A class of Markov processes on a semi-infinite strip

Ivo Adan and Jacques Resing

Department of Mathematics and Computing Science
Eindhoven University of Technology
P.O. Box 513, 5600 MB Eindhoven, The Netherlands

Abstract. In this paper we determine the equilibrium distribution for a general class of Markov processes on a semi-infinite strip. We expose a method to express the equilibrium probabilities of the Markov process as a finite sum of terms, which are geometric in the unbounded variable. The geometric factors are the roots inside the unit circle of a determinantal equation. By using a generating-function technique we are able to determine these roots very efficiently. Because of this, the expression for the equilibrium probabilities becomes numerically very attractive.

1 Introduction

We consider a class of irreducible Markov processes, the state space of which can be partioned into two parts: a finite set V plus a semi-infinite strip of states (i, j) where i ranges from 0 to s and j from 0 to ∞. In the sequel we call the states in V and the states in the strip with $j = 0$ the boundary states. For the states in V we require that no transitions to states (i, j) with $j > 0$ are possible, i.e., the only way to enter the strip from the set V is via the boundary states of the strip. For the states in the strip we impose more structure on the state transitions. More specifically, in state (i, j) the following transitions are possible $(k = -j+1, -j+2, \ldots, 0, 1)$:

- From (i, j) to $(i+1, j+k)$ with rate $a_k(s-i)$.
- From (i, j) to $(i, j+k)$ with rate $b_k(s-i) + c_k i$.
- From (i, j) to $(i-1, j+k)$ with rate $d_k i$.
- From (i, j) to boundary states with total rate $\sum_{k=-\infty}^{-j} [a_k(s-i) + b_k(s-i) + c_k i + d_k i]$.

So transitions from (i, j) are only possible to states (m, n) with $m = i-1, i, i+1$ and $n \leq j+1$. Further, the rates a_k and b_k are scaled by a factor $(s-i)$, and the rates c_k and d_k are scaled by a factor i. The transition rate diagram is shown in Figure 1. Note that we allow in our Markov process large jumps to the left, but no large jumps to the right. An example of a queueing model that fits into this framework is the $M/E_2/s$ queue. The service times in this model consist of two exponential phases. The set V contains those states in which at least one server is idle, while the strip contains the states in which all servers are busy. In this example, i denotes the number of servers working on a second service phase and j denotes the number of jobs waiting in the queue. In Section 2 we present several other queueing models, which fit into this class of Markov processes.

Provided the Markov process is ergodic and denoting by $p(i, j)$ the equilibrium probability of being in state (i, j), we will show that the vectors $p_j = (p(0, j), p(1, j), \ldots, p(s, j))$ can be expressed as a linear combination of $s+1$ geometric terms, i.e.,

$$p_j = \sum_{m=0}^{s} C_m y_m x_m^j, \quad j = 0, 1, \ldots, \tag{1}$$

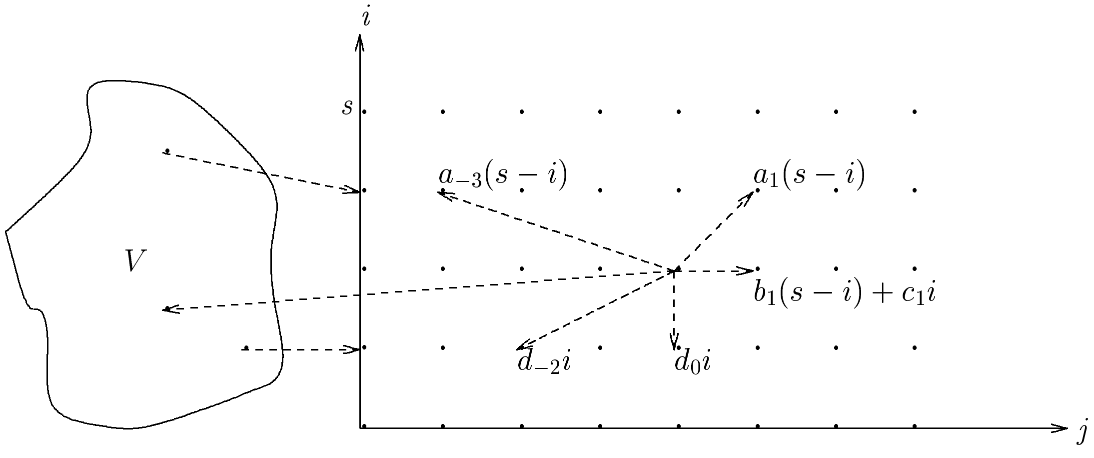

Fig. 1. Transition rate diagram

where C_m is a scalar and y_m is the row vector $(y_m(0), y_m(1), \ldots, y_m(s))$. The geometric factors x_m are the roots inside the unit circle of a determinantal equation. By using a generating-function technique we can reduce the single equation for these $s+1$ roots inside the unit circle to $s+1$ equations for a single root in the interval $(-1, 1)$. This considerably simplifies the determination of the roots, also because the computations can now be restricted to the real domain.

In queueing models, expression (1) can be exploited to obtain similar expressions for important performance measures, such as the moments of the number of jobs in the system and the distribution of the waiting time (see e.g. [4, 1]).

Result (1) may be linked to the (modified) matrix-geometric representation of the equilibrium distribution (see [16]). It is readily seen that it has the form

$$p_j = p_0 R^j, \quad j = 0, 1, \ldots, \tag{2}$$

where the so-called rate matrix R is the minimal nonnegative solution of a nonlinear matrix equation. It is well-known that the determination of the matrix R may cause numerical problems, cf. [7]. Clearly, when R is diagonalizable, the factors x_m in the form (1) are the eigenvalues of R and the row vectors y_m are the associated row eigenvectors.

An approach, closely related to the one in this paper, is the spectral expansion method in [14]. It is based on reducing the equilibrium equations to a vector difference equation with constant coefficients, the solution of which can be expressed in terms of eigenvalues and eigenvectors of the associated characteristic polynomial. The difficult step in this approach is the computation of the eigenvalues. In special cases, particular properties may be exploited to simplify the eigenvalue problem, see e.g. [10]. In fact, the main contribution of this paper is that for a general class of Markov processes it is shown that the relevant eigenvalues can indeed be determined very efficiently.

