Applied Probability Trust (23 July 2014)

BOUNDED TRUNCATION ERROR FOR LONG RUN AVERAGES IN INFINITE MARKOV CHAINS

HENDRIK BAUMANN,[*] *Clausthal University of Technology*

WERNER SANDMANN,[**] *University of Derby*

Abstract

We consider long run averages of additive functionals on infinite discrete-state Markov chains, either continuous or discrete in time. Special cases include long run average costs or rewards, stationary moments of the components of ergodic multi-dimensional Markov chains, queueing network performance measures, and many others. By exploiting Foster-Lyapunov-type criteria involving drift conditions for the finiteness of long run averages we determine suitable finite subsets of the state space such that the truncation error is bounded. Illustrative examples demonstrate the application of this method.

Keywords: Infinite Markov chain; additive functional; long run average; state space truncation; bounded truncation error; Foster-Lyapunov-type criterion; drift condition

2000 Mathematics Subject Classification: Primary 60J22

Secondary 60J10;60J27;60J28

1. Introduction

We consider infinite Markov chains, either continuous or discrete in time, on a countable state space $\mathcal{S}$. In continuous time we denote the Markov chain by $(X_t)_{t\geq 0}$ and its generator matrix by $Q = (q_{ij})_{i,j\in\mathcal{S}}$. In discrete time we denote the Markov chain by $(Y_n)_{n\in\mathbb{N}}$ and its transition probability matrix by $P = (p_{ij})_{i,j\in\mathcal{S}}$.

For irreducible recurrent Markov chains, an invariant measure $\psi = (\psi_i)_{i\in\mathcal{S}}$ exists,

[*] Postal address: Department of Applied Stochastics and Operations Research, Clausthal University of Technology, Erzstr. 1, D–38678 Clausthal–Zellerfeld, Germany

[**] Postal address: School of Computing and Mathematics, University of Derby, Kedleston Road, Derby, DE22 1GB, UK

which is unique up to a multiplicative constant, and for $f^{(1)}, f^{(2)} : \mathcal{S} \to \mathbb{R}$ with $\psi \left| f^{(1)} \right|, \psi \left| f^{(2)} \right| < \infty$ we have

$$\lim_{t\to\infty} \frac{\int\limits_0^t f^{(1)}(X_s)\, ds}{\int\limits_0^t f^{(2)}(X_s)\, ds} = \frac{\psi f^{(1)}}{\psi f^{(2)}}, \quad \text{resp.,} \quad \lim_{N\to\infty} \frac{\sum\limits_{n=0}^N f^{(1)}(Y_n)}{\sum\limits_{n=0}^N f^{(2)}(Y_n)} = \frac{\psi f^{(1)}}{\psi f^{(2)}}, \tag{1}$$

with probability 1, see [8, pp. 85–86,203–209]. Hence, obtaining ψf for functions f on the state space of irreducible recurrent Markov chains is of high practical relevance.

Particularly important special cases of irreducible recurrent Markov chains are ergodic Markov chains where ψ is the unique stationary distribution $\pi = (\pi_i)_{i\in S}$ that coincides with the limiting distribution. It satisfies $\pi Q = 0$ in continuous time and $\pi P = \pi$ in discrete time, respectively. According to the respective ergodic theorems, if the expectation $\mathrm{E}_\pi[|f|]$ for a function $f : \mathcal{S} \to \mathbb{R}$ is finite, then the averages of additive functionals converge (for time approaching infinity) almost surely to the stationary expectation $\mathrm{E}_\pi[f] = \sum_{i\in\mathcal{S}} \pi_i f(i) = \pi f$, that is,

$$\lim_{t\to\infty} \frac{1}{t} \int_0^t f(X_s)\, ds = \mathrm{E}_\pi[f], \quad \text{resp.,} \quad \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^N f(Y_n) = \mathrm{E}_\pi[f], \tag{2}$$

with probability 1, see [2, pp. 52–54], [19, pp. 264–265] for the continuous time case and [2, pp. 16–19], [19, pp. 45–47] for the discrete time case. Hence, $\mathrm{E}_\pi[f]$ is the long run average of an additive functional on the respective Markov chain.

We are interested in approximating ψf for infinite recurrent Markov chains by using finite state truncations, which is important in cases where no analytical solution to the infinite chain is available and the state space must be truncated, e.g. for computational purposes. While corresponding truncation approximations of stationary distributions have been studied quite extensively [9, 11, 13, 14, 16, 24, 25], [18, Chap. 7], there is a lack of similar studies for ψf or its special case of stationary expectations.

The goal is to perform the truncation such that the truncation error is bounded by an a priori specified constant. Obviously, since in general no information on the value of ψf is available in advance we have to bound the relative truncation error. Therefore, provided that ψf is finite, we shall provide a method for determining a finite subset

$\mathcal{C} \subset \mathcal{S}$ of the state space such that for a small prescribed $\epsilon \in (0,1)$:

$$\frac{\sum_{i \in \mathcal{C}} \psi_i f(i)}{\sum_{i \in \mathcal{S}} \psi_i f(i)} \geq 1 - \epsilon. \tag{3}$$

Note that this yields a 'true' a priori truncation error bound in that ϵ indeed bounds the proportion of ψf that is cut off by the finite state truncation. There is no need to compute the left hand side, in particular the numerator, of the inequality (3) since we shall guarantee that $\mathcal{C}$ is chosen such that the truncation error is bounded by ϵ. In other words, we do not aim in computing the truncation error a posteriori, but we start with an a priori fixed maximum truncation error and obtain a suitable truncation.

In Section 2 we establish appropriate 'Foster-Lyapunov-type criteria' involving 'drift conditions' and in Section 3 we show how to use them for determining appropriate finite sets $\mathcal{C} \subset \mathcal{S}$ that meet (3). Subsequently, in Section 4 we give application examples. Finally, Section 5 concludes the paper and outlines further research directions.

