ON TRUNCATIONS FOR SZK MODEL

Alexander Zeifman
Vologda State University, S.Orlova, 6, Vologda, Russia
Institute of Informatics Problems, Russian Academy of Sciences

INTRODUCTION

The problem of existence and construction of limiting characteristics for inhomogeneous continuous-time Markov chains is important for queueing applications, see for instance, (Granovsky and Zeifman 2004, Zeifman et al. 2006). Calculation of the limiting characteristics for birth-death process via truncations was firstly mentioned in (Zeifman 1991) and was considered in details in (Zeifman et al. 2006).

About two decades ago V. Kalashnikov suggested that in some cases one can obtain uniform (in time) error bounds of truncation, and in (Zeifman et al. 2014b) we prove this conjecture for inhomogeneous birth-death processes.

A new class of Markovian time-inhomogeneous queueing models with batch arrivals and group services was introduced and studied in our recent papers (Satin et al. 2013, Zeifman et al. 2014a). Bounds of the rate of convergence, perturbation bounds, and approximations via truncations were studied in these papers.

Here we consider this model and obtain uniform in time error bounds of truncation.

Consider a time-inhomogeneous continuous-time Markovian queueing model (“SZK model”) on the state space $E = \{0, 1, \ldots\}$ with possible batch arrivals and group services.

Let $X = X(t)$, $t \geq 0$ be a queue-length process for the queue, $p_{ij}(s,t) = \Pr \{X(t) = j \mid X(s) = i\}$, $i, j \geq 0$, $0 \leq s \leq t$, be transition probabilities for $X = X(t)$, and $p_i(t) = \Pr \{X(t) = i\}$ be its state probabilities. Throughout the paper we assume that

$$\Pr (X(t + h) = j \mid X(t) = i) = \begin{cases} q_{ij}(t) h + \alpha_{ij}(t, h), & \text{if } j \neq i, \\ 1 - \sum_{k \neq i} q_{ik}(t) h + \alpha_i(t, h), & \text{if } j = i, \end{cases}$$

where all $\alpha_i(t, h)$ are $o(h)$ uniformly in $i$, i.e. $\sup_{t, h} |\alpha_i(t, h)| = o(h)$. We also assume $q_{i,i+k}(t) = \lambda_k(t)$, $\eta_{i,i-k}(t) = \mu_k(t)$ for any $k > 0$. In other words, we suppose that the arrival rates $\lambda_k(t)$ and the service rates $\mu_k(t)$ do not depend on the the length of queue. In addition, we assume that $\lambda_{k+1}(t) \leq \lambda_k(t)$ and $\mu_{k+1}(t) \leq \mu_k(t)$ for any $k$ and almost all $t \geq 0$.

Applying our standard approach (see details in (Granovsky and Zeifman 2004, Zeifman 1995a, Zeifman et al. 2006, 2008)) we suppose in addition, that all intensity functions are linear combinations of a finite number of locally integrable on $[0, \infty)$ nonnegative functions. Moreover we assume

$$\lambda_k(t) \leq \lambda_k, \quad \mu_k(t) \leq m_k,$$

for any $k$ and almost all $t \geq 0$, and

$$L_\lambda = \sum_{i=1}^\infty \lambda_i < \infty, \quad L_\mu = \sum_{i=1}^\infty \mu_i < \infty.$$

Then the probabilistic dynamics of the process is represented by the forward Kolmogorov system:

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

where $A(t)$ is the infinitesimal generator of the process $X(t)$, $p(t)$ is the probability vector, and $A(t)$ is the matrix of transitions rates.
where \( A(t) = \begin{pmatrix} a_{ii}(t) & a_{i1}(t) & a_{i2}(t) & \cdots & a_{in}(t) \\ a_{1i}(t) & a_{11}(t) & a_{12}(t) & \cdots & a_{1n}(t) \\ a_{2i}(t) & a_{21}(t) & a_{22}(t) & \cdots & a_{2n}(t) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{ni}(t) & a_{n1}(t) & a_{n2}(t) & \cdots & a_{nn}(t) \end{pmatrix} \)

and \( a_{ii}(t) = -\sum_{k=1}^{n} \mu_k(t) - \sum_{k=1}^{n} \lambda_k(t) \).