The rest of this paper is organized as follows. In the next section we present a number of queueing models from the literature, which fit into the class of Markov processes described above. Next, we describe in Section 3 some conditions imposed on the rates of the Markov process and we derive the ergodicity condition of the process. In Section 4, we will start the study of the equilibrium distribution. The generating-function approach which is used to determine the roots of the determinantal equation is presented in Section 5. After that, the main result of the paper is formulated in Section 6 and an example of a Markov process with larger jumps to the right is presented in Section 7. Finally, in Section 8 we give some conclusions and final remarks.

2 Examples

The following queueing models known from the literature fit into the class of Markov processes that we study in this paper.

Example 1. The $M/E_2/s$ queue (see [17]).
Jobs arrive according to a Poisson process with rate λ and have service times consisting of two exponential phases with parameter μ. As already mentioned in the introduction, the set V consists of all states in which at least one server is idle and the strip contains all states in which all servers are busy. In this example, i represents the number of servers working on a second service phase and j represents the number of jobs waiting in the queue. We have $a_0 = \mu$, $b_1 = c_1 = \lambda/s$, $d_{-1} = \mu$ and all other coefficients are equal to zero.

Example 2. The $M/C_2/s$ queue (see [6]).
Like the previous example, but now the service times of jobs have a Coxian distribution with two phases. Denoting the parameter of the i-th exponential phase by μ_i, $i = 1, 2$, and the probability of bypassing the second phase by $1 - p$, we have $a_0 = \mu_1 p$, $b_1 = c_1 = \lambda/s$, $b_{-1} = \mu_1(1 - p)$, $d_{-1} = \mu_2$ and all other coefficients are equal to zero.

Example 3. The $M/H_2/s$ queue (see [8, 9]).
Like Example 1, but now the service times of jobs have a hyperexponential distribution of order 2. Denoting the parameters of the two exponentials by μ_1 and μ_2, respectively, the branching probabilities by p_1 and p_2 (with of course $p_1 + p_2 = 1$) and letting i be the number of servers working on an exponential service with parameter μ_1, we have $a_{-1} = \mu_2 p_1$, $b_1 = c_1 = \lambda/s$, $b_{-1} = \mu_2 p_2$, $c_{-1} = \mu_1 p_1$, $d_{-1} = \mu_1 p_2$ and all other coefficients are equal to zero.

Example 4. The $M/M/s$ queue with service interruptions (see [15]).
Jobs arrive according to a Poisson process with rate λ. Each server, when operative, serves jobs with rate μ. However, the servers are subject to breakdowns. The times that servers are operative are exponentially distributed with parameter α. The repair times are exponentially distributed with parameter β. In this case, i represents the number of servers that are operative. We have $a_0 = \beta$, $c_{-1} = \mu$, $b_1 = c_1 = \lambda/s$, $d_0 = \alpha$ and all other coefficients are equal to zero.

Example 5. The $\sum IPP/M/1$ queue (see [10]).
The arrival process is a superposition of s independent, identical, interrupted Poisson processes with exponentially distributed on-times with parameter α, exponentially distributed

off-times with parameter β and arrival rate during on-times equal to λ. Service times of customers are exponentially distributed with parameter μ. In this example the set V is empty, i denotes the number of Poisson processes in the on-state and j denotes the number of customers in the system. We have $a_0 = \beta$, $b_{-1} = c_{-1} = \mu/s$, $c_1 = \lambda$, $d_0 = \alpha$ and all other coefficients are equal to zero.

Example 6. The multi-server queue with locking (see [1]).
Customers arrive according to a Poisson process with rate λ. Service times are exponentially distributed with parameter μ. There are s groups of servers, each group consisting of two servers (see Figure 2). Within a group, one of the servers is called the front server and the

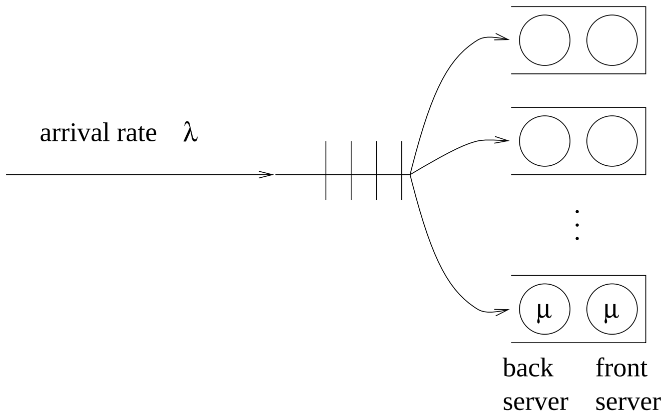

Fig. 2. The multi-server queue with locking

other one is called the back server. Customers are served by front servers as long as these are available. If all front servers are occupied, new customers are served by back servers. If the service of a customer at the front server has been completed and there is another customer in service at the back server, the customer can not leave the system until also the service of the customer at the back server has been completed. During this period the front server is blocked and can not serve a new customer. The model is motivated by a situation encountered at a maintenance facility for trains.

The set V consists of the states in which not all server positions are occupied. The strip contains the states in which all server positions are occupied. In this example, i denotes the number of customers that already completed their service but that are locked in by a customer at the back server and j denotes the number of customers waiting in the queue. We have $a_0 = \mu$, $b_1 = c_1 = \lambda/s$, $b_{-1} = \mu$, $d_{-2} = \mu$ and all other coefficients are equal to zero.

3 Conditions

In this section we will impose some conditions on the rates of the Markov process and we will determine the ergodicity condition of the process. First, let us introduce the functions

$$A(x) := \sum_{k=-\infty}^{1} a_k x^{1-k}, \quad B(x) := \sum_{k=-\infty}^{1} b_k x^{1-k},$$
$$C(x) := \sum_{k=-\infty}^{1} c_k x^{1-k}, \quad D(x) := \sum_{k=-\infty}^{1} d_k x^{1-k}.$$

The conditions that we impose on the rates of the Markov process are the following:

Condition 1

(*i*) *$A(1)$, $B(1)$, $C(1)$ and $D(1)$ are finite.*
(*ii*) *$A(1) > 0$ and $D(1) > 0$.*
(*iii*) *$A'(1)$, $B'(1)$, $C'(1)$ and $D'(1)$ are finite.*

The conditions (*i*) and (*iii*) are obvious, condition (*ii*) is imposed to exclude exceptional cases. The relaxation of condition (*ii*) will be discussed in the final section. Next, we will determine the ergodicity condition of the process. Because the process is a Markov process of the $GI/M/1$ type, we can use Neuts' mean drift condition, i.e. formula (1.7.11) in [16], to obtain the ergodicity condition. Here, the mean drift condition reads

$$\sum_{k=-\infty}^{1} k \sum_{i=0}^{s} \pi_i[(s-i)(a_k + b_k) + i(c_k + d_k)] < 0 \tag{3}$$

where π is the limiting distribution of the Markov process, the transition rate diagram of which is shown in Figure 3.

