

Solving for  $\phi^\varepsilon(\cdot)$ , we obtain

$$\phi^\varepsilon(t) = - \int_0^t \eta^\varepsilon(s)\check{Q}_0 ds.$$

Writing  $\chi^\varepsilon(t) - \overline{\chi}^\varepsilon(t)$  in terms of  $\phi^\varepsilon(t)$  and  $\eta^\varepsilon(t)$  yields,

$$\chi^\varepsilon(t) - \overline{\chi}^\varepsilon(t) = \eta^\varepsilon(t) - \int_0^t \eta^\varepsilon(s)\check{Q}_0 ds.$$

The rest of the proof follows that of Theorem 5.29. □

Similar to Section 5.2, we derive estimates that are analogous to Corollary 5.7 and Corollary 5.8. The details are omitted, however.

### 5.3.3 Asymptotic Distributions

In Section 5.2, we obtained a central limit theorem for a class of Markov chains generated by  $Q^\varepsilon(t) = Q(t)/\varepsilon + \check{Q}(t)$  with a weakly irreducible  $Q(t)$ . In this case for sufficiently small  $\varepsilon > 0$ ,  $Q^\varepsilon(t)$  is weakly irreducible. What, if anything, can be said about the weak and strong interaction models, when  $\check{Q}(t)$  is not weakly irreducible? Is there a central limit theorem for the corresponding occupation measure when one has a singularly perturbed Markov chain with weak and strong interactions? This section deals with

such an issue; our interest lies in the asymptotic distribution as  $\varepsilon \rightarrow 0$ . It is shown that the asymptotic distribution of the corresponding occupation measure can be obtained. However, the limit distribution is no longer Gaussian, but a Gaussian mixture, and the proof is quite different from that of the irreducible case in Section 5.2.

For each  $i = 1, \dots, l, j = 1, \dots, m_i, \alpha \in \mathcal{M}$ , and  $t \geq 0$ , let  $\beta_{ij}(t)$  be a bounded Borel measurable deterministic function. Use  $W_{ij}(t, \alpha)$  defined in (5.61) and the normalized occupation measure

$$n^\varepsilon(t) = (n_{11}^\varepsilon(t), \dots, n_{1m_1}^\varepsilon(t), \dots, n_{l1}^\varepsilon(t), \dots, n_{lm_l}^\varepsilon(t)),$$

with

$$n_{ij}^\varepsilon(t) = \frac{1}{\sqrt{\varepsilon}} \int_0^t W_{ij}(s, \alpha^\varepsilon(s)) ds.$$

We will show in this section that  $n^\varepsilon(\cdot)$  converges weakly to a switching diffusion modulated by  $\bar{\alpha}(\cdot)$ . The procedure is as follows:

- (a) Show that  $(n^\varepsilon(\cdot), \bar{\alpha}^\varepsilon(\cdot))$  is tight;
- (b) verify that the limit of a subsequence of  $(n^\varepsilon(\cdot), \bar{\alpha}^\varepsilon(\cdot))$  is a solution to a martingale problem that has a unique solution;
- (c) characterize the solution of the associated martingale problem;
- (d) construct a switching diffusion that is also a solution to the martingale problem and therefore the limit of  $(n^\varepsilon(\cdot), \bar{\alpha}^\varepsilon(\cdot))$ .

To accomplish our goal, these steps are realized by proving a series of lemmas. Recall that  $\mathcal{F}_t^\varepsilon = \sigma\{\alpha^\varepsilon(s) : 0 \leq s \leq t\}$  denotes the filtration generated by  $\alpha^\varepsilon(\cdot)$ . The lemma below is on the order estimates of the conditional moments, and is useful for getting the tightness result in what follows.

**Lemma 5.35.** *Assume (A5.5) and (A5.6). Then for all  $0 \leq s \leq t \leq T$  and  $\varepsilon$  small enough, the following hold:*

- (a)  $\sup_{s \leq t \leq T} E[n^\varepsilon(t) - n^\varepsilon(s) | \mathcal{F}_s^\varepsilon] = O(\sqrt{\varepsilon});$
- (b)  $\sup_\varepsilon E[|n^\varepsilon(t) - n^\varepsilon(s)|^2 | \mathcal{F}_s^\varepsilon] = O(t - s).$

Proof: First, note that for any fixed  $i, j$ ,

$$E[(n_{ij}^\varepsilon(t) - n_{ij}^\varepsilon(s)) | \mathcal{F}_s^\varepsilon] = \frac{1}{\sqrt{\varepsilon}} \int_s^t E[W_{ij}(r, \alpha^\varepsilon(r)) | \mathcal{F}_s^\varepsilon] dr.$$

Moreover, in view of the definition of  $W_{ij}(t, \alpha)$  and the Markov property, we have, for  $0 \leq s \leq r$ ,

$$\begin{aligned}
 & E[W_{ij}(r, \alpha^\varepsilon(r)) | \mathcal{F}_s^\varepsilon] \\
 &= E \left[ \left( I_{\{\alpha^\varepsilon(r) = s_{ij}\}} - \nu_j^i(r) I_{\{\alpha^\varepsilon(r) \in \mathcal{M}_i\}} \right) \middle| \mathcal{F}_s^\varepsilon \right] \beta_{ij}(r) \\
 &= \left( P(\alpha^\varepsilon(r) = s_{ij} | \mathcal{F}_s^\varepsilon) - \nu_j^i(r) P(\alpha^\varepsilon(r) \in \mathcal{M}_i | \mathcal{F}_s^\varepsilon) \right) \beta_{ij}(r) \\
 &= \left( P(\alpha^\varepsilon(r) = s_{ij} | \alpha^\varepsilon(s)) - \nu_j^i(r) P(\alpha^\varepsilon(r) \in \mathcal{M}_i | \alpha^\varepsilon(s)) \right) \beta_{ij}(r).
 \end{aligned}$$

In view of Lemma 5.24, in particular, similar to (5.55) and (5.56), for all  $i_0 = 1, \dots, l$  and  $j_0 = 1, \dots, m_{i_0}$ ,

$$\begin{aligned}
 P(\alpha^\varepsilon(r) = s_{ij} | \alpha^\varepsilon(s) = s_{i_0 j_0}) - \nu_j^i(r) P(\alpha^\varepsilon(r) \in \mathcal{M}_i | \alpha^\varepsilon(s) = s_{i_0 j_0}) \\
 = O \left( \varepsilon + \exp \left( -\frac{\kappa_0(r-s)}{\varepsilon} \right) \right).
 \end{aligned}$$

Thus owing to Lemma A.42, we have

$$\begin{aligned}
 & \left( P(\alpha^\varepsilon(r) = s_{ij} | \alpha^\varepsilon(s)) - \nu_j^i(r) P(\alpha^\varepsilon(r) \in \mathcal{M}_i | \alpha^\varepsilon(s)) \right) \beta_{ij}(r) \\
 &= \sum_{i_0=1}^l \sum_{j_0=1}^{m_{i_0}} I_{\{\alpha^\varepsilon(s) = s_{i_0 j_0}\}} \left( P(\alpha^\varepsilon(r) = s_{ij} | \alpha^\varepsilon(s) = s_{i_0 j_0}) \right. \\
 &\quad \left. - \nu_j^i(r) P(\alpha^\varepsilon(r) \in \mathcal{M}_i | \alpha^\varepsilon(s) = s_{i_0 j_0}) \right) \beta_{ij}(r) \\
 &= O \left( \varepsilon + \exp \left( -\frac{\kappa_0(r-s)}{\varepsilon} \right) \right).
 \end{aligned}$$

Note also that

$$\frac{1}{\sqrt{\varepsilon}} \int_s^t O \left( \varepsilon + \exp \left( -\frac{\kappa_0(r-s)}{\varepsilon} \right) \right) dr = O(\sqrt{\varepsilon}).$$

This implies (a).

To verify (b), fix and suppress  $i, j$  and define

$$\eta^\varepsilon(t) = E \left[ \left( \int_s^t W_{ij}(r, \alpha^\varepsilon(r)) dr \right)^2 \middle| \mathcal{F}_s^\varepsilon \right].$$

Then by the definition of  $n_{ij}(\cdot)$ ,

$$E \left[ \left( n_{ij}^\varepsilon(t) - n_{ij}^\varepsilon(s) \right)^2 \middle| \mathcal{F}_s^\varepsilon \right] = \frac{\eta^\varepsilon(t)}{\varepsilon}. \quad (5.86)$$

In accordance with the definition of  $\bar{\alpha}^\varepsilon(\cdot)$ ,  $\bar{\alpha}^\varepsilon(t) = i$  iff  $\alpha^\varepsilon(t) \in \mathcal{M}_i$ . In what follows, we use  $\alpha^\varepsilon(t) \in \mathcal{M}_i$  and  $\bar{\alpha}^\varepsilon(t) = i$  interchangeably. Set

$$\begin{aligned} \Psi_1^\varepsilon(t, r) &= I_{\{\alpha^\varepsilon(r)=s_{ij}\}} I_{\{\alpha^\varepsilon(t)=s_{ij}\}} - \nu_j^i(t) I_{\{\alpha^\varepsilon(r)=s_{ij}\}} I_{\{\bar{\alpha}^\varepsilon(t)=i\}}, \\ \Psi_2^\varepsilon(t, r) &= -\nu_j^i(r) I_{\{\bar{\alpha}^\varepsilon(r)=i\}} I_{\{\alpha^\varepsilon(t)=s_{ij}\}} + \nu_j^i(r) \nu_j^i(t) I_{\{\bar{\alpha}^\varepsilon(r)=i\}} I_{\{\bar{\alpha}^\varepsilon(t)=i\}}. \end{aligned}$$

Then as in the proof of Theorem 5.25,

$$\frac{d\eta^\varepsilon(t)}{dt} = 2 \int_s^t E[\Psi_1^\varepsilon(t, r) + \Psi_2^\varepsilon(t, r) | \mathcal{F}_s^\varepsilon] \beta_{ij}(r) \beta_{ij}(t) dr.$$

Using Lemma 5.24, we obtain

$$\begin{aligned} E[\Psi_1^\varepsilon(t, r) | \alpha^\varepsilon(s) = s_{i_0 j_0}] &= O\left(\varepsilon + \exp\left(-\frac{\kappa_0(t-r)}{\varepsilon}\right)\right), \\ E[\Psi_2^\varepsilon(t, r) | \alpha^\varepsilon(s) = s_{i_0 j_0}] &= O\left(\varepsilon + \exp\left(-\frac{\kappa_0(t-r)}{\varepsilon}\right)\right), \end{aligned}$$

for all  $i_0 = 1, \dots, l$  and  $j_0 = 1, \dots, m_{i_0}$ . Then from Lemma A.42, we obtain

$$\begin{aligned} E[\Psi_1^\varepsilon(t, r) | \mathcal{F}_s^\varepsilon] &= O\left(\varepsilon + \exp\left(-\frac{\kappa_0(t-r)}{\varepsilon}\right)\right), \\ E[\Psi_2^\varepsilon(t, r) | \mathcal{F}_s^\varepsilon] &= O\left(\varepsilon + \exp\left(-\frac{\kappa_0(t-r)}{\varepsilon}\right)\right). \end{aligned}$$

As a consequence, we have

$$\frac{d\eta^\varepsilon(t)}{dt} = O(\varepsilon).$$

Integrating both sides over  $[s, t]$  and recalling  $\eta^\varepsilon(s) = 0$  yields

$$\frac{\eta^\varepsilon(t)}{\varepsilon} = O(t - s).$$

This completes the proof of the lemma. □

The next lemma is concerned with the tightness of  $\{(n^\varepsilon(\cdot), \bar{\alpha}^\varepsilon(\cdot))\}$ .

**Lemma 5.36.** *Assume (A5.5) and (A5.6). Then  $\{(n^\varepsilon(\cdot), \bar{\alpha}^\varepsilon(\cdot))\}$  is tight in  $D([0, T]; \mathbb{R}^m \times \overline{\mathcal{M}})$ .*

Proof: The proof uses Lemma A.17. We first verify that the condition given in Remark A.18 holds. To this end, note that  $0 \leq \bar{\alpha}^\varepsilon(t) \leq l$  for all  $t \in [0, T]$ . Moreover, by virtue of Theorem 5.25, for each  $\delta > 0$  and each rational  $t \geq 0$ ,

$$\begin{aligned} \inf_{\varepsilon} P(|n^\varepsilon(t)| \leq K_{t,\delta}) &= \inf_{\varepsilon} [1 - P(|n^\varepsilon(t)| \geq K_{t,\delta})] \\ &\geq \inf_{\varepsilon} \left( 1 - \frac{E|n^\varepsilon(t)|^2}{K_{t,\delta}^2} \right) \\ &\geq 1 - \frac{Kt}{K_{t,\delta}^2}, \end{aligned}$$

where the last inequality is due to Theorem 5.25. Thus if we choose  $K_{t,\delta} > \sqrt{KT/\delta}$ , (A.6) will follow.

It follows from Lemma 5.35 and (5.58) that for all  $t \in [0, T]$ ,

$$\begin{aligned} \lim_{\Delta \rightarrow 0} \left\{ \limsup_{\varepsilon \rightarrow 0} \left( \sup_{0 \leq s \leq \Delta} E \left\{ E \left[ |n_{ij}^\varepsilon(t+s) - n_{ij}^\varepsilon(t)|^2 | \mathcal{F}_t^\varepsilon \right] \right\} \right) \right\} &= 0, \\ \lim_{\Delta \rightarrow 0} \left\{ \limsup_{\varepsilon \rightarrow 0} \left( \sup_{0 \leq s \leq \Delta} E \left\{ E \left[ |\bar{\alpha}^\varepsilon(t+s) - \bar{\alpha}^\varepsilon(t)|^2 | \mathcal{F}_t^\varepsilon \right] \right\} \right) \right\} &= 0. \end{aligned} \tag{5.87}$$

Using (5.86) and (5.87), Theorem A.17 yields the desired result. □

The tightness of  $(n^\varepsilon(\cdot), \bar{\alpha}^\varepsilon(\cdot))$  and Prohorov’s theorem allow one to extract convergent subsequences. We next show that the limit of such a subsequence is uniquely determined in distribution. An equivalent statement is that the associated martingale problem has a unique solution. The following lemma is a generalization of Theorem 5.25 and is needed for proving such a uniqueness property.

**Lemma 5.37.** *Let  $\xi(t, x)$  be a real-valued function that is Lipschitz in  $(t, x) \in \mathbb{R}^{m+1}$ . Then*

$$\sup_{0 \leq \varsigma \leq T} E \left| \int_0^\varsigma W_{ij}(s, \alpha^\varepsilon(s)) \xi(s, n^\varepsilon(s)) ds \right|^2 \rightarrow 0,$$

where  $W_{ij}(t, \alpha) = (I_{\{\alpha = s_{ij}\}} - \nu_j^i(t) I_{\{\alpha \in \mathcal{M}_i\}}) \beta_{ij}(t)$  as defined in (5.61).

**Remark 5.38.** This lemma indicates that the weighted occupation measure (with weighting function  $\xi(t, \alpha^\varepsilon(t))$ ) defined above goes to zero in mean square uniformly in  $t \in [0, \varsigma]$ . If  $\xi(\cdot)$  were a bounded and measurable deterministic function not depending on  $\alpha^\varepsilon(\cdot)$  or  $n^\varepsilon(\cdot)$ , this assertion would follow from Theorem 5.25 easily. In the current situation, it is a function of  $n^\varepsilon(\cdot)$  and therefore a function of  $\alpha^\varepsilon(\cdot)$ , which results in much

of the difficulty. Intuitively, if we can “separate” the functions  $W_{ij}(\cdot)$  and  $\xi(\cdot)$  in the sense treating  $\xi(\cdot)$  as deterministic, then Theorem 5.25 can be applied to obtain the desired limit. To do so, subdivide the interval  $[0, \varsigma]$  into small intervals so that on each of the small intervals, the two functions can be separated. To be more specific, on each partitioned interval, use a piecewise-constant function to approximate  $\xi(\cdot)$ , and show that the error goes to zero. In this process, the Lipschitz condition of  $\xi(t, x)$  plays a crucial role.