Throughout the paper by \( \| \cdot \| \) we denote the \( l_1 \)-norm, i. e. \( \| x \| = \sum |x_i| \), and \( \| B \| = \sup \sum |b_{ij}| \) for \( B = (b_{ij})_{i,j=0}^{\infty} \).

Let \( \Omega \) be a set all stochastic vectors, i. e. \( l_1 \) vectors with nonnegative coordinates and unit norm. Hence we have \( \| A(t) \| = 2 \sum_{k=1}^{n} (\lambda_k(t) + \mu_k(t)) \leq 2 (L_\lambda + L_\mu) \) for almost all \( t \geq 0 \). Hence the operator function \( A(t) \) from \( l_1 \) into itself is bounded for almost all \( t \geq 0 \) and locally integrable on \([0, \infty)\). Therefore we can consider (4) as a differential equation in the space \( l_1 \) with bounded operator.

It is well known, see (Daleckij and Krein 1974), that the Cauchy problem for differential equation (4) has a unique solution for an arbitrary initial condition, and \( p(s) \in \Omega \) implies \( p(t) \in \Omega \) for \( t \geq s \geq 0 \).

Denote by \( E(t,k) = E \{ X(t) | X(0) = k \} \) the (mathematical expectation) of the process at the moment \( t \) under the initial condition \( X(0) = k \).

Recall that a Markov chain \( X(t) \) is called weakly ergodic, if \( \| p^*(t) - p^{**}(t) \| \rightarrow 0 \) as \( t \rightarrow \infty \) for any initial conditions \( p^*(0), p^{**}(0) \), where \( p^*(t) \) and \( p^{**}(t) \) are the corresponding solutions of (4).

A Markov chain \( X(t) \) has the limiting mean \( \varphi(t) \), if \( \lim_{t \rightarrow \infty} \varphi(t) = E(t,k) \).

Let \( \{d_i\}, i = 1, 2, \ldots \) be an increasing sequence of positive numbers, \( d_1 = 1 \). Put

\[
W = \inf_{i \geq 1} \frac{d_i}{i}, \quad g_i = \sum_{n=1}^{i} d_n,
\]

Denote

\[
\alpha_i(t) = -a_{ii}(t) - \sum_{k=1}^{i} \lambda_k(t) \frac{d_{k+1}}{d_i} - \sum_{k=1}^{i-1} (\mu_{i-k}(t) - \mu_i(t)) \frac{d_k}{d_i},
\]

and

\[
\alpha(t) = \inf_{i \geq 1} \alpha_i(t).
\]

Assume that for some positive \( K \)

\[
d_1 \lambda_1 + (d_1 + d_2) \lambda_2 + \cdots \leq K,
\]

By introducing \( p_i(t) = 1 - \sum_{j \geq i} p_j(t) \), from (4) we obtain the equation

\[
\frac{dz}{dt} = B(t)z(t) + f(t),
\]
Therefore
\[ ||V(t, s)||_{1D} \leq e^{-\int_0^t \alpha(u) \, du}. \tag{15} \]

**TRUNCATIONS**

Consider the “truncated” process $X_{N-1}(t)$ on the state space $E_{N-1} = \{0, 1, \ldots, N-1\}$ with the corresponding intensity matrix
\[ A_{N-1} = \begin{pmatrix} b_{00} & \mu_1 & \mu_2 & \mu_3 & \cdots & \mu_{N-1} \\ \lambda_1 & b_{11} & \mu_1 & \mu_2 & \cdots & \mu_{N-2} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \lambda_{N-1} & \lambda_{N-2} & \lambda_{N-3} & \lambda_2 & \cdots & \lambda_1 & b_{11,N-1} \end{pmatrix} \]

where $b_{ij}(t) = -\sum_{k=1}^{N-1} \mu_k(t) - \sum_{k=1}^{N-1} \lambda_{N-1-k}(t)$. Instead of (4), for $X_{N-1}(t)$ we have the following forward Kolmogorov system:
\[ \frac{dp^*}{dt} = A_{N-1}(t)p^*. \tag{17} \]