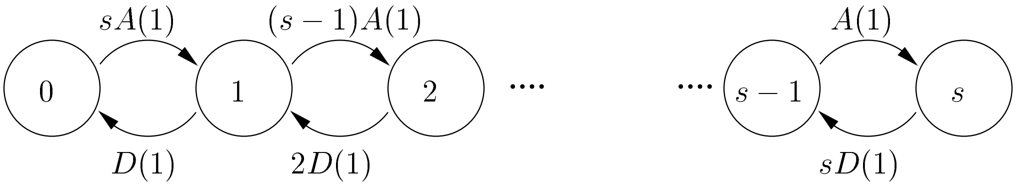

Fig. 3. Transition rate diagram

Clearly, π is equal to a binomial distribution with parameters s and $A(1)/(A(1)+D(1))$. Using this fact, we obtain, after a bit of rewriting, the following result.

Lemma 1. *The Markov process is ergodic if and only if*

$$D(1)(A'(1) - A(1) + B'(1) - B(1)) + A(1)(C'(1) - C(1) + D'(1) - D(1)) > 0. \tag{4}$$

From now on we assume that condition (4) holds.

4 Equilibrium distribution

In this section we will concentrate on the equilibrium distribution of the Markov process. By equating in each state (i,j) with $j>0$ the flow out of and into that state, we obtain the following balance equations for the equilibrium probabilities $p(i,j)$,

$$\begin{aligned}\sum_{k=-\infty}^{1}\Big((a_k+b_k)(s-i)+(c_k+d_k)i\Big)p(i,j) \;=\; &\sum_{k=-\infty}^{1}\Big(d_k(i+1)p(i+1,j-k)\\ &+(b_k(s-i)+c_k i)p(i,j-k)+a_k(s-i+1)p(i-1,j-k)\Big),\\ &i=0,\dots,s, \qquad j=1,2,\dots, \end{aligned} \tag{5}$$

where by definition $p(-1,j)=p(s+1,j)=0$. The balance equations for the states $(i,0)$ and the states in V can be obtained similarly. We will refer to these latter equations as the *boundary equations*. The precise form of the boundary equations is not relevant to the analysis, and therefore, they are omitted. Note that the number of boundary equations is equal to $|V|+s+1$.

We are interested in solutions of (5) for which $\sum\sum p(i,j)$ absolutely converges (so they can be normalized). Since the space of absolutely convergent solutions is linear, we may linearly combine these solutions in order to satisfy the boundary equations and the normalization equation. Because the balance equations are dependent, one of the boundary equations may be omitted and hence the number of independent equations that have to be satisfied is equal to $|V|+s+1$. The unknowns are the coefficients in the linear combination and the equilibrium probabilities $p(v)$ with $v\in V$. The number of unknowns should be equal to the number of equations to be satisfied, which implies that the number of independent solutions needed is equal to $s+1$. The following lemma states that if there indeed exist $s+1$ independent solutions, then the equilibrium distribution can be expressed in terms of these solutions. The proof is based on standard arguments (cf. [2, 5, 14]).

Lemma 2. *If there are $s+1$ independent, absolutely convergent solutions of (5), say $p_0(i,j),\dots,p_s(i,j)$, then the equilibrium probabilities $p(i,j)$ can be expressed as*

$$p(i,j)=\sum_{m=0}^{s}C_m p_m(i,j), \qquad i=0,\dots,s, \quad j=0,1,\dots,$$

where the coefficients C_m are uniquely determined from the boundary equations and the normalization equation.

From Lemma 2 and the uniqueness of the equilibrium distribution we immediately obtain the following.

Corollary 1. *The dimension of the space of absolutely convergent solutions of (5) is at most $s+1$.*

Lemma 2 reduces the determination of the equilibrium distribution to that of finding $s+1$ independent solutions of (5). To find these, we try solutions of the form

$$p(i,j)=y(i)x^j, \tag{6}$$

with $|x| < 1$, which is necessary for absolute convergence. Substitution of the form (6) into (5) and dividing by the common power x^{j-1} yields, for $i = 0, \cdots, s$,

$$0 = (i+1)D(x)y(i+1) + ((s-i)B(x) + iC(x))y(i) - \\ ((s-i)(A(1)+B(1)) + i(C(1)+D(1)))xy(i) + (s-i+1)A(x)y(i-1), \qquad (7)$$

where by convention $y(-1) = y(s+1) = 0$. For given x, this is a system of linear homogeneous equations for $y(0), \ldots, y(s)$. Now we have to find the values of x for which this system has a non-null solution. It has a non-null solution if and only if the determinant of this system is equal to zero. Hence, the desired values of x are the zeros inside the unit circle of this determinant. If it has $s+1$ different zeros, then the corresponding solutions (6) are independent, and thus form a basis, in terms of which we can express the equilibrium distribution. However, it is difficult to prove directly that there are $s+1$ zeros inside the unit circle and to numerically determine these zeros. Therefore we employ an idea of [17, 6, 1] to transform the difference equations (7) into a single differential equation for the generating function of the sequence $y(i)$. This approach, in fact, reduces the determinantal equation for $s+1$ roots inside the unit circle to $s+1$ equations for a single root in the interval $(-1, 1)$.

Remark 1. (Dimensionality).
In fact, the dimension of the space of absolutely convergent solutions of (5) is exactly $s+1$. This follows, e.g., from the matrix-geometric representation (2). It is easily seen that when R is diagonalizable, with eigenvalues x_m and associated row eigenvectors y_m, $m = 0, \ldots, s$, the sequences $y_0 x_0^j, \ldots, y_s x_s^j$, $j = 0, 1, \ldots$, form a basis for the solution space. In general, $s+1$ independent solutions of (5) can be constructed from the Jordan basis of the rate matrix R (cf. [11, 12]).

