J. Appl. Prob. **28**, 146–158 (1991)
Printed in Israel

RATE CONSERVATION FOR STATIONARY PROCESSES

JOSEP M. FERRANDIZ* AND
AUREL A. LAZAR,** *Columbia University*

Abstract

We derive a rate conservation law for distribution densities which extends a result of Brill and Posner. Based on this conservation law, we obtain a generalized Takács equation for the $G/G/m/B$ queueing system that only requires the existence of a stochastic intensity for the arrival process and the residual service time distribution density for the $G/GI/1/B$ queue. Finally, we solve Takács' equation for the $N/GI/1/\infty$ queueing system.

PALM PROBABILITY; STOCHASTIC INTENSITY; TAKÁCS FORMULA; GROUP ARRIVALS

1. Introduction

Most basic results in queueing theory are derived directly or indirectly from rate conservation principles. These include for instance, the global balance equations for equilibrium probabilities of Markovian queueing networks and the arrival theorem. In general, these principles relate the so-called customer and time averages of queueing systems. In the queueing literature, rate conservation has been extensively studied by Franken et al. [6], [9].

In this paper, we present a rate conservation formula which extends a result of Brill and Posner [4] to a non-Markovian setting. This formula relates the stationary distribution density of a process at a point x to the number of upcrossings and downcrossings of level x by this process per time unit.

The rate conservation formula obtained in conjunction with Papangelou's theorem [1], [3], allows a simple derivation of a generalized Takács equation for the $G/G/m/B$ queueing system. For its derivation, we require the existence of a stationary regime for the queueing system and of a stochastic intensity for the arrival process. We also apply the conservation law to obtain the residual service time density for $G/GI/1/B$ queues.

As a further application of the rate conservation formula, we solve, in the transform domain, Takács' equation for the $N/GI/1/\infty$ queueing system [13].

Received 13 April 1989; revision received 12 January 1990.
* Present address: Hewlett-Packard Laboratories, Filton Road, Stoke Gifford, Bristol BS12 6QZ, UK.
** Postal address: Department of Electrical Engineering, 1312 S. W. Mudd Building, Columbia University, New York, NY 10027, USA.

Throughout this paper, we use the notation and terminology of Baccelli and Brémaud [1] for Palm probability. For the notions related to stochastic intensity, the reader is referred to Brémaud's monograph [2].

In Section 2, we prove the rate conservation formula. Section 3 contains the derivation of Takács' equation for the $G/G/m/B$ queue and the residual service time distribution density for the $G/GI/1/B$ queueing system. In Section 4, the rate conservation formula is applied to the study of the workload distribution density of an $N/GI/1$ queue. Finally, the conclusions of this paper are in Section 5.

2. Rate conservation for distribution densities

Let $(Z(t))_{t\in\mathbb{R}}$ be a right-continuous process with left-hand limits and let

$$Z'(t) = \lim_{\varepsilon\downarrow 0^+} \frac{Z(t+\varepsilon) - Z(t)}{\varepsilon},$$

be the right-hand derivative of $Z(\cdot)$. We assume that this derivative exists.

We define a point process $N = (T_n)_{n\in\mathbb{Z}}$ by

$$N_C = \sum_{t\in C} \mathbf{1}[Z(t) \neq Z(t^-)],$$

where C is a Borel-measurable set on the real line. This point process counts the discontinuity jumps of $Z(\cdot)$. We assume that N has a finite rate, i.e. $\lambda = \mathbb{E}[N_{]0,1]}] < \infty$ and denote the Palm probability associated with N by $\mathbb{P}_N^0$.

Suppose that for $x \in \mathbb{R}$

$$\mathbb{P}_N^0(Z(0) = x) = \mathbb{P}_N^0(Z(0^-) = x) = 0, \tag{2.1}$$

and that $(Z(t))_{t\in\mathbb{R}}$ admits a density $h(x)$ with respect to $\mathbb{P}$. Then, the following rate conservation formula holds.

Proposition 2.1. If $(Z(t))_{t\in\mathbb{R}}$ satisfies condition (2.1) at $x\in\mathbb{R}$ and admits a density $h(\cdot)$ with respect to $\mathbb{P}$, then

$$\begin{aligned} &h(x)\mathbb{E}[Z'(0) \mid Z(0) = x] \\ &\quad = \lambda \mathbb{E}_N^0[\mathbf{1}(Z(0^-) > x)\mathbf{1}(Z(0) \leqq x) - \mathbf{1}(Z(0^-) \leqq x)\mathbf{1}(Z(0) > x)]. \end{aligned} \tag{2.2}$$

Proof. Apply the inversion formula [1] between the Palm probability $\mathbb{P}_N^0$ associated with N and the stationary probability $\mathbb{P}$ to $Z'(t)\mathbf{1}_{]x,x+\varepsilon]}(Z(t))$. It yields

$$\mathbb{E}[Z'(0)\mathbf{1}_{]x,x+\varepsilon]}(Z(0))] = \lambda \mathbb{E}_N^0\left[\int_0^{T_1} Z'(t)\mathbf{1}_{]x,x+\varepsilon]}(Z(t))dt\right]. \tag{2.3}$$

The proposition will be proved by evaluating the integral in (2.3), dividing both members of the equation (2.3) by ε and letting $\varepsilon \to 0$. For simplicity we write $Y = \int_0^{T_1} Z'(t)\mathbf{1}_{]x,x+\varepsilon]}(Z(t))dt$.