2. Foster-Lyapunov-type criteria

For discrete-time Markov chains $(Y_n)_{n \in \mathbb{N}}$ the drift function $d_g : \mathcal{S} \to \mathbb{R}$ with respect to a function $g : \mathcal{S} \to \mathbb{R}$ is defined by

$$d_g(i) = \mathrm{E}[g(Y_n) - g(Y_{n-1}) | Y_{n-1} = i] = \sum_{j \in \mathcal{S}} p_{ij} g(j) - g(i), \tag{4}$$

that is, when writing g and d_g in the form of column vectors, $d_g = Pg - g$. Hence, $d_g(i)$ is the (generalized) drift in state i with respect to g.

For continuous-time Markov chains $(X_t)_{t \geq 0}$ the drift function $d_g : \mathcal{S} \to \mathbb{R}$ with respect to a function $g : \mathcal{S} \to \mathbb{R}$ is defined by

$$d_g(i) = \frac{d}{dt} \mathrm{E}[g(X_t) | X_t = i] = \sum_{j \in \mathcal{S}} q_{ij} g(j), \tag{5}$$

that is, when writing g and d_g in the form of column vectors, we have $d_g = Qg$.

For finite $\mathcal{C} \subset \mathcal{S}$, $\gamma > 0$ and $f, g : \mathcal{S} \to \mathbb{R}_{\geq 0}$ we consider the conditions

(C1) $\forall i \in \mathcal{S} \setminus \mathcal{C} : \ d_g(i) \leq -\gamma f(i)$,

(C2) $\forall i \in \mathcal{C} : \ d_g(i) < \infty$,

(C3) $\forall r < \infty : \ |\{i \in \mathcal{S} : \ g(i) \leq r\}| < \infty$.

Conditions of this form are often referred to as Foster-Lyapunov-type criteria since they are generalizations of classical criteria for positive recurrence or ergodicity, respectively, of Markov chains. For discrete-time Markov chains, in the special case where $f(i) = 1$ we have a criterion for positive recurrence, which in fact is very famous. In the case $|\mathcal{C}| = 1$ it is due to Foster [10], in the slightly more general case of arbitrary finite $\mathcal{C}$ it was proven by Pakes [17]. For continuous-time Markov chains, in the special case where $f(i) = 1$ we have a famous criterion for regularity and positive recurrence, which is due to Tweedie [21, Theorem 2.3]. Appropriate functions g with respect to which the drift function d_g is defined are often called Lyapunov functions and the conditions on d_g as (generalized) drift conditions.

Theorem 1. *Let $(Y_n)_{n\in\mathbb{N}}$ be an irreducible discrete-time Markov chain with transition probability matrix $P = (p_{ij})_{i,j\in\mathcal{S}}$, $\mathcal{C} \subset \mathcal{S}$ finite, $\gamma > 0$ and let $f, g : \mathcal{S} \to \mathbb{R}_{\geq 0}$ meet the conditions (C1)–(C3). Then $(Y_n)_{n\in\mathbb{N}}$ is recurrent and for any invariant measure ψ the sum $\psi f = \sum_{i\in\mathcal{S}} \psi_i f(i)$ is finite.*

The recurrence of $(Y_n)_{n\in\mathbb{N}}$ follows from [21, Theorem 3.3], so that an invariant measure ψ exists. The finiteness of ψf follows as a special case of [23, Theorem 1], where Markov chains in discrete-time on a general state space are considered without any irreducibility assumption, which generalizes an earlier result for ergodic discrete-time Markov chains on a general state space [22, Theorem 1].

Theorem 2. *Let $(X_t)_{t\geq 0}$ be an irreducible continuous-time Markov chain with generator matrix $Q = (q_{ij})_{i,j\in\mathcal{S}}$, $\mathcal{C} \subset \mathcal{S}$ finite, $\gamma > 0$ and let $f, g : \mathcal{S} \to \mathbb{R}_{\geq 0}$ meet the conditions (C1)–(C3). Then Q is regular (it uniquely defines $(X_t)_{t\geq 0}$, the Feller process of Q), $(X_t)_{t\geq 0}$ is recurrent and for any invariant measure ψ the sum $\psi f = \sum_{i\in\mathcal{S}} \psi_i f(i)$ is finite.*

The regularity and the recurrence follow from [21, Theorem 2.2], so that an invariant measure ψ exists. The finiteness of ψf can be shown by applying Theorem 1 to the embedded jump chain of $(X_t)_{t\geq 0}$ (cf. [6]). Hence, consider the embedded discrete-time jump chain $(Y_n)_{n\in\mathbb{N}}$ with transition probability matrix $P^* = (p^*_{ij})_{i,j\in\mathcal{S}}$ given by

$$p^*_{ij} = \frac{1}{q_i} q_{ij} + \delta_{ij}, \tag{6}$$

where $q_i = -q_{ii}$. Since our continuous-time Markov chain $(X_t)_{t\geq 0}$ is irreducible recurrent, the jump chain $(Y_n)_{n\in\mathbb{N}}$ is irreducible recurrent, too (see, e.g., [1, pp. 184–188]). An invariant measure ψ^* for the jump chain is given by $\psi_j^* = q_j\psi_j$. The continuous-time drift condition

$$\sum_{j=0}^{\infty} q_{ij} g(j) \leq -\gamma f(i) \tag{7}$$

yields the drift condition

$$\sum_{j=0}^{\infty} p_{ij}^* g(j) - g(i) = \sum_{j=0}^{\infty} \left(\frac{q_{ij}}{q_i} + \delta_{ij}\right) g(j) - g(i) = \frac{1}{q_i} \sum_{j=0}^{\infty} q_{ij} g(j) = \frac{d_g(i)}{q_i} \leq -\frac{\gamma f(i)}{q_i} \tag{8}$$

for the embedded chain. Thus, according to Theorem 1, $\psi^* f^*$ with $f^*(i) = \frac{f(i)}{q_i}$ is finite. Since obviously $\psi^* f^* = \psi f$ the proof is completed.

The following is fundamental for obtaining the state space truncation procedure in the next section.