Proof of Lemma 5.37: For  $0 < \delta < 1$  and  $0 < \varsigma \leq T$ , let  $N = \lceil \varsigma/\varepsilon^{1-\delta} \rceil$ . Use a partition of  $[0, \varsigma]$  given by

$$[t_0, t_1] \cup [t_1, t_2] \cup \cdots \cup [t_N, t_{N+1}]$$

of  $[0, \varsigma]$ , where  $t_k = \varepsilon^{1-\delta}k$  for  $k = 0, 1, \dots, N$  and  $t_{N+1} = \varsigma$ . Consider a piecewise-constant function

$$\tilde{\xi}(t) = \begin{cases} \xi(0, n^\varepsilon(0)), & \text{if } 0 \leq t < t_1, \\ \xi(t_{k-1}, n^\varepsilon(t_{k-1})), & \text{if } t_{k-1} \leq t < t_k, \quad k = 2, \dots, N, \\ \xi(t_{N-1}, n^\varepsilon(t_{N-1})), & \text{if } t = t_N. \end{cases}$$

Let  $W_{ij}^\varepsilon(t) = W_{ij}(t, \alpha^\varepsilon(t))$ . Then

$$\begin{aligned} E \left| \int_0^\varsigma W_{ij}^\varepsilon(t) \xi(t, n^\varepsilon(t)) dt \right|^2 \\ \leq 2E \left| \int_0^\varsigma W_{ij}^\varepsilon(t) |\xi(t, n^\varepsilon(t)) - \tilde{\xi}(t)| dt \right|^2 + 2E \left| \int_0^\varsigma W_{ij}^\varepsilon(t) \tilde{\xi}(t) dt \right|^2. \end{aligned} \tag{5.88}$$

We now estimate the first term on the second line above. In view of the Cauchy inequality and the boundedness of  $W_{ij}^\varepsilon(t)$ , it follows, for  $0 \leq \varsigma \leq T$ , that

$$\begin{aligned} E \left| \int_0^\varsigma W_{ij}^\varepsilon(t) |\xi(t, n^\varepsilon(t)) - \tilde{\xi}(t)| dt \right|^2 &\leq TE \int_0^\varsigma (\xi(t, n^\varepsilon(t)) - \tilde{\xi}(t))^2 dt \\ &= T \int_0^\varsigma E(\xi(t, n^\varepsilon(t)) - \tilde{\xi}(t))^2 dt. \end{aligned}$$

Note that Theorem 5.25 implies

$$E|n^\varepsilon(t)|^2 \leq K,$$

for a positive constant  $K$  and for all  $t \in [0, T]$ . Therefore, in view of the Lipschitz condition of  $\xi(\cdot)$ , we have

$$E|\xi(t, n^\varepsilon(t))| \leq K(1 + E|n^\varepsilon(t)|) \leq K(1 + (E|n^\varepsilon(t)|^2)^{\frac{1}{2}}) = O(1).$$



Noting that  $t_2 = 2\varepsilon^{1-\delta} = O(\varepsilon^{1-\delta})$ , it follows that

$$\begin{aligned} & \int_0^\varsigma E(\xi(t, n^\varepsilon(t)) - \tilde{\xi}(t))^2 dt \\ &= \sum_{k=2}^N \int_{t_k}^{t_{k+1}} E(\xi(t, n^\varepsilon(t)) - \tilde{\xi}(t))^2 dt + O(\varepsilon^{1-\delta}). \end{aligned}$$

Using the definition of  $\tilde{\xi}(t)$ , the Lipschitz property of  $\xi(t, x)$  in  $(t, x)$ , the choice of the partition of  $[0, \varsigma]$ , and Lemma 5.35, we have

$$\begin{aligned} & \sum_{k=2}^N \int_{t_k}^{t_{k+1}} E(\xi(t, n^\varepsilon(t)) - \tilde{\xi}(t))^2 dt \\ &= \sum_{k=2}^N \int_{t_k}^{t_{k+1}} E(\xi(t, n^\varepsilon(t)) - \xi(t_{k-1}, n^\varepsilon(t_{k-1})))^2 dt \\ &\leq 2 \sum_{k=2}^N \int_{t_k}^{t_{k+1}} K \left( (t - t_{k-1})^2 + E|n^\varepsilon(t) - n^\varepsilon(t_{k-1})|^2 \right) dt \\ &\leq 2 \sum_{k=2}^N \int_{t_k}^{t_{k+1}} K \left( (t - t_{k-1})^2 + O(t - t_{k-1}) \right) dt \\ &= 2 \sum_{k=2}^N \int_{t_k}^{t_{k+1}} O(\varepsilon^{1-\delta}) dt = O(\varepsilon^{1-\delta}). \end{aligned}$$

Let us estimate the second term on the second line in (5.88). Set

$$\tilde{\eta}^\varepsilon(t) = E \left( \int_0^t W_{ij}^\varepsilon(s) \tilde{\xi}(s) ds \right)^2.$$

Then the derivative of  $\tilde{\eta}^\varepsilon(t)$  is given by

$$\begin{aligned} \frac{d\tilde{\eta}^\varepsilon(t)}{dt} &= 2E \int_0^t W_{ij}^\varepsilon(s) \tilde{\xi}(s) W_{ij}^\varepsilon(t) \tilde{\xi}(t) ds \\ &= 2 \int_0^t E \left( W_{ij}^\varepsilon(s) \tilde{\xi}(s) W_{ij}^\varepsilon(t) \tilde{\xi}(t) \right) ds. \end{aligned}$$

For  $0 \leq t \leq t_2$ , in view of the Lipschitz property and Theorem 5.25, we obtain

$$\begin{aligned} \int_0^{t_2} E \left( W_{ij}^\varepsilon(s) \tilde{\xi}(s) W_{ij}^\varepsilon(t) \tilde{\xi}(t) \right) ds &\leq \int_0^{t_2} E \left( |\tilde{\xi}(s)| \cdot |\tilde{\xi}(t)| \right) ds \\ &\leq \int_0^{t_2} (E|\tilde{\xi}(s)|^2)^{\frac{1}{2}} (E|\tilde{\xi}(t)|^2)^{\frac{1}{2}} ds \\ &= O(t_2) = O(\varepsilon^{1-\delta}). \end{aligned}$$

If  $t_k \leq t < t_{k+1}$ , for  $k = 2, \dots, N$ , then using the same argument gives us

$$\begin{aligned} & \int_{t_{k-1}}^t E \left( W_{ij}^\varepsilon(s) \tilde{\xi}(s) W_{ij}^\varepsilon(t) \tilde{\xi}(t) \right) ds \\ &= O(t - t_{k-1}) = O(t_{k+1} - t_{k-1}) = O(\varepsilon^{1-\delta}) \end{aligned}$$

and

$$\frac{d\tilde{\eta}^\varepsilon(t)}{dt} = 2 \int_0^{t_{k-1}} E \left( W_{ij}^\varepsilon(s) \tilde{\xi}(s) W_{ij}^\varepsilon(t) \tilde{\xi}(t) \right) ds + O(\varepsilon^{1-\delta}).$$

Recall that  $\mathcal{F}_t^\varepsilon = \sigma\{\alpha^\varepsilon(s) : 0 \leq s \leq t\}$ . For  $s \leq t_{k-1} < t_k \leq t < t_{k+1}$ ,

$$\begin{aligned} & E \left( W_{ij}^\varepsilon(s) \tilde{\xi}(s) W_{ij}^\varepsilon(t) \tilde{\xi}(t) \right) \\ &= E \left( W_{ij}^\varepsilon(s) \tilde{\xi}(s) E[W_{ij}^\varepsilon(t) \tilde{\xi}(t) | \mathcal{F}_{t_{k-1}}] \right). \end{aligned} \tag{5.89}$$

Moreover, in view of the definition of  $\tilde{\xi}(\cdot)$  and the proof of Lemma 5.35, we have for some  $\kappa_0 > 0$ ,

$$\begin{aligned} E[W_{ij}^\varepsilon(t) \tilde{\xi}(t) | \mathcal{F}_{t_{k-1}}] &= \tilde{\xi}(t) E[W_{ij}^\varepsilon(t) | \mathcal{F}_{t_{k-1}}] \\ &= \tilde{\xi}(t) O \left( \varepsilon + \exp \left( -\frac{\kappa_0(t - t_{k-1})}{\varepsilon} \right) \right) \\ &= \tilde{\xi}(t) O \left( \varepsilon + \exp \left( -\frac{\kappa_0(t_k - t_{k-1})}{\varepsilon} \right) \right) \\ &= \tilde{\xi}(t) O \left( \varepsilon + \exp \left( -\frac{\kappa_0}{\varepsilon^\delta} \right) \right) = \tilde{\xi}(t) O(\varepsilon). \end{aligned}$$

Combine this with (5.89) to obtain

$$E \left( W_{ij}^\varepsilon(s) \tilde{\xi}(s) W_{ij}^\varepsilon(t) \tilde{\xi}(t) \right) = O(\varepsilon) E|\tilde{\xi}(s) \tilde{\xi}(t)| = O(\varepsilon).$$

Therefore,

$$\frac{d\tilde{\eta}^\varepsilon(t)}{dt} = O(\varepsilon^{1-\delta})$$

uniformly on  $[0, T]$ , which implies, together with  $\tilde{\eta}^\varepsilon(0) = 0$ , that

$$\sup_{0 \leq s \leq T} \tilde{\eta}^\varepsilon(s) = \sup_{0 \leq s \leq T} \int_0^s \left( \frac{d\tilde{\eta}^\varepsilon(t)}{dt} \right) dt = O(\varepsilon^{1-\delta}).$$

This completes the proof. □

To characterize the limit of  $(n^\varepsilon(\cdot), \bar{\alpha}^\varepsilon(\cdot))$ , consider the martingale problem associated with  $(n^\varepsilon(\cdot), \bar{\alpha}^\varepsilon(\cdot))$ . Note that

$$\frac{dn^\varepsilon(t)}{dt} = \frac{1}{\sqrt{\varepsilon}} W(t, \alpha^\varepsilon(t)) \text{ and } n^\varepsilon(0) = 0,$$

where

$$W(t, \alpha) = (W_{11}(t, \alpha), \dots, W_{1m_1}(t, \alpha), \dots, W_{l1}(t, \alpha), \dots, W_{lm_l}(t, \alpha)).$$

Let  $\mathcal{G}^\varepsilon(t)$  be the operator

$$\begin{aligned} \mathcal{G}^\varepsilon(t)f(t, x, \alpha) &= \frac{\partial}{\partial t}f(t, x, \alpha) + \frac{1}{\sqrt{\varepsilon}}\langle W(t, \alpha), \nabla_x f(t, x, \alpha) \rangle \\ &\quad + Q^\varepsilon(t)f(t, x, \cdot)(\alpha), \end{aligned}$$

for all  $f(\cdot, \cdot, \alpha) \in C^{1,1}$ , where  $\nabla_x$  denotes the gradient with respect to  $x$  and  $\langle \cdot, \cdot \rangle$  denotes the usual inner product in Euclidean space. It is well known that (see Davis [41, Chapter 2])

$$f(t, n^\varepsilon(t), \alpha^\varepsilon(t)) - \int_0^t \mathcal{G}^\varepsilon(s)f(s, n^\varepsilon(s), \alpha^\varepsilon(s))ds \tag{5.90}$$

is a martingale.

We use the perturbed test function method (see Ethier and Kurtz [59] and Kushner [139]) to study the limit as  $\varepsilon \rightarrow 0$ . To begin with, we define a functional space on  $\mathbb{R}^m \times \overline{\mathcal{M}}$

$$\begin{aligned} C_L^2 &= \left\{ f^0(x, i) : \text{with bounded derivatives up to the} \right. \\ &\quad \left. \text{second order such that the second derivative is Lipschitz} \right\}. \end{aligned} \tag{5.91}$$

For any real-valued function  $f^0(\cdot, i) \in C_L^2$ , define

$$\bar{f}(x, \alpha) = \sum_{i=1}^l f^0(x, i)I_{\{\alpha \in \mathcal{M}_i\}} = \begin{cases} f^0(x, 1), & \text{if } \alpha \in \mathcal{M}_1, \\ \vdots & \vdots \\ f^0(x, l), & \text{if } \alpha \in \mathcal{M}_l, \end{cases}$$

and consider the function

$$f(t, x, \alpha) = \bar{f}(x, \alpha) + \sqrt{\varepsilon}h(t, x, \alpha), \tag{5.92}$$

where  $h(t, x, \alpha)$  is to be specified later. The main idea is that by appropriate choice of  $h(\cdot)$ , the perturbation is small and results in the desired cancelation in the calculation.

In view of the block-diagonal structure of  $\tilde{Q}(t)$  and the definition of  $\bar{f}(x, \alpha)$ , it is easy to see that

$$\tilde{Q}(t)\bar{f}(x, \cdot)(\alpha) = 0.$$

Applying the operator  $\mathcal{G}^\varepsilon(t)$  to the function  $f(\cdot)$  defined in (5.92) yields that

$$\begin{aligned} & \bar{f}(n^\varepsilon(t), \alpha^\varepsilon(t)) + \sqrt{\varepsilon}h(t, n^\varepsilon(t), \alpha^\varepsilon(t)) \\ & - \int_0^t \left\{ \frac{1}{\sqrt{\varepsilon}} \langle W(s, \alpha^\varepsilon(s)), \nabla_x \bar{f}(n^\varepsilon(s), \alpha^\varepsilon(s)) + \sqrt{\varepsilon} \nabla_x h(s, n^\varepsilon(s), \alpha^\varepsilon(s)) \rangle \right. \\ & \quad + \sqrt{\varepsilon} \frac{\partial}{\partial s} h(s, n^\varepsilon(s), \alpha^\varepsilon(s)) + \frac{1}{\sqrt{\varepsilon}} \tilde{Q}(s) h(s, n^\varepsilon(s), \cdot)(\alpha^\varepsilon(s)) \\ & \quad \left. + \hat{Q}(s) (\bar{f}(n^\varepsilon(s), \cdot) + \sqrt{\varepsilon} h(s, n^\varepsilon(s), \cdot)(\alpha^\varepsilon(s))) \right\} ds \end{aligned}$$

defines a martingale.

The basic premise of the perturbed test function method is to choose the function  $h(\cdot)$  that cancels the “bad” terms of order  $1/\sqrt{\varepsilon}$ :

$$\tilde{Q}(s)h(s, x, \cdot)(\alpha) = -\langle W(s, \alpha), \nabla_x \bar{f}(x, \alpha) \rangle. \tag{5.93}$$

Note that as mentioned previously,  $\tilde{Q}(t)$  has rank  $m - l$ . Thus the dimension of the null space is  $l$ ; that is,  $N(\tilde{Q}(t)) = l$ . A crucial observation is that in view of the Fredholm alternative (see Lemma A.37 and Corollary A.38), a solution of (5.93) exists iff the matrix  $(\langle W(s, s_{ij}), \nabla_x \bar{f}(x, s_{ij}) \rangle)$  is orthogonal to  $\tilde{\mathbb{1}}_{m_1}, \dots, \tilde{\mathbb{1}}_{m_l}$ , the span of  $N(\tilde{Q}(t))$  (see Remark 4.23 for the notation). Moreover, since  $f^0(\cdot, i)$  is  $C_L^2$ ,  $h(\cdot)$  can be chosen to satisfy the following properties assuming  $\beta_{ij}(\cdot)$  to be Lipschitz on  $[0, T]$ :

- (1)  $h(t, x, \alpha)$  is uniformly Lipschitz in  $t$ ;
- (2)  $|h(t, x, \alpha)|$  and  $|\nabla_x h(t, x, \alpha)|$  are bounded;
- (3)  $\nabla_x h(t, x, \alpha)$  is Lipschitz in  $(t, x)$ .

Such an  $h(\cdot)$  leads to

$$\begin{aligned} & \bar{f}(n^\varepsilon(t), \alpha^\varepsilon(t)) + \sqrt{\varepsilon}h(t, n^\varepsilon(t), \alpha^\varepsilon(t)) \\ & - \int_0^t \left\{ \langle W(s, \alpha^\varepsilon(s)), \nabla_x h(s, n^\varepsilon(s), \alpha^\varepsilon(s)) \rangle \right. \\ & \quad + \sqrt{\varepsilon} \left( \frac{\partial}{\partial s} h(s, n^\varepsilon(s), \alpha^\varepsilon(s)) \right) + \hat{Q}(s) \bar{f}(n^\varepsilon(s), \cdot)(\alpha^\varepsilon(s)) \\ & \quad \left. + \sqrt{\varepsilon} \hat{Q}(s) h(s, n^\varepsilon(s), \cdot)(\alpha^\varepsilon(s)) \right\} ds \end{aligned} \tag{5.94}$$

being a martingale. For each  $s, x, \alpha$ , define

$$g(s, x, \alpha) = \langle W(s, \alpha), \nabla_x h(s, x, \alpha) \rangle. \quad (5.95)$$

With  $f^0 \in C_L^2$ , it is easy to see that  $g(s, x, \alpha)$  is Lipschitz in  $(s, x)$ . This function will be used in defining the operator for the limit problem later.

**Remark 5.39.** Note that the choice of  $h(\cdot)$  in (5.93) is not unique. If  $h_1(\cdot)$  and  $h_2(\cdot)$  are both solutions to (5.93), then the irreducibility of  $\tilde{Q}^i(s)$  implies that, for each  $i = 1, \dots, l$ ,

$$\begin{pmatrix} h_1(s, x, s_{i1}) \\ \vdots \\ h_1(s, x, s_{im_i}) \end{pmatrix} - \begin{pmatrix} h_2(s, x, s_{i1}) \\ \vdots \\ h_2(s, x, s_{im_i}) \end{pmatrix} = h^0(s, x, i) \mathbb{1}_{m_i}$$

for some scalar functions  $h^0(s, x, i)$ . Although the choice of  $h$  is not unique, the resulting function  $g(s, x, \alpha)$  is well defined. As in Remark 4.23, the consistency condition or solvability condition due to Fredholm alternative is in force. Therefore, if  $h_1$  and  $h_2$  are both solutions to (5.93), then

$$\langle W(s, \alpha), \nabla_x h_1(s, x, \alpha) \rangle = \langle W(s, \alpha), \nabla_x h_2(s, x, \alpha) \rangle,$$

for  $\alpha \in \mathcal{M}_i$  and  $i = 1, \dots, l$ .