Let $\tau_0 = 0$ and

$$\tau_n = \inf\{t : \tau_{n-1} < t \text{ and } (Z(t) = x \text{ or } Z(t) = x+\varepsilon)\} \wedge T_1, \qquad n \geqq 1.$$

Thus, $(\tau_n)_{n\in\mathbb{N}}$ is the sequence of crossing times of levels x and $x+\varepsilon$ during $[0, T_1[$ by the process $Z(t)$. Since $Z(\cdot)$ is continuous in $[0, T_1[$ and since $Z(\tau_n)$ is either equal to x or $x+\varepsilon$, it follows that Y will be non-zero only if there is an odd number of crossings of level x or of level $x+\varepsilon$ in $[0, T_1[$ or both. With this observation, one can evaluate Y by considering the different possible positions of $Z(0)$ and $Z(T_1^-)$ with respect to x and $x+\varepsilon$.

(i) If $Z(0)\in]x, x+\varepsilon]$ and $Z(T_1^-)\in]x, x+\varepsilon]$, then $Y = Z(T_1^-) - Z(0)$.

(ii) If $Z(0) > x+\varepsilon$ and $Z(T_1^-)\in]x, x+\varepsilon]$, then $Y = Z(T_1^-) - (x+\varepsilon)$. The symmetric case is obtained when $Z(0)\in]x, x+\varepsilon]$ and $Z(T_1^-) > x+\varepsilon$. It yields $Y = x+\varepsilon - Z(0)$.

(iii) If $Z(0) > x+\varepsilon$ and $Z(T_1^-)\leqq x$, then $Y = -\varepsilon$. The symmetric case is obtained when $Z(0)\leqq x$ and $Z(T_1^-) > x+\varepsilon$. It yields $Y = \varepsilon$.

(iv) If $Z(0)\leqq x$ and $Z(T_1^-)\in]x, x+\varepsilon]$, then $Y = Z(T_1^-) - x$. The symmetric case is obtained when $Z(0)\in]x, x+\varepsilon]$ and $Z(T_1^-)\leqq x$. It yields $Y = x - Z(0)$.

(v) If $Z(0) > x+\varepsilon$ and $Z(T_1^-) > x+\varepsilon$, then $Y = 0$. The symmetric case is obtained when $Z(0)\leqq x$ and $Z(T_1^-)\leqq x$. It yields $Y = 0$.

Note that in all cases $|Y|\leqq\varepsilon$. It follows from Equation (2.3) that

$$
\begin{aligned}
&\mathbb{E}[Z'(0)\mathbf{1}_{]x,x+\varepsilon]}(Z(0))] \\
&\quad = \lambda\mathbb{E}_N^0[(Z(T_1^-) - Z(0))\mathbf{1}_{]x,x+\varepsilon]}(Z(0))\mathbf{1}_{]x,x+\varepsilon]}(Z(T_1^-)) \\
&\qquad + (Z(T_1^-) - (x+\varepsilon))\mathbf{1}(Z(0) > x+\varepsilon)\mathbf{1}_{]x,x+\varepsilon]}(Z(T_1^-)) \\
&\qquad + (x+\varepsilon - Z(0))\mathbf{1}_{]x,x+\varepsilon]}(Z(0))\mathbf{1}(Z(T_1^-) > x+\varepsilon) \qquad (2.4) \\
&\qquad + \varepsilon\{\mathbf{1}(Z(0)\leqq x)\mathbf{1}(Z(T_1^-) > x+\varepsilon) - \mathbf{1}(Z(0) > x+\varepsilon)\mathbf{1}(Z(T_1^-)\leqq x)\} \\
&\qquad + (Z(T_1^-) - x)\mathbf{1}(Z(0)\leqq x)\mathbf{1}_{]x,x+\varepsilon]}(Z(T_1^-)) \\
&\qquad + (x - Z(0))\mathbf{1}_{]x,x+\varepsilon]}(Z(0))\mathbf{1}(Z(T_1^-)\leqq x)].
\end{aligned}
$$

Since in all the cases $|Y|\leqq\varepsilon$, one gets for the first term on the right-hand side of (2.4),

$$
\begin{aligned}
&|\mathbb{E}_N^0[(Z(T_1^-) - Z(0))\mathbf{1}_{]x,x+\varepsilon]}(Z(0))\mathbf{1}_{]x,x+\varepsilon]}(Z(T_1^-))]| \\
&\quad \leqq \varepsilon\mathbb{E}_N^0[\mathbf{1}_{]x,x+\varepsilon]}(Z(0))\mathbf{1}_{]x,x+\varepsilon]}(Z(T_1^-))] \\
&\quad \leqq \varepsilon\mathbb{E}_N^0[\mathbf{1}_{]x,x+\varepsilon]}(Z(0))],
\end{aligned}
$$

and thus, from condition (2.1), we get

$$\lim_{\varepsilon\downarrow 0}\frac{1}{\varepsilon}|\mathbb{E}_N^0[(Z(T_1^-) - Z(0))\mathbf{1}_{]x,x+\varepsilon]}(Z(0))\mathbf{1}_{]x,x+\varepsilon]}(Z(T_1^-))]| = 0.$$

Using the same method one can show that all terms but the third one (iii) on the right-hand side of Equation (2.4) will vanish after dividing by ε and taking limits as $\varepsilon\to 0$. Thus, the limit

$$\lim_{\varepsilon \downarrow 0} \frac{\mathbb{E}[Z'(0)\mathbf{1}_{]x,x+\varepsilon]}(Z(0))]}{\varepsilon},$$

exists and, after a simple conditioning on the left-hand side of (2.4), we have

$$\begin{aligned} &h(x)\mathbb{E}[Z'(0) \mid Z(0)=x] \\ \text{(2.5)} \qquad &= \lambda \mathbb{E}_N^0[\mathbf{1}(Z(0) \leqq x)\mathbf{1}(Z(T_1^-)>x) - \mathbf{1}(Z(0)>x)\mathbf{1}(Z(T_1^-) \leqq x)]. \end{aligned}$$

