SEVERAL TYPES OF ERGODICITY FOR M/G/1-TYPE MARKOV CHAINS AND MARKOV PROCESSES

YUANYUAN LIU * ** AND

ZHENTING HOU,* *** *Central South University, Changsha*

Abstract

In this paper we study polynomial and geometric (exponential) ergodicity for M/G/1-type Markov chains and Markov processes. First, practical criteria for M/G/1-type Markov chains are obtained by analyzing the generating function of the first return probability to level 0. Then the corresponding criteria for M/G/1-type Markov processes are given, using their *h*-approximation chains. Our method yields the radius of convergence of the generating function of the first return probability, which is very important in obtaining explicit bounds on geometric (exponential) convergence rates. Our results are illustrated, in the final section, in some examples.

Keywords: Markov chain; Markov process; polynomial ergodicity; geometric ergodicity; strong ergodicity

2000 Mathematics Subject Classification: Primary 60J10
Secondary 60J27

1. Introduction

We consider homogeneous discrete-time Markov chains and continuous-time Markov processes on a countable state space having an M/G/1-type structure, i.e. the transition matrix of the chains and the infinitesimal generator of the processes are upper block Hessenberg with a repetitive structure. For the M/G/1-type structure, some authors (see, e.g. [9]) have coined the descriptive term *skip-free to the left*. M/G/1-type Markov chains, introduced in [19, pp. 2–5] and [20, Chapter 2], model a very large variety of queueing problems. Continuous-time M/G/1-type Markov processes have also been studied extensively (see, e.g. [10, pp. 268–275] and [21]). Such processes are often the modeling tool of choice for modern computer and communications systems [16, pp. 6–8], [24]. Because of their special structure, it has been possible to derive many elegant results. The condition sufficient and necessary for (ordinary) ergodicity (i.e. the existence of the invariant probability vector) of the chains and processes has been found. Subsequently, based on (ordinary) ergodicity, there appeared many research works devoted to the study of the computation of the invariant probability vector (see, e.g. [14]) and the tail asymptotics of the stationary distribution (see, e.g. [4]). Here, we investigate the rates of convergence of the transition functions to the stationary distribution for M/G/1-type Markov chains and Markov processes.

The criteria of polynomial (see [6] and [25]), geometric, and strong (see [7, Chapter 1] and [15, Chapters 15 and 16]) ergodicity for the embedded M/G/1 queue were given in [5] by analyzing the generating function of the first hitting time, and later, in [12], the largest geometric

---

convergence rate was obtained. Note that the embedded chain is a typical numeric Markov chain of M/G/1 type. Zeifman [27] studied the exponential ergodicity (see [1, Chapter 6] and [3, Chapter 4]) and investigated the bounds on the convergence rates of birth–death processes, including the random walk in continuous time. Note that the random walk is a numeric Markov process of M/G/1 type. As the continuation of [5] and [12], this paper studies these ergodicities for more extensive M/G/1-type Markov chains and Markov processes, respectively, using the matrix-analytic method.

As a preliminary, we first review the definitions and criteria of the several types of ergodicity. Lemmas 2.1, 2.2, and 2.3 reveal the relations between the ergodicity of a Markov process and that of its h-approximation chain. By combining the matrix-analytic method with the usual Foster–Lyapunov method, we obtain the criteria for polynomial and geometric ergodicity in Section 3, mainly by directly analyzing the generating function of the first return probability to level 0, using the matrix-analytic method. Moreover, the radius of convergence of the generating function $\tilde{G}(z)$ is derived in Theorem 3.2. This is of great importance when investigating explicit geometric convergence rates. In Section 4 we obtain the corresponding criteria for M/G/1-type processes, using their h-approximation chains. The results about strong ergodicity are simple: none of the M/G/1-type Markov chains and Markov processes are strongly ergodic. All the criteria are easy to check and widely applicable. To illustrate our results, we present three examples in Section 5, and in each investigate the explicit bounds on the geometric (exponential) convergence rates.

2. Preliminaries

First, we review the definitions of the several types of ergodicity. Let $\mathbb{Z}_+ = \{0, 1, 2, \dots\}$ be the set of nonnegative integers, let $\mathbb{N}_+ = \{1, 2, \dots\}$ be the set of positive integers, and let $\mathbb{R}_+ = [0, \infty)$ be the set of nonnegative real numbers. In this paper we consider only the usual total variation norm: for a signed measure μ, we write $\|\mu\| = \sup_{|g| \leq 1} |\mu(g)|$. Let $\{X_n, n \in \mathbb{Z}_+\}$ be an irreducible aperiodic discrete-time Markov chain on a countable state space $(E, \mathcal{F})$, with stochastic transition matrix $\boldsymbol{P} = (P(i, j), i, j \in E)$. Define $\tau_H = \inf\{n \geq 1 \colon X_n \in H\}$ to be the first hitting time at a nonempty finite set H. Denote by P_i the conditional probability of the chain X_n when starting from i, and by E_i the corresponding conditional expectation. For $n \geq 1$, define

$$P^n(i, j) = \mathrm{P}\{X(n) = j \mid X(0) = i\}.$$

The superscript will be omitted when $n = 1$. The chain X_n is said to be (ordinarily) ergodic if there exists a probability measure π such that

$$\|P^n(i, \cdot) - \pi(\cdot)\| = \sum_{j \in E} |P^n(i, j) - \pi(j)| \to 0, \qquad n \to \infty,$$

for all $i \in E$. An ergodic chain X_n is said to be geometrically ergodic if, for positive constants $\rho_G < 1$ and $D_i < \infty$,

$$\|P^n(i, \cdot) - \pi(\cdot)\| \leq D_i \rho_G^n$$

holds for all $n \in \mathbb{Z}_+$ and $i \in E$, and is said to be strongly ergodic if

$$\sup_{i \in E} \|P^n(i, \cdot) - \pi(\cdot)\| \to 0 \quad \text{as } n \to \infty.$$

For $l \in \mathbb{N}_+$, the chain X_n is called l-ergodic if there exists some finite nonempty set H such that $\mathrm{E}_i[(\tau_H)^l] < \infty$ for all $i \in H$. If X_n is l-ergodic then it is ergodic and, for all $i \in E$,

$$n^{l-1}\|P^n(i, \cdot) - \pi(\cdot)\| \to 0, \qquad n \to \infty,$$

Now we review the criteria for exponential (geometric) and strong ergodicity for Markov chains and Markov processes, which will be used to prove Lemmas 2.2 and 2.3. These criteria are given by, for example, Theorems 4.31 and 4.45 of [3] with a slight change in form.

(i) *Drift criterion for geometric ergodicity.* There exists a finite vector V (with $V_i \geq 1$ for all $i \in E$), some finite set C, and positive constants $\beta_\mathrm{G} < 1$ and $b_\mathrm{G} < \infty$ such that

$$\sum_{j=0}^{\infty} P(i, j)V_j \leq (1 - \beta_\mathrm{G})V_i + b_\mathrm{G}1_C(i), \qquad i \in E. \tag{2.1}$$

Here $1_C(i)$ is the indicator function for set C, which can in fact be any finite set.

(ii) *Drift criterion for strong ergodicity.* There exists a bounded, nonnegative vector V, a finite set C, and a positive constant $b < \infty$ such that

$$\sum_{j=0}^{\infty} P(i, j)V_j \leq V_i - 1 + b1_C(i), \qquad i \in E. \tag{2.2}$$

Let $\{X_t, \ t \in R_+\}$ be an irreducible continuous-time Markov process on a countable state space $(E, \mathcal{F})$, with transition function $P^t(i, j)$ and regular intensity matrix $Q = (q_{ij})$. The process X_t is said to be (ordinarily) ergodic if there exists a probability measure π such that

$$\|P^t(i, \cdot) - \pi(\cdot)\| \to 0, \qquad t \to \infty,$$

for all $i \in E$. An ergodic Markov process is said to be exponentially ergodic if, for positive constants $\rho_\mathrm{E} < 1$ and $D_i < \infty$,

$$\|P^t(i, \cdot) - \pi(\cdot)\| \leq D_i \mathrm{e}^{-\rho_\mathrm{E}t}$$

holds for all $t \geq 0$ and $i \in E$, and is said to be strongly ergodic if

$$\sup_{i \in E} \|P^t(i, \cdot) - \pi(\cdot)\| \to 0 \quad \text{as } t \to \infty.$$

(iii) *Drift criterion for exponential ergodicity.* There exists a finite vector V (with $V_i \geq 1$ for all $i \in E$), some finite set C, and positive constants $\beta_\mathrm{E} < \inf_{i \in E} q_i$ and $b_\mathrm{E} < \infty$ such that

$$\sum_{j=0}^{\infty} q_{ij}V_j \leq -\beta_\mathrm{E}V_i + b_\mathrm{E}1_C(i), \qquad i \in E. \tag{2.3}$$

Here, C can in fact be any finite set.