Using  $g(s, x, \alpha)$  defined above, we obtain

$$\begin{aligned} & \int_0^t \langle W(s, \alpha^\varepsilon(s)), \nabla_x h(s, n^\varepsilon(s), \alpha^\varepsilon(s)) \rangle ds \\ &= \int_0^t g(s, n^\varepsilon(s), \alpha^\varepsilon(s)) ds \\ &= \int_0^t \sum_{i=1}^l \sum_{j=1}^{m_i} I_{\{\alpha^\varepsilon(s)=s_{ij}\}} g(s, n^\varepsilon(s), s_{ij}) ds \\ &= \int_0^t \sum_{i=1}^l \sum_{j=1}^{m_i} (I_{\{\alpha^\varepsilon(s)=s_{ij}\}} - \nu_j^i(s) I_{\{\bar{\alpha}^\varepsilon(s)=i\}}) g(s, n^\varepsilon(s), s_{ij}) ds \\ & \quad + \int_0^t \sum_{i=1}^l \sum_{j=1}^{m_i} I_{\{\bar{\alpha}^\varepsilon(s)=i\}} \nu_j^i(s) g(s, n^\varepsilon(s), s_{ij}) ds. \end{aligned}$$

In view of Lemma 5.37, the term in the fourth line above goes to zero in mean square uniformly in  $t \in [0, T]$ . Let

$$\bar{g}(s, x, i) = \sum_{j=1}^{m_i} \nu_j^i(s) g(s, x, s_{ij}).$$

Then it follows that

$$\begin{aligned} & \int_0^t \sum_{i=1}^l \sum_{j=1}^{m_i} I_{\{\bar{\alpha}^\varepsilon(s)=i\}} \nu_j^i(s) g(s, n^\varepsilon(s), s_{ij}) ds \\ &= \int_0^t \sum_{i=1}^l I_{\{\bar{\alpha}^\varepsilon(s)=i\}} \bar{g}(s, n^\varepsilon(s), i) ds \\ &= \int_0^t \bar{g}(s, n^\varepsilon(s), \bar{\alpha}^\varepsilon(s)) ds. \end{aligned}$$

Therefore, as  $\varepsilon \rightarrow 0$ , we have

$$\begin{aligned} E \left| \int_0^t \langle W(s, \alpha^\varepsilon(s)), \nabla_x h(s, n^\varepsilon(s), \alpha^\varepsilon(s)) \rangle ds \right. \\ \left. - \int_0^t \bar{g}(s, n^\varepsilon(s), \bar{\alpha}^\varepsilon(s)) ds \right|^2 \rightarrow 0 \end{aligned} \tag{5.96}$$

uniformly in  $t \in [0, T]$ .

Furthermore, we have

$$\begin{aligned} & \int_0^t \widehat{Q}(s) \bar{f}(n^\varepsilon(s), \cdot)(\alpha^\varepsilon(s)) ds \\ &= \int_0^t \sum_{i=1}^l \sum_{j=1}^{m_i} I_{\{\alpha^\varepsilon(s)=s_{ij}\}} \widehat{Q}(s) \bar{f}(n^\varepsilon(s), \cdot)(s_{ij}) ds \\ &= \int_0^t \sum_{i=1}^l \sum_{j=1}^{m_i} (I_{\{\alpha^\varepsilon(s)=s_{ij}\}} - \nu_j^i(s) I_{\{\bar{\alpha}^\varepsilon(s)=i\}}) \widehat{Q}(s) \bar{f}(n^\varepsilon(s), \cdot)(s_{ij}) ds \\ & \quad + \int_0^t \sum_{i=1}^l \sum_{j=1}^{m_i} \nu_j^i(s) I_{\{\bar{\alpha}^\varepsilon(s)=i\}} \widehat{Q}(s) \bar{f}(n^\varepsilon(s), \cdot)(s_{ij}) ds. \end{aligned}$$

Again, Lemma 5.37 implies that the third line above goes to 0 in mean square uniformly in  $t \in [0, T]$ . The last term above equals

$$\int_0^t \bar{Q}(s) f^0(n^\varepsilon(s), \cdot)(\bar{\alpha}^\varepsilon(s)) ds,$$

where  $\bar{Q}(s) = \text{diag}(\nu^1(t), \dots, \nu^l(t)) \widehat{Q}(s) \tilde{\mathbb{1}}$ . It follows that as  $\varepsilon \rightarrow 0$ ,

$$\begin{aligned} E \left| \int_0^t \widehat{Q}(s) \bar{f}(n^\varepsilon(s), \cdot)(\alpha^\varepsilon(s)) ds \right. \\ \left. - \int_0^t \bar{Q}(s) f^0(n^\varepsilon(s), \cdot)(\bar{\alpha}^\varepsilon(s)) ds \right| \rightarrow 0 \end{aligned} \tag{5.97}$$

uniformly in  $t \in [0, T]$ .

We next examine the function  $\bar{g}(s, x, i)$  closely. Using the block-diagonal structure of  $\tilde{Q}(s)$ , we can write (5.93) in terms of each block  $\tilde{Q}^j(s)$ . For  $j = 1, \dots, l$ ,

$$\tilde{Q}^j(s) \begin{pmatrix} h(s, x, s_{j1}) \\ \vdots \\ h(s, x, s_{jm_j}) \end{pmatrix} = - \begin{pmatrix} \langle W(s, s_{j1}), \nabla_x f^0(x, j) \rangle \\ \vdots \\ \langle W(s, s_{jm_j}), \nabla_x f^0(x, j) \rangle \end{pmatrix}. \quad (5.98)$$

Note that  $\tilde{Q}^j(s)$  is weakly irreducible so  $\text{rank}(\tilde{Q}^j(s)) = m_j - 1$ . As in Remark 4.9, equation (5.98) has a solution since it is consistent and the solvability condition in the sense of Fredholm alternative is satisfied. We can solve (5.98) using exactly the same technique as in Section 4.2 for obtaining the  $\varphi_i(t)$ , that is, replacing one of the rows of the augmented matrix in (5.98) by  $(1, 1, \dots, 1, 0)$ , which represents the equation  $\sum_{k=1}^{m_j} h(s, x, s_{jk}) = 0$ . The coefficient matrix of the resulting equation then has full rank; one readily obtains a solution. Equivalently, the solution may be written as

$$\begin{pmatrix} h(s, x, s_{j1}) \\ \vdots \\ h(s, x, s_{jm_j}) \end{pmatrix} = - \left[ \begin{pmatrix} \tilde{Q}^j(s) \\ \mathbb{1}'_{m_j} \end{pmatrix} \begin{pmatrix} \tilde{Q}^j(s) \\ \mathbb{1}'_{m_j} \end{pmatrix} \right]^{-1} \begin{pmatrix} \tilde{Q}^j(s) \\ \mathbb{1}'_{m_j} \end{pmatrix} \begin{pmatrix} \langle W(s, s_{j1}), \nabla_x f^0(x, j) \rangle \\ \vdots \\ \langle W(s, s_{jm_j}), \nabla_x f^0(x, j) \rangle \\ 0 \end{pmatrix}.$$

Note that

$$I_{\{\alpha=s_{jk}\}} - \nu_k^j(t) I_{\{\alpha \in \mathcal{M}_j\}} = 0 \text{ if } \alpha \notin \mathcal{M}_j.$$

Recall the notation for the partitioned vector  $x = (x^1, \dots, x^l)$  where  $x^j$  is an  $m_j$ -dimensional vector and  $x^j = (x_1^j, \dots, x_{m_j}^j)$ . For the partial derivatives, use the notation

$$\partial_{j,k} = \frac{\partial}{\partial x_k^j} \text{ and } \partial_{j_1, j_2}^2 = \frac{\partial^2}{\partial x_{j_1}^j \partial x_{j_2}^j}.$$

Then  $h(s, x, s_{jk})$  is a functional of  $\partial_{j,j_1} f^0(x, j), \dots, \partial_{j,m_j} f^0(x, j)$ . It follows that  $g(s, x, s_{jk})$  is a functional of  $\partial_{j_1, j_2}^2 f^0(x, j)$ , for  $j_1, j_2 = 1, \dots, m_j$ , and so is  $\bar{g}(s, x, j)$ . Write

$$\bar{g}(s, x, j) = \frac{1}{2} \sum_{j_1, j_2=1}^{m_j} a_{j_1 j_2}(s, j) \partial_{j_1, j_2}^2 f^0(x, j), \quad (5.99)$$

for some continuous functions  $a_{j_1 j_2}(s, j)$ .

**Lemma 5.40.** *Assume (A5.5) and (A5.6). Suppose  $(n^\varepsilon(\cdot), \bar{\alpha}^\varepsilon(\cdot))$  converges weakly to  $(n(\cdot), \bar{\alpha}(\cdot))$ . Then for  $f^0(\cdot, i) \in C_L^2$ ,*

$$f^0(n(t), \bar{\alpha}(t)) - \int_0^t (\bar{g}(s, n(s), \bar{\alpha}(s)) + \bar{Q}(s)f^0(n(s), \cdot)(\bar{\alpha}(s))) ds$$

is a martingale.

Proof: Define

$$\begin{aligned} H^\varepsilon(t) &= \bar{f}(n^\varepsilon(t), \alpha^\varepsilon(t)) + \sqrt{\varepsilon}h(t, n^\varepsilon(t), \alpha^\varepsilon(t)) \\ &\quad - \int_0^t \left\{ \left\langle W(s, \alpha^\varepsilon(s)), \nabla_x h(s, n^\varepsilon(s), \alpha^\varepsilon(s)) \right\rangle \right. \\ &\quad \left. + \sqrt{\varepsilon} \frac{\partial}{\partial s} h(s, n^\varepsilon(s), \alpha^\varepsilon(s)) + \widehat{Q}(s)\bar{f}(n^\varepsilon(s), \cdot)(\alpha^\varepsilon(s)) \right. \\ &\quad \left. + \sqrt{\varepsilon}\widehat{Q}(s)h(s, n^\varepsilon(s), \cdot)(\alpha^\varepsilon(s)) \right\} ds. \end{aligned}$$

The martingale property implies that

$$E[(H^\varepsilon(t) - H^\varepsilon(s))z_1(n^\varepsilon(t_1), \bar{\alpha}^\varepsilon(t_1)) \cdots z_k(n^\varepsilon(t_k), \bar{\alpha}^\varepsilon(t_k))] = 0,$$

for any  $0 \leq t_1 \leq \cdots \leq t_k \leq s \leq t$  and any bounded and continuous functions  $z_1(\cdot), \dots, z_k(\cdot)$ .

In view of the choice of  $h(\cdot)$ , it follows that all the three terms

$$\begin{aligned} &\sqrt{\varepsilon}h(t, n^\varepsilon(t), \alpha^\varepsilon(t)), \\ &\sqrt{\varepsilon} \left( \frac{\partial}{\partial t} h(t, n^\varepsilon(t), \alpha^\varepsilon(t)) \right), \text{ and} \\ &\sqrt{\varepsilon}\widehat{Q}(t)h(t, n^\varepsilon(t), \cdot)(\alpha^\varepsilon(t)) \end{aligned}$$

converge to 0 in mean square. Recall (5.96), (5.97), and

$$\bar{f}(n^\varepsilon(t), \alpha^\varepsilon(t)) = f^0(n^\varepsilon(t), \bar{\alpha}^\varepsilon(t)).$$

Denote the weak limit of  $H^\varepsilon(\cdot)$  by  $\bar{H}(\cdot)$ . We have

$$E[(\bar{H}(t) - \bar{H}(s))z_1(n(t_1), \bar{\alpha}(t_1)) \cdots z_k(n(t_k), \bar{\alpha}(t_k))] = 0,$$

where  $\bar{H}(\cdot)$  is given by

$$\begin{aligned} \bar{H}(t) &= f^0(n(t), \bar{\alpha}(t)) \\ &\quad - \int_0^t (\bar{g}(r, n(r), \bar{\alpha}(r)) + \bar{Q}(r)f^0(n(r), \cdot)(\bar{\alpha}(r))) dr. \end{aligned}$$

Thus  $(n(\cdot), \bar{\alpha}(\cdot))$  is a solution to the martingale problem. □



**Lemma 5.41.** *Let  $\mathcal{L}$  denote the operator given by*

$$\mathcal{L}f^0(x, j) = \frac{1}{2} \sum_{j_1, j_2=1}^{m_j} a_{j_1 j_2}(s, j) \partial_{j_1 j_2}^2 f^0(x, j) + \bar{Q}(s) f^0(x, \cdot)(j).$$

*Then the martingale problem with operator  $\mathcal{L}$  has a unique solution.*

Proof: In view of Lemma A.14, we need only verify the uniqueness in distribution of  $(n(t), \bar{\alpha}(t))$  for each  $t \in [0, T]$ . Let

$$f(x, j) = \exp(\iota\{\langle \theta, x \rangle + \theta_0 j\}),$$

where  $\theta \in \mathbb{R}^m$ ,  $\theta_0 \in \mathbb{R}$ ,  $j \in \mathcal{M}$ , and  $\iota$  is the pure imaginary number with  $\iota^2 = -1$ .

For fixed  $j_0, k_0$ , let  $F_{j_0 k_0}(x, j) = I_{\{j=j_0\}} f(x, k_0)$ . Then

$$F_{j_0 k_0}(n(t), \bar{\alpha}(t)) = I_{\{\bar{\alpha}(t)=j_0\}} f(n(t), k_0).$$

Moreover, note that

$$\begin{aligned} \bar{g}(s, n(s), \bar{\alpha}(s)) &= \sum_{j=1}^l I_{\{\bar{\alpha}(s)=j\}} \bar{g}(s, n(s), j) \\ &= \frac{1}{2} \sum_{j=1}^l I_{\{\bar{\alpha}(s)=j\}} \sum_{j_1, j_2=1}^{m_j} a_{j_1 j_2}(s, j) \partial_{j_1 j_2}^2 F_{j_0 k_0}(n(s), j) \\ &= \frac{1}{2} I_{\{\bar{\alpha}(s)=j_0\}} \sum_{j_1, j_2=1}^{m_{j_0}} a_{j_1 j_2}(s, j_0) \partial_{j_0}^2 f(n(s), k_0) \\ &= \frac{1}{2} \sum_{j_1, j_2=1}^{m_j} a_{j_1 j_2}(s, j_0) (-\theta_{j_0 j_1} \theta_{j_0 j_2}) (I_{\{\bar{\alpha}(s)=j_0\}} f(n(s), k_0)). \end{aligned} \tag{5.100}$$

Furthermore, we have

$$\begin{aligned} &\bar{Q}(s) F_{j_0 k_0}(n(s), \cdot)(\bar{\alpha}(s)) \\ &= \sum_{j=1}^l I_{\{\bar{\alpha}(s)=j\}} \bar{Q}(s) F_{j_0 k_0}(n(s), \cdot)(j) \\ &= \sum_{j=1}^l I_{\{\bar{\alpha}(s)=j\}} \sum_{k=1}^l \bar{q}_{jk}(s) F_{j_0 k_0}(n(s), k) \\ &= \sum_{j=1}^l I_{\{\bar{\alpha}(s)=j\}} \sum_{k=1}^l \bar{q}_{jk}(s) I_{\{k=j_0\}} f(n(s), k_0) \\ &= \sum_{j=1}^l I_{\{\bar{\alpha}(s)=j\}} \bar{q}_{j j_0}(s) f(n(s), k_0) \\ &= \sum_{j=1}^l \bar{q}_{j j_0}(s) I_{\{\bar{\alpha}(s)=j\}} f(n(s), k_0). \end{aligned} \tag{5.101}$$

Let

$$\phi_{jk}(t) = E(I_{\{\bar{\alpha}(t)=j\}}f(n(t),k)), \text{ for } j,k = 1, \dots, l.$$

Then in view of (5.100) and (5.101),

$$\begin{aligned} \phi_{j_0k_0}(t) - \phi_{j_0k_0}(0) - \int_0^t \left\{ \sum_{j_1, j_2=1}^{m_{j_0}} a_{j_1j_2}(s, j_0)(-\theta_{j_0j_1}\theta_{j_0j_2})\phi_{j_0k_0}(s) \right. \\ \left. + \sum_{j=1}^l \bar{q}_{jj_0}(s)\phi_{jk_0}(s) \right\} ds = 0. \end{aligned} \tag{5.102}$$

Let

$$\phi(t) = (\phi_{11}(t), \dots, \phi_{1m_1}(t), \dots, \phi_{l1}(t), \dots, \phi_{lm_l}(t)).$$

Rewrite (5.102) in terms of  $\phi(\cdot)$  as

$$\phi(t) = \phi(0) + \int_0^t \phi(s)B(s)ds,$$

where  $\phi(0) = (\phi_{jk}(0))$  with  $\phi_{jk}(0) = EI_{\{\bar{\alpha}(0)=j\}}f(0,k)$ , and  $B(t)$  is a matrix-valued function whose entries are defined by the integrand of (5.102). The equation for  $\phi(t)$  is a linear ordinary differential equation. It is well known that such a differential equation has a unique solution. Hence,  $\phi(t)$  is uniquely determined. In particular,

$$\begin{aligned} E \exp(\iota\{\langle \theta, n(t) \rangle + \theta_0\bar{\alpha}(t)\}) \\ = \sum_{j=1}^l E(I_{\{\bar{\alpha}(t)=j\}} \exp(\iota\{\langle \theta, n(t) \rangle + j\theta_0\})) \end{aligned}$$

is uniquely determined for all  $\theta, \theta_0$ , so is the distribution of  $(n(t), \bar{\alpha}(t))$  by virtue of the uniqueness theorem and the inversion formula of the characteristic function (see Chow and Teicher [30]). □

The tightness of  $(n^\varepsilon(\cdot), \bar{\alpha}^\varepsilon(\cdot))$  together with Lemma 5.40 and Lemma 5.41 implies that  $(n^\varepsilon(\cdot), \bar{\alpha}^\varepsilon(\cdot))$  converges weakly to  $(n(\cdot), \bar{\alpha}(\cdot))$ . We will show that  $n(\cdot)$  is a switching diffusion, i.e., a diffusion process modulated by a Markov process such that the covariance of the diffusion depends on the Markov jump process. Precisely, owing to the presence of the jump Markov chains, the limit process does not possess the independent increment property shared by many processes. A moment of reflection reveals that, necessarily, the coefficients in  $\bar{g}(s, x, i)$  must consist of a symmetric nonnegative definite matrix serving as a covariance matrix. The following lemma verifies this assertion.

