

SIMPLE SPECTRAL REPRESENTATIONS FOR THE M/M/1 QUEUE

Joseph ABATE

AT&T Bell Laboratories, LC 2W-E06, 184 Liberty Corner Road, Warren, NJ 07060, U.S.A.

Ward WHITT

AT&T Bell Laboratories, MH 2C-178, 600 Mountain Avenue, Murray Hill, NJ 07974, U.S.A.

Received 3 August 1987

Revised 8 March 1988

Abstract

This paper shows that certain basic descriptions of the time-dependent behavior of the M/M/1 queue have very simple representations as mixtures of exponentials. In particular, this is true for the busy-period density, the probability that the server is busy starting at zero, the expected queue length starting at zero and the autocorrelation function of the stationary queue-length process. In each case the mixing density is a minor modification of a beta density. The last two representations also apply to regulated or reflected Brownian motion (RBM) by virtue of the heavy-traffic limit. Connections are also established to the classical spectral representations of Ledermann and Reuter (1954) and Karlin and McGregor (1958) and the associated trigonometric integral representations of Ledermann and Reuter, Vulot (1954), Morse (1955), Riordan (1961) and Takács (1962). Overall, this paper aims to provide a more unified view of the M/M/1 transient behavior and show how several different approaches are related.

Keywords: M/M/1 queue, Brownian motion, spectral representation, mixtures of exponentials, busy period, autocorrelation function, time-dependent mean, transient behavior, Chebycheff polynomials, duality, the associated process.

1. Introduction

A fundamental result for the M/M/1 queue and more general birth-and-death processes is the spectral representation of the probability transition function $P_{ij}(t)$; see Ledermann and Reuter [23], Karlin and McGregor [17,18] and van Doorn [33]. This spectral representation is an explicit representation as a (not necessarily nonnegative) mixture of exponentials, i.e.,

$$P_{ij}(t) = \pi_j \int_0^\infty e^{-xt} q_i(x) q_j(x) d\Phi(x) \quad (1.1)$$

where $\pi_0 = 1$ and $\pi_j = (\lambda_0 \dots \lambda_{j-1}) / (\mu_1 \dots \mu_j)$ with λ_j the birth rate and μ_j the

death rate in state j , $q_i(x)$ is a recursively defined system of polynomials in x satisfying orthogonality relations

$$\pi_j \int_0^\infty q_i(x) q_j(x) d\Phi(x) = \delta_{ij} \quad (1.2)$$

with $\delta_{ij} = 1$ if $i = j$ and 0 otherwise, and $\Phi(x)$ is a positive measure on $[0, \infty)$ called the spectral measure. However, even for the M/M/1 queue, the explicit representation (1.1) is fairly complicated, yielding orthogonal polynomials $q_i(x)$ related to the Chebycheff polynomials of the second kind. For example, from (1.1) it is not obvious that there is a convenient representation for the mean queue length at time t . One purpose of this paper is to show how the explicit representation (1.1) for the M/M/1 queue can be further exploited. For example, we show that tractable expressions can be obtained for the moments.

A second purpose of this paper is to point out that several M/M/1 quantities of interest have remarkably simple spectral representations. In particular, this is true for the busy-period density, the probability that the server is busy starting at zero, the mean queue length starting at zero and the (auto) correlation function of the stationary queue-length process. If $g(t)$ denotes one of these functions, then it has a representation

$$g(t) = \int_{\tau_1}^{\tau_2} y^{-1} e^{-t/y} w(y) dy, \quad t \geq 0, \quad (1.3)$$

where the mixing density $w(y)$ in (1.3) is a simple modification of a four-parameter beta density ($\beta(p, q) \equiv \beta(p, q, \tau_1, \tau_2)$), i.e., $w(y) = Af(y)/y^k$ for constants A and k where $f(y)$ is the beta density

$$f(y) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} \frac{(y-\tau_1)^{p-1}(\tau_2-y)^{q-1}}{(\tau_2-\tau_1)^{p+q-1}}, \quad \tau_1 \leq y \leq \tau_2, \quad (1.4)$$

with $\Gamma(p)$ being the gamma function; see Ch. 24 of Johnson and Kotz [15]. These simple spectral representations facilitate calculation by numerical integration and they improve our understanding of the M/M/1 transient behavior. Moreover, these simple M/M/1 spectral representations imply corresponding simple spectral representations for the first moment function starting at zero and the correlation function of regulated or reflecting Brownian motion (RBM).

In fact, a simple spectral representation for the M/M/1 busy-period distribution was already obtained by Karlin and McGregor in (6.4) of [18], but it doesn't seem to be sufficiently well appreciated. We indicated how to obtain this particular M/M/1 result directly without getting involved in the full spectral theory (theorem 3.1 below). In fact, our argument coincides with Keilson and Kooharian's analysis of the probability of emptiness in M/M/1 on p. 110 of [21]. The other simple spectral representations follow immediately from the busy-period representation by exploiting relations that we established in [3-5]. (Our recent papers [1-5] do not discuss spectral representations; here we provide connections

between our recent results and the earlier literature.) In fact, all four spectral representations follow immediately from any one, by virtue of these relations. First, the probability that the server is busy at time t when the system starts empty, $1 - P_{00}(t)$, is connected to the busy-period cdf $B(t)$ by the remarkable relation $1 - P_{00}(t) = \rho B(t)$, as shown in corollary 4.2.3 of [4]. (As indicated in section 6, this relation can be explained to a large extent by duality; see Chapter 3 of [33]. From this relation, Keilson and Kooharian's result in [21] can be obtained from Karlin and McGregor [18], while our theorem 3.1 can be obtained from either [18] or [21].) The other two quantities are connected to $B(t)$ via stationary-excess relations, as indicated in section 4 below. We also obtain a new proof of the spectral representation for the transition function $P_{ij}(t)$ in (1.1) for the M/M/1 queue starting with the spectral representation for the busy-period distribution; see (6.5), (6.7) and section 8. We show that our proof of (1.1) starting from the busy-period distribution is very natural by relating it to the associated process introduced by Karlin and McGregor [18]; see section 10.

The RBM spectral representations emerge in the limit as $\rho \rightarrow 1$ where ρ is the traffic intensity. The connection to RBM is made transparent by using an appropriate time scaling. For the RBM correlation function, an equivalent simple spectral representation was already obtained by Ott [28]; also see Woodside, Pagurek and Newell [35]. Woodside et al. show how the simple spectral representation can be exploited to obtain accurate numerical results by numerical integration. In fact, several papers have recently proposed other techniques for performing numerical calculations to describe the transient behavior of the M/M/1 queue [8,12,13,16,31], but it appears that numerical integration (e.g., by Simpson's rule) using integral representations is a better way to proceed. (However, the generalized Q -functions in [12,13] seem useful to obtain very high accuracy.) The advantages of integral representations for generating numbers for the M/M/1 queue seems to have been first discovered and applied by Morse [26].

Given the simple spectral representations developed here, it is evident that considerable unanticipated simplification occurs when you compute the mean directly from the probability transition function in (1.1) via $\sum_{j=1}^{\infty} j P_{0j}(t)$. We show how this simplification and others occur via relationships for the Chebycheff polynomials of the second kind. In other words, we also show how the new simple spectral representation can be obtained directly from the general spectral representation (1.1) by establishing appropriate properties of the Chebycheff polynomials; see theorems 5.3, 6.2 and 7.1.

We began this research to see if we might discover how our recent results in [1-5] are connected to the earlier literature on spectral representations. We found that the nice M/M/1 structure in [1-5] is indeed reflected by a corresponding nice structure in the spectral representations, and that these results can also be deduced relatively directly from the spectral theory. Inevitably our goal expanded to trying to better understand the connections among several different results and methods for the M/M/1 queue. To reflect what we learned, this paper contains a

certain amount of expository material, which we hope will help others understand the literature. To a large extent, the M/M/1 queue is a solved problem, but we believe that a better understanding of the solution is still needed.

The rest of this paper is organized as follows. We provide background on the M/M/1 queue in section 2. We derive the simple spectral representation for the busy-period density in section 3. We derive the associated simple spectral representations for the mean queue length starting at zero and the correlation function in section 4. We also obtain the corresponding results for RBM there.

In section 5 we begin to establish connections to (1.1) by establishing the spectral representation for the first-passage-time density from n to 0, i.e., the n -fold convolution of the busy-period density. In section 6 we derive the spectral representation for the transition function $P_{0n}(t)$ and consider the direct evaluation of the mean $\sum_{n=1}^{\infty} nP_{0n}(t)$ from (1.1) and [18]. In section 7 we obtain relatively simple spectral representations for all moments of the queue length starting at i , giving explicit integral representations for the first two. In section 8 we obtain a new derivation of (1.1), based on a relation between the first passage times up and the orthogonal polynomials $q_i(x)$. In section 9 we established connections to the trigonometric integral representations due to Ledermann and Reuter [23], Vaultot [34] and Riordan [29], Morse [26] and Takács [32]. Finally, in section 10 we discuss the role of the associated birth-and-death process, which is used extensively in [18].