(iv) *Drift criterion for strong ergodicity.* There exists a bounded, nonnegative vector V, a finite set C, and a positive constant $b < \infty$ such that

$$\sum_{j=0}^{\infty} q_{ij}V_j \leq -1 + b1_C(i), \qquad i \in E. \tag{2.4}$$

Assume that X_t is uniformizable, i.e. that $\bar{q} := \sup_{i \in E} q_i < \infty$; then $\boldsymbol{Q}$ is regular. Let $h < \bar{q}^{-1}$ be the length of the time-discretization interval. Then the transition probabilities $P^h(i, j)$ have first-order approximations ${}^{\alpha(h)}P(i, j) = (\boldsymbol{I} + h\boldsymbol{Q})_{ij}$, $i, j \in E$. This first-order approximation defines the h-approximation chain $\{X^{\alpha(h)}(n), n \in \mathbb{Z}_+\}$ of the process X_t. From Proposition 2.10 of [1, Chapter 2], the following relation can be shown to hold:

$$P^t(i, j) = (\mathrm{e}^{t\boldsymbol{Q}})_{ij} = \sum_{n=0}^{\infty} {}^{\alpha(h)}P^n(i, j)\mathrm{e}^{-th^{-1}}\frac{(th^{-1})^n}{n!}.$$

In the following lemmas we derive the relations between these types of ergodicity for the Markov process and its h-approximation chain. These lemmas will be used in Section 4.

Lemma 2.1. *Let $l \in \mathbb{N}_+$. If, for the h-approximation chain,*

$$n^l\|{}^{\alpha(h)}P^n(i, \cdot) - \pi(\cdot)\| \to 0, \qquad n \to \infty, \tag{2.5}$$

for all $i \in E$, then, for the process X_t,

$$t^l\|P^t(i, \cdot) - \pi(\cdot)\| \to 0, \qquad t \to \infty, \tag{2.6}$$

for all $i \in E$.

Proof. From (2.5) we find that, for any $\varepsilon > 0$, there exists an $N \in \mathbb{N}_+$ such that

$$\|{}^{\alpha(h)}P^n(i, \cdot) - \pi(\cdot)\| \leq n^{-l}\frac{\varepsilon}{2[(1+l)h]^l}$$

for all $i \in E$ when $n > N$. Then

$$\begin{aligned}
t^l\|P^t(i, \cdot) - \pi(\cdot)\| &= t^l\left\|\sum_{n=0}^{\infty}({}^{\alpha(h)}P^n(i, \cdot) - \pi(\cdot))\mathrm{e}^{-th^{-1}}\frac{(th^{-1})^n}{n!}\right\| \\
&\leq t^l\sum_{n=0}^{\infty}\|{}^{\alpha(h)}P^n(i, \cdot) - \pi(\cdot)\|\mathrm{e}^{-th^{-1}}\frac{(th^{-1})^n}{n!} \\
&= t^l\sum_{n=0}^{N}\|{}^{\alpha(h)}P^n(i, \cdot) - \pi(\cdot)\|\mathrm{e}^{-th^{-1}}\frac{(th^{-1})^n}{n!} \\
&\quad + t^l\sum_{n=N+1}^{\infty}n^{-l}\frac{\varepsilon}{2[(1+l)h]^l}\mathrm{e}^{-th^{-1}}\frac{(th^{-1})^n}{n!} \\
&=: I_1(t) + I_2(t). \tag{2.7}
\end{aligned}$$

Since $\lim_{t \to \infty} I_1(t) = 0$, there exists some $T > 0$ such that

$$I_1(t) < \frac{\varepsilon}{2} \tag{2.8}$$

for $t > T$. Obviously, we have

$$n^l n! \leq (n+l)! \leq (n+l)^l n! \leq (1+l)^l n^l n!$$

for any $n \in \mathbb{N}_+$, which implies that

$$I_2(t) \le t^l \frac{\varepsilon(1+l)^l}{2[(1+l)h]^l} \sum_{n=N+1}^{\infty} \frac{\mathrm{e}^{-th^{-1}}(th^{-1})^n}{(n+l)!} = \frac{\varepsilon \mathrm{e}^{-th^{-1}}}{2} \sum_{n=N+1}^{\infty} \frac{(th^{-1})^{n+l}}{(n+l)!} \le \frac{\varepsilon}{2}. \qquad (2.9)$$

From (2.7)–(2.9) we see that, for all $t > T$ and $i \in E$,

$$t^l \|P^t(i, \cdot) - \pi(\cdot)\| < \varepsilon.$$

Hence, (2.6) holds. $\qquad \square$

The following lemma is adapted from Lemma 2.7 and Theorem 2.2 of [23]. For the reader's convenience, we nevertheless give the proof.

Lemma 2.2. *The h-approximation chain is geometrically ergodic if and only if the process X_t is also. Furthermore, if $\rho_{\mathrm{G}}(h)$ is a feasible geometric convergence rate of the h-approximation chain, then $(1 - \rho_{\mathrm{G}}(h))/h$ is a feasible exponential convergence rate of the process X_t.*

Proof. Let V be a solution to (2.1) for a finite set C, a positive constant $b_{\mathrm{G}} < \infty$, and some parameter $\beta_{\mathrm{G}} < 1$, for the h-approximation chain. Fix a positive number $\alpha < \inf_{i \in E} q_i$; then V is also the solution to (2.3) for the same set C, a positive constant $b_{\mathrm{E}} = h^{-1}b_{\mathrm{G}}$, and a parameter $\beta_{\mathrm{E}} = \alpha\beta_{\mathrm{G}}$. Thus, if the h-approximation chain is geometrically ergodic then the process X_t is exponentially ergodic. The converse statement follows along the same lines.

Furthermore, if $\rho_{\mathrm{G}}(h)$ is a feasible geometric convergence rate of the h-approximation chain, i.e. if

$$\|^{\alpha(h)}P^n(i, \cdot) - \pi(\cdot)\| \le D_i \rho_{\mathrm{G}}(h)^n,$$

then

$$\|P^t(i, \cdot) - \pi(\cdot)\| \le \sum_{n=0}^{\infty} \|^{\alpha(h)}P^n(i, \cdot) - \pi(\cdot)\| \mathrm{e}^{-th^{-1}} \frac{(th^{-1})^n}{n!}$$

$$\le D_i \exp[-(1 - \rho_{\mathrm{G}}(h))th^{-1}],$$

implying that $\|P^t(i, \cdot) - \pi(\cdot)\| \le D_i \mathrm{e}^{-\rho_{\mathrm{E}}t}$ for $\rho_{\mathrm{E}} = (1 - \rho_{\mathrm{G}}(h))h^{-1}$, i.e. that ρ_{E} is a feasible exponential convergence rate for the process. $\qquad \square$

Lemma 2.3. *The h-approximation chain is strongly ergodic if and only if the process X_t is strongly ergodic.*

Proof. Let V be a solution to (2.2) for a finite set C and a positive constant $b < \infty$, for the h-approximation chain. It is then easy to check that hV is the solution to (2.4) for the same C and b. Thus, if the h-approximation chain is strongly ergodic then so is the process. The converse statement follows along the same lines. $\qquad \square$

3. M/G/1-type Markov chains

We consider an irreducible aperiodic M/G/1-type Markov chain X_n whose transition matrix P is partitioned into block form:

$$P = \begin{pmatrix} B_0 & B_1 & B_2 & B_3 & \cdots \\ C_0 & A_1 & A_2 & A_3 & \cdots \\ 0 & A_0 & A_1 & A_2 & \cdots \\ 0 & 0 & A_0 & A_1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

Here, the matrices $\{A_n, \ n \in \mathbb{Z}_+\}$ are square matrices of order m and the matrix B_0 is a square matrix of order m_1. The matrices $\{B_n, \ n \in \mathbb{N}_+\}$ and C_0 are respectively of dimensions $m_1 \times m$ and $m \times m_1$. We assume that P is a stochastic transition matrix, i.e. that

$$\sum_{n=0}^{\infty} A_n e = e, \qquad \sum_{n=0}^{\infty} B_n e = e, \qquad C_0 e + \sum_{n=1}^{\infty} A_n e = e,$$

where e is a column vector of 1s. The state space of the above block-partitioned Markov chain can be expressed as $E = \{\bigcup_{i=0}^{\infty} L_i\}$, where $L_0 = \{(0, j), \ j = 1, 2, \ldots, m_1\}$ and $L_i = \{(i, j), j = 1, 2, \ldots, m\}$ denote the level 0 and the level i, respectively.

Let $G_{jj'}(k)$ be the conditional probability that the chain X_n, when starting in the state $(i + 1, j)$, $i \geq 1, 1 \leq j \leq m$, first reaches the level i, after exactly $k \in \mathbb{N}_+$ transitions, by hitting the state (i, j'), $i \geq 1, 1 \leq j' \leq m$. The matrices $G(k)$ take the $G_{jj'}(k)$ as entries and are of order m. Define $\tilde{G}(z) = \sum_{k=1}^{\infty} G(k) z^k$, $z \in \mathbb{R}$. The following proposition is just a simple extension of Theorem 2.2.2 of [20]: we enlarge the range of z from $-1 \leq z \leq 1$ to $-\infty < z < \infty$. It plays an important role in our analysis.