5 Generating-function approach

So, we transform the difference equations (7) into a differential equation for the generating function

$$Y(z) = \sum_{i=0}^{s} y(i) z^i$$

of the sequence $y(i)$. Multiplying (7) by z^i and adding with respect to i we obtain

$$D(x)Y'(z) + sB(x)Y(z) + (C(x) - B(x))zY'(z) - (A(1) + B(1))sxY(z) \\ +(A(1) + B(1) - C(1) - D(1))xzY'(z) + sA(x)zY(z) - A(x)z^2Y'(z) = 0,$$

which may be rewritten in the form

$$\frac{Y'(z)}{Y(z)} = \frac{s[A(x)z + B(x) - (A(1) + B(1))x]}{A(x)z^2 - ((A(1) + B(1) - C(1) - D(1))x + C(x) - B(x))z - D(x)} \\ = \frac{E(x)}{z - z_1(x)} + \frac{s - E(x)}{z - z_2(x)}, \qquad (8)$$

where $z_1(x)$ and $z_2(x)$ are the roots ($z_1(x)$: + sign, $z_2(x)$: − sign) of the denominator of (8) and $E(x)$ satisfies the equation

$$2E(x) - s = s \cdot \frac{B(x) + C(x) - x\,(A(1) + B(1) + C(1) + D(1))}{\sqrt{F(x)^2 + 4A(x)D(x)}}, \tag{9}$$

where

$$F(x) = (A(1) + B(1) - C(1) - D(1))x + C(x) - B(x).$$

Note that we have divided numerator and denominator of (8) by $A(x)$, and hence, we implicitly assume here that $A(x) \neq 0$. We come back to this issue later on in Section 6. The general solution of the differential equation (8) is

$$Y(z) = K(z - z_1(x))^{E(x)}(z - z_2(x))^{s-E(x)}, \tag{10}$$

with K a constant. Now the key idea to proceed is that, since $Y(z)$ is a polynomial in z, the exponents $E(x)$ and $s - E(x)$ should be equal to a non-negative integer, i.e. $E(x) = m, m = 0, \ldots, s$. Hence, for each m we get an equation in x, for which we will prove that, under the ergodicity condition (4), there is at least one solution, x_m say, in the interval $(-1, 1)$ (uniqueness will be established later). To prove this, we distinguish between three cases:

(a) $(A(0) = 0$ or $D(0) = 0)$ and $(C(0) = B(0) \neq 0)$,
(b) $(A(0) = 0$ or $D(0) = 0)$ and $(C(0) = B(0) = 0)$,
(c) $(A(0) \neq 0$ and $D(0) \neq 0)$ or $(C(0) \neq B(0))$.

Lemma 3. *In case (a) we have for all $m = 0, \ldots, s$ that the equation*

$$\frac{2m - s}{s} = \frac{B(x) + C(x) - x\,(A(1) + B(1) + C(1) + D(1))}{\sqrt{F(x)^2 + 4A(x)D(x)}}, \tag{11}$$

has a solution $x = x_m$ in the interval $(0, 1)$.

Proof: Let $G(x)$ denote the function on the right-hand side of (11). Then, it is straightforward to check that $G(1) = -1$, $G'(1) > 0$ (which holds if and only if ergodicity condition (4) is satisfied) and $\lim_{x \downarrow 0} G(x) = +\infty$, which follows from the conditions of case (a). The lemma now follows from a continuity argument. □

All examples of Section 2, with exception of Example 5, satisfy the conditions of case (a). To illustrate the result of Lemma 3 we look in more detail at Example 6.

Example 7. (Continuation of Example 6)
In this example we have $A(x) = \mu x$, $B(x) = \lambda/s + \mu x^2$, $C(x) = \lambda/s$, and $D(x) = \mu x^3$. The function $G(x)$ and the roots x_m are depicted in Figure 4 for the case $s = 4$, $\lambda = 2$ and $\mu = 1$. It also shows that the roots in $(0, 1)$ are unique and ordered (see Theorem 2 and Remark 2).

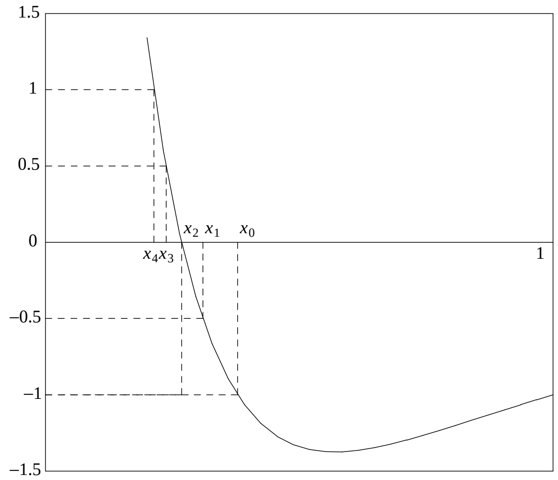

Fig. 4. The function $G(x)$ in Example 6 for $s = 4$, $\lambda = 2$ and $\mu = 1$

Lemma 4. *In case (b) we have for all* $m = 0, \dots, \lceil s/2 \rceil - 1$ *that equation (11) has a solution* $x = x_m$ *in the interval* $(0, 1)$.