Similarly, instead of (10), we obtain the correspondent reduced system for the truncated process in the form:
\[ \frac{dz^*}{dt} = B_{N-1}(t)z^*(t) + f_{N-1}(t), \tag{18} \]

where $f_{N-1}(t) = (\lambda_1, \ldots, \lambda_{N-1}, 0, \cdots)^T$, $z^*(t) = (p_1, p_2, \cdots, p_{N-1})^T$, and
\[ B_{N-1} = (b_{ij}(t))_{i,j=1}^{N-1} = \begin{pmatrix} b_{11} - \lambda_1 & \mu_1 - \lambda_1 & \cdots & \mu_{N-1} - \lambda_1 \\ \lambda_1 - \lambda_2 & b_{22} - \lambda_2 & \cdots & \mu_{N-2} - \lambda_2 \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{N-2} - \lambda_{N-1} & \lambda_{N-3} - \lambda_{N-1} & \cdots & b_{11,N-1} - \lambda_{N-1} \end{pmatrix}. \tag{19} \]

Below we will identify the finite vector with entries, say, $(a_1, \ldots, a_{N-1})^T$ and the infinite vector with the same first $N-1$ coordinates and the others equal to zero.

For bounding of the truncation error we rewrite (18) in the following form:
\[ \frac{dz^*}{dt} = B(t)z^*(t) + f(t) - \hat{B}(t)z^*(t) - \hat{f}(t), \tag{20} \]

where $\hat{B}(t) = B(t) - B_{N-1}(t)$, and $\hat{f}(t) = f(t) - f_{N-1}(t)$.

Then we have
\[ z^*(t) = V(t)z^*(0) + \int_0^t V(t, \tau)f(\tau) \, d\tau - \int_0^t V(t, \tau)\hat{B}(\tau)z^*(\tau) \, d\tau - \int_0^t V(t, \tau)\hat{f}(\tau) \, d\tau, \tag{21} \]

Hence, if $z(0) = z^*(0)$, then the sum of first and second summands gives us $z(t)$, and we obtain in any norm
\[ ||z(t) - z^*(t)|| \leq ||\int_0^t V(t, \tau)\hat{B}(\tau)z^*(\tau) \, d\tau|| + ||\int_0^t V(t, \tau)\hat{f}(\tau) \, d\tau|| \leq \int_0^t ||V(t, \tau)|| ||\hat{B}(\tau)z^*(\tau)|| \, d\tau + \int_0^t ||V(t, \tau)|| ||\hat{f}(\tau)|| \, d\tau. \tag{22} \]

We have
\[ \hat{f}(t) = f(t) - f_{N-1}(t) = (0, \cdots, 0, \lambda_N(t), \lambda_{N+1}(t), \cdots)^T, \tag{23} \]

and
\[ ||\hat{f}(t)||_{1D} = ||D\hat{f}(t)||_1 = (d_1 + \cdots + d_N) \lambda_N(t) + (d_1 + \cdots + d_{N+1}) \lambda_{N+1}(t) + \cdots \leq \lambda_N \sum_{k=1}^N \lambda_N + \lambda_{N+1} \lambda_{N+1+1} + \cdots \rightarrow 0 \text{ as } N \rightarrow \infty, \tag{24} \]

in accordance with assumption (9).

On the other hand,
\[ B(t)z^*(t) = (B(t) - B_{N-1}(t))z^*(t) = \begin{pmatrix} a_1(t) & \cdots & a_N(t) \\ \cdots & \cdots & \cdots \end{pmatrix}z^*(t) \tag{25} \]

and
\[ ||B(t)z^*(t)||_{1D} = ||D(B(t) - B_{N-1}(t))z^*(t)||_1 = \sum_{k=1}^N \lambda_k(t) \left( \sum_{i,j=1}^{N-1} b_{ij} \right) + \sum_{k=N+1}^N \lambda_k(t) \left( \sum_{i,j=1}^{N-1} b_{ij} \right) \tag{26} \]