2. The M/M/1 model with time scaling

Let $Q(t)$ represent the queue length (including the customer service, if any) at time t in the M/M/1 model. Let the service rate be 1, so that the arrival rate coincides with the traffic intensity ρ . Assume that $\rho < 1$, so that the system is stable with $Q(t)$ converging in distribution to $Q(\infty)$ as $t \rightarrow \infty$, where $P(Q(\infty) = k) = (1 - \rho)\rho^k$, $k \geq 0$.

As in [3–6], we further scale time by $2\theta^2 = 2/(1 - \rho)^2$; i.e., we consider $Q(2t/(1 - \rho)^2)$ and let $P_{ij}(t)$ be the time-scaled transition function

$$P_{ij}(t) = P\left(Q\left(2t/(1 - \rho)^2\right) = j \mid Q(0) = i\right). \quad (2.1)$$

The time scaling in (2.1) is a very important part of the story; we use it throughout the paper. It is significant that the first-order effect of the single parameter ρ is captured by this time scaling. As discussed in section 2.2 of [3], the time-scaling captures the heavy-traffic behavior as $\rho \rightarrow 1$. In particular, the family of processes $\{2^{-1}(1 - \rho)Q(2t/(1 - \rho)^2); t \geq 0\}$ indexed by ρ converges to canonical RBM, having drift coefficient -1 and diffusion coefficient 1, as $\rho \rightarrow 1$.

We start by focusing on the busy-period distribution. Let $B(t)$ be the time-scaled busy-period cdf (cumulative distribution function); let $B^c(t) \equiv 1 - B(t)$ be

the complementary busy-period cdf; and let $b(t)$ be the density. Let $\hat{B}^c(s)$ and $\hat{b}(s)$ be the associated Laplace transforms, which are

$$\begin{aligned} \hat{b}(s) &\equiv \int_0^\infty e^{-st}b(t) dt = z_1(s) = [1 - \theta + \theta^2s - \theta\Psi(s)]/\rho \\ \hat{B}^c(s) &= 2\theta/[1 + \theta s + \Psi(s)], \end{aligned} \tag{2.2}$$

where

$$\begin{aligned} \theta &= (1 - \rho)/2, & \Psi(s) &= [1 + 2(1 - \theta)s + (\theta s)^2]^{1/2} \\ r_1(s) &= \Psi + (1 - \theta s), & r_2(s) &= \Psi - (1 - \theta s) \\ \rho z_1 &= 1 - \theta r_1, & \rho z_2 &= 1 + r_2 \\ r_1 r_2 &= 2s, & \rho z_1 z_2 &= 1 \quad \text{and} \quad \rho(1 - z_2)(z_2 - 1) = 2\theta^2 s. \end{aligned} \tag{2.3}$$

The functions $z_1 \equiv z_1(s)$ and $z_2 \equiv z_2(s)$ are the two roots of the basic quadratic equation $\rho z^2 - (1 + \rho + 2\theta^2 s)z + 1 = 0$.

Kendall [22], pp. 168–171, first showed that the Laplace transform $\hat{b}(s)$ in (2.2) can be inverted to obtain the representation (in our time scale)

$$b(t) = \frac{1}{t\sqrt{\rho}} e^{-t/\tau} [e^{-\nu} I_1(\nu)], \quad t \geq 0, \tag{2.4}$$

where $I_1(\nu)$ is a modified Bessel function of the first kind (p. 377 of [7]),

$$\tau = \frac{(1 + \sqrt{\rho})^2}{2} \quad \text{and} \quad \nu = t\theta^{-2}\sqrt{\rho}. \tag{2.5}$$

The parameter τ in (2.5) is the *time-scaled relaxation time*. Interesting probabilistic and algebraic derivations of (2.4) have been given by Champenowne [14] and Massey [25].

3. The spectral and mixing densities of the busy-period density

We start with the busy-period density $b(t)$ and define two representations

$$b(t) = \int_{r_1}^{r_2} x e^{-xt} \phi(x) dx, \quad t \geq 0, \tag{3.1}$$

for $0 < r_1 < r_2 < \infty$ and

$$b(t) = \int_{\tau_1}^{\tau_2} y^{-1} e^{-t/y} w(y) dy, \quad t \geq 0, \tag{3.2}$$

for $0 < \tau_1 < \tau_2 < \infty$. We call $\phi(x)$ in (3.1) the *spectral density* and $w(y)$ in (3.2) the *mixing density*; $\phi(x)$ in (3.1) averages the exponential *rates* x , while $w(y)$ averages the exponential *means* or *times* y . (Since $\rho B(t) = 1 - P_{00}(t)$ and $q_0(x) = 1$, it is easy to see that $\phi(x)$ in (3.1) is the density of the spectral measure $\Phi(x)$ in (1.1), but Φ also has an atom at the origin; see (6.5).)

From what we have done here, we do not yet know that (3.1) and (3.2) are valid, for any $\phi(x)$ and $w(y)$, but we will show that they are. (This also follows from (6.4) of [18].) Given that (3.1) and (3.2) hold, the spectral density $\phi(x)$ and the mixing density $w(y)$ are easily related through the change of variables $y = 1/x$, so that

$$w(y) = y^{-2}\phi(y^{-1}) \quad \text{and} \quad \phi(x) = x^{-2}w(x^{-1}) \quad (3.3)$$

and the limits of integration in (3.1) and (3.2) are related by

$$r_1 = \tau_2^{-1} \quad \text{and} \quad r_2 = \tau_1^{-1}. \quad (3.4)$$

Hence, it suffices to consider only one of (3.1) and (3.2). For interpretation, we prefer (3.2) because it relates directly to the relaxation time. Indeed, the upper limit of integration τ_2 in (3.2) is the relaxation time. Moreover, (3.1) and (3.2) can easily be used to describe the asymptotic behavior as $t \rightarrow \infty$, e.g., via Laplace's method; see p. 86 of Olver [27].

THEOREM 3.1

Representations (3.1) and (3.2) are valid with

$$\tau_1 = \frac{(1 - \sqrt{\rho})^2}{2}, \quad \tau_2 = \frac{(1 + \sqrt{\rho})^2}{2} \quad (3.5)$$

$$\phi(x) = \frac{(1 - \rho)}{2\pi\rho} \frac{\sqrt{(1 - \tau_1 x)(\tau_2 x - 1)}}{x}, \quad \tau_2^{-1} \leq x \leq \tau_1^{-1}, \quad (3.6)$$

$$w(y) = \frac{(1 - \rho)}{2\pi\rho} \frac{\sqrt{(y - \tau_1)(\tau_2 - y)}}{y^2}, \quad \tau_1 \leq y \leq \tau_2. \quad (3.7)$$

The densities ϕ and w in (3.6) and (3.7) have similar properties; e.g., both are unimodal and concave; the derivatives are strictly decreasing from $+\infty$ at the lower limit to $-\infty$ at the upper limit. The modes of $\phi(x)$ and $w(y)$ are $2/(1 + \rho)$ and $(3(1 + \rho) - \sqrt{1 + 34\rho + \rho^2})/4$, respectively. Sample values for the case $\rho = 0.75$ appear in table 1. As noted in [6], $w(y)$ has relatively little mass near its upper limit τ_2 , so that we should not be surprised that simple approximations based on the relaxation time τ_2 do not perform well.

Before proving theorem 3.1, we show how it can be discovered and proved from the moments of the busy-period, which we know (theorem 3.2 of [4] or section 2.5 of [6]). In particular, it is easy to see that the k^{th} moment of $w(y)$ is the k^{th} moment of $b(t)$, say m_k , divided by $k!$ Similarly, it is easy to see that the k^{th} moment of $\phi(x)$ is expressed via the derivatives of $b(t)$ as

$$\int_{r_1}^{r_2} x^k \phi(x) dx = (-1)^{k-1} b^{(k-1)}(0), \quad k \geq 2. \quad (3.8)$$

Table 1
 Sample values of the M/M/1 busy-period spectral density $\phi(x)$ in (3.6) and mixing density $w(y)$ in (3.7) for the case $\rho = 0.75$.

rate x	spectral density $\phi(x)$ in (3.6)	time y	mixing density $w(y)$ in (3.7)
(asymptotic rate)			
$\tau_2^{-1} = 0.574$	0.000	$\tau_1 = 0.009$	0.0
0.60	0.019	0.010	22.1
0.75	0.039	mode = 0.012	26.6
1.00	0.0456	0.015	24.0
mode = $\frac{2}{1+\rho} = 1.14$	0.0461	0.03	11.2
1.25	0.0460	0.10	2.05
1.50	0.045	0.50	0.17
2.00	0.042	1.00	0.05
10.0	0.021	1.50	0.01
100.0	0.002	$\tau_2 = 1.74$	0.00
$\tau_1^{-1} = 111.0$	0.000	(relaxation time)	

However, from theorem 9 of [5], we can express the derivatives $b^{(k)}(0)$ in terms of the moments as

$$(-1)^{k-1} b^{(k-1)}(0) = \frac{m_{k+1}}{(k+1)! \theta^{2k+1}}, \quad k \geq 2. \tag{3.9}$$

Therefore, we have determined both $\phi(x)$ and $w(y)$ via their moment sequences. We can thus identify them by solving this moment problem.