Proof: Again, with $G(x)$ the right-hand side of (11), we have that $G(1) = -1$ and $G'(1) > 0$. The conditions of case (b) imply that $\lim_{x \downarrow 0} G(x) = 0$. To see this, assume $A(0) = 0$ (the proof is similar when $D(0) = 0$). Since also $B(0) = C(0) = 0$, irreducibility of the state space implies that $a_1 > 0$ (otherwise it is impossible to move in the North-East direction on the strip). Hence, it follows that, for x close to zero, the numerator is asymptotically linear in x, whereas the denominator behaves asymptotically like $\sqrt{x}$. Now the lemma follows again from a continuity argument. □

Example 8.
Consider the model with $A(x) = 2x$, $B(x) = C(x) = 2x^2$ and $D(x) = 1 + x^2$. The transition rate diagram in case $s = 4$ is illustrated in Figure 5. In this case $G(x)$ is given by

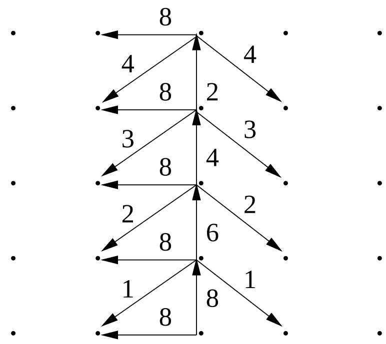

Fig. 5. The transition rate diagram in Example 8

$(4x^2 - 8x)/\sqrt{8x(1+x^2)}$, and it is shown in Figure 6. Note that for $x < 0$ the square root, and thus $G(x)$, is imaginary, so we do not find solutions of equation (11) for $m \geq \lceil s/2 \rceil$.

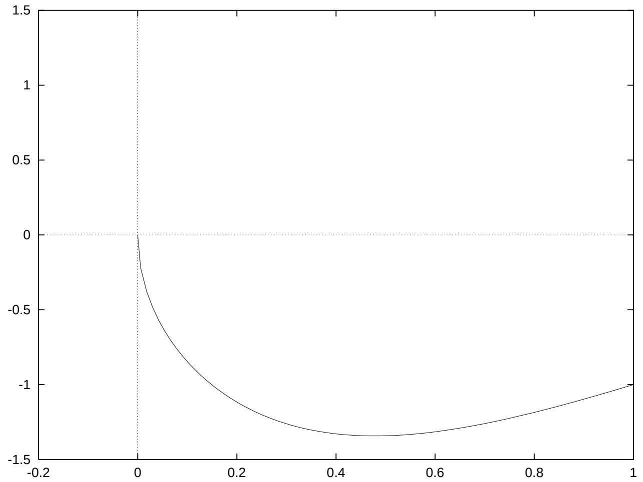

Fig. 6. The function $G(x)$ in Example 8

Lemma 5. *In case (c) we have for all $m = 0, \ldots, s$ that equation (11) has a solution $x = x_m$ in the interval $(-1, 1)$.*

Proof: Again, it holds that $G(1) = -1$ and $G'(1) > 0$. Under the conditions of case (c) we can prove that either there is an $\hat{x} \in [-1, 0)$ such that $\lim_{x \downarrow \hat{x}} G(x) = +\infty$ or that $G(-1) \geq 1$. To see this we evaluate $F(x)^2 + 4A(x)D(x)$ at $x = -1$,

$$\begin{aligned} &F(-1)^2 + 4A(-1)D(-1) \\ &\quad = \Big((D(1) - A(1)) - (B(1) + B(-1)) + (C(1) + C(-1))\Big)^2 + 4A(-1)D(-1) \\ &\quad \leq \Big(|D(1) - A(1)| + (B(1) + B(-1)) + (C(1) + C(-1))\Big)^2 + 4A(1)D(1) \\ &\quad \leq \Big(D(1) + A(1) + B(1) + B(-1) + C(1) + C(-1)\Big)^2. \end{aligned} \tag{12}$$

Hence, either $F(x)^2 + 4A(x)D(x)$ is positive for all $x \in [-1, 0]$, and in that case equation (12) implies that $G(-1) \geq 1$, or there is an $\hat{x} \in [-1, 0)$ such that $F(x)^2 + 4A(x)D(x) > 0$ for $x > \hat{x}$ and $F(\hat{x})^2 + 4A(\hat{x})D(\hat{x}) = 0$, and in that case $\lim_{x \downarrow \hat{x}} G(x) = +\infty$. Now a continuity argument completes the proof of the lemma. There is only a complication when $G(-1) = 1$, since in this case it may be possible that $x = -1$ is the only root of $G(x) = 1$. However, it is readily seen that $G(-1) = 1$ only if both $B(\cdot)$ and $C(\cdot)$ are odd functions and both $A(\cdot)$ and $D(\cdot)$ are even functions. This implies that also $G(\cdot)$ is an odd function and so we have that $G'(-1) > 0$. Hence, $G(x) = 1$ also has a solution in $(-1, 1)$. $\square$

To illustrate the result of Lemma 5 we look at three examples. The first example is the example of Section 2 which satisfies the conditions of case (c). In the second example negative roots appear, and in the third example the complication $G(-1) = 1$ arises.

Example 9. (Continuation of Example 5)
Consider the situation with $\alpha = \beta = \lambda = 1, \mu = 3$ and $s = 4$. In this case we have $A(x) = D(x) = x$, $B(x) = 3x^2/4$ and $C(x) = 1 + 3x^2/4$, which yields that $G(x) = (3x^2 - 9x + 2)/(2\sqrt{5x^2 - 2x + 1})$ (see Figure 7).

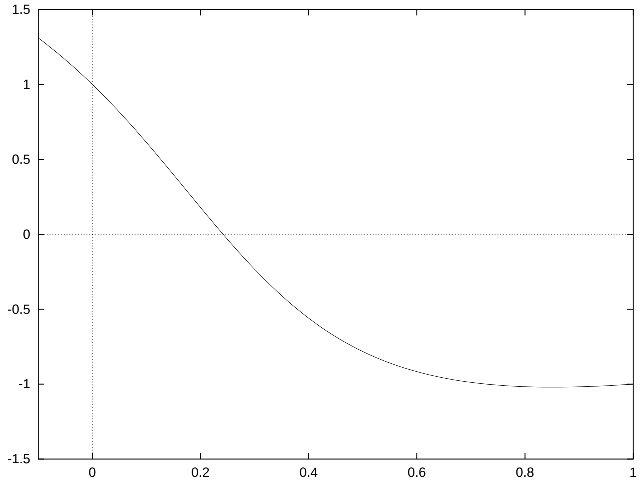

Fig. 7. The function $G(x)$ in Example 5 for $\alpha = \beta = \lambda = 1, \mu = 3$ and $s = 4$

Example 10.
Consider the model with $A(x) = D(x) = 1$ and $B(x) = C(x) = 3x^2/2$. The function $G(x)$ is in this case given by $(3x^2 - 5x)/2$. For $s = 2$ the solutions x_s are given by $x_0 = 2/3$, $x_1 = 0$ and $x_2 = -1/3$. The transition rate diagram is shown in Figure 8.

