8 $G/M/1$ type models with linear rates

In the previous chapter we have seen that there are two problems for the application of the spectral expansion method: (i) proof of the existence of sufficiently many roots of the determinantal equation, and (ii) the computation of these roots. In this chapter we will consider a special (but rich) class of $G/M/1$ type models for which both problems can be solved completely. The models in this class have the property that the transtion rates are linear in one of the state variables.

8.1 The $G/M/1$ model with linear rates

We consider a Markov process 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 ∞ and j from 0 to m. The number of states in V is n. The states on the strip are ordered lexicographically, that is, $(0,0), (0,1), \ldots, (0,m)$, $(1,0), \ldots, (1,m), \ldots$. The set of states $\{(i,0), (i,1), \ldots, (i,m)\}$, $i \geq 1$, will be called *level* i. The states in V are all put together into level -1. We partition the state space according to these levels, and for this partitioning we assume that the generator Q is of the form

$$Q = \begin{pmatrix} B_{-1,-1} & B_{-1,0} & 0 & 0 & 0 & \cdots \\ B_{0,-1} & A_1 & A_0 & 0 & 0 & \cdots \\ B_{1,-1} & A_2 & A_1 & A_0 & 0 & \cdots \\ B_{2,-1} & A_3 & A_2 & A_1 & A_0 & \cdots \\ B_{3,-1} & A_4 & A_3 & A_2 & A_1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \end{pmatrix},$$

where the matrix $B_{-1,-1}$ is of dimension $(n+1)\times(n+1)$, $B_{-1,0}$ of dimension $(n+1)\times(m+1)$, the matrices $B_{i,-1}$, $i \geq 0$, of dimension $(m+1)\times(n+1)$, and A_i, $i \geq 0$, are square matrices of dimension $m+1$. Further we assume that the Markov process Q is irreducible.

This is still the standard set-up for $G/M/1$ type models, as we have seen in the previous chapter. In addition we are now going to impose more structure on the transition rates. More specifically, in state (i,j) with $i \geq 0$ the following transitions are possible ($k = -i, -i+1, \ldots, 0, 1$):

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

Clearly, the rates a_k and b_k are scaled by a factor $(m-j)$, and the rates c_k and d_k are scaled by a factor j. The last requirement has been added to guarantee that the total outflow from state (i, j) is the same for each $i > 1$. The above structure implies that the blocks A_k are tridiagonal; for example, A_0 has the form

$$A_0 = \begin{pmatrix} b_1 m & a_1 m & 0 & & & \\ d_1 & b_1(m-1)+c_1 & a_1(m-1) & & & \\ 0 & d_1 2 & b_1(m-2)+c_1 2 & & & \\ & \ddots & \ddots & \ddots & & \\ & & b_1 2 + c_1(m-2) & a_1 2 & 0 \\ & & d_1(m-1) & b_1 + c_1(m-1) & a_1 \\ & & 0 & d_1 m & c_1 m \end{pmatrix}.$$

The transition rate diagram is shown in figure 1; so, looking in the vertical direction, the upward rates are linear in $(m-j)$ (and thus vanish for $j = m$) and the downward rates are linear in j (and thus vanish for $j = 0$). Further, the size of upward and downward jumps and the size of jumps to the right are at most one, whereas jumps to the left are unlimited.

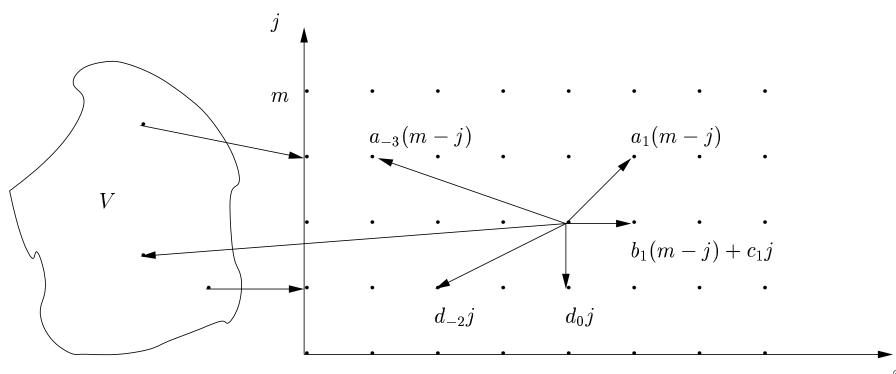

Figure 1: Transition rate diagram for the $G/M/1$ type model with linear rates

To exclude exceptional cases we impose some more conditions on the rates of the Markov process. First we introduce the generating 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 extra conditions on the rates are the following:

Condition 8.1

(*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;*
(*iv*) ($A(0) = 0$ *or* $D(0) = 0$) *and* ($C(0) = B(0) \neq 0$).

The conditions (*i*) and (*iii*) are obvious; the other two are imposed to exclude special cases, but both can be relaxed (see [2] for a more complete picture). Below we present some queueing models that fit into this class of $G/M/1$ type models.

Example 8.2 *The* $M/E_2/m$ *queue (see [5]).*
Jobs arrive according to a Poisson process with rate λ and have service times consisting of two exponential phases with parameter μ. In this example, 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. The variable i represents the number of jobs *waiting in the queue*, and j the number of servers working on a second service phase. We have $a_0 = \mu$, $b_1 = c_1 = \lambda/m$, $d_{-1} = \mu$ and all other coefficients are equal to zero.

Example 8.3 *The* $M/C_2/m$ *queue (see [3]).*
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 8.4 *The* $M/H_2/m$ *queue (see [6, 7]).*
Like Example 8.2, 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 8.5 *The* $M/M/m$ *queue with service interruptions (see [4]).*
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, j represents the number of servers that are operative. We have $a_0 = \beta$, $c_{-1} = \mu$, $b_1 = c_1 = \lambda/m$, $d_0 = \alpha$ and all other coefficients are equal to zero.

Example 8.6 *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 m groups of servers, each group consisting of two servers (see figure 2).

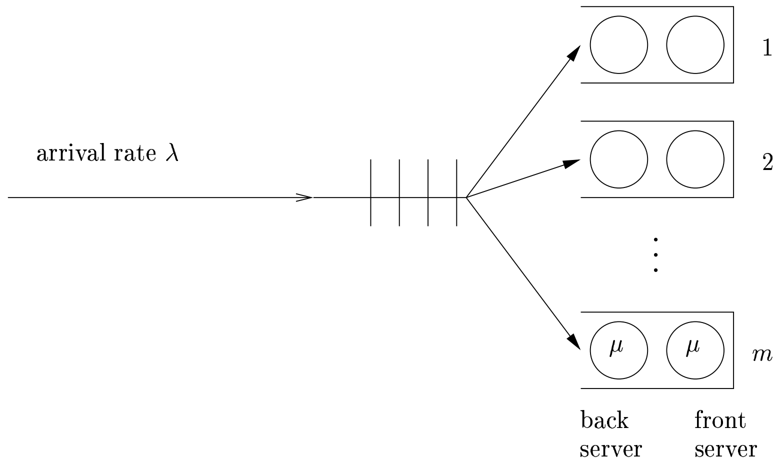

Figure 2: The multi-server queue with locking

Within a group, one of the servers is called the front server and the 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.

In this example, 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. The variable j denotes the number of customers that already completed their service but that are locked in by a customer at the back server and i denotes the number of customers waiting in the queue. We have $a_0 = \mu$, $b_1 = c_1 = \lambda/m$, $b_{-1} = \mu$, $d_{-2} = \mu$ and all other coefficients are equal to zero.

8.2 Stability condition

Now we will determine the stability condition of the process. Because the process is a Markov process of the $GI/M/1$ type, we can use Neuts' mean drift condition to obtain the stability condition. Here, the mean drift condition reads

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

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

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.