Let now a sequence $\{d_i\}$ be such that $1 = d_1 \leq d_2 \leq \cdots$, and the following two assumptions hold:
\[ ||V(t, s)||_{1D} \leq Me^{-\alpha(t-s)} \tag{27} \]

for any $0 \leq s \leq t$, and some positive numbers $M$ and $\alpha$, and
\[ ||V(t, s)||_{1D} \leq M^*e^{-\alpha^*(t-s)} \tag{28} \]

for any $0 \leq s \leq t$, some positive numbers $M^*$ and $\alpha^*$, where
\[ D^* = \begin{pmatrix} d_1 & d_2 & d_3 & \cdots \\ 0 & d_2 & d_2 & \cdots \\ 0 & 0 & d_3 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}. \tag{29} \]

Let, in addition, there exist a positive number $K^*$ such that
\[ d_1^2 \lambda_1 + (d_2^2 + d_3^2) \lambda_2 + \cdots \leq K^*. \tag{30} \]

Now we try to estimate $||\hat{B}(t)z^*(t)||_{1D}$. 

\[ D^* = \begin{pmatrix} D^*_{11} & 0 & 0 & \cdots \\ 0 & D^*_{22} & 0 & \cdots \\ 0 & 0 & D^*_{33} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}. \tag{29} \]
Firstly, 

$$\|z^*(t)\|_{1D^*} \leq \|V(t)\|_{1D^*} \|z^*(0)\|_{1D^*} + \int_0^t \|V(t, \tau)\|_{1D^*} \|f(\tau)\|_{1D^*} \, d\tau \leq M^* e^{-a^*t} \|z^*(0)\|_{1D^*} + \frac{K^* M^*}{a^*},$$  

(31)

because $$\|f(\tau)\|_{1D^*} \leq K^*$$ for almost all $$t \geq 0$$. 

Put $$X(0) = X_{N-1}(0) = 0$$, then $$z^*(0) = 0$$, hence 

$$\|z^*(t)\|_{1D^*} \leq \frac{K^* M^*}{a^*},$$  

(32)

for any $$t \geq 0$$.

Suppose for definiteness that $$N$$ is odd. All $$p^*_i(t) \geq 0$$, then 

$$\|z^*(t)\|_{1D^*} = \sum_{i \geq 1} p^*_i(t) \sum_{k=1}^i d_k^2 \geq \sum_{i \geq \frac{N-1}{2}} \sum_{k=1}^{N-1} d_k^2 p^*_i(t),$$  

(33)

On the other hand we have the following bound:

$$\begin{align*}
(d_1 + d_2) \sum_{k \geq N-1} \lambda_k(t) p^*_i(t) + \\
(d_3 + \cdots + d_{N-1}) \sum_{k \geq 2} \lambda_k(t) p^*_i(t) \leq \\
(d_1 + \cdots + d_{N-1}) \sum_{k \geq 2} \lambda_k(t) \sum_{i=1}^{N-1} p^*_i(t) + \\
\sum_{k \geq 1} \lambda_k(t) \left( (d_1 + \cdots + d_{N-1}) p^*_i(t) + \cdots + (d_1 + \cdots + d_{N-1}) p^*_i(t) \right).
\end{align*}$$  

(34)

Denote by $$\Lambda_K = \sum_{k \geq K} \lambda_k$$.

Then we obtain from (26), (33) and (34):

$$\begin{align*}
\|z^*(t)\|_{1D^*} \leq \frac{g_{N-1} \lambda_{N-1}}{d_{N-1}} \sum_{i=1}^{N-1} p^*_i(t) + \\
L_\lambda \left( g_{N-1} p^*_i(t) + \cdots + g_{N-1} p^*_i(t) \right) \leq \\
g_{N-1} \lambda_{N-1} + L_\lambda \left( g_{N-1} p^*_i(t) + \cdots + g_{N-1} p^*_i(t) \right) \leq \\
\cdots + \sum_{i=1}^{N-1} d_i^2 p^*_i(t) \leq \frac{g_{N-1} \lambda_{N-1}}{d_{N-1}} + L_\lambda \frac{g_{N-1} \lambda_{N-1}}{d_{N-1}} \frac{K^* M^*}{a^*},
\end{align*}$$  

(35)

for any $$t \geq 0$$.