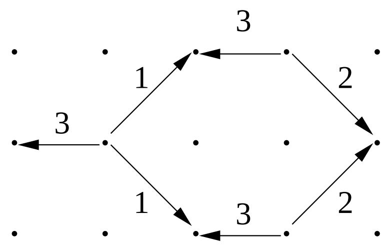

Fig. 8. The transition rate diagram in Example 10

Example 11.
Consider the model with $A(x) = D(x) = 1 + x^2$ and $B(x) = C(x) = x^3$. The function $G(x)$ is in this case given by $(x^3 - 3x)/(1 + x^2)$ (see Figure 9). We see that, besides $x = -1$, also $x = 1 - \sqrt{2}$ is a solution of $G(x) = 1$.

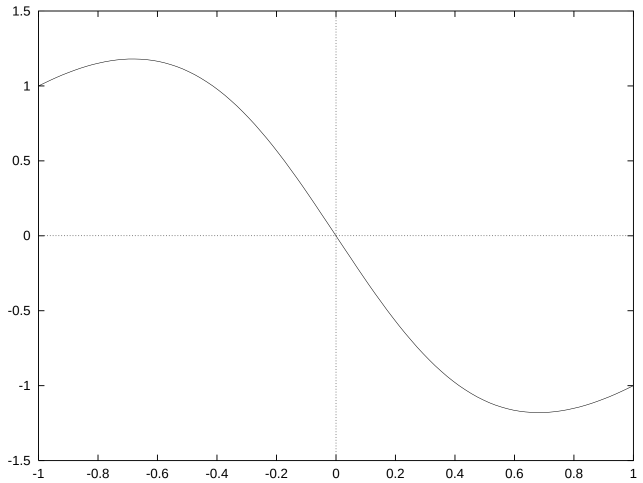

Fig. 9. The function $G(x)$ in Example 11

6 Main result

The three lemmas in the previous section provide the desired independent solutions of (5). In case (a) we have $s + 1$ solutions of the form

$$p(i, j) = y_m(i)x_m^j, \tag{13}$$

where x_m is a root of (11) in $(0, 1)$ and the sequence $y_m(i)$ is a non-null solution of (7) with $x = x_m$. Note that in the derivation of the generating function (10) we assumed that $A(x) \neq 0$. This indeed holds for $x = x_m$, since $x_m > 0$. The solutions (13) are independent, since the roots x_m are different.

In case (c) we also find $s+1$ independent solutions of the form (13). The only difference is that x_m may be zero or negative, in which case it has to be verified that $A(x_m) \neq 0$. If $A(x) = 0$ for some $x \in (-1, 0)$, then

$$G(x) = \frac{(B(x) - xB(1)) + (C(x) - xC(1)) - xA(1) - xD(1)}{|B(x) - xB(1) - (C(x) - xC(1)) - xA(1) + xD(1)|} > 1,$$

so x can not be a solution of equation (11). If $A(0) = 0$, then again $G(0) > 1$, except when $B(0) = 0$ or $C(0) = 0$, in which case $G(0) = 1$. This suggests that $x_s = 0$ produces a useful solution, and indeed, it is easily verified that in case $A(0) = B(0) = 0$ or $A(0) = C(0) = 0$ the equations (7) with $x = 0$ have a non-null solution (see also Example 9).

Case (b) is more involved. Lemma 8 yields only $\lceil s/2 \rceil$ independent solutions. The construction of the remaining independent solutions is described below. Suppose that $A(0) = 0$ (the case $D(0) = 0$ is similar). The assumptions of case (b) imply that $a_1 = b_1 = c_1 = 0$, $d_1 > 0$ and $a_0 > 0$ (cf. Figure 5). For $n \le s/2$ we now construct a solution, $p_n(i,j)$ say, as follows. For $i = 0, \dots, s$ and $j \ge n$ we set

$$p_n(\cdot, n) = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}, \qquad p_n(\cdot, j) = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}, \quad j > n. \tag{14}$$

Note that $p_n(i,j)$ only lives on a finite part of the strip. Due to the special transition structure, $p_n(i,j)$ already satisfies the balance equations (5) for $j > n$. The remaining vectors $p_n(\cdot, n-1), \dots, p_n(\cdot, 0)$ are subsequently determined such that the balance equations for $j = n, \dots, 2, 1$ are satisfied. This only succeeds in case $n \le s/2$. The construction of $p_n(i,j)$ is illustrated in Example 12. Hence, we can construct $\lfloor s/2 \rfloor + 1$ solutions, which together with the $\lceil s/2 \rceil$ solutions of the form (13), make up a set of $s+1$ independent solutions of the balance equations (5).

Example 12. (Continuation of Example 8)
For the example with $s = 4$, it is straightforward to verify (from the transition diagram in Figure 5) that the sequences $p_0(i,j)$, $p_1(i,j)$ and $p_2(i,j)$, given by

$$p_0 = \begin{bmatrix} 0\,0\cdots \\ 0\,0 \\ 0\,0 \\ 0\,0 \\ 1\,0\cdots \end{bmatrix}, \quad p_1 = \begin{bmatrix} 0\,0\,0\cdots \\ 0\,0\,0 \\ -4\,0\,0 \\ 16\,0\,0 \\ 0\,1\,0\cdots \end{bmatrix}, \quad p_2 = \begin{bmatrix} 4 & 0\,0\,0\cdots \\ -53\frac{1}{3} & 0\,0\,0 \\ 128 & -4\,0\,0 \\ -8 & 16\,0\,0 \\ 0 & 0\,1\,0\cdots \end{bmatrix},$$

satisfy (14), and indeed are solutions of the balance equations (5).

From Lemma 2 and Corollary 1 we now obtain the main result of the paper.