If $g(T_1^-) = \mathbf{1}(Z(T_1^-)>x)\mathbb{E}[\mathbf{1}(Z(0) \leqq x) \mid Z(T_1^-)]$,

$$\mathbb{E}_N^0[\mathbf{1}(Z(0) \leqq x)\mathbf{1}(Z(T_1^-)>x)] = \mathbb{E}_N^0[g(T_1^-)] = \mathbb{E}_N^0[g(0^-)],$$

by the θ_{T_1} invariance of $\mathbb{P}_N^0$. Thus,

$$\mathbb{E}_N^0[\mathbf{1}(Z(0) \leqq x)\mathbf{1}(Z(T_1^-)>x)] = \mathbb{E}_N^0[\mathbf{1}(Z(0) \leqq x)\mathbf{1}(Z(0^-)>x)],$$

and the result follows.

Substituting $\mathbf{1}(Z(T_1^-) \leqq x) = 1 - \mathbf{1}(Z(T_1^-)>x)$ and $\mathbf{1}(Z(0) \leqq x) = 1 - \mathbf{1}(Z(0)>x)$ into Equation (2.5), and using the invariance of the Palm probability yields

$$\text{(2.6)} \qquad h(x)\mathbb{E}[Z'(0) \mid Z(0)=x] = \lambda \mathbb{E}_N^0[\mathbf{1}(Z(0^-)>x) - \mathbf{1}(Z(0)>x)].$$

Equation (2.6) has also been derived in [10].

Note that Equations (2.2), (2.5) and (2.6) are all rate conservation formulas. Equation (2.2) gives a conservation formula for the rate of crossings of level x at discontinuity jump times and Equation (2.5) for the rate of crossings of level x between discontinuity jumps.

In an ergodic context, we can interpret $\mathbb{E}_N^0[f(Z(0))]$ and $\mathbb{E}_N^0[f(Z(0^-))]$ respectively as

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} f(Z(T_i))$$

and

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} f(Z(T_i^-)).$$

Hence, Equations (2.2) and (2.5) also provide a practical method to evaluate the density of $(Z(t))_{t \in \mathbf{R}}$ by relating it to the empirical average number of upcrossings and downcrossings at level x per time unit. This interpretation was first pointed out by Brill and Posner [4].

Remark (Brémaud, personal communication). For Proposition 2.1 to hold we have assumed that

(i) $\mathbb{P}_N^0(Z(0)=x) = \mathbb{P}_N^0(Z(0^-)=x) = 0$ holds.

(ii) $Z(t)$ admits a density with respect to $\mathbb{P}$ at x.

These two assumptions have implied the existence of the limit

$$\lim_{\varepsilon \downarrow 0} \frac{1}{\varepsilon} \mathbb{E}[Z'(0)\mathbf{1}_{]x,x+\varepsilon]}(Z(0))].$$

In general, the existence of this limit, does not *a priori* guarantee that $Z(0)$ admits a density with respect to $\mathbb{P}$. However, in most queueing applications, $(Z(t))_{t\in\mathbf{R}}$ evolves according to a differential equation of the form $\dot{Z} = f(Z)$ for some continuous function $f(\cdot)$. In this case

$$\lim_{\varepsilon\downarrow 0} \frac{1}{\varepsilon} \mathbb{E}[Z'(0)\mathbf{1}_{]x,x+\varepsilon]}(Z(0))] = f(x) \lim_{\varepsilon\downarrow 0} \frac{1}{\varepsilon} \mathbb{E}[\mathbf{1}_{]x,x+\varepsilon]}(Z(0))],$$

and thus $(Z(t))_{t\in\mathbf{R}}$ admits a density with respect to $\mathbb{P}$.

3. Applications of the rate conservation formula

We apply the rate conservation formula to derive a generalized Takács equation for the $G/G/m/B$ queueing system and the residual service time distribution density for the $G/GI/1/B$ queue. Further applications can be found in [5] where the joint distribution of the residual service time and the queue length for an $N/GI/1/B$ queueing system is obtained.

3.1. *Takács' equation.* We consider a $G/G/m/B$ queue (m servers and a waiting room of size B) with a work conserving discipline. We assume that its workload $W(\cdot)$ satisfies Equation (2.1) for $x > 0$.

The arrival times to the system will be denoted by $(T_n)_{n\in\mathbf{Z}}$ and we let $A_C = \Sigma_n \mathbf{1}_C(T_n)$, for C Borel measurable. The arrival rate will be denoted by λ. We assume that the arrival process admits a $(\mathbb{P}, \mathscr{F}_t)$-stochastic intensity $(\lambda_t)_{t\in\mathbf{R}}$, where $\mathscr{F}_t \supset \mathscr{F}_t^W$. We define

$$\lambda(W(0)) = \mathbb{E}[\lambda_0 \mid W(0)],$$

and, if $(\sigma_n)_{n\in\mathbf{Z}}$ is the sequence of service times,

$$F(W(0^-), y) = \mathbb{E}_A^0[\mathbf{1}(\sigma_0 \leqq y) \mid W(0^-)].$$

Let $W'(t)$ be the right-hand derivative of $W(t)$ and

$$r(W(0)) = -\mathbb{E}[W'(0) \mid W(0)]$$

be the system service rate (average number of busy servers given the workload). If we denote the workload density by $h(x)$, then Proposition 2.1 gives

$$h(x)r(x) = \lambda \mathbb{E}_A^0[\mathbf{1}(W(0^-) \leqq x)\mathbf{1}(W(0) > x)]. \tag{3.1}$$

Equation (3.1) was obtained by Brill and Posner [4] for a $M/GI/1$ queue.