Proposition 3.1. *For any $z \in (-\infty, \infty)$, $\tilde{G}(z)$ satisfies the following equation:*

$$\tilde{G}(z) = \sum_{v=0}^{\infty} z A_v \tilde{G}^v(z). \tag{3.1}$$

Moreover, when z only takes values in the range $[0, \infty)$, $\tilde{G}(z)$ is the minimal nonnegative solution to (3.1).

Let $L_{jj'}(k)$ be the conditional probability that the chain X_n, when starting in the state $(1, j), 1 \leq j \leq m$, first reaches the level 0, after exactly $k \in \mathbb{N}_+$ transitions, by hitting the state $(0, j'), 1 \leq j' \leq m_1$. The matrices $L(k)$ take the $L_{jj'}(k)$ as entries and are of dimensions $m \times m_1$. Let $K_{jj'}(k)$ be the conditional probability that the chain X_n, when starting in the state $(0, j), 1 \leq j \leq m_1$, first returns to the level 0, after exactly $k \in \mathbb{N}_+$ transitions, by hitting the state $(0, j'), 1 \leq j' \leq m_1$. The matrices $K(k)$ take the $K_{jj'}(k)$ as entries and are of dimensions $m_1 \times m_1$. Define $\tilde{L}(z) = \sum_{k=1}^{\infty} L(k) z^k$ and $\tilde{K}(z) = \sum_{k=1}^{\infty} K(k) z^k$. In the following proposition we derive important relations between $\tilde{G}(z)$, $\tilde{L}(z)$, and $\tilde{K}(z)$; these are the simple extensions of Equations (2.4.2) and (2.4.8) of [20, pp. 107–109] from the range $-1 \leq z \leq 1$ to the range $-\infty < z < \infty$.

Proposition 3.2. *For any $z \in (-\infty, \infty)$, we obtain*

$$\tilde{L}(z) = z C_0 + \sum_{v=1}^{\infty} z A_v \tilde{G}^{v-1}(z) \tilde{L}(z), \tag{3.2}$$

$$\tilde{K}(z) = z B_0 + \sum_{v=1}^{\infty} z B_v \tilde{G}^{v-1}(z) \tilde{L}(z). \tag{3.3}$$

3.1. Polynomial ergodicity

In [20], a basic assumption is that $G := \sum_{k=1}^{\infty} G(k)$ is irreducible. The elements $G_{jj'}$, $1 \leq j, j' \leq m$, of the matrix G are the conditional probabilities that the chain X_n will eventually hit the level i in the state (i, j'), given that it starts in the state $(i + 1, j)$, $i \geq 1$. Let $\beta = \sum_{v=0}^{\infty} v A_v e$ and let μ be the invariant probability vector of $A := \sum_{v=0}^{\infty} A_v$ (i.e. $\mu = \mu A$

and $\boldsymbol{\mu e} = 1$). Note that if $\boldsymbol{G}$ is irreducible, then so is $\boldsymbol{A}$ and that if $\boldsymbol{A}$ is irreducible, then $\boldsymbol{G}$ is stochastic if and only if $\lambda := \boldsymbol{\mu\beta} \le 1$. When $\boldsymbol{G}$ is irreducible, Theorem 3.2.1 of [20] states that X_n is ergodic if and only if $\lambda < 1$ and the matrix $\sum_{v=1}^{\infty} v\boldsymbol{B}_v$ is finite. In our approach, we also assume that $\boldsymbol{G}$ is known and irreducible. This assumption is reasonable because $\boldsymbol{G}$ can be computed by some efficient methods (see [2] and [14]).

For $r \in \mathbb{N}_+$, define $\tilde{\boldsymbol{M}}_r = \tilde{\boldsymbol{G}}^{(r)}(z)|_{z=1}$ to be the value of the rth derivative of $\tilde{\boldsymbol{G}}(z)$ at $z = 1$. Define the sequence $\{\boldsymbol{V}_n^{\langle r \rangle}(\boldsymbol{X}), r \in \mathbb{N}_+, n \ge r\}$, for any square matrix $\boldsymbol{X}$ of the same order as $\boldsymbol{G}$, as follows:

$$\boldsymbol{V}_n^{\langle 1 \rangle}(\boldsymbol{X}) = \sum_{k=0}^{n-1} \boldsymbol{G}^k \boldsymbol{X} \boldsymbol{G}^{n-k-1}, \qquad\qquad n \ge 1,$$

$$\boldsymbol{V}_n^{\langle 2 \rangle}(\boldsymbol{X}) = \sum_{k_1=0}^{n-2} \sum_{k_2=0}^{k_1} \boldsymbol{G}^{k_2} \boldsymbol{X} \boldsymbol{G}^{k_1-k_2} \boldsymbol{X} \boldsymbol{G}^{n-2-k_1}, \qquad\qquad n \ge 2,$$

$$\boldsymbol{V}_n^{\langle r \rangle}(\boldsymbol{X}) = \sum_{k_1=0}^{n-r} \sum_{k_2=0}^{k_1} \cdots \sum_{k_r=0}^{k_1-\sum_{i=2}^{r-1} k_i} \boldsymbol{G}^{k_2} \boldsymbol{X} \boldsymbol{G}^{k_3} \boldsymbol{X} \cdots \boldsymbol{G}^{k_r} \boldsymbol{X} \boldsymbol{G}^{k_1-\sum_{i=2}^{r} k_i} \boldsymbol{X} \boldsymbol{G}^{n-r-k_1}, \quad r \ge 3, n \ge r.$$

Lemma 3.1. *Let $r \in \mathbb{N}_+$. If $\boldsymbol{G}$ is irreducible and $\lambda < 1$, then*

$$\lim_{n\to\infty} n^{-r} \boldsymbol{V}_n^{\langle r \rangle}(\boldsymbol{X}) = (\boldsymbol{g} \boldsymbol{X} \boldsymbol{e})^r \hat{\boldsymbol{G}},$$

where the vector $\boldsymbol{g}$ is the invariant probability vector of $\boldsymbol{G}$ (i.e. $\boldsymbol{g} = \boldsymbol{g}\boldsymbol{G}$ and $\boldsymbol{g}\boldsymbol{e} = 1$) and $\hat{\boldsymbol{G}}$ is a square matrix whose rows are all equal to $\boldsymbol{g}$. Furthermore, $\sum_{n=r}^{\infty} \boldsymbol{D}_n \boldsymbol{V}_n^{\langle r \rangle}(\tilde{\boldsymbol{M}}_1) < \infty$ if and only if $\sum_{n=0}^{\infty} \boldsymbol{D}_n n^r < \infty$, for $\boldsymbol{D}_n = \boldsymbol{A}_n, \boldsymbol{B}_n$.

Proof. By Theorem 6 of [17] we have $\lim_{n\to\infty} n^{-1} \boldsymbol{V}_n^{\langle 1 \rangle}(\boldsymbol{X}) = (\boldsymbol{g}\boldsymbol{X}\boldsymbol{e})\hat{\boldsymbol{G}}$, and by Lemma 1 of [18] we have $\lim_{n\to\infty} n^{-2} \boldsymbol{V}_n^{\langle 2 \rangle}(\boldsymbol{X}) = (\boldsymbol{g}\boldsymbol{X}\boldsymbol{e})^2 \hat{\boldsymbol{G}}$. In fact, when $r = 3$, by interchanging the order of the three summations, we obtain $\lim_{n\to\infty} n^{-3} \boldsymbol{V}_n^{\langle 3 \rangle}(\boldsymbol{X}) = (\boldsymbol{g}\boldsymbol{X}\boldsymbol{e})^3 \hat{\boldsymbol{G}}$. Thus, by induction on r, we obtain $\lim_{n\to\infty} n^{-r} \boldsymbol{V}_n^{\langle r \rangle}(\boldsymbol{X}) = (\boldsymbol{g}\boldsymbol{X}\boldsymbol{e})^r \hat{\boldsymbol{G}}$. From Equation 11 of [18], we have $\boldsymbol{g}\tilde{\boldsymbol{M}}_1 \boldsymbol{e} = 1/(1 - \lambda)$. Since

$$\lim_{n\to\infty} n^{-r} \boldsymbol{V}_n^{\langle r \rangle}(\tilde{\boldsymbol{M}}_1) = (\boldsymbol{g}\tilde{\boldsymbol{M}}_1 \boldsymbol{e})^r \hat{\boldsymbol{G}} = \left(\frac{1}{1-\lambda}\right)^r \hat{\boldsymbol{G}},$$

for any $\varepsilon > 0$ there exists some $N \in \mathbb{N}_+$ such that, for any $n > N$,

$$\sum_{n=N}^{\infty} n^r \boldsymbol{D}_n \left[\left(\frac{1}{1-\lambda}\right)^r \hat{\boldsymbol{G}} - \hat{\boldsymbol{\varepsilon}} \right] < \sum_{n=N}^{\infty} \boldsymbol{D}_n \boldsymbol{V}_n^{\langle r \rangle}(\tilde{\boldsymbol{M}}_1) < \sum_{n=N}^{\infty} n^r \boldsymbol{D}_n \left[\left(\frac{1}{1-\lambda}\right)^r \hat{\boldsymbol{G}} + \hat{\boldsymbol{\varepsilon}} \right],$$

where $\hat{\boldsymbol{\varepsilon}}$ is a square matrix in which each element equals ε. The second part of the assertion follows.