Theorem 2. *In cases (a) and (c), the equilibrium probabilities $p(i,j)$ can be expressed as*

$$p(i,j) = \sum_{m=0}^{s} C_m y_m(i) x_m^j, \qquad i = 0, \dots, s, \quad j = 0, 1, \dots,$$

where x_m is the unique root of equation (11) in the interval $(0,1)$, resp. $(-1,1)$, and the corresponding $y_m(i)$ are a non-null solution of system (7) with $x = x_m$.
In case (b) it holds that

$$p(i,j) = \sum_{m=0}^{\lceil s/2 \rceil - 1} C_m y_m(i) x_m^j + \sum_{n=0}^{\lfloor s/2 \rfloor} D_n p_n(i,j), \qquad i = 0, \dots, s, \quad j = 0, 1, \dots,$$

where x_m is the unique root of (11) in the interval $(0,1)$, and $p_n(i,j)$ is a solution of (5) satisfying (14). In each case, the coefficients are uniquely determined from the boundary equations and the normalization equation.

Remark 2. (Monotonicity of x_m)
A direct consequence of the uniqueness of the roots x_m and the behaviour of the function $G(x)$ in Section 5 is that the roots x_m are decreasing in m, i.e. $x_0 > x_1 > \cdots$.

Remark 3. (Calculation of y_m)
The generating function of the sequence $y_m(i)$ is given by (10) with $x = x_m$ and $E(x) = m$,

$$Y_m(z) = \sum_{i=0}^{s} y_m(i) z^i = K(z - z_1(x_m))^m (z - z_2(x_m))^{s-m}.$$

So, if we set $K = 1$, it follows that

$$y_m(i) = (-1)^{s-i} \sum_{k=0}^{m} \binom{m}{k} \binom{s-m}{i-k} z_1(x_m)^{m-k} z_2(x_m)^{s-m-i+k}, \tag{15}$$

where by definition $\binom{s-m}{i-k} = 0$ if $i - k < 0$ or $i - k > s - m$. For numerical calculations, however, it is more attractive to compute $y_m(i)$ directly from the recursion (7), instead of using the explicit form (15).

Remark 4. (Computational complexity)
The computational effort to compute all roots x_m and their corresponding vectors y_m is proportional to s^2. In comparison, the matrix geometric approach requires the determination of the rate matrix R. This matrix can be computed iteratively from a nonlinear matrix equation and the computational effort of a single iteration is proportional to s^3 (cf. [13]). Both approaches further require the solution of $|V| + s + 1$ (linear) boundary equations.

7 Large jumps to the right

So far we did not consider Markov processes with large jumps to the right (e.g. queueing models with batch arrivals). In this section we discuss the extension to this type of processes by means of an example.

We consider the $M^2/E_2/s$ queue, i.e., the model in Example 1 where jobs arrive in batches of size 2 (instead of one by one). The transition rate diagram for $s = 4$ is shown in Figure 10 where we used the same notation and state description as in Example 1. The balance equations for the states (i, j) with $j > 1$ are given by,

$$\begin{aligned}(\lambda + s\mu)p(i,j) = {} & \lambda p(i, j-2) + (i+1)\mu p(i+1, j+1) \\ & + (s - (i-1))\mu p(i-1, j)\,, \qquad i = 0, \ldots, s, \quad j = 2, 3, \ldots,\end{aligned} \tag{16}$$

where by definition $p(-1, j) = p(s+1, j) = 0$. The balance equations for the states $(i, 0)$ and $(i, 1)$ and the states in V are treated as the boundary equations. In comparison with the $M/E_2/s$ queue we now have $s + 1$ extra boundary equations, namely the balance equations in the states $(i, 1)$. This implies that we need $s + 1$ additional, and thus in total $2(s + 1)$ independent solutions of (16) in order to satisfy the boundary equations and the normalization equation. To find these we again try solutions of the form

$$p(i, j) = y(i) x^j, \qquad i = 0, \ldots, s, \qquad j = 0, 1, \ldots, \tag{17}$$

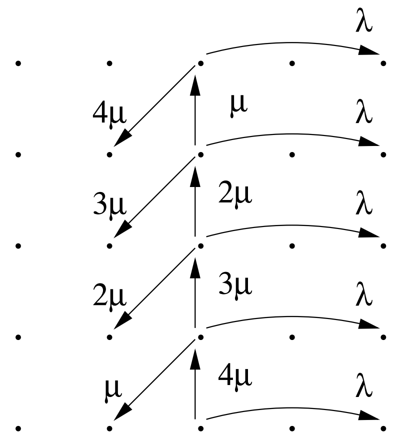

Fig. 10. The transition rate diagram of the $M^2/E_2/s$ queue for $s = 4$

with $|x| < 1$. Following the generating function approach in Section 5 we get for each $m = 0, \ldots, s$ the equation

$$\frac{2m - s}{s} = \frac{\lambda - (\lambda + s\mu)x^2}{\mu s x^2 \sqrt{x}}. \tag{18}$$

For the parameter values $\lambda = 1$, $\mu = 2$ and $s = 4$ the right-hand side of (18) is depicted in Figure 11. It is easy to prove that, under the ergodicity condition $4\lambda/\mu < s$, equation (18) has a unique solution $x = x_m$ in the interval $(0, 1)$. This yields $s + 1$ independent solutions of the form (17).

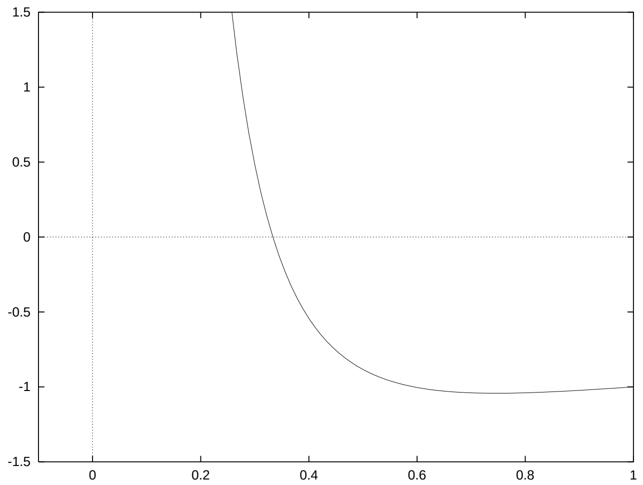

Fig. 11. The right-hand side of (18) for $\lambda = 1$, $\mu = 2$ and $s = 4$

Hence, $s + 1$ solutions are still missing. To find these, we need to consider *complex roots* of equation (18). Numerical experiments suggest that for $m > s/2$ equation (18) has two complex conjugate roots inside the unit circle. Note that $x = -x_m$ also satisfies (18) if