**Lemma 5.42.** For  $s \in [0, T]$  and  $j = 1, \dots, l$ , the matrix

$$A(s, j) = (a_{j_1 j_2}(s, j))$$

is symmetric and nonnegative definite.

Proof: Let  $\eta^j = (\eta_{j_1}, \dots, \eta_{j_{m_j}})'$  and  $x^j = (x_{j_1}, \dots, x_{j_{m_j}})'$ . Define

$$f_j(x) = \frac{1}{2} (\langle \eta^j, x^j \rangle)^2.$$

Then the corresponding  $\bar{g}(\cdot)$  defined in (5.99) has the following form:

$$\bar{g}(s, x, j) = \frac{1}{2} \eta^{j'} A(s, j) \eta^j.$$

Moreover, let  $f_j(x, k) = f_j(x)$ , independent of  $k$ . Then for all  $k = 1, \dots, l$ ,

$$\bar{Q}(s) f_j(n^\varepsilon(s), \cdot)(k) = 0.$$

To verify the nonnegativity of  $A(s, j)$ , it suffices to show that

$$\int_s^t \eta^{j'} A(r, j) \eta^j dr \geq 0,$$

for all  $0 \leq s \leq t \leq T$ . Recall that  $f_j(x)$  is a quadratic function. In view of (5.94) and the proof of Lemma 5.40, it then follows that

$$\frac{1}{2} \int_s^t \eta^{j'} A(r, j) \eta^j dr = \lim_{\varepsilon \rightarrow 0} (E f_j(n^\varepsilon(t)) - E f_j(n^\varepsilon(s))).$$

We are in a position to show that the limit is nonnegative. Let

$$n^{\varepsilon, j}(t) = (n_{j_1}^\varepsilon(t), \dots, n_{j_{m_j}}^\varepsilon(t)).$$

Then

$$\begin{aligned} E(f_j(n^\varepsilon(t)) - f_j(n^\varepsilon(s))) \\ = \frac{1}{2} E \left( \langle \eta^j, n^{\varepsilon, j}(t) \rangle^2 - \langle \eta^j, n^{\varepsilon, j}(s) \rangle^2 \right). \end{aligned}$$

For  $t \geq s \geq 0$ , using

$$\langle \eta^j, n^{\varepsilon, j}(t) \rangle = \langle \eta^j, n^{\varepsilon, j}(s) \rangle + \langle \eta^j, n^{\varepsilon, j}(t) - n^{\varepsilon, j}(s) \rangle,$$

we have

$$\begin{aligned} E \left( \langle \eta^j, n^{\varepsilon, j}(t) \rangle^2 - \langle \eta^j, n^{\varepsilon, j}(s) \rangle^2 \right) \\ = E \left( 2 \langle \eta^j, n^{\varepsilon, j}(s) \rangle \langle \eta^j, n^{\varepsilon, j}(t) - n^{\varepsilon, j}(s) \rangle + \langle \eta^j, n^{\varepsilon, j}(t) - n^{\varepsilon, j}(s) \rangle^2 \right) \\ \geq 2E \left( \langle \eta^j, n^{\varepsilon, j}(s) \rangle \langle \eta^j, n^{\varepsilon, j}(t) - n^{\varepsilon, j}(s) \rangle \right) \\ = 2E \left( \langle \eta^j, n^{\varepsilon, j}(s) \rangle E \left[ \langle \eta^j, n^{\varepsilon, j}(t) - n^{\varepsilon, j}(s) \rangle \middle| \mathcal{F}_s^\varepsilon \right] \right). \end{aligned}$$

We next show that the last term goes to 0 as  $\varepsilon \rightarrow 0$ . In fact, in view of (a) in Lemma 5.35, it follows that

$$E[n^{\varepsilon,j}(t) - n^{\varepsilon,j}(s) | \mathcal{F}_s^\varepsilon] = O(\sqrt{\varepsilon}),$$

and hence

$$\begin{aligned} & E \left[ \langle \eta^j, n^{\varepsilon,j}(t) - n^{\varepsilon,j}(s) \rangle \middle| \mathcal{F}_s^\varepsilon \right] \\ &= \langle \eta^j, E[(n^{\varepsilon,j}(t) - n^{\varepsilon,j}(s)) | \mathcal{F}_s^\varepsilon] \rangle = O(\sqrt{\varepsilon}). \end{aligned}$$

Using (b) in Lemma 5.35, we derive the following inequalities

$$E \langle \eta^j, n^{\varepsilon,j}(s) \rangle^2 \leq |\eta^j|^2 E |n^{\varepsilon,j}(s)|^2 \leq |\eta^j|^2 O(s).$$

The Cauchy–Schwarz inequality then leads to

$$\begin{aligned} & \left| E \left( \langle \eta^j, n^{\varepsilon,j}(s) \rangle E \left[ \langle \eta^j, n^{\varepsilon,j}(t) - n^{\varepsilon,j}(s) \rangle \middle| \mathcal{F}_s^\varepsilon \right] \right) \right| \\ & \leq \left( E \langle \eta^j, n^{\varepsilon,j}(s) \rangle^2 \right)^{\frac{1}{2}} \left( E \left[ \langle \eta^j, n^{\varepsilon,j}(t) - n^{\varepsilon,j}(s) \rangle \middle| \mathcal{F}_s^\varepsilon \right]^2 \right)^{\frac{1}{2}} \\ & = \left( E \langle \eta^j, n^{\varepsilon,j}(s) \rangle^2 \right)^{\frac{1}{2}} O(\sqrt{\varepsilon}) \rightarrow 0, \text{ as } \varepsilon \rightarrow 0. \end{aligned}$$

As a result for some  $K > 0$ , we have

$$E \left( \langle \eta^j, n^{\varepsilon,j}(s) \rangle E \left[ \langle \eta^j, n^{\varepsilon,j}(t) - n^{\varepsilon,j}(s) \rangle \middle| \mathcal{F}_s^\varepsilon \right] \right) \geq -K |\eta^j|_s \sqrt{\varepsilon} \rightarrow 0,$$

as  $\varepsilon \rightarrow 0$ . The nonnegativity of  $A(s, j)$  follows.

To show that  $A(s, j)$  is symmetric, consider

$$f_{j,j_1 j_2}(x) = x_{j j_1} x_{j j_2} \text{ for } j_1, j_2 = 1, \dots, m_j.$$

Then, we have

$$\frac{1}{2} \int_0^t a_{j_1 j_2}(s, j) ds = \lim_{\varepsilon \rightarrow 0} E(n_{j_1}^{\varepsilon,j}(t) n_{j_2}^{\varepsilon,j}(t)) = \frac{1}{2} \int_0^t a_{j_2 j_1}(s, j) ds, \quad (5.103)$$

for all  $t \in [0, T]$ . Thus,  $A(s, j)$  is symmetric. □

Next, we derive an explicit representation of the nonnegative definite matrix  $A(s, j)$  similar to that of Theorem 5.9. Recall that given a function  $f^0(\cdot)$ , one can find  $h(\cdot)$  as in (5.93). Using this  $h(\cdot)$ , one defines  $f(\cdot)$  as in

(5.95) which leads to  $\bar{g}(\cdot)$  given in (5.99). In view of the result in Theorem 5.9 for a single block of the irreducible matrix  $\tilde{Q}^j(t)$  together with the computations of  $\bar{g}(s, x, j)$ , it follows that  $A(s, j) = 2A^0(s, j)$ , where

$$\begin{aligned} A^0(t, j) &= \beta_{\text{diag}}^j(t) \left( \nu_{\text{diag}}^j(t) \int_0^\infty Q_0(r, t, j) dr \right. \\ &\quad \left. + \left( \int_0^\infty Q_0(r, t, j) dr \right) \nu_{\text{diag}}^j(t) \right) \beta_{\text{diag}}^j(t), \end{aligned}$$

with

$$\beta_{\text{diag}}^j(t) = \text{diag}(\beta_{j1}(t), \dots, \beta_{jm_j}(t)),$$

$$\nu_{\text{diag}}^j(t) = \text{diag}(\nu_1^j(t), \dots, \nu_{m_j}^j(t)),$$

and

$$Q_0(r, t, j) = \left[ I - \begin{pmatrix} \nu^j(t) \\ \vdots \\ \nu^j(t) \end{pmatrix} \right] \exp\left(\tilde{Q}^j(t)r\right).$$

Applying Lemma 5.42 to the case of  $\tilde{Q}(s)$  a single block irreducible matrix  $\tilde{Q}^j(s)$ , it follows that  $A^0(s, j)$  is symmetric and nonnegative definite. Hence, standard results in linear algebra yield that there exists an  $m_j \times m_j$  matrix  $\sigma^0(s, j)$  such that

$$\sigma^0(s, j)\sigma^{0'}(s, j) = A^0(s, j). \quad (5.104)$$

Note that the definition of  $\bar{g}(s, x, j)$  is independent of  $\hat{Q}(t)$ , so for determining  $A^0(s, j)$ , we may consider  $\hat{Q}(t) = 0$ . Note also that

$$\tilde{Q}(t) = \text{diag}(\tilde{Q}^1(t), 0, \dots, 0) + \dots + \text{diag}(0, \dots, 0, \tilde{Q}^l(t)).$$

The foregoing statements suggest that in view of (5.104), the desired covariance matrix is given by

$$\begin{aligned} \sigma(s, j) &= \begin{pmatrix} 0_{m_1 \times m_1} & & & & & \\ & 0_{m_2 \times m_2} & & & & \\ & & \ddots & & & \\ & & & \sigma^0(s, j) & & \\ & & & & \ddots & \\ & & & & & 0_{m_l \times m_l} \end{pmatrix} \quad (5.105) \\ &= \text{diag}(0_{m_1 \times m_1}, 0_{m_2 \times m_2}, \dots, \sigma^0(s, j), \dots, 0_{m_l \times m_l}), \end{aligned}$$

where  $0_{m_k \times m_k}$  is the  $m_k \times m_k$  zero matrix. That is, it is a matrix with the  $j$ th block-diagonal submatrix equal to  $\sigma^0(s, j)$  and the rest of its elements equal to zero.

**Theorem 5.43.** *Assume that (A5.5) holds. Suppose  $\tilde{Q}(\cdot)$  is twice differentiable with Lipschitz continuous second derivative and  $\hat{Q}(\cdot)$  is differentiable with Lipschitz continuous derivative. Let  $\beta_{ij}(\cdot)$  be bounded and Lipschitz continuous deterministic functions. Then  $n^\varepsilon(\cdot)$  converges weakly to a switching diffusion  $n(\cdot)$ , where*

$$n(t) = \left( \int_0^t \sigma(s, \bar{\alpha}(s)) dw(s) \right)'$$

and  $w(\cdot)$  is a standard  $m$ -dimensional Brownian motion.

Proof: Let

$$\tilde{n}(t) = \left( \int_0^t \sigma(s, \bar{\alpha}(s)) dw(s) \right)'$$

and  $\bar{\alpha}(\cdot)$  be a Markov chain generated by  $\bar{Q}(t)$ . Then for all  $f^0(\cdot, i) \in C_L^2$ ,

$$f^0(\tilde{n}(t), \bar{\alpha}(t)) - \int_0^t (\bar{g}(s, \tilde{n}(s), \bar{\alpha}(s)) + \bar{Q}(s) f^0(\tilde{n}(s), \cdot)(\bar{\alpha}(s))) ds$$

is a martingale. This and the uniqueness of the martingale problem in Lemma 5.41 yields that  $(\tilde{n}(\cdot), \bar{\alpha}(\cdot))$  has the same probability distribution as  $(n(\cdot), \bar{\alpha}(\cdot))$ . This proves the theorem.  $\square$

**Remark 5.44.** Note that the Lipschitz condition on  $\beta_{ij}(\cdot)$  is not required in analyzing the asymptotic normality in Section 5.3.3. It is needed in this section because the perturbed test function method typically requires smoothness conditions of the associated processes.

It appears that the conditions in (A5.5) and (A5.6) together with the Lipschitz property of  $\beta_{ij}(\cdot)$  are sufficient for the convergence of  $n^\varepsilon(\cdot)$  to a switching diffusion  $n(\cdot)$ . The additional assumptions on further derivatives of  $\tilde{Q}(\cdot)$  and  $\hat{Q}(\cdot)$  are needed for computing the covariance of the limit process  $n(\cdot)$ .

**Remark 5.45.** If  $\bar{\alpha}(\cdot)$  were a deterministic function,  $n(\cdot)$  above would be a diffusion process in the usual sense. However since the limit  $\bar{\alpha}(\cdot)$  is a Markov chain, the diffusion process is modulated by this jump process; the resulting distribution has the features of the “continuous” diffusion process and the “discrete” Markov chain limit.

In this section, we use the perturbed test function method, which is quite different from the approach of Section 5.2. The method used in that section, which might be called a direct approach, is interesting in its own right and makes a close connection between asymptotic expansion and asymptotic normality. It is effective whenever it can be applied. One of the main ingredients is that the direct approach makes use of the mixing properties of the scaled occupation measures heavily. In fact, using asymptotic expansion, it was shown that the scaled sequence of occupation measures is a mixing

process with exponential mixing rate. For the weak and strong interaction cases presented, the mixing condition, and even approximate mixing conditions, no longer hold. To illustrate, consider Example 4.20 with constant jump rates and calculate

$$E[n^{\varepsilon,\prime}(s)(n^\varepsilon(t) - n^\varepsilon(s))].$$

By virtue of the proof of Theorem 5.25, a straightforward but tedious calculation shows that

$$E[n^{\varepsilon,\prime}(s)(n^\varepsilon(t) - n^\varepsilon(s))] \not\rightarrow 0 \text{ as } \varepsilon \rightarrow 0$$

for the weak and strong interaction models because  $E[n^{\varepsilon,\prime}(s)(n^\varepsilon(t) - n^\varepsilon(s))]$  depends on  $P_1(t, s)$ , generally a nonzero function. A direct consequence is that the limit process does *not* have independent increments in general. It is thus difficult to characterize the limit process via the direct approach. The perturbed test function method, on the other hand, can be considered as a combined approach. It uses enlarged or augmented states by treating the scaled occupation measure  $n^\varepsilon(\cdot)$  and the Markov chain  $\alpha^\varepsilon(\cdot)$  together. That is, one considers a new state variable with two components  $(x, \alpha)$ . This allows us to bypass the verification of mixing-like properties such that the limit process is characterized by means of solutions of appropriate martingale problems via perturbed test functions, which underlies the rationale and essence of the approach. As a consequence, the limit process is characterized via the limit of the underlying sequence of operators.

Note that if  $\tilde{Q}(t)$  itself is weakly irreducible (i.e.,  $\tilde{Q}(t)$  consists of only one block), then the covariance matrix is given by (5.30). In this case, since there is only one group of recurrent states, the jump behavior due to the limit process  $\bar{\alpha}(\cdot)$  will disappear. Moreover, owing to the fast transition rate  $\tilde{Q}(t)/\varepsilon$ , the singularly perturbed Markov chain rapidly reaches its quasi-stationary regime. As a result, the jump behavior does not appear in the asymptotic distribution, and the diffusion becomes the dominant factor. Although the method employed in this chapter is different from that of Section 5.2, the result coincides with that of Section 5.2 under irreducibility. We state this in the following corollary.

**Corollary 5.46.** *Assume that the conditions of Theorem 5.43 are fulfilled with  $l = 1$  (i.e.,  $\tilde{Q}(t)$  has only one block). Then  $n^\varepsilon(\cdot)$  converges weakly to the diffusion process*

$$n(t) = \left( \int_0^t \sigma(s)dw(s) \right)',$$

where  $w(\cdot)$  is an  $m$ -dimensional standard Brownian motion with covariance

$$A(t) = \sigma(t)\sigma'(t)$$

given by (5.30).

To further illustrate, consider the following example. This problem is concerned with a singularly perturbed Markov chain with four states divided into two groups. It has been used in modeling production planning problems with failure-prone machines. As was mentioned, from a modeling point of view, it may be used to depict the situation that two machines operate in tandem, in which the operating conditions (the machine capacity) of one of the machines change much faster than the other; see also related discussions in Chapters 7 and 8.

