QUEUES AND RISK MODELS WITH SIMULTANEOUS ARRIVALS

E. S. BADILA,* **

O. J. BOXMA * AND

J. A. C. RESING,* *Eindhoven University of Technology*

E. M. M. WINANDS,*** *University of Amsterdam*

Abstract

We focus on a particular connection between queueing and risk models in a multidimensional setting. We first consider the joint workload process in a queueing model with parallel queues and simultaneous arrivals at the queues. For the case that the service times are ordered (from largest in the first queue to smallest in the last queue), we obtain the Laplace–Stieltjes transform of the joint stationary workload distribution. Using a multivariate duality argument between queueing and risk models, this also gives the Laplace transform of the survival probability of all books in a multivariate risk model with simultaneous claim arrivals and the same ordering between claim sizes. Other features of the paper include a stochastic decomposition result for the workload vector, and an outline of how the two-dimensional risk model with a general two-dimensional claim size distribution (hence, without ordering of claim sizes) is related to a known Riemann boundary-value problem.

Keywords: Queue with simultaneous arrival; workload; stochastic decomposition; duality; multivariate risk model

2010 Mathematics Subject Classification: Primary 60K25
Secondary 91B30

1. Introduction

There are several connections between queueing and risk models. A classical result is that the ruin probability in the Cramér–Lundberg risk model, in which the arrival process of claims is a compound Poisson process, is related to the workload (or waiting time) in an M/G/1 queue with the same compound Poisson input. More precisely, denoting by $(R_t)_{t \geq 0}$ the surplus process in the Cramér–Lundberg risk model, by τ the time of ruin of this process, and by $(V_t)_{t \geq 0}$ the workload process in the corresponding M/G/1 queue, we have $\mathbb{P}(\tau \leq t \mid R_0 = u) = \mathbb{P}(V_t > u \mid V_0 = 0)$; in particular, the probability of ruin ever occurring when starting at u equals the probability that the steady-state workload exceeds u. See, e.g. the nice geometric duality argument on page 46 of [1], or [23].

Other ruin-related performance measures have a counterpart in queueing theory. By interpreting the interarrival times of the claims as service times of the corresponding queue and the

claim sizes as interarrival times of the queue, the standard Cramér–Lundberg model is translated into a G/M/1 queue. The time to ruin in the Cramér–Lundberg model is now related to a busy period of the corresponding queue, the deficit at ruin to an idle period, and the surplus just before ruin to the sojourn time of the last customer in a busy period (see [16] and [19]).

In this paper our focus is on a connection between queueing and risk models in a multidimensional setting. In particular, we look at the joint workload process in a queueing model with parallel queues and simultaneous arrivals at the queues. Under the condition that, with probability 1, the service times of the customers arriving at the same time at the different queues are ordered (i.e. the customer in queue 1 has the largest service time, the customer in queue 2 the second largest service time, and so on), we are able to find the Laplace–Stieltjes transform of the joint stationary workload distribution in the different queues. Using a multivariate duality argument between queueing and risk models, this immediately gives the Laplace transform of the survival (nonruin) probabilities in a multivariate risk model with simultaneous claim arrivals (and the same ordering property for the claim sizes of the simultaneous claims at the different books in the model).

Queueing models with parallel queues and simultaneous arrivals are also often called fork-join queues. These models have many applications in computer, communication, and production systems in which jobs are split among a number of different processors, communication channels, or machines. Clearly, the queues in these models are dependent due to the simultaneous arrivals. In general, this makes an exact analysis of the model very hard. Only in the case of two queues are exact results available (see, e.g. [4], [10], [12], [15], and [25]). We will come back to some of these exact results in Section 6 of the paper, where we give a partial account of these results (for a more detailed discussion of these results, see the extended version of this paper [7]). For the model with more than two servers, no exact analytical results are available in the literature. In this case, bounds and approximations for several performance measures have been developed; see, e.g. [5], [21], and [22].

Multivariate risk models with simultaneous claim arrivals have several applications in the area of ruin theory. One example is provided by reinsurance models in which, whenever a claim arrives, several insurance companies pay a part of the claim. Another example would be a large insurance company with multiple lines of business, where correlated claims arrive at the various business lines. Albeit in a different area of risk management, analysis of the dependence between the stochastic asset processes of several counter parties is also one of the most challenging aspects in the field of credit risk. Specifically, in a two-dimensional setting one has to study the joint asset process of an obligor and a guarantor in credit default swaps.