Finally, we have from (22), (24), and (35) the following bound of truncation error:

$$\begin{align*}
\|z(t) - z^*(t)\| \leq \int_0^t M e^{-a^*(t-\tau)} \left( g_{N-1} \lambda_{N-1} + L_\lambda \frac{g_{N-1} \lambda_{N-1}}{d_{N-1}} \frac{K^* M^*}{a^*} \right) \, d\tau + \\
\int_0^t M e^{-a^*(t-\tau)} (g_{N-1} \lambda_{N-1} + g_{N-1} \lambda_{N-1} + \cdots) \, d\tau \leq \\
\frac{M}{a^*} \left( g_{N-1} \lambda_{N-1} + L_\lambda \frac{g_{N-1} \lambda_{N-1}}{d_{N-1}} \frac{K^* M^*}{a^*} + \\
g_{N-1} \lambda_{N-1} + g_{N-1} \lambda_{N-1} + \cdots \right).
\end{align*}$$  

(36)

Now let $$l_{1E}$$ be the space of sequences,

$$l_{1E} = \left\{ z = (p_1, p_2, \cdots) \mid \|z\|_{1E} \equiv \sum \sum |p_n| < \infty \right\}.$$

Then (36) and well-known inequality $$\|z\|_{1E} \leq W^{-1} \|z\|_{1D}$$ (see, for instance, Zeifman et al 2006), imply the following statement.

**Theorem 1.** Let (9), (27), (28), (30) be fulfilled. Then $$X(t)$$ is exponentially weakly ergodic, has the limiting mean, say, $$E(t, 0)$$, and the following bound of truncation error holds:

$$\begin{align*}
\frac{M}{a W} \left( g_{N-1} \lambda_{N-1} + L_\lambda \frac{g_{N-1} \lambda_{N-1}}{d_{N-1}} \frac{K^* M^*}{a^*} + \\
g_{N-1} \lambda_{N-1} + g_{N-1} \lambda_{N-1} + \cdots \right) \leq \frac{|E(t, 0) - E_{N-1}(t, 0)|}{E_{N-1}(0) = k}
\end{align*}$$  

(37)

for any $$t \geq 0$$, where $$E_{N-1}(t, k) = E\{X_{N-1}(t) | X_{N-1}(0) = k\}$$ is the mean (the mathematical expectation) of the truncated process at the moment $$t$$ under initial condition $$X_{N-1}(0) = k$$.

**EXAMPLES**

1. Consider the simplest analogue of $$M_t/M_t/S$$ queue for a queueing system with group services, see Section 4 in (Zeifman et al. 2014a).

Namely, we suppose that the arrival intensity of a customer to the queue is $$\lambda_i(t) = \lambda(t) = 1 + \sin 2\pi t$$, $$\lambda_i(t) = 0$$ for any $$i > 1$$ and any $$t \geq 0$$. Let the intensity of departure (servicing) of a group of $$k$$ customers be $$\mu_k(t) = \frac{\mu(t)}{k}$$ if $$k \leq S < \infty$$, and $$\mu_k(t) = 0$$ for any $$i > S$$ and any $$t \geq 0$$, where $$\mu(t) = 3 + \cos 2\pi t$$.

Let $$X = X(t), t \geq 0$$ be a queue-length process for the queue. For definiteness, put $$S = 10^{12}$$. For this queue bound (37) looks essentially simpler:

$$\begin{align*}
|E(t, 0) - E_{N-1}(t, 0)| \leq \frac{\lambda M K^* M^*}{a a^* W} \frac{g_{N-1}}{d_{N-1}}.
\end{align*}$$  

(38)
Here $L_\lambda = 2$.
Put $d = \sqrt{2}$, and $d_{k+1} = d^k$. Then $K = K^* = 2$, $W = \sqrt{2}$. Now, in (8) we have
\[
\alpha(t) \geq \mu(t) - (d-1)\lambda(t), \quad \alpha^*(t) \geq \mu(t) - (d^2-1)\lambda(t), \quad (39)
\]
therefore we can apply bound (15) and obtain $a \geq 2$, $a^* \geq 2$, $M \leq 2$, $M^* \leq 2$.
Finally, we have $g_{N-1} \leq 2^N$, $d_{\frac{N-1}{2}}^2 = 2^{N-2}$, and the following estimate for the error of truncation:
\[
|E(t, 0) - E_{N-1}(t, 0)| \leq 16 \cdot 2^{-\frac{N}{2}}, \quad (40)
\]
which does not depend on $t$ in contrast to the estimate (50) in (Zeifman et al. 2014a).
Applying the approach of (Zeifman et al. 2006) we can find the approximation of the mathematical expectation of the length of queue, see the following figures.