**Example 5.47.** Let  $\alpha^\varepsilon(\cdot)$  be a Markov chain generated by

$$Q^\varepsilon(t) = \frac{1}{\varepsilon} \begin{pmatrix} -\lambda_1(t) & \lambda_1(t) & 0 & 0 \\ \mu_1(t) & -\mu_1(t) & 0 & 0 \\ 0 & 0 & -\lambda_1(t) & \lambda_1(t) \\ 0 & 0 & \mu_1(t) & -\mu_1(t) \end{pmatrix} + \begin{pmatrix} -\lambda_2(t) & 0 & \lambda_2(t) & 0 \\ 0 & -\lambda_2(t) & 0 & \lambda_2(t) \\ \mu_2(t) & 0 & -\mu_2(t) & 0 \\ 0 & \mu_2(t) & 0 & -\mu_2(t) \end{pmatrix}.$$

Then

$$\bar{Q}(t) = \begin{pmatrix} -\lambda_2(t) & \lambda_2(t) \\ \mu_2(t) & -\mu_2(t) \end{pmatrix}.$$

Let  $\bar{\alpha}(\cdot)$  be a Markov chain generated by  $\bar{Q}(t)$ ,  $t \geq 0$ . In this example,

$$\sigma^0(s, 1) = 2 \left( \frac{\lambda_1(s)\mu_1(s)}{(\lambda_1(s) + \mu_1(s))^3} \right)^{\frac{1}{2}} \begin{pmatrix} \beta_{11}(s) & 0 \\ -\beta_{12}(s) & 0 \end{pmatrix},$$

$$\sigma^0(s, 2) = 2 \left( \frac{\lambda_1(s)\mu_1(s)}{(\lambda_1(s) + \mu_1(s))^3} \right)^{\frac{1}{2}} \begin{pmatrix} \beta_{21}(s) & 0 \\ -\beta_{22}(s) & 0 \end{pmatrix},$$

$$\sigma(s, 1) = 2 \left( \frac{\lambda_1(s)\mu_1(s)}{(\lambda_1(s) + \mu_1(s))^3} \right)^{\frac{1}{2}} \begin{pmatrix} \beta_{11}(s) & 0 & 0 & 0 \\ -\beta_{12}(s) & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

and

$$\sigma(s, 2) = 2 \left( \frac{\lambda_1(s)\mu_1(s)}{(\lambda_1(s) + \mu_1(s))^3} \right)^{\frac{1}{2}} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \beta_{21}(s) & 0 \\ 0 & 0 & -\beta_{22}(s) & 0 \end{pmatrix}.$$



The limit of  $n^\varepsilon(\cdot)$  is given by

$$n(t) = \left( \int_0^t \sigma(s, \bar{\alpha}(s)) dw(s) \right)',$$

where  $w(\cdot)$  is a standard Brownian motion taking values in  $\mathbb{R}^4$ .

## 5.4 Measurable Generators

In Section 4.2, we considered the asymptotic expansions of probability distributions. A natural requirement of such expansions is that the generator  $Q^\varepsilon(t)$  be smooth enough to establish the desired error bounds. It would be interesting to consider the case in which the generator  $Q^\varepsilon(t)$ ,  $t \geq 0$ , is merely measurable. The method used in this section is very useful in some manufacturing problems; see Sethi and Zhang [192]. Moreover, the results are used in Section 8.6 to deal with a control problem under relaxed control formulation. Given only the measurability of  $Q^\varepsilon(t)$ , there seems to be little hope to obtain an asymptotic expansion. Instead of constructing an asymptotic series of the corresponding probability distribution, we consider the convergence of  $P(\alpha^\varepsilon(t) = s_{ij})$  under the framework of convergence of

$$\int_0^T P(\alpha^\varepsilon(t) = s_{ij}) f(t) dt \quad \text{for } f(\cdot) \in L^2[0, T]; \mathbb{R}.$$

Since the phrase “weak convergence” is reserved throughout the book for the convergence of probability measures, to avoid confusion, we refer to the convergence above as convergence in the weak sense on  $L^2([0, T]; \mathbb{R})$  or convergence under the weak topology of  $L^2([0, T]; \mathbb{R})$ .

### Case I: Weakly Irreducible $\tilde{Q}(t)$

Let  $\alpha^\varepsilon(\cdot) \in \mathcal{M} = \{1, \dots, m\}$  denote the Markov chain generated by

$$Q^\varepsilon(t) = \frac{1}{\varepsilon} \tilde{Q}(t) + \hat{Q}(t),$$

where both  $\tilde{Q}(t)$  and  $\hat{Q}(t)$  are generators.

We assume the following conditions in this subsection.

**(A5.7)**  $\tilde{Q}(t)$  and  $\hat{Q}(t)$  are bounded and Borel measurable. Moreover,  $\tilde{Q}(t)$  is weakly irreducible.

**Remark 5.48.** In fact, both the boundedness and the Borel measurability in (A5.7) are redundant. Recall that our definition of generators (see Definition 2.2) uses the  $q$ -Property, which includes both the Borel measurability and the boundedness. Thus, (A5.7) requires only weak irreducibility.

Nevertheless, we retain both boundedness and measurability for those who read only this section. Similar comments apply to assumption (A5.8) in what follows.

Define the probability distribution vector

$$p^\varepsilon(t) = (P(\alpha^\varepsilon(t) = 1), \dots, P(\alpha^\varepsilon(t) = m))$$

and the transition matrix

$$P^\varepsilon(t, s) = (p_{ij}^\varepsilon(t, s)) = (P(\alpha^\varepsilon(t) = j | \alpha^\varepsilon(s) = i)).$$

Then using the martingale property in Lemma 2.4, we have

$$p^\varepsilon(t) = p^\varepsilon(s) + \int_s^t p^\varepsilon(r)Q^\varepsilon(r)dr \tag{5.106}$$

and

$$P^\varepsilon(t, s) = I + \int_s^t P^\varepsilon(r, s)Q^\varepsilon(r)dr. \tag{5.107}$$

The next two lemmas are concerned with the asymptotic properties of  $p^\varepsilon(t)$  and  $P^\varepsilon(t, s)$ .

**Lemma 5.49.** *Assume (A5.7). Then for each  $i, j$ , and  $T > 0$ ,  $P(\alpha^\varepsilon(t) = i)$  and  $P(\alpha^\varepsilon(t) = i | \alpha^\varepsilon(s) = j)$  both converge weakly to  $\nu_i(t)$  on  $L^2([0, T]; \mathbb{R})$  and  $L^2([s, T]; \mathbb{R})$ , respectively, that is, as  $\varepsilon \rightarrow 0$ ,*

$$\int_0^T [P(\alpha^\varepsilon(t) = i) - \nu_i(t)]f(t)dt \rightarrow 0 \tag{5.108}$$

and

$$\int_s^T [P(\alpha^\varepsilon(t) = i | \alpha^\varepsilon(s) = j) - \nu_i(t)]f(t)dt \rightarrow 0, \tag{5.109}$$

for all  $f(\cdot) \in L^2([0, T]; \mathbb{R})$  and  $L^2([s, T]; \mathbb{R})$ , respectively.

Proof: We only verify (5.108); the proof of (5.109) is similar. Recall that

$$p^\varepsilon(t) = (p_1^\varepsilon(t), \dots, p_m^\varepsilon(t)) = (P(\alpha^\varepsilon(t) = 1), \dots, P(\alpha^\varepsilon(t) = m)).$$

Since  $p^\varepsilon(\cdot) \in L^2([0, T]; \mathbb{R}^m)$  (space of square-integrable functions on  $[0, T]$  taking values in  $\mathbb{R}^m$ ), for each subsequence of  $\varepsilon \rightarrow 0$  there exists (see Lemma A.36) a further subsequence of  $\varepsilon \rightarrow 0$  (still denoted by  $\varepsilon$  for simplicity), and for such  $\varepsilon$ , the corresponding  $\{p^\varepsilon(\cdot)\}$  converges (in the weak sense on  $L^2([0, T]; \mathbb{R}^m)$ ) to some  $p(\cdot) = (p_1(\cdot), \dots, p_m(\cdot)) \in L^2([0, T]; \mathbb{R}^m)$ , that is,

$$\int_0^T p^\varepsilon(r)(f_1(r), \dots, f_m(r))'dr \rightarrow \int_0^T p(r)(f_1(r), \dots, f_m(r))'dr,$$

for any  $(f_1(\cdot), \dots, f_m(\cdot))' \in L^2([0, T]; \mathbb{R}^m)$ . Moreover,

$$0 \leq p_i(t) \leq 1 \quad \text{and} \quad p_1(t) + \dots + p_m(t) = 1 \quad (5.110)$$

almost everywhere. Since  $\tilde{Q}(\cdot) \in L^2([0, T]; \mathbb{R}^{m \times m})$ , we have for  $0 \leq s \leq t \leq T$ ,

$$\int_s^t p^\varepsilon(r) \tilde{Q}(r) dr \rightarrow \int_s^t p(r) \tilde{Q}(r) dr.$$

Thus, using (5.106) we obtain

$$\begin{aligned} \int_s^t p(r) \tilde{Q}(r) dr &= \lim_{\varepsilon \rightarrow 0} \int_s^t p^\varepsilon(r) \tilde{Q}(r) dr \\ &= \lim_{\varepsilon \rightarrow 0} \left( \varepsilon(p^\varepsilon(t) - p^\varepsilon(s)) - \varepsilon \int_s^t p^\varepsilon(r) \hat{Q}(r) dr \right) = 0. \end{aligned}$$

Since  $s$  and  $t$  are arbitrary, it follows immediately that

$$p(t) \tilde{Q}(t) = 0 \text{ a.e. in } t.$$

By virtue of (5.110), the irreducibility of  $\tilde{Q}(t)$  implies  $p(t) = \nu(t)$  almost everywhere. Thus the limit is independent of the chosen subsequence. Therefore,  $p^\varepsilon(\cdot) \rightarrow \nu(\cdot)$  in the weak sense on  $L^2([0, T]; \mathbb{R}^m)$ .  $\square$

**Theorem 5.50.** *Assume (A5.7). Then for any bounded deterministic function  $\beta_i(\cdot)$  and for each  $i \in \mathcal{M}$  and  $t \geq 0$ ,*

$$E \left| \int_0^t (I_{\{\alpha^\varepsilon(s)=i\}} - \nu_i(s)) \beta_i(s) ds \right|^2 \rightarrow 0 \text{ as } \varepsilon \rightarrow 0. \quad (5.111)$$

Proof: Let

$$\eta(t) = E \left| \int_0^t (I_{\{\alpha^\varepsilon(s)=i\}} - \nu_i(s)) \beta_i(s) ds \right|^2.$$

Then as in the proof of Theorem 5.25, we can show that

$$\eta(t) = 2(\eta_1(t) + \eta_2(t)),$$

where

$$\begin{aligned} \eta_1(t) &= \int_0^t \int_0^s (-\nu_i(r)) [P(\alpha^\varepsilon(s) = i) - \nu_i(s)] \beta_i(s) \beta_i(r) dr ds, \\ \eta_2(t) &= \int_0^t \int_0^s P(\alpha^\varepsilon(r) = i) [P(\alpha^\varepsilon(s) = i | \alpha^\varepsilon(r) = i) - \nu_i(s)] \\ &\quad \times \beta_i(s) \beta_i(r) dr ds. \end{aligned}$$

By virtue of Lemma 5.49,  $P(\alpha^\varepsilon(s) = i) \rightarrow \nu_i(s)$  in the weak sense on  $L^2([0, T]; \mathbb{R})$  and therefore as  $\varepsilon \rightarrow 0$ ,

$$\eta_1(t) = \int_0^t [P(\alpha^\varepsilon(s) = i) - \nu_i(s)]\beta_i(s) \left( \int_0^s (-\nu_i(r))\beta_i(r)dr \right) ds \rightarrow 0.$$

Similarly, in view of the convergence of

$$P(\alpha^\varepsilon(s) = i | \alpha^\varepsilon(r) = i) \rightarrow \nu_i(s)$$

under the weak topology of  $L^2([r, t]; \mathbb{R})$ , we have

$$\begin{aligned} \eta_2(t) &= \int_0^t \left[ \int_r^t [P(\alpha^\varepsilon(s) = i | \alpha^\varepsilon(r) = i) - \nu_i(s)]\beta_i(s)ds \right] \\ &\quad \times P(\alpha^\varepsilon(r) = i)\beta_i(r)dr \rightarrow 0. \end{aligned}$$

This concludes the proof of the theorem. □

Case II:  $\tilde{Q}(t) = \text{diag}(\tilde{Q}^1(t), \dots, \tilde{Q}^l(t))$

This subsection extends the preceding result to the cases in which  $\tilde{Q}(t)$  is a block-diagonal matrix with irreducible blocks. We make the following assumptions:

(A5.8)  $\tilde{Q}(t)$  and  $\tilde{Q}^i(t)$ , for  $i = 1, \dots, l$ , are bounded and Borel measurable. Moreover,  $\tilde{Q}^i(t)$ ,  $i = 1, \dots, l$ , are weakly irreducible.

**Lemma 5.51.** *Assume (A5.8). Then the following assertions hold:*

(a) *For each  $i = 1, \dots, l$  and  $j = 1, \dots, m_i$ ,  $P(\alpha^\varepsilon(t) = s_{ij})$  converges in the weak sense to  $\nu_j^i(t)\vartheta^i(t)$  on  $L^2([0, T]; \mathbb{R})$ , that is,*

$$\int_0^T [P(\alpha^\varepsilon(t) = s_{ij}) - \nu_j^i(t)\vartheta^i(t)]f(t)dt \rightarrow 0, \tag{5.112}$$

for all  $f(\cdot) \in L^2([0, T]; \mathbb{R})$ , where

$$(\vartheta^1(t), \dots, \vartheta^l(t)) = p_0 \tilde{\mathbf{1}} + \int_0^t (\vartheta^1(s), \dots, \vartheta^l(s))\tilde{Q}(s)ds.$$

(b) *For each  $i, j, i_1, j_1$ ,  $P(\alpha^\varepsilon(t) = s_{ij} | \alpha^\varepsilon(s) = s_{i_1 j_1})$  converges in the weak sense to  $\nu_j^i(t)\vartheta_{ii}(t, s)$  on  $L^2([s, T]; \mathbb{R})$ , that is,*

$$\int_s^T [P(\alpha^\varepsilon(t) = s_{ij} | \alpha^\varepsilon(s) = s_{i_1 j_1}) - \nu_j^i(t)\vartheta_{ii}(t, s)]f(t)dt \rightarrow 0, \tag{5.113}$$

for all  $f(\cdot) \in L^2([s, T]; \mathbb{R})$ , where  $\vartheta_{ij}(t, s)$  is defined in Lemma 5.24 (see (5.50)).

Proof: We only derive (5.112); the proof of (5.113) is similar. Let

$$p^\varepsilon(t) = (p_{11}^\varepsilon(t), \dots, p_{1m_1}^\varepsilon(t), \dots, p_{l1}^\varepsilon(t), \dots, p_{lm_l}^\varepsilon(t))$$

where  $p_{ij}^\varepsilon(t) = P(\alpha^\varepsilon(t) = s_{ij})$ . Since  $p^\varepsilon(\cdot) \in L^2([0, T]; \mathbb{R}^m)$ , there exists (see Lemma A.36) a subsequence of  $\varepsilon \rightarrow 0$  (still denoted by  $\varepsilon$  for simplicity), such that corresponding to this  $\varepsilon$ ,  $p^\varepsilon(t)$  converges to some  $p(\cdot) \in L^2([0, T]; \mathbb{R}^m)$  under the weak topology. Let

$$p(t) = (p_{11}(t), \dots, p_{1m_1}(t), \dots, p_{l1}(t), \dots, p_{lm_l}(t)).$$

Then  $0 \leq p_{ij}(t) \leq 1$  and  $\sum_{i,j} p_{ij}(t) = 1$  almost everywhere. Similarly as in the proof of Lemma 5.49, for  $0 \leq t \leq T$ ,

$$p(t)\tilde{Q}(t) = 0 \text{ a.e. in } t.$$

The irreducibility of  $\tilde{Q}^k(t)$ ,  $k = 1, \dots, l$ , implies that

$$p(t) = (\vartheta^1(t), \dots, \vartheta^l(t)) \text{diag}(\nu^1(t), \dots, \nu^l(t)), \tag{5.114}$$

for some functions  $\vartheta^1(t), \dots, \vartheta^l(t)$ .

In view of (5.106), we have

$$p^\varepsilon(t)\tilde{\mathbb{I}} = p_0\tilde{\mathbb{I}} + \int_0^t p^\varepsilon(s) \left( \frac{1}{\varepsilon}\tilde{Q}(s) + \hat{Q}(s) \right) \tilde{\mathbb{I}} ds.$$

Since  $\tilde{Q}(s)\tilde{\mathbb{I}} = 0$ , it follows that

$$p^\varepsilon(t)\tilde{\mathbb{I}} = p_0\tilde{\mathbb{I}} + \int_0^t p^\varepsilon(s)\hat{Q}(s)\tilde{\mathbb{I}} ds.$$

Owing to the convergence of  $p^\varepsilon(t) \rightarrow p(t)$  under the weak topology of  $L^2([0, T]; \mathbb{R}^m)$ , we have

$$p(t)\tilde{\mathbb{I}} = p_0\tilde{\mathbb{I}} + \int_0^t p(s)\hat{Q}(s)\tilde{\mathbb{I}} ds.$$

Using (5.114) and noting that

$$\text{diag}(\nu^1(t), \dots, \nu^l(t))\tilde{\mathbb{I}} = I,$$

we have

$$(\vartheta^1(t), \dots, \vartheta^l(t)) = p_0\tilde{\mathbb{I}} + \int_0^t (\vartheta^1(s), \dots, \vartheta^l(s))\hat{Q}(s)ds.$$

The uniqueness of the solution then yields the lemma. □

**Theorem 5.52.** *Assume (A5.8). Then for any  $i = 1, \dots, l, j = 1, \dots, m_i$ , and bounded deterministic function  $\beta_{ij}(t), t \geq 0$ ,*

$$E \left( \int_0^T (I_{\{\alpha^\varepsilon(t)=s_{ij}\}} - \nu_j^i(t)I_{\{\bar{\alpha}^\varepsilon(t)=i\}}) \beta_{ij}(t)dt \right)^2 \rightarrow 0, \text{ as } \varepsilon \rightarrow 0.$$

Proof: Let  $\eta(t)$  be defined as in (5.52). Then we can show similarly as in the proof of Theorem 5.25 that

$$\eta(T) = 2 \int_0^T \int_0^t \Phi^\varepsilon(t, r) \beta_{ij}(t) \beta_{ij}(r) dr dt,$$

where  $\Phi^\varepsilon(t, r) = \Phi_1^\varepsilon(t, r) + \Phi_2^\varepsilon(t, r)$  with  $\Phi_1^\varepsilon(t, r)$  and  $\Phi_2^\varepsilon(t, r)$  defined by (5.53) and (5.54), respectively.