Theorem 3. *Let $(Y_n)_{n\in\mathbb{N}}$ be an irreducible recurrent discrete-time Markov chain with transition probability matrix $P = (p_{ij})_{i,j\in\mathcal{S}}$, let ψ be an invariant measure, and let $d_g = Pg - g \leq h$ for some ψ-integrable function $h \geq 0$. Then $\psi d_g = \psi(Pg - g) \geq 0$.*

Proof. Without loss of generality let $\mathcal{S} = \mathbb{N}$. Define

$$\ell_{ij}^{(n)} := P(Y_n = j,\ Y_{n-1}, \ldots, Y_1 \neq i \mid Y_0 = i) \quad i, j \in \mathbb{N},\ n \geq 1, \tag{9}$$

$$\psi_j := \sum_{n=1}^{\infty} \ell_{0j}^{(n)}, \quad j \in \mathbb{N}. \tag{10}$$

Then for the $\ell_{ij}^{(n)}$, we have the recursion

$$\ell_{ij}^{(1)} = p_{ij}, \quad \ell_{ij}^{(n)} = \sum_{k\neq i} \ell_{ik}^{(n-1)} p_{kj}, \quad n \geq 2, \tag{11}$$

which yields

$$\psi_j = \sum_{n=1}^{\infty} \ell_{0j}^{(n)} = p_{0j} + \sum_{n=2}^{\infty}\sum_{k=1}^{\infty} \ell_{0k}^{(n-1)} p_{kj} = p_{0j} + \sum_{k=1}^{\infty}\sum_{n=1}^{\infty} \ell_{0k}^{(n)} p_{kj} \tag{12}$$

$$= p_{0j} + \sum_{k=1}^{\infty} \psi_k p_{kj} = \sum_{k=0}^{\infty} \psi_k p_{kj}, \tag{13}$$

since ψ_0 is the probability of eventually returning to state 0, and thus, due to recurrence, $\psi_0 = 1$. Therefore, ψ is an invariant measure, and since any other invariant measure is obtained by multiplication with some positive scalar, it is sufficient to consider only this particular invariant measure.

Remark: In the general case (no recurrence required), by $(1, \psi_1, \psi_2, \ldots)$ the minimal subinvariant measure ψ with $\psi_0 = 1$ is defined, see, e.g., [1, pp. 172–174] for more details.

Now, for $N \in \mathbb{N}$ and $j \in \{0, \ldots, N\}$ define

$$\ell_{0j}^{(n,N)} := P(Y_n = j,\ Y_{n-1}, \ldots, Y_1 \in \{1, \ldots, N\} \mid Y_0 = 0), \tag{14}$$

$$\psi_j^{(N)} := \sum_{n=1}^{\infty} \ell_{0j}^{(n,N)}. \tag{15}$$

With similar considerations as above, we have the recursion

$$\ell_{0j}^{(1,N)} = p_{0j}, \quad \ell_{0j}^{(n,N)} = \sum_{k=1}^{N} \ell_{0k}^{(n-1,N)} p_{kj}, \quad n \geq 2, \tag{16}$$

which yields

$$\psi_j^{(N)} = \sum_{n=1}^{\infty} \ell_{0j}^{(n,N)} = p_{0j} + \sum_{n=2}^{\infty} \sum_{k=1}^{N} \ell_{0k}^{(n-1,N)} p_{kj} = p_{0j} + \sum_{k=1}^{N} \sum_{n=1}^{\infty} \ell_{0k}^{(n,N)} p_{kj} \tag{17}$$

$$= p_{0j} + \sum_{k=1}^{N} \psi_k^{(N)} p_{kj}. \tag{18}$$

Defining $\psi_j^{(N)} = 0$ for $j > N$ and $\psi^{(N)} = \left(\psi_j^{(N)}\right)_{j=0}^{\infty}$, we can state that $\psi_j^{(N)}$ increases monotonically in N with $\lim_{N\to\infty} \psi^{(N)} = \psi$ (componentwise, weak convergence). Now, we can consider $\psi^{(N)}(Pg-g)$ instead of $\psi(Pg-g)$. Since $\psi_j^{(N)} = 0$ for almost all $j \in \mathbb{N}$, there is no problem when changing the order of summation. We can write

$$
\begin{aligned}
\psi^{(N)} d_g &= \psi^{(N)}(Pg-g) = \sum_{i=0}^{N} \psi_i^{(N)} \left(\sum_{j=0}^{\infty} p_{ij} g(j) - g(i) \right) \\
&= \psi_0^{(N)} \left(\sum_{j=0}^{\infty} p_{0j} g(j) - g(0) \right) + \sum_{i=1}^{N} \psi_i^{(N)} \left(\sum_{j=0}^{\infty} p_{ij} g(j) - g(i) \right) \\
&= \psi_0^{(N)} \sum_{j=0}^{\infty} p_{0j} g(j) - \psi_0^{(N)} g(0) + \sum_{i=1}^{N} \psi_i^{(N)} \sum_{j=0}^{\infty} p_{ij} g(j) - \sum_{i=1}^{N} \psi_i^{(N)} g(i) \\
&= \psi_0^{(N)} \sum_{j=0}^{\infty} p_{0j} g(j) + \sum_{i=1}^{N} \psi_i^{(N)} \sum_{j=0}^{\infty} p_{ij} g(j) - \sum_{i=0}^{N} \psi_i^{(N)} g(i) \\
&= \psi_0^{(N)} \sum_{j=0}^{N} p_{0j} g(j) + \sum_{i=1}^{N} \psi_i^{(N)} \sum_{j=0}^{N} p_{ij} g(j) - \sum_{i=0}^{N} \psi_i^{(N)} g(i) + \sum_{i=0}^{N} \psi_i^{(N)} \sum_{j=N+1}^{\infty} p_{ij} g(j) \\
&= \psi_0^{(N)} \sum_{j=0}^{N} p_{0j} g(j) + \sum_{j=0}^{N} g(j) \sum_{i=1}^{N} \psi_i^{(N)} p_{ij} - \sum_{i=0}^{N} \psi_i^{(N)} g(i) + \sum_{i=0}^{N} \psi_i^{(N)} \sum_{j=N+1}^{\infty} p_{ij} g(j) \\
&= \psi_0^{(N)} \sum_{j=0}^{N} p_{0j} g(j) + \sum_{j=0}^{N} g(j) \left(\psi_j^{(N)} - p_{0j} \right) - \sum_{i=0}^{N} \psi_i^{(N)} g(i) + \sum_{i=0}^{N} \psi_i^{(N)} \sum_{j=N+1}^{\infty} p_{ij} g(j) \\
&= \left(\psi_0^{(N)} - 1 \right) \sum_{j=0}^{N} p_{0j} g(j) + \sum_{i=0}^{N} \psi_i^{(N)} \sum_{j=N+1}^{\infty} p_{ij} g(j) \\
&\geq \left(\psi_0^{(N)} - 1 \right) \sum_{j=0}^{N} p_{0j} g(j).
\end{aligned}
$$