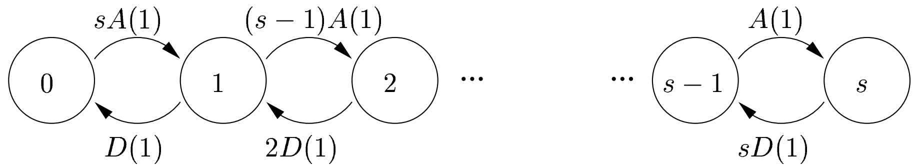

Figure 3: Transition rate diagram

Lemma 8.7 *The Markov process Q is ergodic if and only if*

$$D(1)(A'(1) - A(1) + B'(1) - B(1)) + A(1)(C'(1) - C(1) + D'(1) - D(1)) > 0. \tag{2}$$

From now on we assume that condition (2) holds.

8.3 Equilibrium equations

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

$$\begin{aligned}\sum_{k=-\infty}^{1}\Big((a_k + b_k)(m - j) + (c_k + d_k)j\Big)p(i, j) \;=\; & \sum_{k=-\infty}^{1}\Big(d_k(j + 1)p(i - k, j + 1) \\ & +(b_k(m - j) + c_k j)p(i - k, j) + a_k(m - j + 1)p(i - k, j - 1)\Big), \\ & i = 1, 2, \ldots, \qquad j = 0, 1, \ldots, m, \end{aligned} \tag{3}$$

where by convention $p(i, -1) = p(i, m + 1) = 0$. To solve the equilibrium equations we are going to use the spectral expansion method. That is, we try to find $m + 1$ basis solutions of the form

$$p_i = yx^i, \qquad i = 0, 1, \ldots, \tag{4}$$

where $y = (y(0), y(1), \ldots, y(m)) \neq 0$ and $|x| < 1$. Substitution of this form into (3) and dividing by the common power x^{i-1} yields, for $j = 0, \cdots, m$,

$$\begin{aligned} 0 \;=\; & (j + 1)D(x)y(j + 1) + ((m - j)B(x) + jC(x))y(j) - \\ & ((m - j)(A(1) + B(1)) + j(C(1) + D(1)))xy(j) + (m - j + 1)A(x)y(j - 1), \end{aligned} \tag{5}$$

where by convention $y(-1) = y(m + 1) = 0$. For given x, this is a system of linear homogeneous equations for $y(0), \ldots, y(m)$. 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 $m + 1$ different zeros, then the corresponding solutions (4) 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 [5, 3, 1] to transform the difference equations (5) into a single differential equation for the generating function of the sequence $y(i)$. This approach, in fact, reduces the determinantal equation for $m+1$ roots inside the unit circle to $m+1$ equations for a single root in the interval $(0,1)$.

8.4 Generating function approach

We transform the difference equations (5) into a differential equation for the generating function

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

of the sequence $y(j)$. Multiplying (5) by z^j and adding with respect to j we obtain

$$\begin{aligned} D(x)Y'(z) + mB(x)Y(z) + (C(x) - B(x))zY'(z) - (A(1) + B(1))mxY(z) & \\ +(A(1) + B(1) - C(1) - D(1))xzY'(z) + mA(x)zY(z) - A(x)z^2Y'(z) &= 0, \end{aligned}$$

which may be rewritten in the form

$$\begin{aligned} \frac{Y'(z)}{Y(z)} &= \frac{m[A(x)z + B(x) - (A(1) + B(1))x]}{A(x)z^2 - ((A(1) + B(1) - C(1) - D(1))x + C(x) - B(x))z - D(x)} \\ &= \frac{E(x)}{z - z_1(x)} + \frac{m - E(x)}{z - z_2(x)}, \end{aligned} \tag{6}$$

where $z_1(x)$ and $z_2(x)$ are the roots ($z_1(x)$: $+$ sign, $z_2(x)$: $-$ sign) of the denominator of (6) and $E(x)$ satisfies the equation

$$2E(x) - m = m \cdot \frac{B(x) + C(x) - x\,(A(1) + B(1) + C(1) + D(1))}{\sqrt{F(x)^2 + 4A(x)D(x)}}, \tag{7}$$

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 (6) by $A(x)$, and hence, we implicitly assume here that $A(x) \neq 0$; this indeed holds, since, as we will see later on, the desired roots x are positive. The general solution of the differential equation (6) is

$$Y(z) = K(z - z_1(x))^{E(x)}(z - z_2(x))^{m-E(x)}, \tag{8}$$

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 $m - E(x)$ should be equal to a non-negative integer, i.e. $E(x) = k, k = 0, \ldots, m$. Hence, for each k we get an equation in x, for which we will prove that, under the ergodicity condition (2), there is exactly one solution, x_k say, in the interval $(0,1)$.