Note that by changing the order of integration,

$$\begin{aligned} & \int_0^T \int_0^t \Phi_1^\varepsilon(t, r) \beta_{ij}(t) \beta_{ij}(r) dr dt \\ &= \int_0^T P(\alpha^\varepsilon(r) = s_{ij}) \beta_{ij}(r) \left\{ \int_r^T [P(\alpha^\varepsilon(t) = s_{ij} | \alpha^\varepsilon(r) = s_{ij}) \right. \\ & \quad \left. - \nu_j^i(t)P(\alpha^\varepsilon(t) \in \mathcal{M}_i | \alpha^\varepsilon(r) = s_{ij})] \beta_{ij}(t) dt \right\} dr. \end{aligned}$$

Since the  $\beta_{ij}(\cdot)$  are bounded uniformly on  $[0, T]$ ,  $\beta_{ij}(\cdot) \in L^2([0, T]; \mathbb{R})$ . As a result, Lemma 5.51 implies that

$$\begin{aligned} & \int_r^T [P(\alpha^\varepsilon(t) = s_{ij} | \alpha^\varepsilon(r) = s_{ij}) \\ & \quad - \nu_j^i(t)P(\alpha^\varepsilon(t) \in \mathcal{M}_i | \alpha^\varepsilon(r) = s_{ij})] \beta_{ij}(t) dt \rightarrow 0. \end{aligned}$$

Hence as  $\varepsilon \rightarrow 0$ ,

$$\int_0^T \int_0^t \Phi_1^\varepsilon(t, r) \beta_{ij}(t) \beta_{ij}(r) dr dt \rightarrow 0.$$

Similarly,

$$\int_0^T \int_0^t \Phi_2^\varepsilon(t, r) \beta_{ij}(t) \beta_{ij}(r) dr dt \rightarrow 0, \text{ as } \varepsilon \rightarrow 0.$$

The proof is complete. □

**Theorem 5.53.** *Assume (A5.8). Then  $\bar{\alpha}^\varepsilon(\cdot)$  converges weakly to  $\bar{\alpha}(\cdot)$  on  $D([0, T]; \mathcal{M})$ , as  $\varepsilon \rightarrow 0$ .*

Proof: Recall that  $\chi^\varepsilon(t)$  denotes the vector of indicator functions

$$\left( I_{\{\alpha^\varepsilon(t)=s_{11}\}}, \dots, I_{\{\alpha^\varepsilon(t)=s_{1m_1}\}}, \dots, I_{\{\alpha^\varepsilon(t)=s_{l1}\}}, \dots, I_{\{\alpha^\varepsilon(t)=s_{lm_l}\}} \right),$$

and let

$$\bar{\chi}^\varepsilon(t) = (\bar{\chi}_1^\varepsilon(t), \dots, \bar{\chi}_l^\varepsilon(t)) = \chi^\varepsilon(t) \tilde{\mathbb{1}}.$$

Then  $\bar{\chi}_i^\varepsilon(t) = I_{\{\bar{\alpha}^\varepsilon(t)=i\}}$  for  $i = 1, \dots, l$ .

We show that  $\bar{\chi}^\varepsilon(\cdot)$  is tight in  $D^l[0, T]$  first. Let  $\mathcal{F}_t^\varepsilon = \sigma\{\alpha^\varepsilon(r) : r \leq t\}$ . Then in view of the martingale property associated with  $\alpha^\varepsilon(\cdot)$ , we have, for  $0 \leq s \leq t$ ,

$$E \left[ \chi^\varepsilon(t) - \chi^\varepsilon(s) - \int_s^t \chi^\varepsilon(r) Q^\varepsilon(r) dr \middle| \mathcal{F}_s^\varepsilon \right] = 0.$$

Right multiplying both sides of the equation by  $\tilde{\mathbb{1}}$  and noting that  $\tilde{Q}(r)\tilde{\mathbb{1}} = 0$ , we obtain

$$E \left[ \bar{\chi}^\varepsilon(t) - \bar{\chi}^\varepsilon(s) - \int_s^t \chi^\varepsilon(r) \hat{Q}(r) \tilde{\mathbb{1}} dr \middle| \mathcal{F}_s^\varepsilon \right] = 0. \quad (5.115)$$

Note that

$$\left| \int_s^t \chi^\varepsilon(r) \hat{Q}(r) \tilde{\mathbb{1}} dr \right| = O(t - s).$$

It follows from (5.115) that

$$E \left[ I_{\{\bar{\alpha}^\varepsilon(t)=i\}} \middle| \mathcal{F}_s^\varepsilon \right] = I_{\{\bar{\alpha}^\varepsilon(s)=i\}} + O(t - s). \quad (5.116)$$

Note also that  $(I_A)^2 = I_A$  for any set  $A$ . We have, in view of (5.116),

$$\begin{aligned} & E \left[ \left( I_{\{\bar{\alpha}^\varepsilon(t)=i\}} - I_{\{\bar{\alpha}^\varepsilon(s)=i\}} \right)^2 \middle| \mathcal{F}_s^\varepsilon \right] \\ &= E \left[ I_{\{\bar{\alpha}^\varepsilon(t)=i\}} - 2I_{\{\bar{\alpha}^\varepsilon(t)=i\}} I_{\{\bar{\alpha}^\varepsilon(s)=i\}} + I_{\{\bar{\alpha}^\varepsilon(s)=i\}} \middle| \mathcal{F}_s^\varepsilon \right] \\ &= E \left[ I_{\{\bar{\alpha}^\varepsilon(t)=i\}} \middle| \mathcal{F}_s^\varepsilon \right] - 2E \left[ I_{\{\bar{\alpha}^\varepsilon(t)=i\}} \middle| \mathcal{F}_s^\varepsilon \right] I_{\{\bar{\alpha}^\varepsilon(s)=i\}} + I_{\{\bar{\alpha}^\varepsilon(s)=i\}} \\ &= I_{\{\bar{\alpha}^\varepsilon(s)=i\}} + O(t - s) \\ &\quad - 2 \left( I_{\{\bar{\alpha}^\varepsilon(s)=i\}} + O(t - s) \right) I_{\{\bar{\alpha}^\varepsilon(s)=i\}} + I_{\{\bar{\alpha}^\varepsilon(s)=i\}} \\ &= O(t - s), \end{aligned}$$

for each  $i = 1, \dots, l$ . Hence,

$$\lim_{t \rightarrow s} \lim_{\varepsilon \rightarrow 0} E \left\{ E \left[ \left( I_{\{\bar{\alpha}^\varepsilon(t)=i\}} - I_{\{\bar{\alpha}^\varepsilon(s)=i\}} \right)^2 \middle| \mathcal{F}_s^\varepsilon \right] \right\} = 0.$$

Therefore, by Lemma A.17,  $\bar{\chi}^\varepsilon(\cdot)$  is tight.

The tightness of  $\bar{\chi}^\varepsilon(\cdot)$  implies that for any sequence  $\varepsilon_k \rightarrow 0$ , there exists a subsequence of  $\{\varepsilon_k\}$  (still denoted by  $\{\varepsilon_k\}$ ) such that  $\bar{\chi}^{\varepsilon_k}(\cdot)$  converges weakly. We next show that the limit of such a subsequence is uniquely determined by  $\bar{Q}(\cdot) := \text{diag}(\nu^1(\cdot), \dots, \nu^l(\cdot))\widehat{Q}(\cdot)\mathbb{1}$ .

Note that

$$\begin{aligned} \int_s^t \chi^\varepsilon(r)\widehat{Q}(r)\mathbb{1}dr &= \int_s^t \bar{\chi}^\varepsilon(r)\bar{Q}(r)dr \\ &+ \int_s^t (\chi^\varepsilon(r) - \bar{\chi}^\varepsilon(r)\text{diag}(\nu^1(r), \dots, \nu^l(r)))\widehat{Q}(r)\mathbb{1}dr. \end{aligned}$$

In view of Theorem 5.52, we have, as  $\varepsilon \rightarrow 0$ ,

$$E \left| \int_s^t [\chi^\varepsilon(r) - \bar{\chi}^\varepsilon(r)\text{diag}(\nu^1(r), \dots, \nu^l(r))] \widehat{Q}(r)\mathbb{1}dr \right| \rightarrow 0. \quad (5.117)$$

Now by virtue of (5.115),

$$E \left[ \left( \bar{\chi}^\varepsilon(t) - \bar{\chi}^\varepsilon(s) - \int_s^t \chi^\varepsilon(r)\widehat{Q}(r)\mathbb{1}dr \right) z_1(\bar{\chi}^\varepsilon(t_1)) \cdots z_j(\bar{\chi}^\varepsilon(t_j)) \right] = 0,$$

for  $0 \leq t_1 \leq \dots \leq t_j \leq s \leq t$  and bounded and continuous functions  $z_1(\cdot), \dots, z_j(\cdot)$ .

Let  $\bar{\chi}(\cdot)$  denote the limit in distribution of  $\bar{\chi}^{\varepsilon_k}(\cdot)$ . Then in view of (5.117) and the continuity of  $\int_s^t \eta(r)\bar{Q}(r)dr$  with respect to  $\eta(\cdot)$  (see Lemma A.40), we have  $\bar{\chi}^\varepsilon(\cdot) \rightarrow \bar{\chi}(\cdot)$  as  $\varepsilon_k \rightarrow 0$ , and  $\bar{\chi}(\cdot)$  satisfies

$$E \left[ \left( \bar{\chi}(t) - \bar{\chi}(s) - \int_s^t \bar{\chi}(r)\bar{Q}(r)dr \right) z_1(\bar{\chi}(t_1)) \cdots z_j(\bar{\chi}(t_j)) \right] = 0.$$

It is easy to see that  $\bar{\chi}(\cdot) = (\bar{\chi}_1(\cdot), \dots, \bar{\chi}_l(\cdot))$  is an  $l$ -valued measurable process having sample paths in  $D^l[0, T]$  and satisfying  $\bar{\chi}_i(t) = 0$  or 1 and  $\bar{\chi}_1(\cdot) + \dots + \bar{\chi}_l(\cdot) = 1$  w.p.1. Let

$$\bar{\alpha}(t) = \sum_{i=1}^l iI_{\{\bar{\chi}_i(t)=1\}},$$

or in an expanded form,

$$\bar{\alpha}(t) = \begin{cases} 1, & \text{if } \bar{\chi}_1(t) = 1, \\ 2, & \text{if } \bar{\chi}_1(t) = 0, \bar{\chi}_2(t) = 1, \\ \vdots & \vdots \\ l, & \text{if } \bar{\chi}_i(t) = 0, \text{ for } i \leq l-1, \bar{\chi}_l(t) = 1. \end{cases}$$



Then  $\bar{\alpha}(\cdot)$  is a process with sample paths in  $D([0, T]; \overline{\mathcal{M}})$  and

$$\bar{\chi}(t) = (I_{\{\bar{\alpha}(t)=1\}}, \dots, I_{\{\bar{\alpha}(t)=l\}}) \text{ w.p.1.}$$

Therefore,  $\bar{\alpha}(\cdot)$  is a Markov chain generated by  $\bar{Q}(\cdot)$ . As a result, its distribution is uniquely determined by  $\bar{Q}(\cdot)$ . It follows that  $\bar{\alpha}^\varepsilon(\cdot)$  converges weakly to  $\bar{\alpha}(\cdot)$ .  $\square$

**Remark 5.54.** Note that Theorem 5.53 gives the same result as Theorem 5.27 under weaker conditions. The proofs are quite different. The proof of Theorem 5.53 is based on martingale properties associated with the Markov chain, whereas the proof of Theorem 5.27 follows the traditional approach, i.e., after the tightness is verified, the convergence of finite-dimensional distributions is proved.

**Remark 5.55.** In view of the development in Chapter 4, apart from the smoothness conditions, one of the main ingredients is the use of the Fredholm alternative. One hopes that this will carry over (under suitable conditions) to the measurable generators. A possible approach is the utilization of the formulation of weak derivatives initiated in the study of partial differential equations (see Hutson and Pym [90]).

Following the tactics of the weak sense formulation, for some  $T < \infty$  and for given  $g(\cdot) \in L^2([0, T]; \mathbb{R})$ , a function  $f(\cdot) \in L^2([0, T]; \mathbb{R})$  is a weak solution of  $(d/dt)f = g$  if

$$\int_0^T f(t) \left( \frac{d\phi(t)}{dt} \right) dt = \int_0^T g(t)\phi(t) dt$$

for any  $C^\infty$ -functions on  $[0, T]$  vanishing on the boundary together with their derivatives (denoted by  $\phi \in C_0^\infty([0, T]; \mathbb{R})$ ). Write the weak solution as  $(d/dt)f \stackrel{w}{=} g$ .

Recall that  $L_{loc}^2$  is the set of functions that lie in  $L^2(S; \mathbb{R})$  for every closed and bounded set  $S \subset (0, T)$ . A function  $f(\cdot) \in L_{loc}^2$  has a  $j$ th-order weak derivative if there is a function  $g(\cdot) \in L_{loc}^2$  such that

$$\int_0^T g(t)\phi(t) dt = (-1)^j \int_0^T f(t) \frac{d^j \phi(t)}{dt^j} dt$$

for all  $\phi \in C_0^\infty([0, T]; \mathbb{R})$ . The function  $g(\cdot)$  above is called the  $j$ th-order weak derivative of  $f(\cdot)$ , and is denoted by  $D^j f = g$ .

To proceed, define the space of functions  $H^n$  as

$$H^n = \{f \text{ on } [0, T]; \text{ for } 0 \leq j \leq n, D^j f \text{ exist and are in } L^2([0, T]; \mathbb{R})\}.$$

Equip  $H^n$  with an inner product and a norm as

$$(f, g)_n = \sum_{j \leq n} \int_0^T D^j f D^j g dt,$$

$$|f|_n^2 = (f, f)_n = \sum_{j \leq n} \int_0^T |D^j f|^2 dt.$$

One can then work under such a framework and proceed to obtain the asymptotic expansion of the probability distribution. It seems that the conditions required are not much different from those in the case of smooth generators; we will not pursue this issue further.

## 5.5 Remarks on Inclusion of Transient and Absorbing States

So far, the development in this chapter has focused on Markov chains with only recurrent states (either a single weakly irreducible class or a number of weakly irreducible classes). This section extends the results obtained to the case that a transient class or a group of absorbing states is included.

### 5.5.1 Inclusion of Transient States

Consider the Markov chain  $\alpha^\varepsilon(\cdot) \in \mathcal{M}$ , where its generator is still given by (5.47) and the state space of  $\alpha^\varepsilon(t)$  is given by

$$\mathcal{M} = \mathcal{M}_1 \cup \mathcal{M}_2 \cup \dots \cup \mathcal{M}_l \cup \mathcal{M}_*, \tag{5.118}$$

with  $\mathcal{M}_i = \{s_{i1}, \dots, s_{im_i}\}$  and  $\mathcal{M}_* = \{s_{*1}, \dots, s_{*m_*}\}$ . In what follows, we present results concerning the asymptotic distributions of scaled occupation measures and properties of measurable generators. While main assumptions and results are provided, the full proofs are omitted. The interested reader can derive the results using the ideas presented in the previous sections.

To proceed, assume that  $\tilde{Q}(t)$  is a generator of a Markov chain satisfying

$$\tilde{Q}(t) = \begin{pmatrix} \tilde{Q}^1(t) & & & & \\ & \ddots & & & \\ & & \tilde{Q}^l(t) & & \\ \tilde{Q}_*^1(t) & \dots & \tilde{Q}_*^l(t) & \tilde{Q}_*(t) & \end{pmatrix} \tag{5.119}$$

such that for each  $t \in [0, T]$  and each  $i = 1, \dots, l$ ,  $\tilde{Q}^i(t)$  is a generator with dimension  $m_i \times m_i$ ,  $\tilde{Q}_*(t)$  is an  $m_* \times m_*$  matrix,  $\tilde{Q}_*^i(t) \in \mathbb{R}^{m_* \times m_i}$ , and  $m_1 + m_2 + \dots + m_l + m_* = m$ . We impose the following conditions.

(A5.9) For all  $t \in [0, T]$ , and  $i = 1, \dots, l$ ,  $\tilde{Q}^i(t)$  are weakly irreducible, and  $\tilde{Q}_*(t)$  is Hurwitz (i.e., all of its eigenvalues have negative real parts). Moreover,  $\tilde{Q}(\cdot)$  is differentiable on  $[0, T]$  and its derivative is Lipschitz;  $\hat{Q}(\cdot)$  is Lipschitz continuous on  $[0, T]$ .