Due to finiteness of $d_g(0)$ and recurrence, we have

$$
\lim_{N\to\infty} \sum_{j=0}^{N} p_{0j} g(j) = d_g(0) + g(0) < \infty \qquad \text{and} \qquad \lim_{N\to\infty} \psi_0^{(N)} = 1,
$$

and thus, we obtain

$$
\limsup_{N\to\infty} \psi^{(N)} d_g \geq \lim_{N\to\infty} \left(\psi_0^{(N)} - 1 \right) \sum_{j=0}^{N} p_{0j} g(j) = 0.
$$

Due to $d_g \leq h$, we have $\psi_j(h(j) - d_g(j)) \geq 0$ for all $j \in \mathbb{N}$, and Fatou's Lemma yields

$$\begin{aligned}
\psi(h - d_g) &= \sum_{j=0}^{\infty} \psi_j(h(j) - d_g(j)) \\
&= \sum_{j=0}^{\infty} \lim_{N\to\infty} \psi_j^{(N)}(h(j) - d_g(j)) \\
&\leq \liminf_{N\to\infty} \sum_{j=0}^{\infty} \psi_j^{(N)}(h(j) - d_g(j)) \\
&= \liminf_{N\to\infty} \sum_{j=0}^{\infty} \psi_j^{(N)} h(j) - \limsup_{N\to\infty} \sum_{j=0}^{\infty} \psi_j^{(N)} d_g(j).
\end{aligned}$$

Since h is ψ-integrable, by monotone convergence, we obtain

$$\lim_{N\to\infty} \sum_{j=0}^{\infty} \psi_j^{(N)} h(j) = \sum_{j=0}^{\infty} \psi_j h(j) = \psi h,$$

and hence

$$0 \leq \psi(h - d_g) \leq \psi h - \limsup_{N\to\infty} \psi^{(N)} d_g \leq \psi h.$$

From these inequalities, we obtain that $h - d_g$, and thus, d_g is ψ-integrable with $\psi d_g \geq 0$.

Now, we give an analogous theorem for continuous time Markov chains.

Theorem 4. *Let $(X_t)_{t\geq 0}$ be an irreducible recurrent continuous-time Markov chain with generator matrix $Q = (q_{ij})_{i,j\in\mathcal{S}}$, $d_g = Qg \leq h$ for some ψ-integrable function h. Then $\psi d_g = \psi Qg \geq 0$ for any invariant measure ψ.*

Proof. Consider again the embedded discrete-time jump chain $(Y_n)_{n\in\mathbb{N}}$ with transition probability matrix $P^* = (p^*_{ij})_{i,j\in\mathcal{S}}$ given by (6), i.e. $p^*_{ij} = \frac{1}{q_i} q_{ij} + \delta_{ij}$, invariant measure ψ^* given by $\psi^*_j = q_j \psi_j$, drift d^*_g where $d^*_g(i) = \frac{d_g(i)}{q_i}$ (cf. (8)), and upper drift bound h^* given by $h^*(i) = \frac{h(i)}{q_i}$. Obviously, we have $\psi^* d^*_g = \psi d_g$, and since h^* is ψ^*-integrable, Theorem 3 yields $\psi^* d^*_g \geq 0$.

Before we apply these results to the task of truncating the state space, we make some remarks concerning the drift bound h.

- Under the conditions of Theorem 1 or Theorem 2, respectively, we always have $d_g(j) \leq 0$ for all $j \in \mathcal{S} \setminus \mathcal{C}$, and since $\mathcal{C}$ is finite, $h = d_g 1_{\mathcal{C}} \geq d_g$ is trivially ψ-integrable.

- A special case of our theorems appears in [12] where positive recurrence and thus the summability of ψ is assumed, and the corresponding drift condition is $\sup d_g(j) < \infty$. In this case $h = C = \sup d_g(j)$ is of course ψ-integrable. For a finite state space $\mathcal{S}$ we trivially could state '$= 0$', and '≥ 0' is guaranteed by [12, Theorem 1 (ii)] in a quite general context, where the continuous-time case is given as a special case by [12, Proposition 1].

3. Truncation of the state space

Now we exploit the theorems of the previous section in order to determine finite subsets of the state space such that the truncation error is bounded as stipulated by (3). The drift conditions provided by the aforementioned criterion by Tweedie [21, Theorem 2.3] for regularity and positive recurrence have been exploited in [9] in order to obtain bounds of the form $\sum_{i\in\mathcal{C}} \pi_i \geq 1-\epsilon$ when approximating the stationary distributions of infinite ergodic continuous-time level dependent quasi-birth-and-death (LDQBD) processes. We will use a similar approach for deriving the desired bounds (3) for recurrent continuous-time and discrete-time Markov chains on countable state spaces. The main idea is to choose $\gamma > 0$ and g such that the finite set $\mathcal{C}$ is appropriate for truncation. This procedure is based on the following result:

Theorem 5. *Let $f, g, \gamma, \mathcal{C}$ meet the conditions of Theorem 1 or Theorem 2, respectively. Furthermore, let*

$$f(j) > 0 \quad \textit{for all} \quad j \in \mathcal{C}_0 := \{i \in \mathcal{S} :\ d_g(i) > 0\}, \tag{19}$$

and let $f(j_1) > 0$ for some $j_1 \in \overline{\mathcal{C}}_0 := \mathcal{S} \setminus \mathcal{C}_0$. Then we have

$$\frac{\sum\limits_{j\notin\mathcal{C}} \psi_j f(j)}{\sum\limits_{j\in\mathcal{S}} \psi_j f(j)} \leq \frac{c}{c+\gamma}, \tag{20}$$

where

$$c = \max_{j\in\mathcal{C}_0} \frac{d_g(j)}{f(j)} > 0. \tag{21}$$

Proof. First note that $\mathcal{C}_0 := \{j \in \mathcal{S} :\ d_g(j) > 0\}$ has finitely many elements due to condition (C1). Furthermore, this condition guarantees that $d_g(j_1) < 0$, yielding $\psi_{j_1} d_g(j_1) < 0$ since the invariant measure ψ has no zero-entry. As pointed out above, by Theorem 3 and Theorem 4, under the conditions of Theorem 1 or Theorem 2 respectively, we have $\psi d_g \geq 0$, and therefore, there is some $j_0 \in \mathcal{C}_0$ with $d_g(j_0) > 0$, that is $\mathcal{C}_0 \neq \emptyset$. Assumption (19) guarantees that $c > 0$ is well-defined by (21).

Now we scale g and thus d_g by $\frac{1}{c+\gamma} > 0$, that is

$$g^*(j) := \frac{g(j)}{c+\gamma}, \qquad d_{g^*}(j) = \frac{d_g(j)}{c+\gamma}. \tag{22}$$

This yields $d_{g^*}(j) \leq \frac{cf(j)}{c+\gamma}$ for $j \in \mathcal{C}$ and $d_{g^*}(j) \leq -\frac{\gamma f(j)}{c+\gamma}$ for $j \notin \mathcal{C}$, or, written in concise form,

$$d_{g^*}(j) \leq \left(\frac{c}{c+\gamma} - 1_{\overline{\mathcal{C}}}(j) \right) f(j). \tag{23}$$

Summation of $\psi_j d_{g^*}(j)$ yields

$$0 \leq \psi d_{g^*} = \sum_{j \in \mathcal{S}} \psi_j d_{g^*}(j) \leq \frac{c}{c+\gamma} \sum_{j \in \mathcal{S}} \psi_j f(j) - \sum_{j \notin \mathcal{C}} \psi_j f(j), \tag{24}$$

Due to the assumptions, $\psi f \neq 0$, which immediately implies (20) and completes the proof.

Remark: If $f(j) > 0$ for infinitely many $j \in \mathcal{S}$, due to finiteness of $\mathcal{C}$, there will always be some $j_1 \in \overline{\mathcal{C}}$ with $f(j_1) > 0$. If $f(j) > 0$ holds only for finitely many $j \in \mathcal{S}$, the truncation of the state space is quite easy, we can simply choose the finite set $\mathcal{C} = \{j : \ f(j) > 0\}$. However, Theorem 5 and its proof can be extended to this case via the obvious inequality

$$0 = \sum_{j \notin \mathcal{C}} \psi_j f(j) \leq \frac{c}{c+\gamma} \sum_{j \in \mathcal{S}} \psi_j f(j),$$

where we define $c = 0$ for $\mathcal{C}_0 = \emptyset$ (which is possible in this situation).

With $\epsilon = \frac{c}{c+\gamma}$, Theorem 5 yields the desired bound for our procedure of determining a finite set $\mathcal{C}$ meeting (3). When a Lyapunov function g is given, c is determined by the corresponding drift function d_g and thus we can only vary γ. For guaranteeing $\frac{c}{c+\gamma} = \epsilon$ we choose $\gamma = \frac{c}{\epsilon} - c$. Since $\epsilon < 1$ and $c > 0$ we have $\gamma > 0$. If for this choice of γ the set $\mathcal{C} = \{j \in \mathcal{S} : \ d_g(j) > -\gamma f(j)\}$ is finite, we have an appropriate truncation of the state space. Otherwise we have to find a new Lyapunov function. A simple scaling does not help in this case since by definition of γ and c, $\mathcal{C}$ is invariant with respect to scaling of g.

TABLE 1: State transitions of the gene expression example

From State	To State	Rate
(x_1, x_2)	$(x_1 + 1, x_2)$	λ
(x_1, x_2)	$(x_1, x_2 + 1)$	μx_1
(x_1, x_2)	$(x_1 - 1, x_2)$	$\delta_1 x_1$
(x_1, x_2)	$(x_1, x_2 - 1)$	$\delta_2 x_2$

The results just derived provide the basis for a method of finding an appropriate set $\mathcal{C}_0$ for truncating the sum ψf as follows:

1. Choose a Lyapunov function g
2. Compute the drift d_g.
3. Determine

$$\begin{aligned}
\mathcal{C}_0 &= \{i \in \mathcal{S} :\ d_g(i) > 0\}, && (25)\\
c &= \max_{j \in \mathcal{C}_0} \frac{d_g(j)}{f(j)}, && (26)\\
\gamma &= \frac{c}{\epsilon} - c, && (27)\\
\mathcal{C} &= \{j \in \mathcal{S} :\ d_g(j) > -\gamma f(j)\}. && (28)
\end{aligned}$$

4. If $\mathcal{C}$ is finite, (3) holds. Otherwise choose a new Lyapunov function and restart with 2.

4. Examples

Now, we demonstrate our state space truncation approach by two illustrative examples, where we restrict ourselves to continuous-time Markov chains, as the truncation procedure works similarly in the discrete-time case.