In what follows we derive Takács' equation for the $G/G/m/B$ queueing system. From Equation (3.1),

$$h(x)r(x) = \lambda \mathbb{E}_A^0[\mathbf{1}(W(0^-) \leqq x)(1 - F(W(0^-), x - W(0^-)))], \tag{3.2}$$

and from Papangelou's theorem [1],

$$\begin{aligned} h(x)r(x) &= \mathbb{E}[\lambda_0 \mathbf{1}(W(0) \leqq x)(1 - F(W(0), x - W(0)))] \\ &= \mathbb{E}[\lambda(W(0))\mathbf{1}(W(0) \leqq x)(1 - F(W(0), x - W(0)))]. \end{aligned}$$

From the work conserving discipline assumption, $r(x)>0$ for $x>0$ and thus the last equation can be rewritten as

$$
\begin{aligned}
h(x) &= (1-\rho)\frac{\lambda(0)}{r(x)}(1-F(0,x)) \\
&\quad + \frac{1}{r(x)}\int_0^x \lambda(w)(1-F(w,x-w))h(w)\,dw,
\end{aligned}
\tag{3.3}
$$

where $1-\rho=\mathbb{P}(W(0)=0)$. Equation (3.3) is the generalization of Takács' equation to the $G/G/m/B$ queueing system.

3.2. *Residual service times.* Let $R(t)$ be the residual service time at time t of a $G/GI/1/B$ queueing system with a work conserving discipline. If the system is empty, $R(t)=0$. Otherwise, $R(t)$ is the remaining service time for the customer being served at time t. $R(t)$ has discontinuity jumps at departure times that do not leave behind an empty system and at the arrival times that find an empty system. These jump times form a point process N. Hence, if λ is the arrival rate of non-blocked customers, $R(t)$ has discontinuity jumps at a rate λ. If we take $Z(t)=R(t)$ in Proposition 2.1, it follows that the density $h(\cdot)$ of $R(t)$ exists and is given by

$$h(x)=\lambda\mathbb{E}_N^0[\mathbf{1}(R(0^-)\leqq x)\mathbf{1}(R(0)>x)].$$

But if $t=0$ is a discontinuity of $R(t)$, it must be that $R(0^-)=0$ and $R(0)=\sigma_0$ where σ_0 is the total amount of service required by the customer starting service at time 0. Hence, if $F(x)$ is the distribution of σ_0, we get

$$h(x)=\lambda(1-F(x)), \tag{3.4}$$

as expected.

4. Rate conservation for the $N/G/1/\infty$ queueing system

The $N/GI/1/\infty$ queueing system has been studied by Ramaswami in [13]. We begin by defining N-processes [11] by time changes on a Markov chain [5], [7].

4.1. *Construction of* N*-processes via time changes.* Let $(X_t)_{t\in\mathbf{R}}$ be a positive recurrent Markov process in equilibrium taking values in $\{1,\cdots,L,L+1\}$ and let $E=\{1,\cdots,L\}$. The intensity matrix of $(X_t)_{t\in\mathbf{R}}$ partitioned according to E is given by

$$
\begin{array}{cc}
 & \begin{array}{cc} E & \{L+1\} \end{array} \\
\begin{array}{c} E \\ \{L+1\} \end{array} & \begin{pmatrix} \boldsymbol{Q} & \boldsymbol{\alpha} \\ \boldsymbol{\beta} & -1 \end{pmatrix},
\end{array}
\tag{4.1}
$$

where $\boldsymbol{Q}=(q_{lm})$ is an $L\times L$ substochastic intensity matrix, i.e., such that $\boldsymbol{Q1}<\mathbf{0}$ and $q_{lm}\geqq 0$ for $l\neq m$. Since $(X_t)_{t\in\mathbf{R}}$ is positive recurrent, the $L\times 1$ column vector $\boldsymbol{\alpha}=(\alpha_l)_{1\leq l\leq L}$ is such that $\boldsymbol{Q1}+\boldsymbol{\alpha}=\mathbf{0}$ and the $1\times L$ row vector $\boldsymbol{\beta}=(\beta_l)_{1\leq l\leq L}$ is such that $\boldsymbol{\beta 1}=1$. Thus, if we let $q_l=\Sigma_{m\neq l}q_{lm}$, then the rate out of state l is

$q_l + \alpha_l$, i.e., the lth diagonal element in Equation (4.1), is equal to $-(q_l + \alpha_l)$, $1 \leqq l \leqq L$.