Lemma 3.2. *Let $r \in \mathbb{N}_+$. If $\boldsymbol{G}$ is irreducible and $\lambda < 1$, then $\tilde{\boldsymbol{M}}_r < \infty$ if and only if $\sum_{v=0}^{\infty} v^r \boldsymbol{A}_v < \infty$. Here ∞ is a matrix all of whose entries are infinite, and matrix inequalities apply componentwise.*

Proof. By differentiating both sides of (2.2) r times and then taking $z = 1$, we have

$$\tilde{\boldsymbol{M}}_r = \boldsymbol{C}_1(r) + \sum_{v=1}^{\infty} \boldsymbol{A}_v \sum_{k=0}^{v-1} \boldsymbol{G}^k \tilde{\boldsymbol{M}}_r \boldsymbol{G}^{v-k-1}, \qquad\qquad (3.4)$$

where $C_1(r)$ is a sum of finite items each involving the moment matrices $\tilde{M}_i$ of order $i < r$. We can easily show, by induction on r, that the finiteness of $C_1(r)$ can be reduced to the convergence of $\sum_{n=r}^{\infty} A_n V_n^{\langle r \rangle}(\tilde{M}_1)$. By taking $D_n = A_n$ in Lemma 3.1, we know that $C_1(r)$ is finite if and only if $\sum_{v=0}^{\infty} v^r A_v < \infty$. Since $\lambda < 1$, it follows from Theorem 3.1.1 of [20] that $[I - \sum_{v=1}^{\infty} A_v \sum_{k=0}^{v-1} G^k]^{-1}$ exists. It then follows from (3.4) that

$$\tilde{M}_r e = \left[I - \sum_{v=1}^{\infty} A_v \sum_{k=0}^{v-1} G^k \right]^{-1} C_1(r) e.$$

Therefore, $\tilde{M}_r < \infty$ if and only if $C_1(r) < \infty$, and the assertion holds.

For $r \in \mathbb{N}_+$, define $\tilde{L}_r = \tilde{L}^{(r)}(z)|_{z=1}$ to be the rth derivative of $\tilde{L}(z)$ at $z = 1$ and $\tilde{K}_r = \tilde{K}^{(r)}(z)|_{z=1}$ to be the rth derivative of $\tilde{K}(z)$ at $z = 1$.

Theorem 3.1. *Let $r \in \mathbb{N}_+$. If G is irreducible and $\lambda < 1$, then X_n is r-ergodic if and only if $\sum_{v=0}^{\infty} v^r A_v < \infty$ and $\sum_{v=0}^{\infty} v^r B_v < \infty$.*

Proof. In analogy with the proof of Lemma 3.2, by differentiating both sides of (3.2) we have $\tilde{L}_r < \infty$ if and only if $\sum_{v=0}^{\infty} v^r A_v < \infty$. By differentiating both sides of (3.3) r times and then taking $z = 1$, we have

$$\tilde{K}_r = \sum_{v=1}^{\infty} B_v G^{v-1} \tilde{L}_r + \sum_{v=2}^{\infty} B_v \sum_{k=0}^{v-2} G^k \tilde{M}_r G^{v-k-2} \tilde{L}_1 + C_2(r),$$

where $C_2(r)$ is a sum of finite items each involving the moment matrices $\tilde{M}_i$ and $\tilde{L}_i$ of order $i < r$. Hence, $\tilde{K}_r < \infty$ if and only if $\sum_{v=0}^{\infty} v^r A_v < \infty$ and $\sum_{v=0}^{\infty} v^r B_v < \infty$. Let H be the level L_0. It thus follows from $\tilde{K}_r e < \infty$ that $\mathrm{E}_{(0,j)}[(\tau_H)^r] < \infty$ for all $1 \le j \le m_1$. Hence, X_n is r-ergodic.

3.2. Geometric ergodicity

In this section we deal with geometric ergodicity. From Theorem 4.30 of [3], we know that a Markov chain is geometrically ergodic if and only if there exists some $\gamma > 1$ such that $\mathrm{E}_i[\gamma^{\tau_H}] < \infty$ for all $i \in H$. For $H = L_0$, X_n is geometrically ergodic if and only if $\tilde{K}(z)e < \infty$ for some $z > 1$. Let $\tilde{A}(z) = \sum_{k=0}^{\infty} A_k z^k$. Our idea is to analyze the radius of convergence of $\tilde{K}(z)$ via the Perron–Frobenius eigenvalues of $\tilde{G}(z)$ and $\tilde{A}(z)$, using the matrix-analytic method. Furthermore, the radius of convergence of $\tilde{G}(z)$ is given in computable form, which is of great importance in finding the explicit geometric convergence rates.

Lemma 3.3. *Suppose that G is irreducible and $\lambda \le 1$. Let $\chi(z)$ be the Perron–Frobenius eigenvalue of $\tilde{G}(z)$ and let $\rho(z)$ be the Perron–Frobenius eigenvalue of $\tilde{A}(z)$.*

(i) If $\lambda \le 1$ and $\tilde{G}(\tilde{z}) < \infty$ for some $\tilde{z} > 1$, then $\chi(z)$ is a continuously differentiable function of z for $0 < z \le \tilde{z}$, and the right eigenvector $u(z)$ corresponding to $\chi(z)$ may be chosen to be positive. Furthermore, $\chi(z) = 1$ when $z = 1$ and $\chi(z) > 1$ when $1 < z \le \tilde{z}$.

(ii) If $\tilde{A}(\tilde{z}) < \infty$ for some $\tilde{z} > 1$, then $\rho(z)$ is a continuously differentiable function of z for $0 < z \le \tilde{z}$, and the right eigenvector $v(z)$ corresponding to $\rho(z)$ may be chosen to be positive. Furthermore, $\rho(z) = 1$ when $z = 1$ and $\rho(z) > 1$ when $1 < z \le \tilde{z}$.

Proof. Since $\tilde{G}(1) = G$ is irreducible and $\tilde{G}(\tilde{z}) < \infty$, it follows that $\tilde{G}(\tilde{z})$ is also irreducible and continuously differentiable, for $0 < z \leq \tilde{z}$. It has a uniquely defined Perron–Frobenius eigenvalue $\chi(z)$, which is a continuously differentiable function of z for $0 < z \leq \tilde{z}$, since it is a simple root of a polynomial equation with continuously differentiable coefficients. The corresponding right eigenvector $u(z)$ of $\chi(z)$ is strictly positive and can be chosen so that it is also continuously differentiable for $0 < z < \tilde{z}$. Since G is irreducible and $\lambda \leq 1$, it follows that G is stochastic. Hence, $Ge = e$ and $\tilde{G}(z)e > e$ for $z > 1$, and the second statement of part (i) holds.

We know that A is stochastic. Since G is irreducible, it follows that A is also irreducible. The proof of part (ii) is therefore similar to that of part (i).

To prove our next lemma, we first define the sequence $\{T_v(k), v, k \in \mathbb{N}_+\}$, introduced in [20], by

$$T_0(k) = \mathbf{0}, \quad k \in \mathbb{N}_+, \qquad T_1(1) = A_0, \qquad T_1(k) = \mathbf{0}, \quad k \geq 2,$$

$$T_{v+1}(k) = \sum_{n=1}^{\infty} A_n T_v^{(n)}(k-1), \qquad v \in \mathbb{Z}_+,$$

where $T_v^{(n)}(k)$ is the nth matrix convolution of $T_v(k)$. It is known that $T_v(k)$ is nondecreasing and converges to $G(k)$ when v tends to infinity. In the following lemma we derive a relation between the Perron–Frobenius eigenvalues $\chi(z)$ and $\rho(z)$.

For a sequence of square matrices $\{D_k, k \in \mathbb{N}_+\}$, we denote the radius of convergence of $\sum_{k=1}^{\infty} z^k D_k$ by $\phi_D = \sup\{z \geq 1 : \sum_{k=1}^{\infty} z^k D_k < \infty\}$.

Lemma 3.4. *If G is irreducible then the Perron–Frobenius eigenvalue $\chi(z)$ of $\tilde{G}(z)$ is the smallest positive root of the equation*

$$s(z) = z\rho(s(z)), \tag{3.5}$$

for $z \in \{z : 0 < s(z) \leq \phi_A\}$ (the equality holds only when $\tilde{A}(\phi_A) < \infty$).

Proof. First, we prove that the Perron–Frobenius eigenvalue $\chi(z)$, $z \in \{z : 0 < \chi(z) \leq \phi_A\}$, is a solution to (3.5). (As above, in the domain of z, equality holds only when $\tilde{A}(\phi_A) \leq \infty$.) By postmultiplying by its positive eigenvector $u(z)$ in (3.1), we have

$$\tilde{G}(z)u(z) = \sum_{v=0}^{\infty} zA_v \tilde{G}^v(z)u(z) = z\tilde{A}(\chi(z))u(z) = \chi(z)u(z).$$

It follows from Perron–Frobenius theory that $\chi(z)$ is the Perron–Frobenius eigenvalue of $z\tilde{A}(\chi(z))$, i.e. that $\chi(z) = z\rho(\chi(z))$.