Example 1. We start with an example of a two-dimensional continuous-time Markov chain $(X_t) = \left(X_t^{(1)}, X_t^{(2)}\right)$ with state space $\mathcal{S} = \mathbb{N} \times \mathbb{N}$ and transitions according to Table 1 with parameters $\lambda, \mu, \delta_1, \delta_2 > 0$, which describes a stochastic gene expression model [20] and has been also considered in [9]. As a concrete numerical case we choose

$\lambda = 60$, $\mu = \delta_2 = 0.01$ and $\delta_1 = 0.2$. In [9], the authors looked for a set $\mathcal{C}$ fulfilling

$$\frac{\sum\limits_{(x_1,x_2)\in\mathcal{C}} \pi_{(x_1,x_2)}}{\sum\limits_{(x_1,x_2)\in\mathcal{S}} \pi_{(x_1,x_2)}} \geq 1 - \epsilon, \tag{29}$$

where π is the stationary distribution. The method used in [9] is the above method with $f(x_1, x_2) = 1$, the Lyapunov function g was defined by

$$g(x_1, x_2) = (x_1 - 300)^2 + (x_2 - 300)^2, \tag{30}$$

yielding the drift function

$$d_g(x_1, x_2) = -0.4x_1^2 - 0.02x_2^2 + 0.02x_1x_2 + 234.21x_1 + 6.01x_2 - 35940. \tag{31}$$

Obviously, this drift function is negative up to finitely many values (x_1, x_2), the maximum according to (26) is $c = 126$, and for $\epsilon = 0.05$, by (27), we have to choose $\gamma = 2394$. Thus, from (28) we obtain

$$\mathcal{C} = \{(x_1, x_2) : \; d_g(x_1, x_2) \geq -2394\} \tag{32}$$

as a finite subset of the state space that meets the desired truncation error bound.

It is clear that our truncation procedure does not require a specific transition structure or a specific numbering of the states. In many applications, however, the above characterization of $\mathcal{C}$ might be relatively unpractical, in particular when a specific numbering of the states is given and the generator matrix of the Markov chain must be truncated to render numerical computations possible. Think for example of infinite LDQBD processes, where the states are ordered according to the chosen level definition and the block structured generator matrix is truncated at certain blocks corresponding to high (or low) level numbers such as, e.g., in [3, 4, 5, 7, 9, 15]. Then it is often more convenient to consider an appropriate finite superset of $\mathcal{C}$ rather than to work directly with $\mathcal{C}$. For instance, simple algebra yields

$$\mathcal{C} \subset \{(x_1, x_2) : \; 221 \leq \max\{x_1, x_2\} \leq 657\} \tag{33}$$

$$\mathcal{C} \subset \{(x_1, x_2) : \; 250 \leq x_1 + x_2 \leq 975\}, \tag{34}$$

where the first superset can be found in [9] too. It contains 384123 states, the second one contains 445401 states.

The Lyapunov function defined above is still appropriate when considering the stationary moments of the first or the second component respectively, that is, we consider $\mathrm{E}_\pi[f^{(1)}]$ and $\mathrm{E}_\pi[f^{(2)}]$ with $f^{(1)}(x_1,x_2)=x_1$ and $f^{(2)}(x_1,x_2)=x_2$, respectively. It is easy to see that

$$\mathcal{C}_1=\{(x_1,x_2):\ d_g(x_1,x_2)\geq -\gamma x_1\} \text{ and } \mathcal{C}_2=\{(x_1,x_2):\ d_g(x_1,x_2)\geq -\gamma x_2\} \quad (35)$$

are finite for arbitrary $\gamma>0$. By simple algebra, we obtain $d_g(x_1,0)<0$ and $d_g(0,x_2)<0$ for any $x_1,x_2\in\mathbb{N}$, yielding that $\mathcal{C}_0=\{(x_1,x_2):\ d_g(x_1,x_2)>0\}$ does not contain any point (x_1,x_2) with $f^{(j)}(x_1,x_2)=0$ for $j=1$ or $j=2$. Thus, there is no problem when defining the value c according to (26), for $f^{(1)}$ we have $c=0.42$ and for $\epsilon=0.05$, by (27), we obtain $\gamma=7.98$. For simplicity and means of comparison, we give supersets for $\mathcal{C}_1$ similar to those given above, we have

$$\mathcal{C}_1 \subset \{(x_1,x_2):\ 231\leq \max\{x_1,x_2\}\leq 665\}, \quad (36)$$

$$\mathcal{C}_1 \subset \{(x_1,x_2):\ 261\leq x_1+x_2\leq 993\}, \quad (37)$$

where the first superset contains 390195 states and the second one contains 460324 states.

Similarly, for $f^{(2)}$, we have $c\approx 0.4272$ and $\gamma\approx 8.1176$ (for $\epsilon=0.05$), yielding

$$\mathcal{C}_2 \subset \{(x_1,x_2):\ 230\leq \max\{x_1,x_2\}\leq 920\}, \quad (38)$$

$$\mathcal{C}_2 \subset \{(x_1,x_2):\ 377\leq x_1+x_2\leq 1246\}, \quad (39)$$

where the first superset contains 795341 states and the second one contains 706875 states.

Example 2. We continue with a simple but extremely instructive example that demonstrates the applicability of the state space truncation procedure to non-ergodic recurrent Markov chains and shows some peculiarities with regard to the specific choice of a Lyapunov function.