Consider the additive functional

$$B(s) = \int_0^s \mathbf{1}(X_u \neq L + 1)du.$$

The above functional represents the time spent outside of state $L + 1$ by $(X_t)_{t \in \mathbf{R}_+}$ up to time $s, s > 0$. Its right- and left-continuous inverses are respectively denoted by τ_t and σ_t, i.e.,

$$\tau_t = \inf\{s : B(s) > t\}$$

and

$$\sigma_t = \inf\{s : B(s) \geqq t\}.$$

It follows then that the process $(Y_t)_{t \in \mathbf{R}_+}$ defined by $Y_t = X_{\tau_t}$, $t \geqq 0$, is a positive recurrent Markov process on E with intensity matrix $\boldsymbol{Q} + \boldsymbol{\alpha\beta}$ [5]. The equilibrium distribution of $(Y_t)_{t \geqq 0}$ will be denoted by $\boldsymbol{\pi} = (\pi_l)_{l \in E}$. The process $(Y_t)_{t \in \mathbf{R}_+}$ will be referred to as the phase process (or simply the phase). The jumps of $(Y_t)_{t \in \mathbf{R}_+}$ will be called phase transitions. Phase transitions can be of two types depending on whether or not they correspond to a stopping and restarting of the clock of the process $(X_t)_{t \in \mathbf{R}_+}$. If a phase transition corresponds to a stopping and restarting of the clock, we say that it is a renewal transition. Otherwise, we say that it is a pure phase transition. In what follows, we define the N-process as a marked point process $(T_n, U_n)_{n \in \mathbf{Z}}$. We begin by giving the probabilistic description of its jump times.

Consider L independent Poisson processes $N^1, \cdots, N^L$ of respective rates $\lambda_1, \cdots, \lambda_L$. We define the following point processes:

$$A_C^p = \sum_{l=1}^{L} \int_C \mathbf{1}(Y_s = l) N^l(ds), \tag{4.2}$$

$$A_C^q = \sum_{t \in C} \mathbf{1}(Y_{t^-} \neq Y_t)\mathbf{1}(\sigma_t = \tau_t), \tag{4.3}$$

$$A_C^r = \sum_{t \in C} \mathbf{1}(\sigma_t \neq \tau_t), \tag{4.4}$$

where C is a Borel set of $\mathbb{R}_+$. The points of A^q correspond to pure phase transition times of $(Y_t)_{t \in \mathbf{R}_+}$ and those of A^r to renewal phase transition times. The process A^p is a doubly stochastic Poisson process. Its points are called *Poisson arrivals*. Note that A^p, A^q and A^r have no common jumps.

Let A be the point process defined by

$$A_C = A_C^p + A_C^q + A_C^r,$$

where C is a Borel set of $\mathbb{R}_+$. The extension of A over the entire real line can be easily achieved. Thus, if $(T_n)_{n \in \mathbf{Z}}$ is the sequence of jump times of A then $A_C = \Sigma_n \mathbf{1}(T_n \in C)$, where C is now a Borel set of the real line and $(T_n)_{n \in \mathbf{Z}}$ is obtained by merging the jumps of A^p, A^q and A^r. Next, we define the sequence of marks of an N-process.

Let $(U_n)_{n\in\mathbb{Z}}$ be a sequence of random variables taking values in $\mathbb{N}$ with the following conditional distributions:

$$
\begin{aligned}
b_l(n) &= \mathbb{E}^0_{A^p}[\mathbf{1}(U_0=n) \mid Y_{0^-}=l], \\
a^q_{lm}(n) &= \mathbb{E}^0_{A^q}[\mathbf{1}(U_0=n) \mid Y_{0^-}=l,\, Y_0=m], \quad \text{if } l\neq m, \\
a^q_{ll}(n) &= 0, \qquad\qquad\qquad\qquad\qquad\qquad\quad \text{if } n>0, \\
a^q_{ll}(0) &= 1, \\
a^r_{lm}(n) &= \mathbb{E}^0_{A^r}[\mathbf{1}(U_0=n) \mid Y_{0^-}=l,\, Y_0=m].
\end{aligned}
\tag{4.5}
$$

Thus, $b_l(n)$, $1\leqq l\leqq L$, is the probability that the mark of a Poisson arrival when the phase is l is equal to $n, n\geqq 0$. We can assume without loss of generality that $b_l(0)=0$. If $1\leqq l,\ m\leqq L$, then $a^r_{lm}(n)$ is the probability that the mark of an arrival at a renewal phase transition from l into m is equal to $n, n\geqq 0$. Similarly, for $l\neq m$, $a^q_{lm}(n)$ is the probability that the mark of an arrival at a pure phase transition from l into m is equal to $n, n\geqq 0$.

We can now formally define the N-process.

Definition 4.1. The marked point process $(T_n, U_n)_{n\in\mathbb{Z}}$ is called an N-process.

Intuitively, an N-process is obtained as follows. When $Y_t=l$, there are group arrivals at a Poisson rate λ_l. Furthermore, there will also be a group arrival at every phase transition. In all cases, the group size of an arrival is conditionally independent of the arrival process given the phase transition and the arrival time.

4.2. *A basic result for* N-*processes*. In [11] it is shown that the moment generating function of $N_{]0,t]}$ (number of arrivals in $]0,t]$) is of the form

$$P(t,z)=\sum_{n\geqq 0} z^n\mathbb{E}[\mathbf{1}(N_{]0,t]}=n)]=\boldsymbol{\pi}\exp\{\boldsymbol{R}(z)t\}\mathbf{1}. \tag{4.6}$$

The matrix $\boldsymbol{R}(z)$ will be called the generating matrix of the N-process. In what follows we characterize this generating matrix.