Use the partition

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

where

$$\begin{aligned} \hat{Q}^{11}(t) &\in \mathbb{R}^{(m-m_*) \times (m-m_*)}, \quad \hat{Q}^{12}(t) \in \mathbb{R}^{(m-m_*) \times m_*}, \\ \hat{Q}^{21}(t) &\in \mathbb{R}^{m_* \times (m-m_*)}, \quad \text{and} \quad \hat{Q}^{22}(t) \in \mathbb{R}^{m_* \times m_*}, \end{aligned}$$

and write

$$\begin{aligned} \bar{Q}_*(t) &= \text{diag}(\nu^1(t), \dots, \nu^l(t))(\hat{Q}^{11}(t)\tilde{\mathbb{1}} + \hat{Q}^{12}(t)(a_{m_1}(t), \dots, a_{m_l}(t))) \\ \bar{Q}(t) &= \text{diag}(\bar{Q}_*(t), 0_{m_* \times m_*}), \end{aligned} \tag{5.120}$$

where

$$\tilde{\mathbb{1}} = \text{diag}(\mathbb{1}_{m_1}, \dots, \mathbb{1}_{m_l}), \quad \mathbb{1}_{m_j} = (1, \dots, 1)' \in \mathbb{R}^{m_j \times 1},$$

and

$$a_{m_i}(t) = -\tilde{Q}_*^{-1}(t)\tilde{Q}_*^i(t)\mathbb{1}_{m_i}, \quad \text{for } i = 1, \dots, l. \tag{5.121}$$

In what follows, if  $a_{m_i}(t)$  is time independent, we will simply write it as  $a_{m_i}$ . The requirement on  $\tilde{Q}_*(t)$  in (A5.9) implies that the corresponding states are transient. The Hurwitzian property also has the following interesting implication: For each  $t \in [0, T]$ , and each  $i = 1, \dots, l$ ,  $a_{m_i}(t) = (a_{m_i,1}(t), \dots, a_{m_i,m_*}(t))' \in \mathbb{R}^{m_* \times 1}$ . Then

$$a_{m_i,j}(t) \geq 0 \quad \text{and} \quad \sum_{i=1}^l a_{m_i,j}(t) = 1 \tag{5.122}$$

for each  $j = 1, \dots, m_*$ . That is, for each  $t \in [0, T]$  and each  $j = 1, \dots, l$ ,  $(a_{m_1,j}(t), \dots, a_{m_l,j}(t))$  can be considered a probability row vector. To see this, note that

$$\int_0^\infty \exp(\tilde{Q}_*(t)s)ds = -\tilde{Q}_*^{-1}(t),$$

which has nonnegative components. It follows from the definition that  $a_{m_i}(t) \geq 0$ . Furthermore,

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

Thus (5.122) follows. Similar to the development in the section for the case of weak and strong interactions, we can derive the following results.

**Theorem 5.56.** *Define*

$$\chi_{ij}^\varepsilon(t) = \begin{cases} \int_0^t (I_{\{\alpha^\varepsilon(s)=s_{ij}\}} - \nu_j^i(s)I_{\{\alpha^\varepsilon(s)\in\mathcal{M}_i\}}) ds, & \text{for } i = 1, \dots, l, \\ \int_0^t I_{\{\alpha^\varepsilon(s)=s_{*j}\}} ds, & \text{for } i = *, \end{cases} \tag{5.123}$$

and assume (A5.9). Then for each  $j = 1, \dots, m_i$ ,

$$\sup_{t \in [0, T]} E|\chi_{ij}^\varepsilon(t)|^2 = \begin{cases} O(\varepsilon), & \text{for } i = 1, \dots, l, \\ O(\varepsilon^2), & \text{for } i = *. \end{cases} \tag{5.124}$$

Next, for each fixed  $t \in [0, T]$ , let  $\xi$  be a random variable uniformly distributed on  $[0, 1]$  that is independent of  $\alpha^\varepsilon(\cdot)$ . For each  $j = 1, \dots, m_*$ , define an integer-valued random variable  $\xi_j(t)$  by

$$\begin{aligned} \xi_j(t) = & I_{\{0 \leq \xi \leq a_{m_1, j}(t)\}} + 2I_{\{a_{m_1, j}(t) < \xi \leq a_{m_1, j}(t) + a_{m_2, j}(t)\}} \\ & + \dots + lI_{\{a_{m_1, j}(t) + \dots + a_{m_{l-1}, j}(t) < \xi \leq 1\}}. \end{aligned}$$

Now redefine the aggregated process  $\bar{\alpha}^\varepsilon(\cdot)$  by

$$\bar{\alpha}^\varepsilon(t) = \begin{cases} i, & \text{if } \alpha^\varepsilon(t) \in \mathcal{M}_i, \\ \xi_j(t), & \text{if } \alpha^\varepsilon(t) = s_{*j}(t). \end{cases} \tag{5.125}$$

Note that the state space of  $\bar{\alpha}^\varepsilon(t)$  is  $\bar{\mathcal{M}} = \{1, \dots, l\}$ , and that  $\bar{\alpha}^\varepsilon(\cdot) \in D([0, T]; \bar{\mathcal{M}})$ . Similar to the weak and strong interaction case, but with more effort, we can obtain the following result.

**Theorem 5.57.** *Under conditions (A5.9),  $\bar{\alpha}^\varepsilon(\cdot)$  converges weakly to  $\bar{\alpha}(\cdot)$ , a Markov chain generated by  $\bar{Q}_*(\cdot)$  given by (5.120).*

Next, for  $t \geq 0$ , and  $\alpha \in \mathcal{M}$ , let  $\beta_{ij}(t)$  be bounded Borel measurable deterministic functions, and let

$$W_{ij}(t, \alpha) = \begin{cases} (I_{\{\alpha=s_{ij}\}} - \nu_j^i(t)I_{\{\alpha \in \mathcal{M}_i\}})\beta_{ij}(t), & \text{if } i = 1, \dots, l, j = 1, \dots, m_i, \\ I_{\{\alpha=s_{*j}\}}\beta_{ij}(t), & \text{if } i = *, j = 1, \dots, m_*. \end{cases} \tag{5.126}$$

Consider the normalized occupation measure

$$n^\varepsilon(t) = (n_{11}^\varepsilon(t), \dots, n_{1m_1}^\varepsilon(t), \dots, n_{*1}^\varepsilon(t), \dots, n_{*m_*}^\varepsilon(t)),$$

where

$$n_{ij}^\varepsilon(t) = \frac{1}{\sqrt{\varepsilon}} \int_0^t W_{ij}(s, \alpha^\varepsilon(s)) ds.$$

We can then proceed to obtain the asymptotic distribution.

**Theorem 5.58.** *Assume (A5.9), and suppose  $\tilde{Q}(\cdot)$  is twice differentiable with Lipschitz continuous second derivative. Moreover,  $\hat{Q}(\cdot)$  is differentiable with Lipschitz continuous derivative. Let  $\beta_{ij}(\cdot)$  (for  $i = 1, \dots, l$ ,  $j = 1, \dots, m_i$ ) be bounded and Lipschitz continuous deterministic functions. Then  $n^\varepsilon(\cdot)$  converges weakly to a switching diffusion  $n(\cdot)$ , where*

$$n(t) = \left( \int_0^t \sigma(s, \bar{\alpha}(s)) dw(s) \right)', \tag{5.127}$$

where  $\sigma(s, i)$  is similar to (5.105) with the following modifications:

$$\sigma(s, i) = \text{diag}(0_{m_1 \times m_1}, \dots, \sigma^0(s, i), \dots, 0_{m_l \times m_l}, 0_{m_* \times m_*}) \tag{5.128}$$

and  $w(\cdot)$  is a standard  $m$ -dimensional Brownian motion.

Finally, we confirm that the case of the generators being merely measurable can be treated as well. We state this as the following theorem.

**Theorem 5.59.** *Assume the generator is given by (5.47) with  $\tilde{Q}(\cdot)$  given by (5.120) such that  $\tilde{Q}$  and  $\hat{Q}$  are measurable and bounded and that  $\tilde{Q}^i(t)$  is weakly irreducible for each  $i = 1, \dots, l$ . Then the following assertions hold:*

- For any  $i = 1, \dots, l$ ,  $j = 1, \dots, m_i$ , and bounded deterministic function  $\beta_{ij}(t)$ ,  $t \geq 0$ ,

$$E \left( \int_0^T (I_{\{\alpha^\varepsilon(t)=s_{ij}\}} - \nu_j^i(t) I_{\{\bar{\alpha}^\varepsilon(t)=i\}}) \beta_{ij}(t) dt \right)^2 \rightarrow 0, \text{ as } \varepsilon \rightarrow 0.$$

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

### 5.5.2 Inclusion of Absorbing States

Consider the Markov chain  $\alpha^\varepsilon(\cdot) \in \mathcal{M}$ , where the generator of  $\alpha^\varepsilon(\cdot)$  is still given by (5.47) with

$$\tilde{Q}(t) = \text{diag}(\tilde{Q}^1(t), \dots, \tilde{Q}^l(t), 0_{m_a \times m_a}), \tag{5.129}$$

where  $0_{m_a \times m_a}$  is the  $m_a \times m_a$  zero matrix, the state space of  $\alpha^\varepsilon(t)$  is given by

$$\mathcal{M} = \mathcal{M}_1 \cup \mathcal{M}_2 \cup \dots \cup \mathcal{M}_l \cup \mathcal{M}_a, \tag{5.130}$$

with  $\mathcal{M}_i = \{s_{i1}, \dots, s_{im_i}\}$  and  $\mathcal{M}_a = \{s_{a1}, \dots, s_{am_a}\}$ , and  $m_1 + m_2 + \dots + m_l + m_a = m$ . Assume the following conditions.

(A5.10) For all  $t \in [0, T]$  and  $i = 1, \dots, l$ ,  $\tilde{Q}^i(t)$  is weakly irreducible. Furthermore,  $\tilde{Q}(\cdot)$  is differentiable on  $[0, T]$  and its derivative is Lipschitz. Moreover,  $\hat{Q}(\cdot)$  is Lipschitz continuous on  $[0, T]$ .

Define

$$\begin{aligned} \tilde{\mathbb{1}} &= \text{diag}(\mathbb{1}_{m_1}, \dots, \mathbb{1}_{m_l}) \quad \text{and} \quad \tilde{\mathbb{1}}_a = \text{diag}(\tilde{\mathbb{1}}, I_{m_a}) \\ \overline{Q}(t) &= \text{diag}(\nu^1(t), \nu^2(t), \dots, \nu^l(t), I_{m_a}) \hat{Q}(t) \tilde{\mathbb{1}}_a. \end{aligned} \tag{5.131}$$

Assume that the conditions in (A5.10) are satisfied. Then we can prove the following:

(a) As  $\varepsilon \rightarrow 0$ ,

$$p^\varepsilon(t) = (\vartheta(t), \vartheta^a(t)) \text{diag}(\nu^1(t), \dots, \nu^l(t), I_{m_a}) + O(\varepsilon + \exp(-\kappa_0 t / \varepsilon)),$$

where

$$\vartheta(t) = (\vartheta^1(t), \dots, \vartheta^l(t)) \in \mathbb{R}^{1 \times l} \quad \text{and}$$

$$\vartheta^a(t) = (\vartheta_1^a(t), \dots, \vartheta_{m_a}^a(t)) \in \mathbb{R}^{1 \times m_a},$$

satisfying

$$\frac{d(\vartheta(t), \vartheta^a(t))}{dt} = (\vartheta(t), \vartheta^a(t)) \overline{Q}(t), \quad (\vartheta(0), \vartheta^a(0)) = p^\varepsilon(0) \tilde{\mathbb{1}}_a$$

where  $\overline{Q}(t)$  is given in (5.131) and  $p^\varepsilon(0) = (p^{\varepsilon,1}(0), \dots, p^{\varepsilon,l}(0), p^{\varepsilon,a}(0))$  with  $p^{\varepsilon,i}(0) \in \mathbb{R}^{1 \times m_i}$  and  $p^{\varepsilon,a}(0) \in \mathbb{R}^{1 \times m_a}$ .

(b) For the transition probability  $P^\varepsilon(t, t_0)$ , we have

$$P^\varepsilon(t, t_0) = P^0(t, t_0) + O(\varepsilon + \exp(-\kappa_0(t - t_0) / \varepsilon)), \tag{5.132}$$

for some  $\kappa_0 > 0$ , where

$$P^0(t, t_0) = \tilde{\mathbb{1}}_a \Theta(t, t_0) \text{diag}(\nu^1(t), \dots, \nu^l(t), I_{m_a}),$$

and

$$\frac{d\Theta(t, t_0)}{dt} = \Theta(t, t_0) \overline{Q}(t), \quad \Theta(t_0, t_0) = I.$$

To proceed, we aggregate the states in  $\mathcal{M}_i$  for  $i = 1, \dots, l$  as one state leading to the definition of the following process:

$$\bar{\alpha}^\varepsilon(t) = \begin{cases} i, & \text{if } \alpha^\varepsilon(t) \in \mathcal{M}_i, \\ \alpha^\varepsilon(t), & \text{if } \alpha^\varepsilon(t) \in \mathcal{M}_a. \end{cases} \tag{5.133}$$

For each  $j = 1, \dots, m_i$ , we also define a sequence of centered occupation measures by

$$\chi_{ij}^\varepsilon(t) = \begin{cases} \int_0^t (I_{\{\alpha^\varepsilon(s)=s_{ij}\}} - \nu_j^i(s)I_{\{\bar{\alpha}^\varepsilon(s)=i\}}) ds, & \text{for } i = 1, \dots, l, \\ \int_0^t (I_{\{\alpha^\varepsilon(s)=s_{aj}\}} - \vartheta_j^a(s)) ds. & \end{cases} \quad (5.134)$$

For  $t \geq 0$  and  $\alpha \in \mathcal{M}$ , let

$$W_{ij}(t, \alpha) = \begin{cases} I_{\{\alpha=s_{ij}\}} - \nu_j^i(t)I_{\{\alpha \in \mathcal{M}_i\}}, & \text{for } i = 1, \dots, l, j = 1, \dots, m_i, \\ I_{\{\alpha=s_{aj}\}} - \vartheta_j^a(t), & \text{for } j = 1, \dots, m_a. \end{cases} \quad (5.135)$$

Consider the normalized occupation measure

$$n^\varepsilon(t) = (n_{11}^\varepsilon(t), \dots, n_{1m_1}^\varepsilon(t), \dots, n_{l1}^\varepsilon(t), \dots, n_{lm_l}^\varepsilon(t), n_{a1}^\varepsilon(t), \dots, n_{am_a}^\varepsilon(t)),$$

where

$$n_{ij}^\varepsilon(t) = \begin{cases} \frac{1}{\sqrt{\varepsilon}} \int_0^t W_{ij}(s, \alpha^\varepsilon(s)) \beta_{ij}(s) ds, & i = 1, \dots, l, j = 1, \dots, m_i, \\ \int_0^t W_{aj}(s, \alpha^\varepsilon(s)) \beta_{aj}(s) ds, & j = 1, \dots, m_a. \end{cases}$$

Note that

$$\frac{dn^\varepsilon(t)}{dt} = \begin{cases} \frac{1}{\sqrt{\varepsilon}} W^r(t, \alpha^\varepsilon(t)), & \text{for } \alpha^\varepsilon(t) \in \mathcal{M}_1 \cup \dots \cup \mathcal{M}_l, \\ W^a(t, \alpha^\varepsilon(t)), & \text{for } \alpha^\varepsilon(t) \in \mathcal{M}_a, \end{cases}$$

$$n^\varepsilon(0) = 0,$$

where

$$W^r(t, \alpha) = (W_{11}(t, \alpha), \dots, W_{1m_1}(t, \alpha), \dots, W_{l1}(t, \alpha), \dots, W_{lm_l}(t, \alpha)),$$

$$W^a(t, \alpha) = (W_{a1}(t, \alpha), \dots, W_{am_a}(t, \alpha)), \text{ and}$$

$$W(t, \alpha) = (W^r(t, \alpha), W^a(t, \alpha)).$$

$$\langle W^a(u, \alpha), \nabla_x^\alpha f^0(x, \alpha) \rangle = \sum_{j=1}^{m_a} b_j(u, \alpha) \frac{\partial}{\partial a_j} f^0(x, \alpha).$$

We can obtain the following results.

**Theorem 5.60.** *Assume (A5.10). Then the following assertions hold.*

- (a) For all  $i = 1, \dots, l$  and  $j = 1, \dots, m_i$ , corresponding to the recurrent states,  $\sup_{t \in [0, T]} E|O_{ij}^\varepsilon(t)|^2 = O(\varepsilon)$ .
- (b)  $\bar{\alpha}^\varepsilon(\cdot)$  converges weakly to  $\bar{\alpha}(\cdot)$ , a Markov chain generated by  $\bar{Q}(\cdot)$ .
- (c) Define the generator  $\mathcal{L}$  by

$$\begin{aligned} \mathcal{L}f^0(x, \alpha) &= \frac{1}{2} \sum_{j_1, j_2=1}^{m_\alpha} a_{j_1 j_2}(s, \alpha) \partial_{\alpha, j_1 j_2}^2 f^0(x, \alpha) \\ &\quad + \sum_{j=1}^{m_a} b_j(s, \alpha) \partial_{a, j} f^0(x, \alpha) + \bar{Q}(s) f^0(x, \cdot)(\alpha). \end{aligned}$$

Then the sequence  $Y^\varepsilon(\cdot) = (n^\varepsilon(\cdot), \bar{\alpha}^\varepsilon(\cdot))$  converges weakly to  $\bar{Y}(\cdot) = (n(\cdot), \bar{\alpha}(\cdot))$  that is a solution of the martingale problem with operator  $\mathcal{L}$ .