Consider a birth-death-process $(X_t)_{t\geq 0}$ with birth rate λ and death rate λ, that is a Markov chain with state space $\mathbb{N}$ and the generator matrix

$$Q = \begin{pmatrix} -\lambda & \lambda & & & \\ \lambda & -2\lambda & \lambda & & \\ & \lambda & -2\lambda & \lambda & \\ & & \ddots & \ddots & \ddots \end{pmatrix}.$$

Obviously, $\psi = (1, 1, \ldots)$ is an invariant measure. Consider the computation of

$$H = \lim_{t\to\infty} \frac{\int_0^t \frac{1}{(X_s+1)^2}\, ds}{\int_0^t \frac{1}{2^{X_s}}\, ds} = \frac{\psi f^{(1)}}{\psi f^{(2)}}, \quad f^{(1)}(j) = \frac{1}{(j+1)^2},\ f^{(2)}(j) = \frac{1}{2^j}. \tag{40}$$

We want to use our method for finding finite sets $\mathcal{C}_1$ and $\mathcal{C}_2$ such that

$$(1-\epsilon_1)H \le H^* \le \frac{1}{1-\epsilon_2} H, \tag{41}$$

where

$$H^* = \frac{\sum_{j\in\mathcal{C}_1} \psi_j f^{(1)}(j)}{\sum_{j\in\mathcal{C}_2} \psi_j f^{(2)}(j)} \tag{42}$$

is the approximation obtained by finite summation. Obviously,

$$\frac{\sum_{j\in\mathcal{C}_i} \psi_j f^{(i)}(j)}{\sum_{j=0}^{\infty} \psi_j f^{(i)}(j)} \ge 1-\epsilon_i, \qquad i = 1, 2, \tag{43}$$

is sufficient. We start with considering $f^{(1)}(j) = \frac{1}{(j+1)^2}, j \in \mathbb{N},$. Since $f^{(1)}(j) > 0$ for all $j \in \mathbb{N}$ there are no problems when defining $c = \max_{j\in\mathcal{C}_0} \frac{d_g(j)}{f(j)}$ according to (26), independent of the Lyapunov function g. When directly using Theorem 2, an appropriate choice for g is

$$g(j) = \sum_{k=1}^{j+1} \frac{1}{k}, \quad j \in \mathbb{N}. \tag{44}$$

Then we have

$$d_g(0) = \frac{\lambda}{2}, \quad d_g(j) = -\frac{\lambda}{(j+1)(j+2)},\ j = 1, 2, \ldots. \tag{45}$$

For sufficiently large $j \in \mathbb{N}$ there exists $\gamma > 0$ such that $d_g(j) \le -\gamma f(j)$. Therefore, by Theorem 2, $\psi f^{(1)} < \infty$. Now, we choose γ and $\mathcal{C}$ according to (27) and (28). Hence,

$\gamma = \frac{c}{\epsilon_1} - c$, where $c = d_g(0) = \frac{\lambda}{2}$, implying

$$\mathcal{C}_1 = \left\{ j \in \mathcal{S} : \; -\frac{\lambda}{(j+1)(j+2)} > -\left(\frac{\lambda}{2\epsilon_1} - \frac{\lambda}{2}\right) \frac{1}{(j+1)^2} \right\}. \tag{46}$$

It is straightforward to see that $\mathcal{C}_1$ is a finite set if and only if $\epsilon_1 > \frac{1}{3}$, which for reasonable ϵ_1 is of course not true. Therefore, we have to find a new Lyapunov function. Choose

$$g(j) = \sum_{k=1}^{j+1} \frac{1}{\sqrt{k}}, \quad j \in \mathbb{N}. \tag{47}$$

Then we have $c = d_g(0) = \frac{\lambda}{\sqrt{2}}$ and

$$d_g(j) = -\frac{\lambda}{\sqrt{(j+1)(j+2)}\left(\sqrt{j+1} + \sqrt{j+2}\right)}, \quad j = 1, 2, \ldots. \tag{48}$$

Since $d_g(j) \approx \frac{\lambda}{2j^{\frac{3}{2}}}$ for large j,

$$\mathcal{C}_1 = \left\{ j : \; -\frac{\lambda}{\sqrt{(j+1)(j+2)}\left(\sqrt{j+1} + \sqrt{j+2}\right)} > -\left(\frac{\lambda}{2\epsilon_1} - \frac{\lambda}{2}\right) \frac{1}{(j+1)^2} \right\} \tag{49}$$

is finite for any $\epsilon_1 > 0$. For $\epsilon_1 = 0.05$ we obtain $\mathcal{C}_1 = \{0, 1, \ldots, 361\}$.

Now, consider $f^{(2)}(j) = \frac{1}{2^j}, j \in \mathbb{N}$. Our first Lyapunov function g, defined by

$$g(j) = \sum_{k=1}^{j+1} \frac{1}{k}, \quad j \in \mathbb{N}, \tag{50}$$

is appropriate for $\psi f^{(2)} < \infty$ by Theorem 2, too. In this case, this Lyapunov function can be used for defining $\mathcal{C}_2$ since

$$\mathcal{C}_2 = \left\{ j \in \mathcal{S} : \; -\frac{\lambda}{(j+1)(j+2)} > -\left(\frac{\lambda}{2\epsilon_2} - \frac{\lambda}{2}\right) \frac{1}{2^j} \right\} \tag{51}$$

is finite for all $\epsilon_2 > 0$. For $\epsilon_2 = 0.05$ we obtain $\mathcal{C}_2 = \{0, \ldots, 10\}$.

Note that since for this example we know the exact invariant measure $\psi = (1, 1, 1, \ldots)$ and $\psi f^{(1)} = \frac{\pi^2}{6}$, we can easily determine the 'best choice' for $\mathcal{C}_1$, namely $\mathcal{C}_1 = \{0, 1, \ldots, 11\}$. Similarly, from $\psi f^{(2)} = 2$ we know that $\mathcal{C}_2 = \{0, \ldots, 4\}$ would be the best choice.

Hence, the example demonstrates that there are Lyapunov functions that meet the conditions of Theorem 2 but are not suitable for our state space truncation procedure. Additionally, we see that the truncations we obtain are quite conservative. This implies that the truncation errors are actually much smaller than requested. This can be interpreted as an advantage, but we also have to consider that usually we have to solve for ψ and/or ψf numerically, implying that conservative truncations imply higher effort. Of course, tight bounds are desirable.