We now prove that $\chi(z)$ is the smallest root. Let $\beta(z)$ be the smallest positive root of (3.5) and denote by $v(\beta(z))$ the positive eigenvector of $\tilde{A}(\beta(z))$ corresponding to the Perron–Frobenius eigenvalue $\rho(\beta(z))$. Let $\tilde{T}_v(z) = \sum_{k=1}^{\infty} z^k T_v(k)$; then $\tilde{T}_0(z)v(\beta(z)) = \mathbf{0} < \beta(z)v(\beta(z))$. Furthermore, suppose that $\tilde{T}_v(z)v(\beta(z)) \leq \beta(z)v(\beta(z))$; then

$$\tilde{T}_{v+1}(z)v(\beta(z)) = z\sum_{n=0}^{\infty} A_n \tilde{T}_v^n(z)v(\beta(z)) \leq z\tilde{A}(\beta(z))v(\beta(z)) = \beta(z)v(\beta(z)).$$

By induction, we see that $\tilde{T}_v(z)v(\beta(z)) \leq \beta(z)v(\beta(z))$ for any $v \in \mathbb{Z}_+$. The matrices $\tilde{T}_v(z)$ converge monotonically to the matrix $\tilde{G}(z)$, and therefore

$$\tilde{G}(z)v(\beta(z)) \leq \beta(z)v(\beta(z)).$$

Since $\boldsymbol{v}(\beta(z))$ is positive, this implies that $\chi(z) \leq \beta(z)$. That $\beta(z) = \chi(z)$ follows from the fact that $\chi(z)$ is also a positive root of (3.5). Thus, we have finished the proof.

Theorem 3.2. *Suppose that $\lambda < 1$ and $\boldsymbol{G}$ is irreducible. If $\phi_A > 1$ then $\tilde{\boldsymbol{G}}(z) < \infty$ for some $z > 1$. Moreover, the radius of convergence of $\tilde{\boldsymbol{G}}(z)$ is $z_0 = s_0/\rho(s_0)$, where s_0 is the unique root of $s\rho'(s) = \rho(s)$.*

Proof. Since $\phi_A > 1$, it follows that $\tilde{\boldsymbol{A}}(s) < \infty$ and $\rho(s) < \infty$ for $0 < s \leq \phi_A$. Thus, $y = z\rho(s)$ can be defined for $0 < z < \infty$ and $0 < s \leq \phi_A$. From Lemma 2.3.3 of [20], we see that $\rho'(1-) = \lambda < 1$. Since $\rho(1) = 1$ and $\rho'(1-) < 1$, it follows that the curves $y = \rho(s)$ and $y = s$ have only one intersection point, at $s = 1$, for $0 < s \leq 1$, and two intersection points for $1 \leq s \leq \phi_A$. For some z-values greater than, but sufficiently close to, 1, we know that the curves $y = z\rho(s)$ and $y = s$ have two intersection points. The left-hand intersection point can be written as $(\theta(z), \theta(z))$, where $\theta(z)$, $1 < \theta(z) < \phi_A$, is the smallest positive root of (3.5). From Lemma 3.4 we know that $\chi(z) = \theta(z) < \infty$; hence, $\tilde{\boldsymbol{G}}(z) < \infty$.

When z increases gradually to its maximal value, denoted by z_0, the curves $y = z\rho(s)$ and $y = s$ will have only one intersection point, which we write as (s_0, s_0) with $s_0 < \phi_A$. Hence, $(s = s_0, z = z_0)$ is the unique solution of the following system of equations:

$$z\rho(s) = s, \qquad z\rho'(s) = 1. \tag{3.6}$$

In (3.6), the first equation guarantees that the curves $y = s$ and $y = z\rho(s)$ have intersection points. The second equation ensures that there is only one intersection point.

We now prove that z_0 is the radius of convergence of $\tilde{\boldsymbol{G}}(z)$, i.e. that $\tilde{\boldsymbol{G}}(z) = \infty$ when $z > z_0$, and $\tilde{\boldsymbol{G}}(z) < \infty$ when $z \leq z_0$. When $z \leq z_0$, the curves $y = s$ and $y = z\rho(s)$ have either one or two intersection points, i.e. (3.5) has either one or two solutions. Then $\chi(z) < \infty$ and, so $\boldsymbol{G}(z) < \infty$. Suppose that $\boldsymbol{G}(z) < \infty$ for some $z > z_0$; then $\chi(z) < \infty$ for some $z > z_0$. From Lemma 3.4 we see that (3.5) has at least one solution, i.e. the curves $y = s$ and $y = z\rho(s)$ have at least one intersection point, which contradicts the fact that when $z > z_0$ the curves have no intersection points.

Remark 3.1. If a Markov chain is stochastically ordered, then, relative to those for general Markov chains, it is possible to obtain much more explicit or even tighter convergence rates (see [13] and [22]) in terms of the radius of convergence of the generating function of the first hitting time of the state 0. Theorem 3.2 can thus be used to derive explicit convergence rates for M/G/1-type Markov chains, as will be illustrated in Section 5.

For technical reasons, we introduce a simpler M/G/1-type Markov chain, $\{X'_n, \ n \in N_+\}$, with transition matrix

$$\boldsymbol{P}' = \begin{pmatrix} \boldsymbol{B}'_0 & \boldsymbol{B}_1 & \boldsymbol{B}_2 & \boldsymbol{B}_3 & \cdots \\ \boldsymbol{A}_0 & \boldsymbol{A}_1 & \boldsymbol{A}_2 & \boldsymbol{A}_3 & \cdots \\ \boldsymbol{0} & \boldsymbol{A}_0 & \boldsymbol{A}_1 & \boldsymbol{A}_2 & \cdots \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{A}_0 & \boldsymbol{A}_1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix};$$

$\boldsymbol{P}'$ is the same as $\boldsymbol{P}$ except for a change of two elements, $\boldsymbol{B}_0$ and $\boldsymbol{C}_0$, in the first column. Owing to the relation between $\boldsymbol{P}$ and $\boldsymbol{P}'$, we may also make $\boldsymbol{P}'$ stochastic and X'_n irreducible and aperiodic. For $\boldsymbol{P}'$, we have $\tilde{\boldsymbol{G}}(z) = \tilde{\boldsymbol{L}}(z)$ and it follows from (3.1)–(3.3) that

$$\tilde{\boldsymbol{G}}(z) = \sum_{v=0}^{\infty} z\boldsymbol{A}_v \tilde{\boldsymbol{G}}^v(z), \qquad \tilde{\boldsymbol{K}}(z) = z\boldsymbol{B}'_0 + \sum_{v=1}^{\infty} z\boldsymbol{B}_v \tilde{\boldsymbol{G}}^v(z). \tag{3.7}$$

In the following, we first give the criterion for X'_n to be geometrically ergodic. Then we prove that X_n is geometrically ergodic if and only if X'_n is also.

Theorem 3.3. *If G is irreducible and $\lambda < 1$, then X'_n is geometrically ergodic if and only if* $\min\{\phi_A, \phi_B\} > 1$.

Proof. First, we prove the necessity. If X'_n is geometrically ergodic then $\tilde{K}(z) < \infty$ and $\tilde{G}(z) < \infty$ for some $z > 1$. Thus, from (3.7), we have

$$\sum_{v=1}^{\infty} B_v \tilde{G}^v(z) < \infty, \qquad \sum_{v=1}^{\infty} A_v \tilde{G}^v(z) < \infty. \tag{3.8}$$

From Lemma 3.3, we know that the Perron–Frobenius eigenvalue $\chi(z)$ of $G(z)$ satisfies $1 < \chi(z) < \infty$. By postmultiplying by the positive eigenvector $u(z)$, corresponding to $\chi(z)$, in both equations in (3.8), we obtain

$$\sum_{v=1}^{\infty} B_v \chi^v(z) u(z) < \infty, \qquad \sum_{v=1}^{\infty} A_v \chi^v(z) u(z) < \infty,$$

which implies that $\min\{\phi_A, \phi_B\} > 1$.

We now prove the sufficiency. If $\min\{\phi_A, \phi_B\} > 1$ then $\phi_A > 1$ and $\phi_B > 1$. Since $\phi_A > 1$ and $\lambda < 1$, it follows from Theorem 3.2 that there exists some $s > 1$ such that $\tilde{G}(s) < \infty$. By Lemma 3.3, we know that the Perron–Frobenius eigenvalue $\chi(z)$ of $\tilde{G}(z)$ is continuously differentiable when $0 < z \leq s$. We can thus choose an appropriate s_1, with $1 < s_1 \leq s$, such that $\chi(s_1) < \phi_B$. By postmultiplying by the eigenvector $u(s_1)$ in the second equation of (3.7), we obtain

$$\tilde{K}(s_1) u(s_1) = s_1 B'_0 u(s_1) + s_1 \sum_{v=1}^{\infty} B_v \chi^v(s_1) u(s_1) < \infty.$$

Since $u(s_1)$ is positive, it follows that $\tilde{K}(s_1) e < \infty$; thus, the chain X'_n is geometrically ergodic.