In fact, we first discovered the simple spectral representation for the RBM correlation function by precisely this moment matching approach. As we show in section 4, the mixing density for the RBM correlation function is a beta $\beta(\frac{3}{2}, \frac{3}{2}, 0, 2)$ density symmetric around the mode 1. All moments for this beta distribution are given in (7) on p. 40 of [15]. We first discovered theorem 4.2 below by recognizing that the moments of the mixing density coincide with the moments of the beta $\beta(\frac{3}{2}, \frac{3}{2}, 0, 2)$ density.

Proof of theorem 3.1

We determine $\phi(x)$ in (3.6) by a judicious partial evaluation of the Bromwich integral for the inversion of the Laplace transform $\hat{b}(s)$ in (2.2), i.e.,

$$b(t) = \frac{1}{2\pi i} \int_{-i\infty}^{+i\infty} \hat{b}(s) e^{st} ds, \tag{3.10}$$

as in (7) on p. 289 on Bailey [9]. First note that $\Psi(s)$ in (2.3) can immediately be expressed as

$$\Psi(s) = \sqrt{(\tau_1 s + 1)(\tau_2 s + 1)}. \tag{3.11}$$

Because the singularities of $\hat{b}(s)$ are on the negative real axis, we can replace the contour by a loop which starts at infinity on the negative real axis, stays below the negative real axis, goes round the origin counter-clockwise and returns above the axis to its starting point. The singularities of $\hat{b}(s)$ are a branch point at $s_1 = -\tau_1^{-1}$, a branch point at $s_2 = -\tau_2^{-1}$ and a branch cut from s_1 to s_2 . Therefore, we can further deform the loop contour to two lines which coincide with the upper and lower edges of the branch cut. Hence we have

$$b(t) = \frac{1}{2\pi i} \int_{s_1}^{s_2} \frac{\theta \Psi(s)}{\rho} e^{st} ds + \frac{1}{2\pi i} \int_{s_2}^{s_1} \frac{\theta \Psi(s)}{\rho} e^{st} ds. \quad (3.12)$$

Let $s = e^{-i\pi}x$ in the first integral and let $s = e^{+i\pi}x$ in the second integral. Then we find

$$b(t) = \int_{\tau_2^{-1}}^{\tau_1^{-1}} \frac{\theta}{\pi\rho} \sqrt{(1-\tau_1x)(\tau_2x-1)} e^{-xt} dx. \quad \blacksquare$$

Second proof of theorem 3.1

Starting with (3.1), note that the Laplace transform $\hat{b}(s)$ coincides with the Stieltjes transform of the spectral density $\phi(x)$, say $\bar{\phi}(s)$, see p. 15 of [30]; i.e.,

$$\hat{b}(s) \equiv \int_0^\infty e^{-st} b(t) dt = \bar{\phi}(s) \equiv \int_0^\infty \frac{x}{s+x} \phi(x) dx \quad (3.13)$$

which can be inverted by

$$\begin{aligned} x\phi(x) &= \lim_{\epsilon \rightarrow 0} \frac{1}{2\pi i} [\bar{\phi}(-x - i\epsilon) - \bar{\phi}(-x + i\epsilon)] \\ &= \frac{\theta}{\rho} \lim_{\epsilon \rightarrow 0} \frac{1}{2\pi i} [\bar{\phi}(-x + i\epsilon) - \bar{\phi}(-x - i\epsilon)] \\ &= \frac{\theta}{\rho\pi} \sqrt{(1-\tau_1x)(\tau_2x-1)}. \quad \blacksquare \end{aligned} \quad (3.14)$$

4. Related quantities via the stationary-excess relations

As in [3–5], let $H_1(t)$ be the time-scaled mean queue length starting at zero, divided by the steady-state limit, and let $c_q(t)$ be the time-scaled (auto) correlation function of the stationary M/M/1 queue-length process. From theorems 1 and 5 of [5], we know that $c_q(t)$ coincides with the time-scaled second moment of the M/M/1 virtual-waiting time process starting at zero, divided by its steady-state limit, denoted by $V_2(t)$. The functions $H_1(t)$ and $V_2(t)$ are bonafide cdf's (cumulative distribution functions) with proper density functions, denoted by

$h_1(t)$ and $\nu_2(t)$. The key stationary-excess relations established in corollary 3.1.3 of [3] and theorem 1 of [5] are

$$h_1(t) = \frac{B^c(t)}{\theta} = \theta^{-1} \int_t^\infty b(s) ds, \quad t \geq 0, \tag{4.1}$$

$$\nu_2(t) = c'_q(t) = 2H_1^c(t) = 2 \int_t^\infty h_1(s) ds, \quad t \geq 0. \tag{4.2}$$

From (4.1), (4.2) and theorem 3.1 we immediately obtain spectral representations for $h_1(t)$ and $\nu_2(t)$.

THEOREM 4.1

The M/M/1 densities $h_1(t)$ and $\nu_2(t)$ have the representations

$$(a) \quad h_1(t) = \int_{r_1}^{r_2} x e^{-xt} \phi_1(x) dx = \int_{\tau_1}^{\tau_2} y^{-1} e^{-t/y} w_1(y) dy \tag{4.3}$$

where

$$\begin{aligned} \phi_1(x) &= \frac{\phi(x)}{\theta x} = \frac{1}{\pi \rho} \frac{\sqrt{(1 - \tau_1 x)(\tau_2 x - 1)}}{x^2}, \quad r_1 \leq x \leq r_2, \\ w_1(y) &= \frac{y w(y)}{\theta} = \frac{1}{\pi \rho} \frac{\sqrt{(y - \tau_1)(\tau_2 - y)}}{y}, \quad \tau \leq y \leq \tau_2, \end{aligned} \tag{4.4}$$

for r_1, r_2, τ_1, τ_2 in (3.4) and (3.5);

$$(b) \quad \nu_2(t) = \int_{r_1}^{r_2} x e^{-xt} \phi_2(x) dx = \int_{\tau_1}^{\tau_2} y^{-1} e^{-t/y} w_2(y) dy \tag{4.5}$$

where

$$\begin{aligned} \phi_2(x) &= \frac{2\phi_1(x)}{x} = \frac{2\phi(x)}{\theta x^2} = \frac{2}{\pi \rho} \frac{\sqrt{(1 - \tau_1 x)(\tau_2 x - 1)}}{x^3}, \quad r_1 \leq x \leq r_2, \\ w_2(y) &= 2y w_1(y) = \frac{2y^2 w(y)}{\theta} = \frac{2}{\pi \rho} \sqrt{(y - \tau_1)(\tau_2 - y)}, \quad \tau_1 \leq y \leq \tau_2. \end{aligned} \tag{4.6}$$

REMARKS

(4.1) The mixing density $w_2(y)$ in (4.6) is a four-parameter beta density $\beta(\frac{3}{2}, \frac{3}{2}, \tau_1, \tau_2)$ as in (1.4) symmetric around $(1 + \rho)/2$. Thus $w_2(y)$ is a beta density, while $w_1(y) = w_2(y)/2y$ and $w(y) = 2\theta w_2(y)/y^2$.

(4.2) Bailey obtained something close to the simple spectral representation for the mean queue length in (31) of [10]: After time scaling, $G(\lambda, 0, t)$ there is $\rho H_1^c(t)/2\theta$.

(4.3) Additional links can easily be added to the stationary-excess chain. Obviously all integrals and derivatives of $b(t)$ have simple spectral representations obtained directly from (3.1), (3.2) and theorem 3.1. In fact, the associated

process considered in section 10 adds a meaningful fourth link to $V_2(t)$, $H_1(t)$ and $B(t)$; see (10.11). ■

We immediately obtain the corresponding RBM spectral representations from theorem 4.1 by letting $\rho \rightarrow 1$. The scaling in (2.1) makes the limit canonical RBM with drift coefficient -1 and diffusion coefficient 1. Note that $\tau_1 \rightarrow 0$ and $\tau_2 \rightarrow 2$ as $\rho \rightarrow 1$ in (3.5). Convergence to proper limits as $\rho \rightarrow 1$ can be rigorously justified in several ways; it suffices to apply corollary 5.2.2 (a) of [4]. Direct derivations from [1] are also not difficult.