Lemma 8.8 *For each $k = 0, \ldots, m$ the equation*

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

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

Proof: Let $G(x)$ denote the function on the right-hand side of (9). Then, it is straightforward to check that $G(1) = -1$, $G'(1) > 0$ (which holds if and only if ergodicity condition (2) is satisfied) and $\lim_{x \downarrow 0} G(x) = +\infty$, which follows from the condition 8.1(iv). The existence of a root now follows from a continuity argument; its uniqueness follows from the observation that, since the equilibrium distribution is unique, there can be at most $m + 1$ basis solutions of the form (4). □

Example 8.9 *The $M/M/m$ queue with service interruptions.*
Let us consider the multi-server queue with unrelaible servers, described in example 8.5. For this model we have

$$A(x) = \theta x, \qquad B(x) = \frac{\lambda}{m}, \qquad C(x) = \frac{\lambda}{m} + \mu x^2, \qquad D(x) = \eta x.$$

Thus $F(x)$ and $G(x)$ are given by

$$\begin{aligned} F(x) &= x(\theta - \eta + \mu(x-1)), \\ G(x) &= \frac{2\dfrac{\lambda}{m} + \mu x^2 - x(\theta - \eta + \mu(x-1))}{x\sqrt{(\theta - \eta + \mu(x-1))^2 + 4\theta\eta}}. \end{aligned}$$

In figure 4 we show the function $G(x)$ for the parameter values $\lambda = 1$, $\mu = 2$, $\eta = 3$, $\theta = 4$ and $m = 4$. The intersections of $G(x)$ with the horizontal lines provide the desired x_k, $k = 0, 1, 2, 3, 4$.

Hence, according to the lemma above, there are $m + 1$ basis solutions of the form (4). The equilibrium probabilities $p(i, j)$ can be expressed as a linear combination of these basis solutions, the coefficients of which follow from the equilibrium equations at the levels 0 and 1 (the so-called boundary equations) and the normalization equation. These findings are summarized in the following theorem.

Theorem 8.10 *The equilibrium probabilities $p(i, j)$ can be expressed as*

$$p(i,j) = \sum_{k=0}^{m} c_k y_k(j) x_k^i, \qquad i = 0, 1, \ldots, \qquad j = 0, 1, \ldots, m,$$

where x_k is the unique root of equation (9) in the interval $(0, 1)$, and the corresponding $y_k(j)$ are a nonnull solution of system (5) with $x = x_k$. The coefficients c_k are uniquely determined from the boundary equations and the normalization equation.