Theorem 3.4. *If G is irreducible and $\lambda < 1$, then X_n is geometrically ergodic if and only if* $\min\{\phi_A, \phi_B\} > 1$.

Proof. Let $C = L_0 \cup L_1$. Suppose that X_n is geometrically ergodic; then there exists a finite solution V with $V_i \geq 1$, $i \in E$, to (2.1) for the set C, and some positive constant $b_G < \infty$. It is easy to check that, when P' is used in place of P, V is also a solution of (2.1) for the same set C and some positive constant $b'_G < \infty$. Hence, X'_n is also geometrically ergodic. Similarly, we can prove that if X'_n is geometrically ergodic, then X_n is also. Thus, it follows from Theorem 3.3 that our assertion holds.

Remark 3.2. An M/G/1-type Markov chain is called a quasi-birth–death chain if $B_k = 0$ for $k \geq 2$ and $A_k = 0$ for $k \geq 3$. Obviously $\min\{\phi_A, \phi_B\} > 1$, so by Theorem 3.4 we can easily show that if G is irreducible and $\lambda < 1$, the quasi-birth–death chain is geometrically ergodic.

Remark 3.3. From Propositions 2.2 and 2.3 of [5], we see that a Markov chain with the transition matrix $P = (P(i, j))$ is not strongly ergodic if $\lim_{i \to \infty} P(i, j) = 0$ for each j. According to this result, X_n is not strongly ergodic.

4. M/G/1-type Markov processes

In this section, we consider an irreducible M/G/1-type Markov process X_t whose intensity matrix $\boldsymbol{Q} = (q_{ij})$ is partitioned into block form:

$$\boldsymbol{Q} = \begin{pmatrix} \hat{\boldsymbol{L}} & \hat{\boldsymbol{F}}^{(1)} & \hat{\boldsymbol{F}}^{(2)} & \hat{\boldsymbol{F}}^{(3)} & \hat{\boldsymbol{F}}^{(4)} & \cdots \\ \hat{\boldsymbol{B}} & \boldsymbol{L} & \boldsymbol{F}^{(1)} & \boldsymbol{F}^{(2)} & \boldsymbol{F}^{(3)} & \cdots \\ \boldsymbol{0} & \boldsymbol{B} & \boldsymbol{L} & \boldsymbol{F}^{(1)} & \boldsymbol{F}^{(2)} & \cdots \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{B} & \boldsymbol{L} & \boldsymbol{F}^{(1)} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

Here the matrices $\boldsymbol{B}$, $\boldsymbol{L}$, and $\{\boldsymbol{F}^{(n)}, n \in \mathbb{N}_+\}$ are square matrices of order m, the matrix $\hat{\boldsymbol{L}}$ is a square matrix of order m_1, and the matrices $\{\hat{\boldsymbol{F}}^{(n)}, n \in N_+\}$ and $\hat{\boldsymbol{B}}$ are respectively of dimensions $m_1 \times m$ and $m \times m_1$. We suppose that $\boldsymbol{Q}$ is both conservative, i.e. that

$$\hat{\boldsymbol{L}}e + \sum_{n=1}^{\infty} \hat{\boldsymbol{F}}^{(n)}e = \boldsymbol{0}, \qquad \hat{\boldsymbol{B}}e + \boldsymbol{L}e + \sum_{n=1}^{\infty} \boldsymbol{F}^{(n)}e = \boldsymbol{0}, \qquad \boldsymbol{B}e + \boldsymbol{L}e + \sum_{n=1}^{\infty} \boldsymbol{F}^{(n)}e = \boldsymbol{0},$$

and totally stable, i.e. $\sup_{i \in E} q_i < \infty$. Thus, $\boldsymbol{Q}$ is regular. As above, the state space of the process is $E = \{\bigcup_{i=0}^{\infty} L_i\}$, with $L_0 = \{(0, j), j = 1, 2, \ldots, m_1\}$ and $L_i = \{(i, j), j = 1, 2, \ldots, m\}$. Let $P^t(i, j)$ be the transition function of X_ε. We investigate the rates of convergence of $P^t(i, j)$ to the stationary distribution by comparing X_t with its h-approximation chain.

The h-approximation chain of X_t is an irreducible, aperiodic M/G/1-type Markov chain with the following transition matrix:

$$\boldsymbol{P} = \begin{pmatrix} h\hat{\boldsymbol{L}} + \boldsymbol{I}_{m_1} & h\hat{\boldsymbol{F}}^{(1)} & h\hat{\boldsymbol{F}}^{(2)} & h\hat{\boldsymbol{F}}^{(3)} & h\hat{\boldsymbol{F}}^{(4)} & \cdots \\ h\hat{\boldsymbol{B}} & h\boldsymbol{L} + \boldsymbol{I}_m & h\boldsymbol{F}^{(1)} & h\boldsymbol{F}^{(2)} & h\boldsymbol{F}^{(3)} & \cdots \\ \boldsymbol{0} & h\boldsymbol{B} & h\boldsymbol{L} + \boldsymbol{I}_m & h\boldsymbol{F}^{(1)} & h\boldsymbol{F}^{(2)} & \cdots \\ \boldsymbol{0} & \boldsymbol{0} & h\boldsymbol{B} & h\boldsymbol{L} + \boldsymbol{I}_m & h\boldsymbol{F}^{(1)} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

Here $\boldsymbol{I}_{m_1}$ and $\boldsymbol{I}_m$ are identity matrices of orders m_1 and m, respectively.

Let $\hat{\boldsymbol{G}}$ be the minimal nonnegative solution to the following equation:

$$\boldsymbol{B} + \boldsymbol{L}\hat{\boldsymbol{G}} + \sum_{k=1}^{\infty} \boldsymbol{F}^{(k)}\hat{\boldsymbol{G}}^{k+1} = \boldsymbol{0}.$$

The matrix $\hat{\boldsymbol{G}}$ has the same probabilistic interpretation as $\boldsymbol{G}$ for the discrete-time chains: the (k, l)th entry of $\hat{\boldsymbol{G}}$ expresses the conditional probability of the process first entering L_{j-1} through state l, given that it starts from state k of L_j. The auxiliary matrix can be computed by an efficient iterative method, e.g. the cyclic reduction algorithm [2].

Let $\hat{\boldsymbol{\mu}}$ be the invariant probability vector of the matrix $\boldsymbol{B} + \boldsymbol{L} + \sum_{k=1}^{\infty} \hat{\boldsymbol{F}}^{(k)}$, i.e.

$$\hat{\boldsymbol{\mu}}\left(\boldsymbol{B} + \boldsymbol{L} + \sum_{k=1}^{\infty} \boldsymbol{F}^{(k)}\right) = \boldsymbol{0}.$$

Obviously, $\hat{G}$ is also the minimal nonnegative solution to

$$h\boldsymbol{B} + (h\boldsymbol{L} + \boldsymbol{I}_m)\hat{\boldsymbol{G}} + h\sum_{k=1}^{\infty} \boldsymbol{F}^{(k)}\hat{\boldsymbol{G}}^{k+1} = \hat{\boldsymbol{G}}, \tag{4.1}$$

and $\hat{\boldsymbol{\mu}}$ is also the invariant probability vector of $h\boldsymbol{B} + (h\boldsymbol{L} + \boldsymbol{I}_m) + h\sum_{k=1}^{\infty} \boldsymbol{F}^{(k)}$, i.e.

$$\hat{\boldsymbol{\mu}} = \hat{\boldsymbol{\mu}}\left[h\boldsymbol{B} + (h\boldsymbol{L} + \boldsymbol{I}_m) + h\sum_{k=1}^{\infty} \boldsymbol{F}^{(k)}\right]. \tag{4.2}$$

Let $\hat{\boldsymbol{\beta}} = h[\boldsymbol{L} + \sum_{k=1}^{\infty}(k+1)\boldsymbol{F}^{(k)}]\boldsymbol{e} + \boldsymbol{e}$ and $\hat{\lambda} = \hat{\boldsymbol{\mu}}\hat{\boldsymbol{\beta}}$.

Theorem 4.1. *Let $l \in \mathbb{N}_+$. Suppose that $\hat{G}$ is irreducible and $\hat{\lambda} < 1$. If $\sum_{n=1}^{\infty} n^l \hat{\boldsymbol{F}}^{(n)} < \infty$ and $\sum_{n=1}^{\infty} n^l \boldsymbol{F}^{(n)} < \infty$, then*

$$t^{l-1}\|P^t(i, \cdot) - \pi(\cdot)\| \to 0, \qquad t \to \infty, \tag{4.3}$$

for all $i \in E$.

Proof. Since $\hat{G}$ is irreducible and $\hat{\lambda} < 1$, it follows from (4.1) and (4.2) that $\boldsymbol{G} = \hat{\boldsymbol{G}}$ is irreducible and that $\lambda = \hat{\lambda} < 1$ for the h-approximation chain. From Theorem 3.1, we then know that the h-approximation chain is l-ergodic, and thus that

$$n^{l-1}\|P^n(i, \cdot) - \pi(\cdot)\| \to 0, \qquad n \to \infty,$$

for all $i \in E$. Hence, by Lemma 2.1 we know that (4.3) holds.