THEOREM 4.2

The RBM densities $h_1(t)$ and $\nu_2(t)$ can be represented as in (4.3) and (4.5) with

$$\begin{aligned}\phi_1(x) &= \frac{\sqrt{2x-1}}{\pi x^2}, \quad \frac{1}{2} \leq x < \infty, \\ w_1(y) &= \frac{\sqrt{2-y}}{\pi\sqrt{y}}, \quad 0 \leq y \leq 2, \quad (\beta(\frac{1}{2}, \frac{3}{2})) \\ \phi_2(x) &= \frac{2\sqrt{2x-1}}{\pi x^3}, \quad \frac{1}{2} \leq x < \infty \\ w_2(y) &= \frac{2}{\pi} \sqrt{y(2-y)}, \quad 0 \leq y \leq 2, \quad (\beta(\frac{3}{2}, \frac{3}{2})).\end{aligned}$$

REMARKS

(4.4) The representation for $\nu_2(t)$ in (4.5) for RBM via the mixing density $w_2(y)$ was obtained previously by Ott [28] and Woodside et al. [35]: Since the variance of equilibrium RBM is $1/4$, the RBM covariance function $C(t)$ is represented by

$$\begin{aligned}C(t) &= (1/4) \int_t^\infty \nu_2(s) ds = (1/4) \int_0^2 e^{-t/y} \frac{2}{\pi} \sqrt{y(2-y)} dy \\ &= (2/\pi) \int_0^1 e^{-t/2y} \sqrt{y(1-y)} dy, \quad t \geq 0,\end{aligned}\tag{4.7}$$

with the last step due to the change of variables $z = y/2$.

(4.5) The gamma density with mean and shape parameter $1/2$,

$$\gamma(t) = (2\pi t)^{-1/2} e^{-t/2}, \quad t \geq 0,\tag{4.8}$$

and its stationary-excess density $\gamma_e(t)$, which are fundamental building blocks for RBM, appearing in (4.2) and (4.4) of [1] and theorem 9.1 (e) of [4], also can be represented as mixtures of exponentials with beta mixing densities. For $\gamma(t)$ and $\gamma_e(t)$ the mixing densities are $1/\pi\sqrt{y(2-y)}$ which is $\beta(1/2, 1/2)$ and $\sqrt{y}/\pi\sqrt{2-y}$ which is $\beta(3/2, 1/2)$, respectively.

5. Spectral representation for convolutions of busy-period densities

Let $f(t; n, 0)$ represent the density of the first-passage time from n to 0 in the time-scaled M/M/1 queue, i.e., the n -fold convolution of the busy-period density $b(t)$. Since convolution does not preserve representation as a probabilistic mixture of exponentials [20], we should not expect to get such a representation for $f(t; n, 0)$ and indeed we do not.

We first establish a recursion for $f(t; n, 0)$.

THEOREM 5.1

The first-passage-time densities satisfy

$$\begin{aligned} \rho f(t; n+1, 0) &= (1 + \rho)f(t; n, 0) + 2\theta^2 f'(t; n, 0) - f(t, n-1, 0), \\ n &\geq 1, \end{aligned} \tag{5.1}$$

where $f(t, 0, 0) \equiv 0$.

Proof

The recursion (5.1) follows from the basic law of motion. From the basic quadratic equation after (2.3),

$$\rho z_1^2 - (1 + \rho + 2\theta^2 s)z_1 + 1 = 0 \tag{5.2}$$

or

$$\rho z_1^{n+1} = (1 + \rho + 2\theta^2 s)z_1^n - z_1^{n-1}, \quad n \geq 1, \tag{5.3}$$

with $z^0 = 1$. Since z_1^n is the Laplace transform of $f(t; n, 0)$ by (2.2) and sz_1^n is the Laplace transform of $f'(t; n, 0)$, the result follows. ■

We now obtain the spectral representation for $f(t; n, 0)$; this constitutes a new proof of (6.6) on p. 103 of Karlin and McGregor [18]. The representation is expressed in terms of the *Chebysheff polynomials of the second kind*, denoted by $U_n(\alpha)$; see p. 256–259 of Magnus et al. [24]; e.g., $U_{-1}(\alpha) = 0$, $U_0(\alpha) = 1$, $U_1(\alpha) = 2\alpha$, $U_2(\alpha) = 4\alpha^2 - 1$, $U_3(\alpha) = 8\alpha^3 - 4\alpha$ and

$$U_{n+1}(\alpha) = 2\alpha U_n(\alpha) - U_{n-1}(\alpha), \quad n \geq 1. \tag{5.4}$$

THEOREM 5.2

The first-passage-time density can be represented as

$$f(t; n, 0) = \int_{\tau_2^{-1}}^{\tau_1^{-1}} \phi(x, n) x e^{-xt} dx, \quad t \geq 0, \tag{5.5}$$

where τ_1 and τ_2 are given in (3.5) and

$$\phi(x, n+1) = \rho^{-n/2} U_n(\alpha(x)) \phi(x) \tag{5.6}$$

with $\phi(x)$ the spectral density in (3.6),

$$\alpha(x) = \frac{1 + \rho - 2\theta^2 x}{2\sqrt{\rho}} \quad (5.7)$$

and $U_n(\alpha)$, $n \geq 1$, are the Chebycheff polynomials of the second kind in (5.4).

Proof

Assume (5.5) and apply the recursion in theorem 5.1 to get

$$\rho\phi(x, n+1) = (1 + \rho - 2\theta^2 x)\phi(x, n) - \phi(x, n-1), \quad n \geq 1, \quad (5.8)$$

with $\phi(x, 1) = \phi(x)$. Solving the second-order difference equation (5.8) for fixed x gives (5.6), as can be checked by substitution, using (5.4). ■

Note that

$$\phi(x, 2) = \frac{(1 + \rho - 2\theta^2 x)}{\rho} \phi(x) \quad (5.9)$$

which is negative for $x > (1 + \rho)/2\theta^2$, so indeed negative weights occur when $n \geq 1$.

We now indicate how to apply theorem 5.2 to obtain the spectral representation for $h_1(t)$ in theorem 4.1(a). By corollary 3.1.3 of [3],

$$h_1(t) = \sum_{n=1}^{\infty} \frac{(1 - \rho)\rho^n}{\rho} f(t; n, 0), \quad (5.10)$$

which implies that

$$\phi_1(x) = \frac{\phi(x)}{\theta x} = \sum_{n=1}^{\infty} 2\theta\rho^{n-1}\phi(x, n). \quad (5.11)$$

Given theorems 4.1 and 5.2, (5.11) is equivalent to the following relation among the Chebycheff polynomials of the second kind, which we establish directly.

THEOREM 5.3

For the Chebycheff polynomials $U_n(\alpha)$ in (5.4) with $\alpha(x)$ in (5.7),

$$\sum_{n=1}^{\infty} \rho^{(n-1)/2} U_{n-1}(\alpha(x)) = \frac{1}{2\theta^2 x}. \quad (5.12)$$

Proof

Apply the generating functions for $U_n(\alpha)$, p. 259 of Magnus et al. [24]. Since

$$\sum_{n=0}^{\infty} U_n(\alpha) z^n = 1/(1 - 2\alpha z + z^2),$$

$$\sum_{n=1}^{\infty} \rho^{(n-1)/2} U_{n-1}(\alpha(x)) = \frac{1}{1 - 2\alpha(x)\sqrt{\rho} + \rho} = \frac{1}{2\theta^2 x}. \quad \blacksquare$$

6. The spectral representations for $P_{0n}(t)$ and $P_{n0}(t)$

In [4] we saw that the M/M/1 transition functions $P_{0n}(t)$ and $P_{n0}(t)$ have nice structure not present for the general case $P_{in}(t)$. In this section we relate $P_{0n}(t)$ to Karlin and McGregor [18] and use it to obtain a direct evaluation of the mean $m_1(t, 0) \equiv \sum_{n=1}^{\infty} n P_{0n}(t)$, and thus an alternate proof of theorem 4.1(a). (Recall that $H_1(t) = m_1(t, 0)/m_1(\infty, 0)$.)

Since $P_{ij}(t) = \rho^{j-i} P_{ji}(t)$ by reversibility as in (4.3) of [4], results for $P_{0n}(t)$ also translate immediately into results for $P_{n0}(t)$. From corollary 4.2.2 of [4],

$$P_{0n}(t) = (1 - \rho)\rho^n - \rho^n F^c(t; n, 0) + \rho^{n+1} F^c(t; n + 1, 0) \tag{6.1}$$

where $F^c(t; n, 0)$ is the complementary cdf associated with $f(t; n, 0)$.

REMARK 6.1

It is significant that (6.1) is a minor modification of what can be obtained in general from the *dual process*; see chapter 3 of van Doorn [33]. The dual birth-and-death process, denoted by an asterisk, is constructed from the original by defining new birth and death rates λ_n^* and μ_n^* in terms of the old ones by $\lambda_n^* = \mu_{n+1}$ and $\mu_n^* = \lambda_n$. Since $\lambda_0 > 0$, $\mu_0^* > 0$ so that the dual process has an absorbing state at -1 . In general, the original process and dual are related by

$$P_{0n}(t) = F^*(t; n - 1, -1) - F^*(t; n - 1). \tag{6.2}$$

The special M/M/1 structure (homogeneity) then enables us to convert (6.2) into (6.1). First, by symmetry, obviously

$$F^*(t; n, -1) = F_u(t; 0, n + 1) \tag{6.3}$$

where $F_u(t; 0, n)$ is the first-passage-time cdf for the extension of the original M/M/1 process to all integers by removing the barrier at 0. Next, by reversibility and homogeneity,

$$F_u(t; 0, n) = \rho^n F(t; n, 0), \tag{6.4}$$

as shown in the proof of theorem 1.4 in [1]. ■

From (6.1) and theorem 5.2, we immediately obtain a spectral representation for $P_{0n}(t)$.