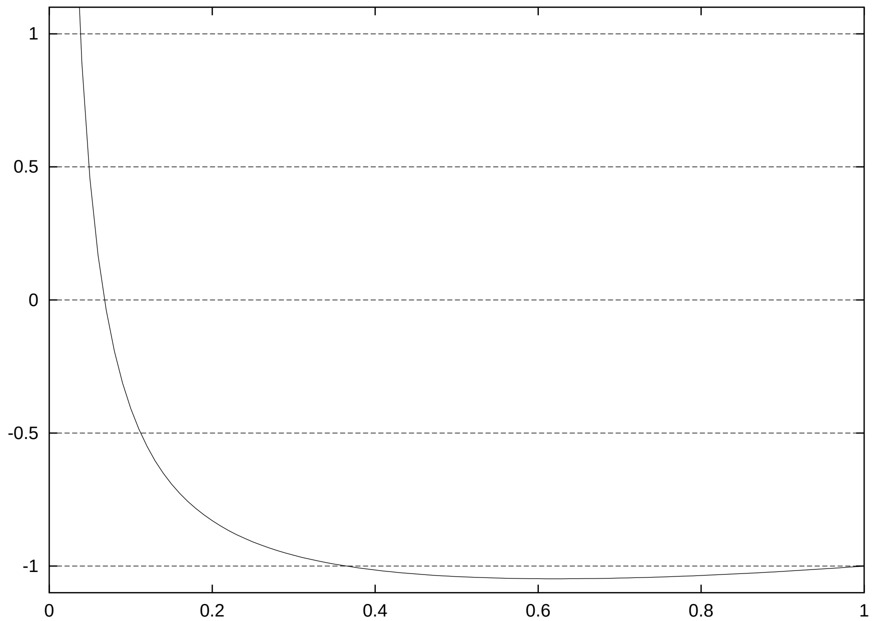

Figure 4: The function $G(x)$ for the multi-server queue with service interruptions, for $\lambda = 1$, $\mu = 2$, $\eta = 3$, $\theta = 4$ and $m = 4$

8.5 Absorption times

In this section we consider processes for which A_0 is of the special form

$$A_0 = \lambda I, \tag{10}$$

for some λ. In fact, in all examples 8.2 up to 8.6 the transition matrix A_0 is of this form, where λ is the arrival rate of the Poisson stream of customers. For these processes we are going to study the time until absorption in the set V, given that at time $t = 0$, the process starts in state (i, j) and there are no new arrivals for $t \geq 0$; this means that for $t \geq 0$ transitions from state (i, j) to $(i + 1, j)$ are not possible. Thus we consider a Markov process with generator

$$\begin{pmatrix} B_{-1,-1} & B_{-1,0} & 0 & 0 & 0 & \cdots \\ B_{0,-1} & A_1 + A_0 & 0 & 0 & 0 & \cdots \\ B_{1,-1} & A_2 & A_1 + A_0 & 0 & 0 & \cdots \\ B_{2,-1} & A_3 & A_2 & A_1 + A_0 & 0 & \cdots \\ B_{3,-1} & A_4 & A_3 & A_2 & A_1 + A_0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \end{pmatrix}.$$

The motivation for studying the absorption time is that in each of the examples 8.2 up to 8.6 this time is exactly the waiting time of a customer, arriving in state (i, j); see also example 8.11. Let $F_{ij}(t)$ denote the probability that the time to reach the set V is geater than t, given that the process starts in state (i, j) at time $t = 0$.

Example 8.11 Consider the $M/E_2/m$ system described in example 8.2. Let W denote the waiting time of a customer. By conditioning on the state seen on arrival and using PASTA and theorem 8.10 we obtain

$$\begin{aligned} P(W > t) &= \sum_{i=0}^{\infty} \sum_{j=0}^{m} p(i,j) F_{ij}(t) \\ &= \sum_{k=0}^{m} c_k \sum_{j=0}^{m} y_k(j) \sum_{i=0}^{\infty} x_k^i F_{ij}(t). \end{aligned}$$

Hence, once $F_{ij}(t)$ is known, we can determine the waiting time distribution.

Below we derive a set of differential equations for the probabilities $F_{ij}(t)$. For small $\Delta t \geq 0$ it holds that

$$F_i(t + \Delta t) = F_i(t) + \Delta t (A_1 + A_0) F_i(t) + \Delta t \sum_{j=0}^{i-1} A_{i+1-j} F_j(t),$$

where $F_i(t) = (F_{i0}(t), F_{i1}(t), \ldots, F_{im}(t))'$. Dividing these equations by Δt and letting Δt tend to zero, yields

$$F_i'(t) = (A_1 + A_0) F_i(t) + \sum_{j=0}^{i-1} A_{i+1-j} F_j(t), \qquad i = 0, 1, 2, \ldots, \tag{11}$$

with initial condition $F_i(0) = 1$. To solve these differential equations we are going to use Laplace transforms. Let

$$F_i^*(s) = \int_{t=0}^{\infty} F_i(t) e^{-st} dt, \qquad s \geq 0.$$

Transforming the differential equations (11) for $F_i(t)$ and using that

$$\int_{t=0}^{\infty} F_i'(t) e^{-st} dt = -e + s F_i^*(s),$$

gives

$$(sI - A_1 - A_0) F_i^*(s) = e + \sum_{j=0}^{i-1} A_{i+1-j} F_j^*(s). \tag{12}$$