Theorem 4.2. *If $\hat{G}$ is irreducible and $\hat{\lambda} < 1$, then $X(t)$ is geometrically ergodic if and only if $\min\{\phi_F, \phi_{\hat{F}}\} > 1$.*

Proof. The assertion follows directly from Theorem 3.4 and Lemma 2.2.

Remark 4.1. An M/G/1-type Markov process is called a quasi-birth–death process if $\hat{\boldsymbol{F}}^{(k)} = \boldsymbol{0}$ and $\boldsymbol{F}^{(k)} = \boldsymbol{0}$ for $k \geq 2$. Obviously $\min\{\phi_F, \phi_{\hat{F}}\} > 1$, so by Theorem 4.2 we can easily see that if $\hat{G}$ is irreducible and $\hat{\lambda} < 1$, the quasi-birth–death process is geometrically ergodic.

Remark 4.2. From Remark 3.3, we know that the h-approximation chain is not strongly ergodic. It thus follows from Lemma 2.3 that $X(t)$ is not strongly ergodic.

5. Applications: explicit geometric and exponential convergence rates

In this section, we give three examples to illustrate our results. In the first example, we obtain the best possible geometric convergence rate for a strictly pathwise Markov chain. In the other examples, we derive explicit lower bounds on the best possible exponential convergence rates for two Markov processes whose h-approximation chains are all stochastically monotone.

We first state the definitions of stochastic ordering and pathwise ordering, taken from [13]. We say that a random variable Y_1 is stochastically larger than a random variable Y_2 if, for all real x, $\mathrm{P}\{Y_1 \leq x\} \leq \mathrm{P}\{Y_2 \leq x\}$. Let X_n^1 and X_n^2 be two copies of the chain X_n, with the possibly random initial states X_0^1 and X_0^2. The chain X_n is said to be stochastically ordered if X_n^1 is stochastically larger than X_n^2 for all $n > 0$, whenever X_0^1 is stochastically larger than X_0^2. Obviously, if a Markov chain is stochastically monotone (i.e. $\sum_{j \geq k} P(i, j) \leq \sum_{j \geq k} P(m, j)$

for any i and m with $i < m$, and for every k), then it is stochastically ordered. It then follows, from [11, pp. 132–136], that stochastic ordering can be extended to pathwise ordering: $X_n^1 \geq X_n^2$ for all n whenever $X_0^1 > X_0^2$. Suppose that the chain X_n is ergodic and has stationary distribution π. The pathwise ordering is said to be strict if there exists some $c > 0$, with $\pi((c, \infty)) > 0$, such that $X_n^1 > X_n^2$ whenever $n \leq \inf\{k > 0 : X_k^2 = 0\}$ and $x > y + c$.

Generally speaking, the value of r in the generating function $E_i[r^{\tau_0}]$ fails to provide information about the geometric convergence rate. For stochastically ordered Markov chains, we know from [22] that if $E_i[r^{\tau_0}] < \infty$ for some $i > 0$ and $r > 1$, then

$$\sum_{n=0}^{\infty} r^n \| P^n(i, \cdot) - \pi(\cdot) \| \leq D_i \tag{5.1}$$

for every i, where

$$D_i = \frac{1}{r-1} E_i[r^{\tau_0}] + \frac{r}{(r-1)^2} (E_0[r^{\tau_0}] - 1) < \infty.$$

For stochastically monotone Markov chains, the value of r can thus provide a lower bound for the maximal geometric convergence rate. Furthermore, if the ordering is still strict, from Theorem 6.2 of [13] we have

$$\limsup_{n \to \infty} r^n \| P^n(0, \cdot) - \pi(\cdot) \| = \infty$$

for $r > r_0$, where r_0 is the radius of convergence of $E_0[r^{\tau_0}]$. Hence, under the condition that the chain X_n be strictly pathwise, the radius of convergence of $E_i[r^{\tau_0}]$ is the best possible geometric convergence rate.

Example 5.1. Define a Markov chain X_n on a countable state space $E = \mathbb{Z}_+$ by

$$X_{n+1} = \begin{cases} X_n - 1 + A_n, & X_n \geq 1, \\ A_n, & X_n = 0, \end{cases} \tag{5.2}$$

where A_n is a sequence of independent, identically distributed discrete random variables distributed according to $P\{A_1 = i\} = a_i$, $i \in \mathbb{Z}_+$, and whose transition matrix is

$$P = \begin{pmatrix} a_0 & a_1 & a_2 & a_3 & \cdots \\ a_0 & a_1 & a_2 & a_3 & \cdots \\ 0 & a_0 & a_1 & a_2 & \cdots \\ 0 & 0 & a_0 & a_1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix},$$

where $\{a_n, n \in \mathbb{Z}_+\}$ is a sequence of nonnegative real numbers such that $\sum_{k=0}^{\infty} a_k = 1$. Suppose that X_n is irreducible and aperiodic. It is easy to show that X_n is ergodic if and only if $\sum_{k=1}^{\infty} k a_k < 1$; and if this holds it follows from Theorem 3.1 that, for some l with $l \in \mathbb{N}_+$, X_n is l-ergodic if and only if $\sum_{n=0}^{\infty} a_n n^l < \infty$.

Let X_n^1 and X_n^2 be two copies of the chain X_n driven by the same sequence $\{A_n\}$, with the initial conditions $X_0^1 = x$ and $X_0^2 = y$. Suppose that $x > y$. Then, from (5.2), we know that $X_n^1 \geq X_n^2$ for every $n \geq 0$, and that the chain X_n is therefore pathwise ordered.

Furthermore, suppose that $x \geq y + 2$. Then, by again using (5.2), we have $X_n^1 - X_n^2 \geq 1$ for $n \leq \inf\{k > 0 \colon X_k^2 = 0\}$, so the ordering is still strict.

Now suppose that $\sum_{k=1}^{\infty} k a_k < 1$ and $\phi_A > 1$. It then follows from Theorem 3.4 that X_n is geometrically ergodic. Because of the special structure of $\boldsymbol{P}$, we have

$$\mathrm{E}_0[z^{\tau_0}] = \mathrm{E}_1[z^{\tau_0}] \quad \text{and} \quad \mathrm{E}_i[z^{\tau_0}] = (\mathrm{E}_0[z^{\tau_0}])^i.$$

By Theorem 3.2, we know that the radius of convergence of $\mathrm{E}_0[z^{\tau_0}]$ is z_0, with $z_0 = s_0/A(s_0)$ and $\mathrm{E}_0[z_0^{\tau_0}] = s_0$, where s_0 is the unique root of $s A'(s) = A(s)$. Thus, from the discussion preceding the example, we know that (5.1) holds for positive constants

$$r = z_0 \quad \text{and} \quad D_i = \frac{1}{z_0 - 1} s_0^i + \frac{z_0}{(z_0 - 1)^2} (s_0 - 1) < \infty,$$

and that z_0 is the best possible geometric convergence rate.

Remark 5.1. For general Markov chains, it is difficult to give the best possible geometric convergence rate. Owing to the facts that the chain is strictly pathwise and that the radius of convergence of $\mathrm{E}_0[r^{\tau_0}]$ can be given explicitly, we obtain the sharpest geometric convergence rate for the chain in Example 5.1. In particular, if we take

$$a_k = \int_{0-}^{\infty} \frac{(\lambda x)^k}{k!} e^{-\lambda x} \, dB(x), \qquad k \in \mathbb{Z}_+,$$

in the example, then X_n becomes the embedded chain of the queue length of the classical M/G/1 queue. The best possible geometric convergence rate for the embedded chain was given in [12] using the same method.

Example 5.2. Let X_t be an irreducible Markov process with intensity matrix $\boldsymbol{Q} = (q_{ij})$, as follows:

$$\boldsymbol{Q} = \begin{pmatrix} -\sum_{k=1}^{\infty} b_k & b_1 & b_2 & b_3 & \cdots \\ b_0 & -b_0 - \sum_{k=2}^{\infty} b_k & b_2 & b_3 & \cdots \\ 0 & b_0 & -b_0 - \sum_{k=2}^{\infty} b_k & b_2 & \cdots \\ 0 & 0 & b_0 & -b_0 - \sum_{k=2}^{\infty} b_k & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

Here $\{b_i, \ i \in \mathbb{Z}_+\}$ is a sequence of nonnegative real numbers with $\sum_{k=0}^{\infty} b_k < \infty$. Since $\boldsymbol{Q}$ is uniformly bounded, by Lemma 4.40 of [3] we know that X_t is ergodic if and only if its embedding chain is also. The transition matrix $\boldsymbol{P} = (P(i, j))$ of the embedding chain is

$$P(i, j) = \frac{q_{ij}(1 - \delta_{ij})}{q_i}, \qquad \delta_{ij} = \begin{cases} 1, & i = j, \\ 0, & i \neq j. \end{cases}$$

The embedding chain is ergodic if and only if

$$\frac{\sum_{k=2}^{\infty} k b_k}{b_0 + \sum_{k=2}^{\infty} b_k} < 1, \quad \text{i.e.} \quad \sum_{k=1}^{\infty} k b_{k+1} < b_0.$$