THEOREM 6.1

$$P_{0n}(t) = (1 - \rho)\rho^n - \int_{\tau_2^{-1}}^{\tau_1^{-1}} \psi(x, n) x e^{-xt} dx, \text{ where } \tau_1, \tau_2 \text{ are given in (3.5),}$$

$$\psi(x, n) = [\phi(x, n) - \phi(x, n + 1)]/x$$

and $\phi(x, n)$ is given in (5.6).

It is interesting to relate theorem 6.1 to Karlin and McGregor [18]. From (6.1) on p. 101 and 5.4 on p. 98 of [18] or (6.4.2) on p. 65 and theorem 6.2.5 on p. 54 of

[33], we obtain

$$P_{in}(t) = (1 - \rho)\rho^n + \int_{\tau_2^{-1}}^{\tau_1^{-1}} \rho^{n+1} q_i(x) q_n(x) \phi(x) e^{-xt} dx \quad (6.5)$$

where $q_n(x)$ are the orthogonal polynomials for M/M/1, so that

$$-q_n(x)\phi(x) = \frac{1}{\rho}\phi(x, n) - \phi(x, n+1) \quad (6.6)$$

and

$$q_n(x) = \rho^{-n/2} U_n(\alpha(x)) - \rho^{-(n+1)/2} U_{n-1}(\alpha(x)) \quad (6.7)$$

for $\alpha(x)$ in (5.7), which is equivalent to (5.5) on p. 99 of [18], but in somewhat cleaner form.

REMARK 6.2

Formulas (6.5) and (6.7) may not seem to be convenient for numerical calculation because $\alpha(x)$ in (5.7) depends on x . At first glance, it might seem that we have to apply (6.7) for each of the uncountably many x in the interval $[\tau_2^{-1}, \tau_1^{-1}]$, but this is not necessary, because we can represent $U_n(\alpha)$ as a polynomial in α of the form $c_{n0} + c_{n1}\alpha + \dots + c_{nn}\alpha^n$ and use (5.4) to obtain a finite recursion for the coefficients, namely, $c_{00} = 1$, $c_{10} = 0$, $c_{11} = 2$, $c_{20} = -1$, $c_{21} = 0$, $c_{22} = 4$ and $c_{n+1,j} = 2c_{n,j-1} - c_{n-1,j}$ for $0 \leq j \leq n+1$ and $n \geq 3$. Alternatively, for any birth-and-death process we can use the basic recursion for the polynomials $q_i(x)$ to obtain a recursion for their coefficients. Indeed, for the M/M/1 queue in our time scaling, it is easy to see that

$$q_n(-x) = 1 + \rho^{-n} \sum_{k=1}^n a_{nk} (2\theta^2 x)^k \quad (6.8)$$

where $a_{nn} = 1$, $a_{n1} = \sum_{k=1}^n k\rho^{k-1}$ and

$$a_{n+1,j} = (1 + \rho)a_{nj} + a_{n,j-1} - \rho^{-1}a_{n-1,j}. \quad \blacksquare \quad (6.9)$$

We now consider a direct evaluation of the mean starting at zero. Given theorem 4.1(a), what we want to establish is

$$\begin{aligned} m_1(t, 0) &\equiv \sum_{n=1}^{\infty} nP_{0n}(t) \\ &= \frac{\rho}{1-\rho} H_1(t) = \frac{\rho}{1-\rho} - \int_{\tau_2^{-1}}^{\tau_1^{-1}} \frac{\rho}{1-\rho} \frac{\phi(x)}{\theta(x)} e^{-xt} dx. \end{aligned} \quad (6.10)$$

Given (6.5), it suffices to establish the following relation for the orthogonal polynomials

$$\sum_{n=1}^{\infty} n\rho^{n+1} q_n(x) = \frac{-\rho}{2\theta^2 x}. \quad (6.11)$$

We establish (6.11) directly via the following result.

THEOREM 6.2

$$\sum_{n=1}^{\infty} n \rho^n q_n(x) = -1/2\theta^2 x.$$

Proof

Apply (6.7) and theorem (5.3) to get

$$\begin{aligned} \sum_{n=1}^{\infty} n \rho^n q_n(x) &= \sum_{n=1}^{\infty} n \rho^n [\rho^{-n/2} U_n(\alpha) - \rho^{-(n+1)/2} U_{n-1}(\alpha)] \\ &= -\sum_{n=0}^{\infty} \rho^{n/2} U_n(\alpha) = -1/2\theta^2 x. \quad \blacksquare \end{aligned}$$

7. More moment spectral representations

We now show that the simplification provided by theorem 6.2 extends to all higher moments. As in [3], it is convenient to work with factorial moments. For this purpose, let $n_{(r)} = n(n-1) \cdots (n-r+1)$. The following result is proved just like theorem 6.2.

THEOREM 7.1

$$\begin{aligned} \sum_{n=r}^{\infty} n_{(r)} \rho^n q_n(x) &= -r \sum_{n=r-1}^{\infty} n_{(r-1)} \rho^{n/2} U_n(\alpha) \\ &= -r \rho^{(r-1)/2} \frac{d^{r-1}}{dz^{r-1}} (1 - 2\alpha z + z^2)^{-2} \end{aligned}$$

for $\alpha \equiv \alpha(x)$ in (5.7) and $z = \sqrt{\rho}$. For $r = 2$,

$$\sum_{n=2}^{\infty} n(n-1) \rho^n q_n(x) = \frac{4\sqrt{\rho}(\alpha - z)}{(1 - 2\alpha z + z^2)^2} = \frac{-1 + \theta x}{\theta^3 x^2}. \tag{7.1}$$

Let $H_2(t)$ be the second moment starting at zero divided by the limit as $t \rightarrow \infty$. We can apply theorem 7.1 to obtain the following representations in terms of the moment cdf's $H_1(t)$ and $V_2(t)$ treated in section 4. (These are equivalent to (3.14) of [3].)

COROLLARY 7.1

$$\begin{aligned} \sum_{n=1}^{\infty} n(n-1) P_{0n}(t) &= \frac{2\rho}{(1-\rho)^2} V_2(t) - \frac{2\rho}{1-\rho} H_1(t) \\ m_2(t, 0) \equiv \sum_{n=1}^{\infty} n^2 P_{0n}(t) &= \frac{2\rho}{(1-\rho)^2} V_2(t) - \frac{\rho}{1-\rho} H_1(t), \end{aligned}$$

and

$$H_2(t) = \frac{2}{1+\rho} V_2(t) - \frac{1-\rho}{1+\rho} H_1(t).$$

From (6.3) and theorems 6.2 and 7.1, we also obtain finite integral representations for the first two moments starting in any state i . For computational purposes, we can perform the integration either directly after determining the polynomial $q_i(x)$ by a finite recursion on the coefficients as indicated in remark 6.1 above or indirectly after making a change of variables to obtain

COROLLARY 7.2

$$\begin{aligned} m_1(t, i) &\equiv \sum_{n=1}^{\infty} n P_{in}(t) = \frac{\rho}{1-\rho} - \rho \int_{\tau_2^{-1}}^{\tau_1^{-1}} q_i(x) \frac{\phi(x) e^{-xt}}{2\theta^2 x} dx \\ &= \frac{\rho}{1-\rho} \left[1 - \int_{\tau_2^{-1}}^{\tau_1^{-1}} q_i(x) \phi_1(x) e^{-xt} dx \right] \end{aligned}$$

and

$$\begin{aligned} m_2(t, i) &\equiv \sum_{n=1}^{\infty} n^2 P_{in}(t) \\ &= \frac{\rho(1+\rho)}{(1-\rho)^2} - \rho \int_{\tau_2^{-1}}^{\tau_1^{-1}} q_i(x) \left[\frac{1}{\theta^3 x^2} - \frac{1}{2\theta^2 x} \right] \phi(x) e^{-xt} dx \\ &= \frac{\rho(1+\rho)}{(1-\rho)^2} \left(1 - \int_{\tau_2^{-1}}^{\tau_1^{-1}} q_i(x) \left[\frac{2\phi_2(x) + (1-\rho)\phi_1(x)}{1+\rho} \right] e^{-xt} dx \right). \end{aligned}$$

For the first moment starting in i , we need not apply the representation for $P_{in}(t)$ in (6.3). Instead we can apply the decomposition from section 6 of [4], namely,

$$m_1(t, i) = m_1(t, 0) + iG_1^c(t, i) \quad (7.2)$$

where

$$iG_1^c(t, i) = \sum_{n=1}^i F^c(t; n, 0) \quad (7.3)$$

with $F^c(t; n, 0)$ being the complementary cdf associated with the density $f(t; n, 0)$ in section 5. From theorem 5.2,

$$F^c(t; n, 0) = \int_{\tau_2^{-1}}^{\tau_1^{-1}} \rho^{-(n-1)/2} U_{n-1}(\alpha(x)) \phi(x) e^{-xt} dx. \quad (7.4)$$

On the other hand, from corollary 7.2,

$$iG_1^c(t, i) = \rho \int_{\tau_2^{-1}}^{\tau_1^{-1}} [1 - q_i(x)] \frac{\phi(x)}{2\theta^2 x} e^{-xt} dx. \quad (7.5)$$

We connect (7.4) and (7.5) by the following relation, which we again establish directly.