5. Conclusion

With regard to long run averages of additive functionals in infinite recurrent Markov chains, we have exploited Foster-Lyapunov-type drift conditions in order to obtain finite subsets of the infinite state space such that at most a prescribed (small) portion of the long run average lies outside this finite set. This can be taken as a state space truncation method with bounded truncation error, which is extremely useful for, e.g., numerically computing long run averages, where a state space truncation is inevitable. The approach is independent of specific ways of computing long run averages. In either case, it provides a bound on the approximation error due to the state space truncation. Error bounds for long run averages rather than for probabilities are particularly valuable when we have a method available that computes long run averages without explicitly relying on the stationary distribution (if it exists) or an invariant measure. In particular, the state space truncation method solves the open issue that the memory-efficient matrix-analytic method presented in [5] for computing stationary expectations in LDQBD processes without at first explicitly computing the stationary distribution was lacking an accuracy measure. Now, in conjunction with the state space truncation method of the present paper, [5] constitutes a powerful matrix-analytic method for numerically approximating long run averages of additive functionals in infinite recurrent LDQBD processes, where an approximation error bound can be specified a priori. This enormously advances the state of the art in matrix-analytic computations and their applicability to, e.g., performance analysis of complex networks with infinite multi-dimensional state spaces. Moreover, as the state space truncation method is not restricted to Markov chains with a specific transition structure, it provides many new options for the analysis of a large class of stochastic models.

A couple of further research issues arise. We have considered nonnegative functions

f, which makes sense, since for applying generalized ergodic theorems we have to guarantee the finiteness of $\psi|f|$. In many applications, nonnegative functions are indeed sufficient to model the problem at hand. Nevertheless, further research on generalizations to arbitrary functions f is desirable and currently ongoing. For the tightness of the approximation error bounds, the chosen Lyapunov function is a crucial factor. Not all Lyapunov functions that guarantee the finiteness of the long run average under consideration are suitable for our state space truncation method, some lead to infinite subsets. Furthermore, even if suitable, different Lyapunov functions yield different finite subsets corresponding to different tightness of the respective bounds. Hence, the systematic derivation of Lyapunov functions that are good in the sense of yielding as tight bounds as possible deserves further attention. For instance, restricted function classes might be considered as candidate Lyapunov functions and their properties with regard to the state space truncation method are to be studied.

References

[1] Anderson, W. J. (1991). *Continuous-time Markov Chains.* Springer-Verlag.

[2] Asmussen, S. (2003). *Applied Probability and Queues* 2nd ed. Springer-Verlag.

[3] Baumann, H. and Sandmann, W. (2010). Numerical solution of level dependent quasi-birth-and-death processes. *Procedia Computer Science* **1,** 1561–1569.

[4] Baumann, H. and Sandmann, W. (2012). Steady state analysis of level dependent quasi-birth-and-death processes with catastrophes. *Computers & Operations Research* **39,** 413–423.

[5] Baumann, H. and Sandmann, W. (2013). Computing stationary expectations in level-dependent QBD processes. *Journal of Applied Probability* **50,** 151–165.

[6] Baumann, H. and Sandmann, W. (2014). On finite long run costs and rewards in infinite Markov chains. *Statistics and Probability Letters* **91,** 41–46.

[7] Bright, L. and Taylor, P. G. (1995). Calculating the equilibrium distribution in level dependent quasi-birth-and-death processes. *Stochastic Models* **11,** 497–525.

[8] Chung, K. L. (1960). *Markov Chains with Stationary Transition Probabilities.* Springer-Verlag.

[9] Dayar, T., Sandmann, W., Spieler, D. and Wolf, V. (2011). Infinite level–dependent QBD processes and matrix-analytic solutions for stochastic chemical kinetics. *Advances in Applied Probability* **43,** 1005–1026.

[10] Foster, F. G. (1953). On the stochastic matrices associated with certain queueing processes. *Ann. Math. Statist.* **24,** 355–360.

[11] Gibson, S. and Seneta, E. (1987). Augmented truncation of infinite stochastic matrices. *Journal of Applied Probability* **24,** 600–608.

[12] Glynn, P. W. and Zeevi, A. (2008). Bounding stationary expectations of Markov processes. In *Markov Processes and Related Topics: A Festschrift for Thomas G. Kurtz.* vol. 4 of *IMS Collections.* Institute of Mathematical Statistics, Beachwood, OH pp. 195–214.

[13] Golub, G. H. and Seneta, E. (1973). Computation of the stationary distribution of an infinite Markov matrix. *Bull. Austral. Math. Soc.* **8,** 333–341.

[14] Golub, G. H. and Seneta, E. (1974). Computation of the stationary distribution of an infinite stochastic matrix of special form. *Bull. Austral. Math. Soc.* **10,** 255–261.

[15] Hanschke, T. (1999). A matrix continued fraction algorithm for the multiserver repeated order queue. *Mathematical and Computer Modelling* **30,** 159–170.

[16] Latouche, G. and Taylor, P. G. (2002). Truncation and augmentation of level-independent QBD processes. *Stochastic Processes and Their Applications* **99,** 53–80.

[17] Pakes, A. G. (1969). Some conditions for ergodicity and recurrence of Markov chains. *Operations Research* **17,** 1058–1061.

[18] Seneta, E. (1981). *Non-negative Matrices and Markov chains.* Springer-Verlag.

[19] Serfozo, R. (2009). *Basics of Applied Stochastic Processes.* Springer-Verlag.

[20] Thattai, M. and van Oudenaarden, A. (2001). Intrinsic noise in gene regulatory networks. *Proc. Nat. Acad. of Sci. USA* **98,** 8614–8619.

[21] Tweedie, R. L. (1975). Sufficient conditions for regularity, recurrence and ergodicity of Markov processes. *Math. Proc. Camb. Phil. Soc.* **78,** 125–136.

[22] Tweedie, R. L. (1983). The existence of moments for stationary Markov chains. *Journal of Applied Probability* **20,** 191–196.

[23] Tweedie, R. L. (1988). Invariant measures for Markov chains with no irreducibility assumptions. *Journal of Applied Probability* **25,** A Celebration of Applied Probability, 275–285.

[24] Tweedie, R. L. (1998). Truncation approximation of invariant measures for Markov chains. *Journal of Applied Probability* **35,** 517–536.

[25] Zhao, Y. Q. and Liu, D. (1996). The censored Markov chain and the best augmentation. *Journal of Applied Probability* **33,** 623–629.