If $\sum_{k=1}^{\infty} k b_{k+1} < b_0$ and $\sum_{n=0}^{\infty} b_n n^l < \infty$, then it follows from Theorem 4.1 that, for every i,

$$t^{l-1} \| P^t(i, \cdot) - \pi(\cdot) \| \to 0, \qquad t \to \infty.$$

Let $h = 1 / \sum_{k=0}^{\infty} b_k$. Then $h < \bar{q}^{-1}$ and the h-approximation chain is stochastically monotone. Define

$$B_h(z) = h b_0 + \left(1 - h b_0 - h \sum_{k=2}^{\infty} b_k\right) z + h \sum_{k=2}^{\infty} b_k z^k.$$

Suppose that $\sum_{k=1}^{\infty} k b_{k+1} < b_0$ and that the radius of convergence of $\sum_{k=0}^{\infty} b_k z^k$ is greater than 1. Then the h-approximation chain is geometrically ergodic. From Example 5.1 and Lemma 2.2, we know that, for every i,

$$\| P^t(i, \cdot) - \pi(\cdot) \| \leq D_i e^{-\rho_E t}$$

for positive constants

$$\rho_E = \frac{1 - \rho_G(h)}{h} = \left(1 - \frac{1}{z_0}\right) \sum_{k=0}^{\infty} b_k \quad \text{and} \quad D_i = \frac{1}{z_0 - 1} s_0^i + \frac{z_0}{(z_0 - 1)^2} (s_0 - 1),$$

where $z_0 = s_0 / B_h(s_0)$ and s_0 is the unique root of the equation $s B_h'(s) = B_h(s)$.

Example 5.3. (*Random walk in continuous time.*) Let $X(t)$ be an irreducible regular birth–death process with birth rates $\lambda_0 = \sigma$ and $\lambda_n = \lambda$, $n \geq 1$, and death rates $\mu_0 = 0$ and $\mu_n = \mu$, $n \geq 1$. Let $h = 1/(\mu + \lambda + \sigma)$. Then the process's h-approximation chain is also stochastically monotone. If $\lambda/\mu < 1$ then we can easily show that the h-approximation chain is geometrically ergodic and that

$$\sum_{n=0}^{\infty} z_0^n \| P^n(i, \cdot) - \pi(\cdot) \| \leq D_i$$

for every i, for positive constants $z_0 = 1/(1 - h(\sqrt{\lambda} - \sqrt{\mu})^2)$ and

$$D_i = \begin{cases} \dfrac{2z_0 - 1}{(z_0 - 1)^2} [z_0(1 - h\sigma) + z_0 h \sigma s_0] - \dfrac{z_0}{(z_0 - 1)^2}, & i = 0, \\[4mm] \dfrac{s_0^i}{z_0 - 1} + \dfrac{z_0}{(z_0 - 1)^2} [z_0(1 - h\sigma) + z_0 h \sigma s_0 - 1], & i \geq 1, \end{cases}$$

where $s_0 = \sqrt{\mu/\lambda}$. Again by Lemma 2.2, we obtain

$$\| P^t(i, \cdot) - \pi(\cdot) \| \leq D_i e^{-(\sqrt{\lambda} - \sqrt{\mu})^2 t}.$$

Remark 5.2. For the processes in Examples 5.2 and 5.3, we can derive explicit lower bounds on the best possible exponential convergence rate, but our methods fail to give corresponding upper bounds. Fortunately, the explicit upper bound for a random walk in continuous time was obtained in [27]; we introduce the result in the following remark.

Remark 5.3. Many authors (see, e.g. [8], [26], [27], and [28]) have investigated the exponential convergence rates of birth–death processes. In [8] and [26], the conditions for exponential ergodicity and bounds for the decay parameter were obtained by the method of spectral representation. Explicit lower and upper bounds on exponential convergence rates for homogenous and inhomogenous birth–death processes were respectively obtained in [27] and [28], using methods from the theory of differential equations. Moreover, for a random walk in continuous time, in [27] it was proved that the upper bound equals the lower bound, i.e. that the best possible exponential convergence rate of the process is $(\sqrt{\lambda} - \sqrt{\mu})^2$.

Acknowledgements

References

[1] ANDERSON, W. J. (1991). *Continuous-Time Markov Chains. An Applications-Oriented Approach*. Springer, New York.

[2] BINI, D. A., MEINI, B. AND RAMASWAMI, V. (2000). Analyzing M/G/1 paradigms through QBDs: the role of the block structure in computing the matrix *G*. In *Advances in Matrix-Analytic Methods for Stochastic Models*, eds G. Latouche and P. Taylor, Notable Publications, NJ, pp. 73–86.

[3] CHEN, M. F. (1992). *From Markov Chains to Non-Equilibrium Particle Systems*. World Scientific, River Edge, NJ.

[4] FALKENBERG, E. (1994). On the asymptotic behaviour of the stationary distribution of Markov chains of $M/G/1$-type. *Stoch. Models* **10,** 75–97.

[5] HOU, Z. AND LIU, Y. (2004). Explicit criteria for several types of ergodicity of the embedded M/G/1 and GI/M/*n* queues. *J. Appl. Prob.* **41,** 778–790.

[6] JARNER, S. F. AND ROBERTS, G. O. (2002). Polynomial convergence rates of Markov chains. *Ann. Appl. Prob.* **12,** 224–247.

[7] KARTASHOV, N. V. (1996). *Strong Stable Markov Chains*. VSP, Utrecht.

[8] KIJIMA, M. (1992). Evaluation of the decay parameter for some specialized birth–death processes. *J. Appl. Prob.* **29,** 781–791.

[9] KIJIMA, M. (1993). Quasi-limiting distributions of Markov chains that are skip-free to the left in continuous time. *J. Appl. Prob.* **30,** 509–517.

[10] LATOUCHE, G. AND RAMASWAMI, V. (1999). *Introduction to Matrix Analytic Methods in Stochastic Modeling* (ASA-SIAM Ser. Stat. Appl. Prob.). SIAM, Philadelphia.

[11] LINDVALL, T. (1992). *Lectures on the Coupling Method*. John Wiley, New York.

[12] LIU, Y. AND HOU, Z. (2005). Explicit convergence rates of the embedded M/G/1 queue. To appear in *Acta. Math. Sin. (Engl. Ser.)*.

[13] LUND, R. B. AND TWEEDIE, R. L. (1996). Geometric convergence rates for stochastically ordered Markov chains. *Math. Operat. Res.* **21,** 182–194.

[14] MEINI, B. (1998). Solving $M/G/1$ type Markov chains: recent advances and applications. *Commun. Statist. Stoch. Models* **14,** 479–496.

[15] MEYN, S. P. AND TWEEDIE R. L. (1993). *Markov Chains and Stochastic Stability*. Springer, London.

[16] NELSON, R. (1995). *Probability, Stochastic Processes, and Queueing Theory*. Springer, New York.

[17] NEUTS, M. F. (1976). Moments formulas for the Markov renewal branching process. *Adv. Appl. Prob.* **8,** 690–711.

[18] NEUTS, M. F. (1978). The second moments of the absorption times in the Markov renewal branching process. *J. Appl. Prob.* **15,** 707–714.

[19] NEUTS, M. F. (1981). *Matrix-Geometric Solutions in Stochastic Models. An Algorithmic Approach*. Johns Hopkins University Press, Baltimore, MD.

[20] NEUTS, M. F. (1989). *Structured Stochastic Matrices of M/G/*1 *Type and their Applications*. Marcel Dekker, New York.

[21] RAMASWAMI, V. AND LATOUCHE, G. (1986). A general class of Markov processes with explicit matrix-geometric solutions. *OR Spektrum* **8,** 209–218.

[22] SCOTT, D. J. AND TWEEDIE, R. L. (1996). Explicit rates of convergence of stochastically ordered Markov chains. In *Athens Conf. Appl. Prob. Time Ser. Anal.*, Vol. 1, Springer, pp. 176–191.

[23] SPIEKSMA, F. M. (1990). Geometrically ergodic Markov chains and the optimal control of queues. Doctoral Thesis, University of Leiden.

[24] SQUILLANTE, M. S. (2000). Matrix-analytic methods: applications, results and software tools. In *Advances in Matrix-Analytic Methods for Stochastic Models*, eds G. Latouche and P. Taylor, Notable Publications, NJ.

[25] TUOMINEN, P. AND TWEEDIE, R. L. (1994). Subgeometric rates of convergence of f-ergodic Markov chains. *Adv. Appl. Prob.* **26,** 775–798.

[26] VAN DOORN, E. A. (1985). Conditions for exponential ergodicity and bounds for the decay parameter of a birth–death process. *Adv. Appl. Prob.* **17,** 514–530.

[27] ZEIFMAN, A. I. (1991). Some estimates of the rate of convergence for birth and death processes. *J. Appl. Prob.* **28,** 268–277.

[28] ZEIFMAN, A. I. (1995). Upper and lower bounds on the rate of convergence for nonhomogeneous birth and death processes. *Stoch. Process. Appl.* **59,** 157–173.