THEOREM 7.2

$$1 - q_i(x) = 2\theta^2 x \sum_{n=1}^i \rho^{-(n+1)/2} U_{n-1}(\alpha(x)).$$

Proof

Use (5.7), (5.4) and then (6.5) to express the right side as

$$\begin{aligned} & \left(1 + \frac{1}{\rho} - \frac{2\alpha}{\sqrt{\rho}}\right) \sum_{n=1}^i \frac{U_{n-1}(\alpha(x))}{\rho^{(n-1)/2}} \\ &= \sum_{n=1}^i \left(\frac{U_{n-1}(\alpha)}{\rho^{(n-1)/2}} + \frac{U_{n-1}(\alpha)}{\rho^{(n+1)/2}} - \frac{U_n(\alpha)}{\rho^{n/2}} - \frac{U_{n-2}(\alpha)}{\rho^{n/2}} \right) \\ &= 1 - \frac{U_i(\alpha)}{\rho^{i/2}} + \frac{U_{i-1}(\alpha)}{\rho^{(i+1)/2}} = 1 - q_i(x). \quad \blacksquare \end{aligned}$$

We conclude this section by noting that integral representations for the moments $m_k(t, i)$ can also be obtained from section 8 of [4]. There it is shown that $m_k(t, i)$ can be expressed as a polynomial in terms of $P_{i0}(t)$ and its first $(k - 1)$ derivatives.

8. The probability transition function and the first passage times up

In theorem 6.1 we directly obtained the spectral representation for $P_{in}(t)$ in (6.5) for the special case $i = 0$. We now show how we can treat the general case. We start with an elementary relation among Laplace transforms given in theorem 4.3 of [4], namely,

$$\hat{P}_{in}(s) = \hat{P}_{0n}(s) / \hat{f}(s; 0, i) \quad \text{for } i < n; \tag{8.1}$$

i.e., to be in state n starting at 0, you must pass through state i for a first time.

The key connection is provided by the following expression for the Laplace transform of the first passage times up in terms of the orthogonal polynomials.

THEOREM 8.1

$$\hat{f}(s; 0, i) = 1/q_i(-s).$$

Proof

By theorem 3.4 of [4],

$$\hat{f}(s; 0, n) = \frac{r_1 + r_2}{r_1 z_1^n + r_2 z_2^n} \tag{8.2}$$

for r_i and z_i in (2.3). Thus,

$$\frac{1}{\hat{f}(s; 0, n)} = A(s)z_1^n + B(s)z_2^n \quad (8.3)$$

where $A(s)$ and $B(s)$ are independent of n . Since z_1 and z_2 are roots of the basic quadratic equation (5.2), their powers satisfy the basic recursion (5.3), which can be expressed as

$$z^n = \frac{2\alpha(-s)}{\sqrt{\rho}} z^{n-1} - \frac{z^{n-2}}{\rho}, \quad n \geq 2, \quad (8.4)$$

for $z^0 = 1$ and $\alpha(x)$ in (5.7). From (5.4), we see that the Chebychev polynomials divided by $\rho^{n/2}$ satisfy the same recursion, i.e.,

$$\frac{U_{n+1}(\alpha)}{\rho^{(n+1)/2}} = \frac{2\alpha}{\sqrt{\rho}} \frac{U_n(\alpha)}{\rho^{n/2}} - \frac{1}{\rho} \frac{U_{n-1}(\alpha)}{\rho^{(n-1)/2}}, \quad n \geq 0. \quad (8.5)$$

However, since

$$\begin{aligned} \frac{1}{\hat{f}(s; 0, 1)} &= \frac{\rho + 2\theta^2 s}{\rho} = \frac{2\alpha(-s)}{\sqrt{\rho}} - \frac{1}{\rho} = \frac{U_1(\alpha)}{\rho^{1/2}} - \frac{1}{\rho} \frac{U_0(\alpha)}{\rho^{0/2}}, \\ \frac{1}{\hat{f}(s; 0, n)} &= \frac{U_n(\alpha)}{\rho^{n/2}} - \frac{1}{\rho} \frac{U_{n-1}(\alpha)}{\rho^{(n-1)/2}} = q_n(-s). \quad \blacksquare \end{aligned}$$

Reasoning as in (8.1), we have

$$\hat{f}(s; 0, n) = \hat{f}(s; 0, i) \hat{f}(s; i, n) \quad \text{for } 1 \leq i \leq n-1, \quad (8.6)$$

so that we can deduce the following additional consequence.

COROLLARY 8.1

For $1 \leq i \leq n-1$, $\hat{f}(s; i, n) = q_i(-s)/q_n(-s)$.

We now apply theorem 8.1 to give a new proof of the M/M/1 spectral representation (6.5) for the case $i \neq 0$. Combining (8.1) and theorem 8.1, we obtain

$$\hat{P}_{in}(s) = q_i(-s) \hat{P}_{0n}(s) \quad \text{for } i < n. \quad (8.7)$$

Since $q_i(-s)$ is a polynomial of the form $a_0 + a_1 s + a_2 s^2 + \dots + a_k s^k$ and multiplication by s^k with transforms is tantamount to taking the k^{th} derivative, we can write

$$P_{in}(t) = q_i(-D) P_{0n}(t) \quad (8.8)$$

where $D \equiv d/dt$ is the differential operator. (Necessarily $a_0 = 1$ and $a_k = 0$ for $k > i$, as can be proved by induction.) Since $q_i(-D)e^{-xt} = q_i(x)e^{-xt}$, we obtain the desired (6.5).

We treat $i > n$ by reversibility: $P_{ni}(t) = \rho^{i-n} P_{in}(t)$. For $i = n$, we use the fact that $\sum_{n=0}^{\infty} \rho^{n+1} q_n(x) = 1$ (reasoning as in theorem 6.2) to obtain

$$\begin{aligned} P_{ii}(t) &= 1 - \sum_{\substack{n=0 \\ n \neq i}}^{\infty} P_{in}(t) = \int q_i(x) \left[\sum_{\substack{n=0 \\ n \neq i}}^{\infty} \rho^{n+1} q_n(x) \right] \phi(x) e^{-xt} dx \\ &= 1 - \int q_i(x) [1 - \rho^{i+1} q_i(x)] \phi(x) e^{-xt} dx \\ &= \int \rho^{i+1} q_i(x)^2 \phi(x) e^{-xt} dx. \end{aligned} \tag{8.9}$$

REMARK 8.1

We have just used theorem 8.1 to establish the spectral representation (1.1). If we take the spectral representation (1.1) as having been established for general birth-and-death processes, then we can reverse the argument to deduce theorem 8.1 and (8.7)–(8.9) for general birth-and-death processes. In fact, this was done by Karlin and McGregor; see p. 378 of [19]. Theorem 8.1 can also be shown to be a special case of a result by Brown and Shao [11]. They show that the distribution of the first passage time from 0 to n can be expressed solely in terms of the eigenvalues of the $n \times n$ submatrix of the infinitesimal generator matrix associated with states 0, 1, ..., $n - 1$ (without a separate calculation of the eigenvalues).



9. Trigonometric integral representations

In this section we relate the spectral and Bessel function representations for the busy period distribution in sections 2 and 3 to trigonometric integral representations. The first M/M/1 trigonometric integral representations were obtained in 1954 by Ledermann and Reuter [23] for the busy-period density $b(t)$ and associated quantities; see (4.10) on p. 365. At the same time, Vaultot [34] independently obtained a trigonometric integral representation for the complementary waiting-time cdf in the M/M/1 queue with the last-come first-served (LCFS) service discipline. In 1961 Riordan [29] observed that this M/M/1-LCFS waiting-time complementary cdf coincides with the M/M/1 busy-period complementary cdf; also see pp. 64, 108 and 114 of [30]. In our time scale, the representation is

$$B^c(t) = \frac{1}{\pi} \int_0^{\pi} \frac{(\sin x)^2}{\gamma(x)} e^{-\gamma(x)t/\theta^2} dx \tag{9.1}$$

where

$$\gamma(x) = (1 + \rho - 2\sqrt{\rho} \cos x)/2. \tag{9.2}$$

It should be apparent that, just like (3.1) and (3.2), (9.1) is convenient for obtaining numerical results by numerical integration. (Indeed a simple trapezoidal rule with 200 points produces an absolute error of order 10^{-7} for $\rho \leq 0.85$. For $\rho > 0.85$ you need more points near the mode.)

From section 4, it is evident that the mean queue length and the correlation function also have trigonometric integral representations. Indeed trigonometric integral representations for the probability transition functions $P_{ij}(t)$ in (1.1) and the correlation function were discovered in 1955 by Morse [26]. Trigonometric integral representations for (1.1) and the mean queue length are given in chapter 1 of Takács [32]. Ledermann and Reuter [23] and Takács [32] obtain their trigonometric representations by considering the truncated M/M/1 model with a finite waiting room. It is significant that the associated integral representations for the truncated model constitute approximating Riemann sums for the integral representations in the standard M/M/1 model. Thus, to consider an approximation with a finite waiting room, as Stern [31] does, is tantamount to performing a certain numerical integration of the standard M/M/1 integral representations.