$m = s/2$. The complex roots, and this negative one, provide the missing $s+1$ solutions. In Table 1 we list for each m the roots inside the unit circle of (18) for the example $\lambda = 1$, $\mu = 2$ and $s = 4$. It may be difficult to prove that for each m equation (18) has

m	x_m
0	0.5940
1	0.3924
2	0.3333
	−0.3333
3	0.2990
	−0.3234 − 0.0397i
	−0.3234 + 0.0397i
4	0.2753
	−0.3018 − 0.0669i
	−0.3018 + 0.0669i

Table 1. All roots inside the unit circle of (18) for the example $\lambda = 1$, $\mu = 2$ and $s = 4$

sufficiently many roots inside the unit circle, since e.g. this number depends on the value of m (see Table 1) and the square root in (11) is not analytic inside the unit circle. These complications can be avoided by considering the square of equation (18) (i.e. by combining the equations for m and $s - m$). This yields for $m = 0, \ldots, \lfloor s/2 \rfloor$, the polynomial equation

$$\mu^2 s^2 x^5 \left(\frac{2m - s}{s}\right)^2 = \left(\lambda - (\lambda + s\mu)x^2\right)^2 . \tag{19}$$

Using Rouché's theorem it is easy to prove that for each m equation (19) has exactly 4 roots inside the unit circle, and these roots are simple (except when $m = s/2$). Clearly, these roots produce the desired $2(s+1)$ independent solutions of (16).

The important conclusion of this section is that for processes with large jumps to the right, the analysis can not be restricted to the real domain anymore. We have also seen that, for the $M^2/E_2/s$, the single equation for the geometric factors x_m can again be reduced to multiple equations for fewer roots, although the reduction is less than in the model with batches of size one. We expect that a similar reduction will be possible for more general processes with large jumps to the right.

8 Conclusions and comments

In this paper we showed, for a class of Markov processes on a semi-infinite strip, that the equilibrium distribution can be expressed as a finite sum of terms, which are geometric in the unbounded variable. By using a generating function technique we proved that each of the geometric factors can be characterized as the unique root in the interval $(-1, 1)$ of a certain equation. Due to this property, the result becomes numerically very attractive. In many queueing applications, it is possible to deduce, from the expression for the equilibrium probabilities, similar expressions for global performance measures.

The Markov processes considered in the paper are skip-free to the right, i.e., only jumps of size 1 to the right are allowed. It is likely that a similar characterization of the equilibrium probabilities and the geometric factors holds for Markov processes with larger jumps to the right. But we have seen in the previous section that the analysis can no longer be restricted to the real domain.

To exclude exceptional cases we assumed that on the strip the Markov process can move both in upward and downward directions, see Condition $1(ii)$. In fact, the analysis becomes much easier if this condition does not hold. If $A(1) = 0$ and $D(1) > 0$, then it is only possible to move from level i to level $i-1$ on the strip, where level i is the set of states $\{(i,0),(i,1),\ldots\}$. Hence, the balance equations can be solved recursively, from level s down to level 0. An example of such a Markov process can be found in [3]. The analysis is similar in case $A(1) > 0$ and $D(1) = 0$. If both $A(1)$ and $D(1)$ are zero, then the Markov process remains at the same level on the strip (it can only move to another level via the states in V). This implies that the balance equations can be solved at each level separately.

References

1. I.J.B.F. Adan, A.G. de Kok and J.A.C. Resing, A multi-server queueing model with locking. *EJOR*, **116** (1999), 16–26.
2. I.J.B.F. Adan, W.A. van de Waarsenburg and J. Wessels, Analyzing $E_k|E_r|c$ queues. *EJOR*, **92** (1996), 112–124.
3. I.J.B.F. Adan, E.A. van Doorn, J.A.C. Resing and W.R.W. Scheinhardt, Analysis of a single-server queue interacting with a fluid reservoir. *QUESTA*, **29** (1998), 313–336.
4. D. Bertsimas, An exact FCFS waiting time analysis for a class of G/G/s queueing systems. *QUESTA*, **3** (1988), 305–320.
5. D. Bertsimas, An analytic approach to a general class of $G/G/s$ queueing systems. *Opns. Res.*, **38** (1990), 139–155.
6. D. Bertsimas and X.A. Papaconstantinou, Analysis of the stationary $E_k/C_2/s$ queueing system. *EJOR*, **37** (1988), 272–287.
7. J.N. Daigle and D.M. Lucantoni, Queueing systems having phase-dependent arrival and service rates, in: W.J. Stewart (ed.), *Numerical solution of Markov chains*, Marcel Dekker, New York (1991), 161–202.
8. J.H.A. De Smit, The queue $GI/M/s$ with customers of different types or the queue $GI/H_m/s$, *Adv. Appl. Probab.*, **15** (1983), 392–419.
9. J.H.A. De Smit, A numerical solution for the multi-server queue with hyper-exponential service times, *Oper. Res. Lett.*, **2** (1983), 217–224.
10. A.I. Elwalid, D. Mitra and T.E. Stern, A theory of statistical multiplexing of Markovian sources: spectral expansions and algorithms, in: W.J. Stewart (editor), *Numerical solution of Markov chains*, Marcel Dekker Inc, New York (1991), 223–238.
11. P.R. Halmos, *Finite-dimensional vector spaces.* Van Nostrand, London, 1958.
12. S. Lang, *Linear algebra.* Addison-Wesley, New York, 1978.
13. G. Latouche and V. Ramaswami, A logarithmic reduction algorithm for quasi-birth-death processes. *J. Appl. Prob.*, **30** (1993), 650–674.
14. I. Mitrani and D. Mitra, A spectral expansion method for random walks on semi-infinite strips, in: R. Beauwens and P. de Groen (eds.), *Iterative methods in linear algebra*, North-Holland, Amsterdam (1992), 141–149.
15. I. Mitrani and B. Avi-Itzhak, A many server queue with service interruptions. *Opns. Res.*, **16** (1968), 628–638.
16. M.F. Neuts, *Matrix-geometric solutions in stochastic models.* Johns Hopkins University Press, Baltimore, MD, 1981.
17. S. Shapiro, The M-server queue with Poisson input and Gamma-distributed service of order two. *Opns. Res.*, **14** (1966), 685–694.