The z-transforms of $(b_l(n))_{n\geqq 0}$ and $(a^k_{lm}(n))_{n\geqq 0}$ are respectively denoted by $\tilde{b}_l(z)$ and $\tilde{a}^k_{lm}(z)$, $k=q,r$. Let $\tilde{\boldsymbol{b}}(z)$ be the diagonal matrix with entries $\tilde{b}_1(z),\cdots,\tilde{b}_L(z)$. Similarly, for $k=q,r$, let $\tilde{\boldsymbol{a}}^k(z)$ be the matrix with entries $\tilde{a}^k_{lm}(z)$, $1\leqq l,\ m\leqq L$. In [11] it is shown that the generating matrix $\boldsymbol{R}(z)$ is given by

$$\boldsymbol{R}(z)=\boldsymbol{\lambda}(\tilde{\boldsymbol{b}}(z)-\boldsymbol{I})+\boldsymbol{Q}\circ\tilde{\boldsymbol{a}}^q(z)+\boldsymbol{\alpha\beta}\circ\tilde{\boldsymbol{a}}^r(z), \tag{4.7}$$

where $\circ$ denotes the entrywise matrix product.

4.3. *The $N/GI/1/\infty$ queueing system*. The arrival process to an $N/GI/1/\infty$ queue is an N-process $(T_n, U_n)_{n\in\mathbb{Z}}$, where $(T_n)_{n\in\mathbb{Z}}$ and $(U_n)_{n\in\mathbb{Z}}$ are, respectively, the sequences of arrival times and their corresponding group size. The packet service times are independent of the arrivals and of each other and have a distribution $F(\cdot)$. Group arrivals are processed on a FCFS basis, with random order within a group.

The system workload is denoted by $W(t)_{t\in\mathbf{R}}$. Takács' equation characterizes the system workload distribution density as the solution of an integral equation. Its solution in the transform domain for the $M/GI/1/\infty$ system is a classical result in queueing theory [8]. A detailed study of $N/GI/1/\infty$ queueing systems has been carried out in [13] using matrix geometric methods [12].

4.4. *The solution to Takács' equation.* In this section we derive Takács' equation using rate conservation arguments for the $N/G/1/\infty$ queueing system. We achieve this by applying Equation (2.6) to the process

$$Z(t) = W(t)\mathbf{1}(Y_t = m).$$

Finally, we obtain the Takács equation in the transform domain.

Let Δ be the point process generated by the discontinuity jumps of $(Z(t))_{t\in\mathbf{R}}$. Its rate will be denoted by λ_Δ. If

$$H(x, m) = \mathbb{P}(W(0) \leqq x, Y_0 = m), \qquad x > 0,$$

we have from Equation (2.6)

$$\frac{dH(x,m)}{dx} = \lambda_\Delta \mathbb{E}_\Delta^0[\mathbf{1}(W(0) > x)\mathbf{1}(Y_0 = m) - \mathbf{1}(W(0^-) > x)\mathbf{1}(Y_{0^-} = m)]. \tag{4.8}$$

Since

$$\mathbf{1}(W(0) > x)\mathbf{1}(Y_0 = m) = \sum_{n \geqq 0} \mathbf{1}(W(0) > x)\mathbf{1}(Y_0 = m)\mathbf{1}(U_0 = n),$$

we have to evaluate the expression $\lambda_\Delta \mathbb{E}_\Delta^0[\mathbf{1}(W(0) > x)\mathbf{1}(Y_0 = m)\mathbf{1}(U_0 = n)]$.

Let $F_n(\cdot)$ denote the n-fold convolution of $F(\cdot)$. We have

$$\begin{aligned} &\lambda_\Delta \mathbb{E}_\Delta^0[\mathbf{1}(W(0) > x)\mathbf{1}(Y_0 = m)\mathbf{1}(U_0 = n)] \\ &\quad = \lambda_\Delta \mathbb{E}_\Delta^0[\mathbf{1}(Y_0 = m)\mathbf{1}(U_0 = n)] \\ &\quad\quad - \lambda_\Delta \mathbb{E}_\Delta^0[F_n(x - W(0^-))\mathbf{1}(Y_0 = m)\mathbf{1}(U_0 = n)]. \end{aligned} \tag{4.9}$$

Furthermore, from the exchange formula [1], the second term on the right-hand side of Equation (4.9) becomes

$$\begin{aligned} &\lambda_\Delta \mathbb{E}_\Delta^0[F_n(x - W(0^-))\mathbf{1}(Y_0 = m)\mathbf{1}(U_0 = n)] \\ &\quad = \lambda_{A^p} \mathbb{E}_{A^p}^0[F_n(x - W(0^-))\mathbf{1}(Y_0 = m)\mathbf{1}(U_0 = n)] \\ &\quad\quad + \lambda_{A^q} \mathbb{E}_{A^q}^0[F_n(x - W(0^-))\mathbf{1}(Y_0 = m)\mathbf{1}(U_0 = n)] \\ &\quad\quad + \lambda_{A^r} \mathbb{E}_{A^r}^0[F_n(x - W(0^-))\mathbf{1}(Y_0 = m)\mathbf{1}(U_0 = n)]. \end{aligned} \tag{4.10}$$

We compute each one of the terms on the right-hand side of Equation (4.10). For the first term, applying Papangelou's theorem to A^p,

$$\begin{aligned}\lambda_{A^p}\mathbb{E}^0_{A^p}[F_n(x-W(0^-))\mathbf{1}(Y_0=m)\mathbf{1}(U_0=n)]\\ =\lambda_{A^p}b_m(n)\mathbb{E}^0_{A^p}[F_n(x-W(0^-))\mathbf{1}(Y_{0^-}=m)]\\ =\lambda_m b_m(n)\mathbb{E}[F_n(x-W(0))\mathbf{1}(Y_0=m)].\end{aligned}\tag{4.11}$$