To go from the Kendall Bessel-function representation (2.4) to the Ledermann-Reuter-Vaulot trigonometric integral representation (9.1) we can exploit a basic trigonometric property of the Chebychev polynomials

$$U_n(\cos y) = \frac{\sin(n+1)y}{\sin y}; \quad (9.3)$$

see p. 257 of [24] and p. 98 of [18]. Since $\alpha(x)$ in (5.7) appears as an argument of U_n , we let $\cos y = \alpha(x)$, which is equivalent to $x = (1 + \rho - 2\sqrt{\rho} \cos y)/2\theta^2$.

From (2.4), we obtain

$$B^c(t) = \int_t^\infty \frac{1}{\sqrt{\rho} x} e^{-x(1+\rho)/2\theta^2} I_1\left(\frac{\sqrt{\rho} x}{\theta^2}\right) dx \quad (9.4)$$

by applying (3.5) and $\tau_1\tau_2 = \theta^2$. By 9.6.18 in [7],

$$I_1\left(\frac{\sqrt{\rho} x}{\theta^2}\right) = \frac{\sqrt{\rho} x}{\pi\theta^2} \int_0^\pi (\sin z)^2 e^{\sqrt{\rho} x \cos z / \theta^2} dz, \quad (9.5)$$

so that

$$\begin{aligned} B^c(t) &= \frac{1}{\pi} \int_0^\pi (\sin z)^2 \left[\int_t^\infty \exp\left[-\frac{x}{2\theta^2}(1+\rho-\sqrt{\rho} \cos z)\right] \frac{dx}{\theta^2} \right] dz \\ &= \frac{1}{\pi} \int_0^\pi (\sin z)^2 \left[\frac{e^{-t(1+\rho-2\sqrt{\rho} \cos z)/2\theta^2}}{(1+\rho-2\sqrt{\rho} \cos z)/2} \right] dz, \end{aligned} \quad (9.6)$$

which agrees with (9.1).

Now we apply (9.1) to establish theorem 3.1. In (9.1) make the change of variables

$$y = \frac{2\theta^2}{1 + \rho - \sqrt{\rho} \cos x} \tag{9.7}$$

We use the following relation

LEMMA 9.1

$$\sqrt{(y - \tau_1)(\tau_2 - y)} = \frac{2\theta\sqrt{\rho} \sin x}{1 + \rho - 2\sqrt{\rho} \cos x}$$

Proof

$$\begin{aligned} \sqrt{\rho} \sin x &= \sqrt{\rho(1 - \cos^2 x)} = \sqrt{\rho(1 - \cos x)(1 + \cos x)} \\ &= \sqrt{(\sqrt{\rho} - \sqrt{\rho} \cos x)(\sqrt{\rho} + \sqrt{\rho} \cos x)} \\ &= \frac{\sqrt{(2\sqrt{\rho} - (1 + \rho) + 2\theta^2 y^{-1})(2\sqrt{\rho} + 1 + \rho - 2\theta^2 y^{-1})}}{2} \\ &= \frac{\theta}{y} \sqrt{(y - \tau_1)(\tau_2 - y)} \\ &= \left(\frac{1 + \rho - 2\sqrt{\rho} \cos x}{2\theta} \right) \sqrt{(y - \tau_1)(\tau_2 - y)}. \quad \blacksquare \end{aligned}$$

Using (9.7) and lemma 9.1, we obtain

$$\begin{aligned} B^c(t) &= \frac{2}{\rho\pi} \int_0^\pi e^{-\gamma(x)t/\theta^2} \frac{(\sqrt{\rho} \sin x)(\sqrt{\rho} \sin x)}{(1 + \rho - 2\sqrt{\rho} \cos x)} dx \\ &= \frac{1}{\pi\rho} \int_{\tau_1}^{\tau_2} e^{-t/y} \frac{\sqrt{(y - \tau_1)(\tau_2 - y)}}{\theta} \frac{\theta^2}{y^2} dy \end{aligned}$$

which is the integral form of theorem 3.1.

10. The associated process

Karlin and McGregor's [18] treatment of the M/M/1 queue and the multi-server M/M/s extension is largely based on an *associated* absorbing birth-and-death process with state zero removed. The fact that the associated process is easier to analyze is closely related to our finding it easier to start with the busy

period density in section 3 and its n -fold convolution in section 5. However, the value of the associated process seems pretty much limited to the M/M/1 and M/M/2 special cases, as we will show.

To understand the role of the associated process, recall that the spectral representation (1.1) has two parts: *existence* and *construction*. For general birth-and-death processes (subject to appropriate regularity conditions), we know that a spectral representation of the form (1.1) exists. The difficulty is the spectral measure Φ . In general, we know that a spectral measure Φ exists, but it remains to identify it.

However, if the orthogonal polynomials $q_i(x)$ in (1.1) turn out to be well-studied classical polynomials, then the spectral measure may have already been determined. For example, for the Chebychev polynomials of the second kind $U_i(y)$, from p. 258 of Magnus et al. [24], we know that

$$\frac{2}{\pi} \int_{-1}^1 U_i(y) U_j(y) \sqrt{1-y^2} \, dy = \delta_{ij}. \quad (10.1)$$

An important observation made by Karlin and McGregor is that in general the spectral measure can be identified through the transition probability $P_{00}(t)$. Since $\pi_0 = q_0(x) = 1$, $P_{00}(t)$ in (1.1) is the Laplace-Stieltjes transform of the spectral measure. Moreover, the Laplace transform $\hat{P}_{00}(s)$ of $P_{00}(t)$ is the Stieltjes transform of Φ , say $\tilde{\Phi}(s)$, i.e.,

$$\hat{P}_{00}(s) = \int_0^\infty e^{-st} P_{00}(t) \, dt = \int_0^\infty \frac{d\Phi(x)}{x+s} \equiv \tilde{\Phi}(s). \quad (10.2)$$

For the M/M/1 queue, the associated process turns out to be ideally suited to identify $P_{00}(t)$ and thus the spectral measure Φ . Let ${}^0P_{ij}(t)$ be the probability of being in state j at time t without ever visiting state 0, starting in state i . The associated process is the birth-and-death process on the positive integers with probability transition function ${}^0P_{ij}(t)$. As on p. 90 of [18], its generator matrix is obtained from the generator matrix of the original process by deleting the first row and column. For the M/M/1 queue, $\mu_0 = 0$ so that all diagonal elements of the generator matrix are identical except the first. When we go to the associated process, this asymmetry is removed: all diagonal elements in the associated generator matrix are identical. This additional symmetry is the key to the nice orthogonal polynomials for the associated process. Thus, we can easily identify ${}^0P_{11}(t)$ and its spectral measure, say ${}^0\Phi$. In particular, in this case the orthogonal polynomials, say ${}^0q_i(x)$, are expressed directly in terms of the Chebychev polynomials of the second kind by

$${}^0q_i(x) = \frac{\phi(x, i)}{\phi(x)} = \rho^{-(i-1)/2} U_{i-1}(\alpha(x)), \quad i \geq 1, \quad (10.3)$$

for $\phi(x)$ in (3.6) and $\phi(x, i)$, $U_i(\alpha)$ and $\alpha(x)$ in section 5. (As in [18], we obtain (10.3) by relating the recursion for ${}^0q_i(x)$ to the recursion for $U_i(\alpha)$ in (5.4).) By

making the change of variables $y = \alpha(x)$ for $\alpha(x)$ in (5.7), we get the spectral measure ${}^0\Phi$ and the full spectral representation (1.1) for ${}^0P_{ij}(t)$ from (10.1). In particular, with our time scaling,

$${}^0\phi(x) \equiv {}^0\Phi'(x) = \frac{\theta^3}{\pi\rho} \sqrt{(\tau_2 x - 1)(1 - \tau_1 x)} = 2\theta^2 x \phi(x) \tag{10.4}$$

and

$${}^0P_{ij}(t) = \rho^{j-1} \int_{\tau_2^{-1}}^{\tau_1^{-1}} e^{-x\rho} q_i(x) {}^0q_j(x) {}^0\phi(x) dx. \tag{10.5}$$

For general birth-and-death processes, we express $P_{00}(t)$ in terms of ${}^0P_{11}(t)$ by decomposing the event of transition from 0 to 0 into the cases of no transition and transition upward plus a return for the first time at time u , $0 \leq u \leq t$. Hence, for the Laplace transforms we have the relation

$$\hat{P}_{00}(s) = \frac{1}{\lambda_0 + \mu_0 + s} + \frac{\lambda_0}{\lambda_0 + \mu_0 + s} {}^0\hat{P}_{11}(s) \mu_1 \hat{P}_{00}(s) \tag{10.6}$$

or

$$\begin{aligned} \tilde{\Phi}(s) \equiv \int_0^\infty \frac{d\Phi(x)}{s+x} &= \hat{P}_{00}(s) = \frac{1}{\lambda_0 + \mu_0 + s - \lambda_0 \mu_1 {}^0\hat{P}_{11}(s)} \\ &= \frac{1}{\lambda_0 + \mu_0 + s - \lambda_0 \mu_1 {}^0\tilde{\Phi}(s)}. \end{aligned} \tag{10.7}$$