From these equations, the Laplace transforms $F_i^*(s)$ may be solved recursively. Example 8.11 suggests that in many applications the function

$$F(t) = \sum_{k=0}^{m} c_k y_k \sum_{i=0}^{\infty} x_k^i F_i(t), \qquad t \geq 0, \tag{13}$$

corresponds to the complementary waiting time distribution, i.e., $P(W > t)$. Below we show that $F(t)$ can be determined explicitly. Define for $s \geq 0$,

$$\begin{aligned} G_k^*(s) &= \sum_{j=0}^{\infty} x_k^j F_j^*(s), \qquad k = 0, \ldots, m \\ F^*(s) &= \int_{t=0}^{\infty} F(t) e^{-st} dt \\ &= \sum_{k=0}^{m} c_k y_k G_k^*(s). \end{aligned}$$

Multiplying (12) with $y_k x_k^i$ and adding over all i leads to

$$\begin{aligned} y_k \left(sI - A_1 - A_0 \right) G_k^*(s) &= \frac{y_k e}{1 - x_k} + \sum_{i=0}^{\infty} \sum_{j=0}^{i-1} y_k A_{i+1-j} x_k^i F_j^*(s) \\ &= \frac{y_k e}{1 - x_k} + \sum_{j=0}^{\infty} \left(\sum_{i=2}^{\infty} y_k A_i x_k^{i-1} \right) x_k^j F_j^*(s). \end{aligned}$$

Since

$$0 = \sum_{i=0}^{\infty} y_k A_i x_k^i = y_k (A_0 + A_1 x_k) + x_k \sum_{i=2}^{\infty} y_k A_i x_k^{i-1},$$

we get

$$y_k \left(sI - A_1 - A_0 \right) G_k^*(s) = \frac{y_k e}{1 - x_k} - y_k \left(\frac{1}{x_k} A_0 + A_1 \right) G_k^*(s),$$

and thus

$$y_k \left(sI - (1 - 1/x_k) A_0 \right) G_k^*(s) = \frac{y_k e}{1 - x_k}.$$

Now, by substituting the special form (10), we obtain

$$y_k G_k^*(s) = \frac{y_k e}{1 - x_k} \cdot \frac{1}{s - \lambda + \lambda / x_k}.$$

Multiplying by c_k and adding over all k finally yields

$$F^*(s) = \sum_{k=0}^{m} \frac{c_k y_k e}{1 - x_k} \cdot \frac{1}{s - \lambda + \lambda / x_k}.$$

The inverse of this transform is readily obtained; the result is summarized in the following theorem.

Theorem 8.12 *The function $F(t)$, defined by (13), is given by*

$$F(t) = \sum_{k=0}^{m} \frac{c_k y_k e}{1 - x_k} e^{\lambda (1 - 1/x_k) t}, \qquad t \geq 0.$$

Example 8.13 Let us again consider the $M/E_2/m$ queueing model in example 8.11. Since $F(t) = P(W > t)$ we may immediately conclude from Theorem 8.12 that the waiting time distribution for the $M/E_2/m$ is a mixture of exponentials.

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), pp. 16–26.

[2] I.J.B.F. Adan, J.A.C. Resing, *A class of Markov processes on a semi-infinite strip*. 3rd International Meeting on the Numerical Solution of Markov Chains, B. Plateau, W.J. Stewart, M. Silva (eds.), 1999, pp. 41–57.

[3] D. Bertsimas and X.A. Papaconstantinou, *Analysis of the stationary* $E_k/C_2/s$ *queueing system*. EJOR, 37 (1988), pp. 272–287.

[4] I. Mitrani and B. Avi-Itzhak, *A many server queue with service interruptions*. Opns. Res., 16 (1968), pp. 628–638.

[5] S. Shapiro, *The* M*-server queue with Poisson input and Gamma-distributed service of order two*. Opns. Res., 14 (1966), pp. 685–694.

[6] 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), pp. 392–419.

[7] J.H.A. De Smit, *A numerical solution for the multi-server queue with hyper-exponential service times*, Oper. Res. Lett., 2 (1983), pp. 217–224.