For the second term, applying Papangelou's theorem to A^q,

$$\begin{aligned}&\lambda_{A^q}\mathbb{E}^0_{A^q}[F_n(x-W(0^-))\mathbf{1}(Y_0=m)\mathbf{1}(U_0=n)]\\ &=\lambda_{A^q}\sum_{l\neq m}a^q_{lm}(n)\mathbb{E}^0_{A^q}[F_n(x-W(0^-))\mathbf{1}(Y_{0^-}=l)\mathbf{1}(Y_0=m)]\\ &=\sum_{l\neq m}q_{lm}a^q_{lm}(n)\mathbb{E}[F_n(x-W(0))\mathbf{1}(Y_0=l)].\end{aligned}\tag{4.12}$$

The third term on the right-hand side of (4.10) is computed in a similar fashion:

$$\begin{aligned}&\lambda_{A^r}\mathbb{E}^0_{A^r}[F_n(x-W(0^-))\mathbf{1}(Y_0=m)\mathbf{1}(U_0=n)]\\ &=\lambda_{A^r}\sum_{l}a^r_{lm}(n)\mathbb{E}^0_{A^r}[F_n(x-W(0^-))\mathbf{1}(Y_{0^-}=l)\mathbf{1}(Y_0=m)]\\ &=\sum_{l}\alpha_l\beta_m a^r_{lm}(n)\mathbb{E}[F_n(x-W(0))\mathbf{1}(Y_0=l)].\end{aligned}\tag{4.13}$$

The first term of the right-hand side of Equation (4.9) can be obtained by letting $x\to\infty$ in Equations (4.11), (4.12) and (4.13). One gets, after summing over n,

$$\lambda_\Delta\mathbb{E}^0_\Delta[\mathbf{1}(Y_0=m)]=\pi_m\lambda_m+\sum_{l\neq m}\pi_l q_{lm}+\sum_l\pi_l\alpha_l\beta_m.\tag{4.14}$$

Therefore, combining Equations (4.11), (4.12), (4.13) and (4.14),

$$\begin{aligned}\lambda_\Delta\mathbb{E}^0_\Delta[\mathbf{1}(W(0)>x)\mathbf{1}(Y_0=m)]&=\pi_m\lambda_m+\sum_{l\neq m}\pi_l q_{lm}+\sum_l\pi_l\alpha_l\beta_m\\ &\quad-\sum_{n\ge 0}\Big(\lambda_m b_m(n)\mathbb{E}[F_n(x-W(0))\mathbf{1}(Y_0=m)]\\ &\quad+\sum_{l\neq m}q_{lm}a^q_{lm}(n)\mathbb{E}[F_n(x-W(0))\mathbf{1}(Y_0=l)]\\ &\quad+\sum_l\alpha_l\beta_m a^r_{lm}(n)\mathbb{E}[F_n(x-W(0))\mathbf{1}(Y_0=l)]\Big).\end{aligned}\tag{4.15}$$

Similarly,

$$\begin{aligned}\lambda_\Delta\mathbb{E}^0_\Delta[\mathbf{1}(W(0^-)>x)\mathbf{1}(Y_{0^-}=m)]&=\lambda_{A^p}\mathbb{E}^0_{A^p}[\mathbf{1}(W(0^-)>x)\mathbf{1}(Y_{0^-}=m)]\\ &\quad+\lambda_{A^q}\mathbb{E}^0_{A^q}[\mathbf{1}(W(0^-)>x)\mathbf{1}(Y_{0^-}=m)]\\ &\quad+\lambda_{A^r}\mathbb{E}^0_{A^r}[\mathbf{1}(W(0^-)>x)\mathbf{1}(Y_{0^-}=m)]\\ &=(\lambda_m+q_m+\alpha_m)\{\pi_m-\mathbb{E}[\mathbf{1}(W(0)\le x)\mathbf{1}(Y_0=m)]\}.\end{aligned}\tag{4.16}$$

Note that since $q_{mm}=-(q_m+\alpha_m)$,

$$\pi_m\lambda_m + \sum_{l\neq m} \pi_l q_{lm} + \sum_l \pi_l \alpha_l \beta_m - \pi_m(\lambda_m + q_m + \alpha_m) = 0,$$

and thus, Equation (4.8) finally becomes

$$\begin{aligned} \frac{dH(x,m)}{dx} &= (\lambda_m + q_m + \alpha_m)H(x,m) \\ &+ \sum_{n\geq 0} \int_{\mathbf{R}_+} F_n(x-w)\Big(\lambda_m b_m(n)H(dw,m) + \sum_{l\neq m} q_{lm}a^q_{lm}(n)H(dw,l) \\ &\qquad + \sum_l \alpha_l\beta_m a^r_{lm}(n)H(dw,l)\Big). \end{aligned} \tag{4.17}$$

Equation (4.17) is in fact Takács' equation. A more compact form is obtained taking transforms on both sides of Equation (4.17). For that purpose we define

$$\tilde{H}(u,m) = \int_{\mathbf{R}_+} e^{-uw}H(dw,m),$$

$$\tilde{F}(u) = \int_{\mathbf{R}_+} e^{-uw}F(dw).$$

Then, one gets

$$\begin{aligned} u[\tilde{H}(u,m) - H(0,m)] = &-\tilde{H}(u,m)\lambda_m[\tilde{b}_m(\tilde{F}(u)) - 1] \\ &- \sum_l \tilde{H}(u,l)[q_{lm}\tilde{a}^q_{lm}(\tilde{F}(u)) + \alpha_l\beta_m\tilde{a}^r_{lm}(\tilde{F}(u))]. \end{aligned} \tag{4.18}$$