Figure 1: First example, approximation of the mean $E(t, 0)$ on $[0, 10]$ with an error $10^{-3}$.

Figure 2: First example, approximation of the mean $E(t, 0)$ on $[9, 10]$ with an error $10^{-3}$.

2. Consider a simplest queuing system with one server and batch arrivals, see Section 5 in (Zeifman et al. 2014a). Namely, let $\mu_i(t) = \mu(t) = 3 + \cos 2\pi t$ be the service rate of a customer, $\mu_i(t) = 0$ for any $i > 1$ and any $t \geq 0$. Let $\lambda_k(t) = \lambda(t)$, where $\lambda(t) = 1 + \sin 2\pi t$ be the arrival intensity of $k$ customers to the queue. We have $L_\lambda < 1$. Put also $d = \sqrt{2}$, and $d_{k+1} = d^k$. Then $K \leq 2$, $K^* = 2$, $W = \sqrt{2}$. Now, in (8) we have
\[
\alpha(t) \geq \mu(t) - (d-1)\lambda(t), \quad \alpha^*(t) \geq \mu(t) - (d^2-1)\lambda(t), \quad (41)
\]
therefore we can apply bound (15) and obtain $a \geq \frac{1}{2}$, $a^* \geq \frac{1}{2}$, $M \leq 2$, $M^* \leq 2$, $A_K \leq 4^{-K}$. We have also $g_{N-1} \leq 2^N$, $d_{\frac{N-1}{2}}^2 = 2^{N-2}$ For this queue we have instead of (37) the following estimate for the error of truncation:
\[
|E(t, 0) - E_{N-1}(t, 0)| \leq 10^2 \cdot 2^{-\frac{N}{2}}, \quad (42)
\]
which does not depend on $t$ in contrast to the estimate (52) in (Zeifman et al. 2014a).
Now, applying the approach of (Zeifman et al. 2006) we can find the approximation of the mathematical expectation of the length of queue, see the following figures.

Figure 3: Second example, approximation of the mean $E(t, 0)$ on $[0, 10]$ with an error $10^{-3}$.

Figure 4: Second example, approximation of the mean $E(t, 0)$ on $[9, 10]$ with an error $10^{-3}$.

Acknowledgement. This work was supported by the Russian Foundation for Basic Research, projects no. 14-07-00041, 13-07-00223, 12-07-00115, 12-07-00109.
REFERENCES


AUTHOR BIOGRAPHIES

ALEXANDER ZEIFMAN Doctor of Science in physics and mathematics; professor, Head of Department of Applied Mathematics, Vologda State University; senior scientist, Institute of Informatics Problems, Russian Academy of Sciences; principal scientist, Institute of Territories Socio-Economic Development, Russian Academy of Sciences. His email is a_zeifman@mail.ru and his personal webpage at http://uni-vologda.ac.ru/ zai/eng.html.

YAKOV SATIN is Candidate of Science (PhD) in physics and mathematics, associate professor, Vologda State University. His email is yacovi@mail.ru.

GALINA SHILOVA is Candidate of Science (PhD) in physics and mathematics, associate professor, Vologda State University. Her email is shgn@mail.ru.

VICTOR KOROLEV is Doctor of Science in physics and mathematics, professor, Head of Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, M.V. Lomonosov Moscow State University; leading scientist, Institute of Informatics Problems, Russian Academy of Sciences. His email is victoryukorolev@yandex.ru.

VLADIMIR BENING is Doctor of Science in physics and mathematics; professor, Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University; senior scientist, Institute of Informatics Problems, Russian Academy of Sciences. His email is bening@yandex.ru.

SERGEY YA. SHORGIN is Doctor of Science in physics and mathematics, professor, Deputy Director, Institute of Informatics Problems, Russian Academy of Sciences. His email is sshorgin@ipiran.ru.