Equation (10.7) expresses the Stieltjes transform $\tilde{\Phi}(s)$ of the spectral measure Φ in terms of the Stieltjes transform ${}^0\tilde{\Phi}(s)$ of the associated spectral measure ${}^0\Phi$. In general, this relation seems to be of little use because in general the associated process is just as complicated as the original process. However, for M/M/1 we can solve the associated process. In particular, since the associated process of the associated process is again the associated process, we obtain a quadratic equation for ${}^0\tilde{\Phi}(s)$; i.e.,

$${}^0\tilde{\Phi}(s) = \frac{1}{\lambda + s - \lambda \mu {}^0\tilde{\Phi}(s)}, \tag{10.8}$$

see p. 96 of [18], so that

$${}^0\tilde{\Phi}(s) = \frac{\lambda + \mu + s - \sqrt{(\lambda + \mu + s)^2 - 4\lambda\mu}}{2\lambda\mu} \tag{10.9}$$

and

$$\tilde{\Phi}(s) = \frac{2(\mu - \lambda - s) + 2\sqrt{(\lambda + \mu + s)^2 - 4\lambda\mu}}{-4\mu s}; \tag{10.10}$$

see (5.7) and (5.3) of [18].

It turns out that it is even easier to go from the associated process directly to the first passage times down. It is easy to see that in general the first-passage-time density $f(t; i, 0)$ of the original process can be expressed as

$$f(t; i, 0) = {}^0P_{i1}(t)\mu_1, \quad (10.11)$$

so that for the M/M/1 model with our scaling

$$\begin{aligned} f(t; i, 0) &= \frac{1}{2\theta^2} {}^0P_{i1}(t) = \frac{1}{2\theta^2} \int_{\tau_2^{-1}}^{\tau_1^{-1}} e^{-xt} {}^0q_i(x) \phi(x) dx \\ &= \frac{1}{2\theta^2} \int_{\tau_2^{-1}}^{\tau_1^{-1}} e^{-xt} \frac{\phi(x, i)}{\phi(x)} 2\theta^2 x \phi(x) dx = \int_{\tau_2^{-1}}^{\tau_1^{-1}} x e^{-xt} \phi(x, i) dx \end{aligned} \quad (10.12)$$

as given in theorem 5.2. Since the associated process is easier to analyze and is directly related to the first passage times down, we should thus expect that the first passage times down are easier to analyze than the probability transition function $P_{ij}(t)$, as we have found to be the case. Also note that by virtue of (10.12) ${}^0P_{11}(t)/2\theta^2$ is a fourth link in the stationary-excess chain; see Remark 4.3: Paralleling (4.1), ${}^0P_{11}(t)/2\theta^2$ is a complementary cdf with $b(t) = {}^0P_{11}(t)/2\theta^2$, $t \geq 0$.

Acknowledgment

We thank Erik van Doorn for pointing out that theorem 8.1 was already deduced from (1.1) by Karlin and McGregor [19], as indicated in remark 8.1.

References

- [1] J. Abate and W. Whitt, Transient behavior of regulated Brownian motion, I: starting at the origin, *Adv. Appl. Prob.* 19 (1987) 560–598.
- [2] J. Abate and W. Whitt, Transient behavior of regulated Brownian motion, II: non-zero initial conditions, *Adv. Appl. Prob.* 19 (1987) 599–631.
- [3] J. Abate and W. Whitt, Transient behavior of the M/M/1 queue starting at the origin, *Queueing Systems* 2 (1987) 41–65.
- [4] J. Abate and W. Whitt, Transient behavior of the M/M/1 queue via Laplace transforms, *Adv. Appl. Prob.* 20 (1988) 145–178.
- [5] J. Abate and W. Whitt, The correlation functions of RBM and M/M/1, *Stochastic Models* 4, (1988) to appear.
- [6] J. Abate and W. Whitt, Approximations for the M/M/1 busy-period distribution, *Queueing Theory and its Applications—Liber Amicorum for J.W. Cohen* (North-Holland, Amsterdam, 1988).
- [7] M. Abramowitz and I.A. Stegun (eds.), *Handbook of Mathematical Functions* (Dover, New York, 1972).
- [8] M.H. Ackroyd, M/M/1 transient state occupancy probabilities via the discrete Fourier transform, *IEEE Trans. Commun.* COM-30 (1982) 557–559.

- [9] N.T.J. Bailey, A continuous-time treatment of a simple queue using generating functions, *J. Roy. Stat. Soc., Ser. B*, 16 (1954) 288–291.
- [10] N.T.J. Bailey, Some further results in the non-equilibrium theory of a simple queue, *J. Roy. Stat. Soc., Ser. B*, 19 (1957) 326–333.
- [11] M. Brown and Y. Shao, Identifying coefficients in the spectral representation for first passage time distributions, *Prob. in Engr. and Info. Sciences* 1 (1987) 69–74.
- [12] P.E. Cantrell, Computation of the transient M/M/1 queue cdf, pdf, and mean with generalized Q -functions, *IEEE Trans. Commun.*, COM3-4 (1986) 814–817.
- [13] P.E. Cantrell and G.R. Beall, Transient M/M/1 queue variance computation using generalized Q -functions, *IEEE Trans. Commun.*, to appear.
- [14] D.G. Champervowne, An elementary method of solution of the queueing problem with a single server and constant parameters, *J. Roy. Stat. Soc. B*, 18 (1956) 125–128.
- [15] N.L. Johnson and S. Kotz, *Continuous Univariate Distributions - 2* (Houghton Mifflin, Boston, 1970).
- [16] S.K. Jones, R.K. Cavin III and D.A. Johnston, An efficient computation procedure for the evaluation of M/M/1 transient state occupancy probabilities, *IEEE Trans. Commun.* COM-28 (1980) 2019–2020.
- [17] S. Karlin and J.L. McGregor, The differential equations of birth-and-death processes, and the Stieltjes moment problem, *Trans. Amer. Math Soc.* 85 (1957) 489–546.
- [18] S. Karlin and J.L. McGregor, Many server queueing processes with Poisson input and exponential service times, *Pacific J. Math.* 8 (1958) 87–118.
- [19] S. Karlin and J.L. McGregor, A characterization of birth and death processes, *Proc. Nat. Acad. Sci. U.S.A.* 45 (1959) 375–379.
- [20] J. Keilson, *Markov Chain Models - Rarity and Exponentially* (Springer-Verlag, New York, 1979).
- [21] J. Keilson and A. Kooharian, On time dependent queueing processes, *Ann. Math. Statist.* (1960) 104–112.
- [22] D.G. Kendall, Some problems in the theory of queues, *J. Roy. Stat. Soc. Ser. B*, 13 (1951) 151–171.
- [23] W. Ledermann and G.E.H. Reuter, Spectral theory for the differential equations of simple birth and death processes, *Phil. Trans. Roy. Soc. London, Ser. A*, 246 (1954) 321–369.
- [24] W.F. Magnus, F. Oberhettinger and R.P. Soni, *Formulas and Theorems for the Special Functions of Mathematical Physics* (Springer-Verlag, New York, 1966).
- [25] W.A. Massey, A transient analysis of the M/M/1 queue without transforms, 1986, submitted for publication.
- [26] P.M. Morse, Stochastic properties of waiting lines, *Opns. Res.* 3 (1955) 255–261.
- [27] F.W.J. Olver, *Asymptotics and Special Functions* (Academic Press, New York, 1974).
- [28] T.J. Ott, The stable M/G/1 queue in heavy traffic and its covariance function, *Adv. Appl. Prob.* 9 (1977) 169–186.
- [29] J. Riordan, Delays for last-come first-served service and the busy period, *Bell System Tech. J.* 40 (1961) 785–793.
- [30] J. Riordan, *Stochastic Service Systems* (Wiley, New York, 1962).
- [31] T.E. Stern, Approximations of queue dynamics and their application to adaptive routing in communication networks, *IEEE Trans. Commun.* COM-27 (1979) 1331–1335.
- [32] L. Takács, *Introduction to the Theory of Queues* (Oxford University Press, New York, 1962).
- [33] E. van Doorn, *Stochastic Monotonicity and Queueing Applications of Birth-Death Processes*, *Lecture Notes in Statistics* 4 (Springer-Verlag, New York, 1980).
- [34] E. Vulot, Délais d'attente des appels téléphoniques dans l'ordre inverse de leur arrivée, *C.R. Acad. Sci. Paris*, 238 (1954) 1188–1189.
- [35] C.M. Woodside, B. Pagurek and G.F. Newell, A diffusion approximation for correlation in queues, *J. Appl. Prob.* 17 (1980) 1033–1047.