If we let $\tilde{\boldsymbol{H}}(u) = (\tilde{H}(u,1),\cdots,\tilde{H}(u,L))$ and $\boldsymbol{H}(0) = (H(0,1),\cdots,H(0,L))$, Equation (4.18) becomes

$$u[\tilde{\boldsymbol{H}}(u) - \boldsymbol{H}(0)] = -\tilde{\boldsymbol{H}}(u)[\boldsymbol{\lambda}(\tilde{\boldsymbol{b}}(\tilde{F}(u)) - \boldsymbol{I}) + \boldsymbol{Q}\circ\tilde{\boldsymbol{a}}^q(\tilde{F}(u)) + \boldsymbol{\alpha\beta}\circ\tilde{\boldsymbol{a}}^r(\tilde{F}(u))]. \tag{4.19}$$

Finally, from Equation (4.7),

$$\tilde{\boldsymbol{H}}(u)[u\boldsymbol{I} + \boldsymbol{R}(\tilde{F}(u))] = u\boldsymbol{H}(0),$$

which implies $\tilde{\boldsymbol{H}}(0) = \boldsymbol{H}(0)$ and for $u > 0$,

$$\tilde{\boldsymbol{H}}(u) = u\boldsymbol{H}(0)[u\boldsymbol{I} + \boldsymbol{R}(\tilde{F}(u))]^{-1}. \tag{4.20}$$

Equation (4.20) is the transform form of the Takács' formula for the $N/G/1/\infty$ system. It has been derived in [13] using a more elaborate approach.

In particular, letting $\alpha_j = 0$, $a^k_{ij}(n) = 0$ for $n > 0$, and $k = q, r$ and $\tilde{b}_l(z) = z$, i.e. the Markov modulated case, we get

$$\tilde{\boldsymbol{H}}(u) = u\boldsymbol{H}(0)[u\boldsymbol{I} + \boldsymbol{Q} - (1 - \tilde{F}(u))\boldsymbol{\lambda}]^{-1}.$$

5. Conclusion

The rate conservation formula we derived in this paper has a simple intuitive interpretation. The rate out of a state of a real stochastic process with jump discontinuities is equal to the average number of upcrossings minus the average number of downcrossings of that state by the process per time unit.

The practical importance of this result can be seen in the context of monitoring integrated telecommunication networks. For instance, assume that one wishes to monitor the packet delay at a network node. This delay can be studied by considering the node workload. For voice and video connections, average delays are not sufficient for an accurate quality assessment. Thus, one needs to monitor delay distributions. The rate conservation formula introduced provides an empirical method for the computation of the system delay distribution density by counting the upcrossings and downcrossings of the load levels per time unit.

This result is particularly useful for processes that leave any given state at a constant rate ($Z'(0) =$ constant). For instance, explicit results could be obtained for the residual service time density of a $G/GI/1/B$ queueing system because $R'(0) = -1$. We derived a generalized Takács formula for the $G/G/m/B$ system but no explicit solution was provided. The difficulty here lies in the evaluation of the function $r(\cdot)$. Finally, the rate conservation formula was applied for a compact derivation of the waiting time distribution of the $N/GI/1/\infty$ queueing system.

Acknowledgements

The authors would like to thank Professor Pierre Brémaud for very useful discussions throughout this work and for his comments on an earlier version of this paper, and Professor Karl Sigman for providing us with reference [10]. The research reported here was supported by the Office of Naval Research under Contract Number 00014-85-K-0371.

References

[1] Baccelli, F. and Brémaud, P. (1987) *Palm Probabilities and Stationary Queues*. Lecture Notes in Statistics 41, Springer-Verlag, Heidelberg.

[2] Brémaud, P. (1981) *Point Processes and Queues: Martingale Dynamics*. Springer-Verlag, New York.

[3] Brémaud, P. (1989) Characteristics of queueing systems observed at events and the connections between stochastic intensity and Palm probability. *QUESTA* **5**, 99–112.

[4] Brill, P. and Posner, M. (1977) The system point method in exponential queues: A level crossing approach. *Math. Operat. Res.* **6**, 31–39.

[5] Ferrandiz, J. M. (1989) Point Processes in Modeling, Analysis and Control of Integrated Networks. Ph.D. thesis, Columbia University, Department of Electrical Engineering.

[6] Franken, P., König, D., Arndt, U. and Schimdt, V. (1982) *Queues and Point Processes*. Wiley, Chichester.

[7] Karatzas, I. and Shreve, S. (1988) *Brownian Motion and Stochastic Calculus*. Springer-Verlag, New York.

[8] Kleinrock, L. (1975) *Queueing Systems*, Volume 1: *Theory*. Wiley, New York.

[9] König, D. and Schmidt, V. (1981) Relationships between time and customer stationary characteristics of service systems. In *Point Processes and Queueing Problems*, eds P. Bartfai and J. Tomko, North-Holland, Amsterdam, pp. 181–225.

[10] Miyazawa, M. (1985) The intensity conservation law for queues with randomly changed service rate. *J. Appl. Prob.* **22**, 408–418.

[11] NEUTS, M. F. (1979) A versatile markovian point process. *J. Appl. Prob.* **16**, 764–769.

[12] NEUTS, M. F. (1981) *Matrix-Geometric Solutions in Stochastic Models*. Johns Hopkins University Press, Baltimore.

[13] RAMASWAMI, V. (1980) The $N/G/1$ queue and its detailed analysis. *Adv. Appl. Prob.* **12**, 222–261.