Next, assume that  $\tilde{Q}(\cdot)$  and  $\hat{Q}(\cdot)$  are bounded and measurable and  $\tilde{Q}^i(t)$  for each  $i = 1, \dots, l$  is weakly irreducible. Then

$$p^\varepsilon(t) = (p_{i1}^\varepsilon(t), \dots, p_{i m_i}^\varepsilon(t), \dots, p_{i1}^\varepsilon(t), \dots, p_{i m_i}^\varepsilon(t), p_{a1}^\varepsilon(t), \dots, p_{a m_a}^\varepsilon(t))$$

converges in the weak topology of  $L^2([0, T]; \mathbb{R}^m)$  (with  $m = \sum_{i=1}^l m_i + m_a$ ) to

$$p(t) = (\vartheta_1(t)\nu^1(t), \dots, \nu^1(t)\vartheta_l(t), p^{0,a}),$$

where  $p^{0,a}$  is the subvector in the initial data  $p^0$  corresponding the absorbing state.

Note that in deriving the asymptotic distribution of the scaled occupation measures, we need to compute the asymptotic covariance of the limit process. That is, we need to evaluate the limit of

$$\begin{aligned} E \int_0^t \left( \frac{1}{\sqrt{\varepsilon}} (W^r(s, \alpha^\varepsilon(s)))' \right) \left( \frac{1}{\sqrt{\varepsilon}} W^r(s, \alpha^\varepsilon(s)), \quad W^a(s, \alpha^\varepsilon(s)) \right) ds \\ \stackrel{\text{def}}{=} \begin{pmatrix} W_\varepsilon^{rr}(t) & W_\varepsilon^{ra}(t) \\ W_\varepsilon^{ar}(t) & W_\varepsilon^{aa}(t) \end{pmatrix}, \end{aligned} \tag{5.136}$$

where

$$\begin{aligned} W_\varepsilon^{rr}(t) &= \frac{1}{\varepsilon} \int_0^t E(W^r(s, \alpha^\varepsilon(s)))' W^r(s, \alpha^\varepsilon(s)) ds \\ W_\varepsilon^{ra}(t) &= \frac{1}{\sqrt{\varepsilon}} \int_0^t E(W^r(s, \alpha^\varepsilon(s)))' W^a(s, \alpha^\varepsilon(s)) ds \\ W_\varepsilon^{ar}(t) &= \frac{1}{\sqrt{\varepsilon}} \int_0^t E(W^a(s, \alpha^\varepsilon(s)))' W^r(s, \alpha^\varepsilon(s)) ds \\ W_\varepsilon^{aa}(t) &= \int_0^t E(W^a(s, \alpha^\varepsilon(s)))' W^a(s, \alpha^\varepsilon(s)) ds. \end{aligned}$$



It can be shown that

$$W_\varepsilon^{rr}(t) \rightarrow \overline{W}^r(t), \quad W_\varepsilon^{ra}(t) \rightarrow 0, \quad W_\varepsilon^{ar}(t) \rightarrow 0, \quad \text{and} \quad W_\varepsilon^{aa}(t) \rightarrow \overline{W}^a(t),$$

as  $\varepsilon \rightarrow 0$ , where for  $i = 1, \dots, l$ ,  $\overline{W}^r(t) = \overline{W}^r(t, i) = \int_0^t \widehat{W}^r(s, i) ds$  with

$$\widehat{W}^r(s, i) = \text{diag}(0_{m_1 \times m_1}, \dots, \sigma(s, i), \dots, 0_{m_l \times m_l}) \tag{5.137}$$

with  $\sigma(s, i)$  the  $m_i \times m_i$  matrix such that  $\sigma(s, i)\sigma'(s, i) = A(s, i)$  for  $i = 1, \dots, l$ , and

$$\overline{W}^a(t) = (\overline{W}_{jk}^a(t)) \quad \text{with} \quad \overline{W}_{jk}^a(t) = \int_0^t (\delta_{jk} \vartheta_j^a(s) - \vartheta_j^a(s) \vartheta_k^a(s)) ds, \tag{5.138}$$

where  $\delta_{jk} = 1$  if  $j = k$ ,  $\delta_{jk} = 0$  if  $j \neq k$ . The detailed proof of Theorem 5.60 can be found in Yin, Zhang, and Badowski [241].

## 5.6 Remarks on a Stability Problem

So far, our study has been devoted to systems with two time scales in a finite interval. In many problems arising in networked control systems, stability is often a main concern. A related problem along this line is in Badowski and Yin [5].

It is interesting to note that intuitive ideas sometimes are not necessarily true for systems with switching, for example, if one put together two stable systems by using, for instance, Markovian switching. Our intuition may lead to the conclusion that the combined systems should also be stable. Nevertheless, this is, in fact, not true. Such an idea was illustrated in Wang, Khargonekar, and Beydoun [212] for deterministically switched systems; see also Chapter 1 of this book concerning this matter.

As a variation of the system in [212], we consider the following example. Suppose that  $\alpha^\varepsilon(\cdot)$  is a continuous-time Markov chain with state space

$$\mathcal{M} = \{1, 2\} \quad \text{and} \quad \text{generator} \quad Q^\varepsilon = Q/\varepsilon, \quad \text{where} \quad Q = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}. \quad \text{Consider}$$

a controlled system

$$\dot{x} = A(\alpha^\varepsilon(t))x + B(\alpha^\varepsilon(t))u(t),$$

with state feedback  $u(t) = K(\alpha^\varepsilon(t))x(t)$ . Then we obtain the equivalent representation

$$\dot{x} = [A(\alpha^\varepsilon(t)) - B(\alpha^\varepsilon(t))K(\alpha^\varepsilon(t))]x. \tag{5.139}$$

Suppose that

$$G(1) = A(1) - B(1)K(1) = \begin{pmatrix} -100 & 20 \\ 200 & -100 \end{pmatrix},$$

$$G(2) = A(2) - B(2)K(2) = \begin{pmatrix} -100 & 200 \\ 20 & -100 \end{pmatrix}.$$

Note that both matrices are Hurwitz (i.e., their eigenvalues have negative real parts). A question of interest is this: Is system (5.139) stable? The key to understanding the system is to examine

$$\dot{x}^\varepsilon(t) = G(\alpha^\varepsilon(t))x^\varepsilon(t), \quad (5.140)$$

where both  $G(1)$  and  $G(2)$  are stable matrices.

Since  $Q$  is irreducible, the stationary distribution associated with  $Q$  is given by  $(1/2, 1/2)$ . As a result, as  $\varepsilon \rightarrow 0$ , using our weak convergence result,  $x^\varepsilon(\cdot)$  converges weakly to  $x(\cdot)$ , which is a solution of the system

$$\dot{x}(t) = \bar{G}x(t), \quad \text{where}$$

$$\bar{G} = \frac{1}{2}(G(1) + G(2)) = \begin{pmatrix} -100 & 110 \\ 110 & -100 \end{pmatrix}. \quad (5.141)$$

In addition, for any  $T < \infty$ , using the large deviations result obtained in He, Yin, and Zhang [84], we can show that for any  $\delta > 0$ , there is a  $c_1 > 0$  such that

$$P(\rho_{0,T}(x^\varepsilon(t), x(t)) \geq \delta) \leq \exp(-c_1/\varepsilon), \quad (5.142)$$

where  $\rho_{0,T}(x, y) = \sup_{0 \leq t \leq T} |x(t) - y(t)|$ .

Note that  $\bar{G}$  is an unstable matrix with eigenvalues  $-210$  and  $10$ . Thus for (5.141), the critical point  $(0, 0)'$  is a saddle point. But why should the stability of the averaged system dominate that of the original system? To see this, from a result of differential equations, there is a nonsingular matrix  $H$  such that  $H\bar{G}H^{-1} = \Lambda = \text{diag}(-210, 10)$ . Clearly, the stability of (5.141) is equivalent to that of

$$\dot{y}(t) = \Lambda y(t) = H \sum_{i=1}^2 \nu_i G(i) H^{-1} y(t) = \text{diag}(-210, 10) y(t), \quad (5.143)$$

where  $y = Hx = (y_1, y_2)'$ . The stability of (5.141) is equivalent to that of (5.143), which is completely decoupled and  $y_1(t) = \exp(-210t)y_1(0) \rightarrow 0$

and  $y_2(t) = \exp(10t)y_2(0) \rightarrow \infty$ . To see how the original system (5.140) behaves, we apply the same transformation to get

$$\dot{y}^\varepsilon(t) = H \sum_{i=1}^2 I_{\{\alpha^\varepsilon(t)=i\}} G(i) H^{-1} y^\varepsilon(t). \tag{5.144}$$

For the transformed system (5.143), by choosing  $V(y) = y_2^2/2$ , we obtain  $\dot{V}(y(t)) = 10y_2^2 > 0$  for all  $y_2 \neq 0$ . Define  $L^\varepsilon z(t) = \lim_{\delta \downarrow 0} E_t^\varepsilon[z(t + \delta) - z(t)]/\delta$  for a real-valued function  $z(t)$  that is continuously differentiable, where  $E_t^\varepsilon$  denotes the conditioning on the  $\mathcal{F}_t^\varepsilon = \sigma\{\alpha^\varepsilon(s) : s \leq t\}$ . With  $V(y) = y_2^2/2$ , we have

$$L^\varepsilon V(y^\varepsilon(t)) = 10(y_2^\varepsilon(t))^2 + V'_y(y^\varepsilon(t)) H \sum_{i=1}^2 [I_{\{\alpha^\varepsilon(t)=i\}} - \nu_i] G(i) H^{-1} y^\varepsilon(t),$$

where  $V'_y(y) = (0, y_2) \in \mathbb{R}^{1 \times 2}$ . Using perturbed Liapunov function techniques as done in Badowski and Yin [5], define a perturbation

$$V_2^\varepsilon(y, t) = E_t^\varepsilon \int_t^\infty e^{t-s} V'_y(y) H \sum_{i=1}^2 [I_{\{\alpha^\varepsilon(s)=i\}} - \nu_i] G(i) H^{-1} y.$$

It can be shown that  $V_2^\varepsilon(y, t) = O(\varepsilon)V(y)$ . In addition,

$$\begin{aligned} L^\varepsilon V_2^\varepsilon(y^\varepsilon(t), t) &= -V'_y(y^\varepsilon(t)) H \sum_{i=1}^2 [I_{\{\alpha^\varepsilon(t)=i\}} - \nu_i] G(i) H^{-1} y^\varepsilon(t) \\ &\quad + O(\varepsilon)V(y^\varepsilon(t)). \end{aligned}$$

Define

$$V^\varepsilon(y, t) = V(y) + V_2^\varepsilon(y, t).$$

Evaluate  $L^\varepsilon V^\varepsilon(y^\varepsilon(t), t)$ . Upon cancelation, for sufficiently small  $\varepsilon$ , we can make

$$O(\varepsilon)V(y^\varepsilon(t)) \geq -(y_2^\varepsilon(t))^2.$$

It then follows that

$$\begin{aligned} L^\varepsilon V^\varepsilon(y^\varepsilon(t), t) &= 10(y_2^\varepsilon(t))^2 + O(\varepsilon)V(y^\varepsilon(t)) \\ &\geq 9(y_2^\varepsilon(t))^2. \end{aligned}$$

Taking expectation of the left- and right-hand sides above leads to

$$\frac{d}{dt} E|y_2^\varepsilon(t)|^2 \geq 9E|y_2^\varepsilon(t)|^2,$$

which in turn yields that

$$E(y_2^\varepsilon(t))^2 \geq E(y_2^\varepsilon(0))^2 \exp(9t) \rightarrow \infty \text{ as } t \rightarrow \infty.$$

Similar to the previous development, choose  $V(y) = y_1^2/2$ , define

$$V_1^\varepsilon(y, t) = E_t^\varepsilon \int_t^\infty e^{t-s} V_y'(y) H \sum_{i=1}^2 [I_{\{\alpha^\varepsilon(s)=i\}} - \nu_i] G(i) H^{-1} y,$$

and redefine

$$V^\varepsilon(y, t) = V(y) + V_1^\varepsilon(y, t).$$

Using the upper bound  $O(\varepsilon)V(y^\varepsilon(t)) \leq (y_1^\varepsilon(t))^2$  this time and calculating  $L^\varepsilon V^\varepsilon(y^\varepsilon(t), t)$ , we obtain

$$\frac{d}{dt} E|y_1^\varepsilon(t)|^2 \leq -209E|y_1^\varepsilon(t)|^2,$$

which in turn yields that

$$E(y_1^\varepsilon(t))^2 \leq E(y_1^\varepsilon(0))^2 \exp(-209t) \rightarrow 0 \text{ as } t \rightarrow \infty.$$

This yields that (5.144) and hence (5.140) are unstable in probability (see Yin and Zhu [244, p. 220] for a definition). In fact, it can be seen that the trivial solution of the original system is also a saddle.

In the same spirit of the last example, consider a system given by

$$\begin{aligned} \dot{x}^\varepsilon(t) &= G(\alpha^\varepsilon(t))x^\varepsilon(t), \quad \alpha^\varepsilon(t) \sim Q/\varepsilon, \quad \text{where} \\ G(1) &= \begin{pmatrix} -\frac{7}{3} & -1 \\ 0 & 1 \end{pmatrix}, \quad G(2) = \begin{pmatrix} 1 & 0 \\ -1 & -\frac{7}{3} \end{pmatrix}, \end{aligned} \tag{5.145}$$

where  $Q$  is as in the last example. Then it can be shown that  $x^\varepsilon(\cdot)$  converges weakly to  $x(\cdot)$  that is a solution of the following system

$$\begin{aligned} \dot{x}(t) &= \overline{G}x(t), \\ \overline{G} &= \begin{pmatrix} -\frac{4}{3} & -1 \\ -1 & -\frac{4}{3} \end{pmatrix}. \end{aligned} \tag{5.146}$$

Neither  $G(1)$  nor  $G(2)$  is a stable matrix, but the system (5.146) is a stable one. The stability analysis is again carried out using perturbed Liapunov function methods. Here exactly the same kind of argument as in [5] can be applied. Using the techniques of perturbed Liapunov functions, we can

show that the stability of the averaged system “implies” that of the original system.

These two examples illustrate that one can combine two stable systems using Markovian switching to produce an unstable limit system. Likewise, one can combine two unstable systems to produce a limit stable systems. More importantly, using our weak convergence result of this chapter and the large deviations results in He, Yin, and Zhang [84], combined with the perturbed Liapunov function argument, we can give the reason why such a thing can happen.

## 5.7 Notes

This chapter concerns sequences of functional occupation measures. It includes convergence of an unscaled sequence (in probability) and central-limit-type results for suitably scaled sequences. For a general introduction to central limit theorems, we refer to the book by Chow and Teicher [30] and the references therein. In the stationary case, that is,  $Q(t) = Q$ , a constant matrix, the central limit theorem may be obtained as in Friedlin and Wentzell [67]. Some results of central limit type for discrete Markov chains are in Dobrushin [50] (see also the work of Linnik on time-inhomogeneous Markov chains [147]). Work in the context of random evolution, with primary concern the central limit theorem involving a singularly perturbed Markov chain, is in Pinsky [176]; see also Kurtz [135, 137] for related discussions and the martingale problem formulation. Exponential bounds for Markov processes are quite useful in analyzing the behavior of the underlying stochastic processes. Some results in connection with diffusions can be found in Kallianpur [102]. Corollary 5.8 can be viewed as a large deviations result. For extensive treatment of large deviations, see Varadhan [207].

The central theme here is limit results of unscaled as well as scaled sequences of occupation measures, which include the law of large numbers for an unscaled sequence, exponential upper bounds, and asymptotic distribution of a suitably scaled sequence of occupation times. Results in Section 5.2 are based on the paper of Zhang and Yin [252]; however, a somewhat different approach to the central limit theorem was used in [252]. Some of the results in Section 5.3 are based on Zhang and Yin [253]. The result on exponential error bound in Section 5.3 is a natural extension for the irreducible generators. Such result holds uniformly in  $t \in [0, T]$  for fixed but otherwise arbitrary  $T > 0$ . The main motivation for treating  $T$  as a parameter stems from various control and optimization problems with discounted cost over the infinite horizon. In such a situation, the magnitude of the bound counts. Thus detailed information on the bounding constant is helpful for dealing with the near optimality of the underlying problem. Section 5.3 also presents a characterization of the limit process using martingale problem formulations. Much of the foundation of this useful approach is

in the work of Stroock and Varadhan [203]. Using perturbed operators to study limit behavior may be traced back to Kurtz [135]. The general idea of perturbed test functions was used in Blankenship and Papanicolaou [16], and Papanicolaou, Stroock, and Varadhan [168]. It was further developed and extended by Kushner [139] for various stochastic systems, and singularly perturbed systems in Kushner [140]; see also Kushner and Yin [145] for related stochastic approximation problems, and Ethier and Kurtz [59] and Kurtz [137] for related work in stochastic processes. The results of this section have benefited from the discussion with Thomas Kurtz, who suggested treating the pair of processes  $(n^\varepsilon(\cdot), \alpha^\varepsilon(\cdot))$  together, which led to the current version. Earlier treatment of a pair of processes may be found in the work of Kesten and Papanicolaou [110] for stochastic acceleration.

The results on asymptotic properties for the inclusion of transient states can be found in Yin, Zhang, and Badowski [239]; the results for the case of generators being measurable can be found in the work of Yin, Zhang, and Badowski [240]; the results on asymptotic properties of occupation measures with absorbing states can be found in Yin, Zhang, and Badowski [241].