Avram *et al.* [2], [3] studied the joint ruin problem for the special case of two insurance companies that divide between them both claims and premia in some specific proportions. In particular, they derived the double Laplace transform with respect to the two initial reserves of the survival probabilities of the two companies. Proportional claims are a special case of our ordered claims, and we show in Section 4 that their survival result is indeed a special case of (7). One of the key observations in [2] and [3] is that, due to the fact that the companies divide the claims in some specific proportions, the two-dimensional ruin problem may be viewed as a one-dimensional crossing problem over a piecewise-linear barrier. Badescu *et al.* [6] extended the two-dimensional model of Avram *et al.* [2], [3] by allowing, next to the arrivals of claims for which the two insurers divide the claim in some specific proportions, extra arrivals of claims which are fully paid by one of the insurers (e.g. insurer 1). They showed that, under some conditions that also hold in this model, the previously mentioned reduction to a one-dimensional problem still holds. However, in [6] the authors did not consider the double

Laplace transform with respect to the two initial reserves of the survival probabilities of the two companies (their main focus was on the Laplace transform of the time until ruin of at least one insurer).

The remainder of the paper is organized as follows. In Section 2 we present our model in detail and we provide the multivariate duality argument. This duality argument allows a translation between results for the queueing model and results for the multivariate risk model. Section 3 is dedicated to the analysis of the two-dimensional queueing model with ordered service times. After introducing the assumptions, we derive the Laplace–Stieltjes transform of the joint stationary workloads in the two queues and present a decomposition theorem for the stationary workload in the two queues. In Section 5 we extend the results of Section 3 to the K-dimensional queueing model. Section 4 is dedicated to relations to other models. We present connections with tandem and priority queues, but also with a reinsurance problem with proportional claim sizes. In Section 6 we discuss the case of a general two-dimensional service time (or claim size) distribution. We indicate that the two-dimensional workload problem has been solved in the queueing literature. The solution is very complicated; our ordered service times case is a degenerate case, but a case which has the advantage of a much more explicit solution which offers more probabilistic insight—and a case that can be generalized to higher dimensions. Finally, in Section 7 we outline possible further research directions.

Among the main contributions of our paper, we mention an explicit result for the transform of the joint workload (and of the joint survival probability) and its extension to the K-dimensional model. In addition, we mention the workload decomposition result. It seems to be new in this setting, although similar results—under the assumption of independent inputs—were obtained for parallel queues (cf. [18]). From a more abstract perspective, another contribution of our paper is that it strengthens the links between queueing and risk models, pointing out that certain results and methods in the literature (and in the present paper) for queues with simultaneous arrivals are of immediate use in the risk setting, and vice versa.

2. Multivariate duality

We consider a K-dimensional risk process in which claims arrive simultaneously in the K branches, according to a Poisson process with rate λ. The claim sizes in the K books are independent, identically distributed random vectors $(B_n^{(1)}, \ldots, B_n^{(K)})$, $n \geq 1$. In the sequel we denote by $(B^{(1)}, \ldots, B^{(K)})$ a random vector with the same distribution as $(B_1^{(1)}, \ldots, B_1^{(K)})$.

For the nth arriving claim vector, denote by A_n the time elapsed since the arrival of the previous claim vector, so that the A_n are independent and have an identical exponential distribution with parameter λ.

Let $R_t^{(i)}$, $i = 1, \ldots, K$, be K risk reserve processes with initial capitals u_i, premium rates $c^{(i)}$, and the *same* arrival instants σ_n, $n \geq 1$. We have $A_n = \sigma_n - \sigma_{n-1}$ and $\sigma_0 = 0$ (no delay). Then

$$R_t^{(i)} = u_i + \sum_{j=1}^{n(t)} (c^{(i)} A_j - B_j^{(i)}) + c^{(i)}(t - \sigma_{n(t)}), \tag{1}$$

where $n(t)$ is the number of arrivals before t. Let $\tau^{(i)}(u_i) = \inf\{t > 0 : R_t^{(i)} < 0\}$ be the times to ruin.

In connection with the ruin process, we consider K parallel M/G/1 queues with simultaneous (coupled) arrivals and correlated service requirements. As in the ruin setting, the A_n are the interarrival times of customers in the K queues and the vector $(B^{(1)}, \ldots, B^{(K)})$ denotes the

generic service requirements. The speed of server i is denoted by $c^{(i)}$, meaning that server i handles $c^{(i)}$ units of work per time unit, $i = 1, \ldots, K$.

Furthermore, we denote by $\rho_i := \lambda \mathbb{E}(B^{(i)})$ the load of queue i, $i = 1, \ldots, K$, and we assume that $\rho_i < c^{(i)}$, to ensure that all queues can handle the offered traffic. These conditions imply positive safety loading in the ruin setting.

From the queueing perspective, let $(V_t^{(1)}, \ldots, V_t^{(K)})$ be the workload vector at time t in the system or, if we consider the nth arrival epoch, this is the workload vector $(V_n^{(1)}, \ldots, V_n^{(K)})$ seen by the customers of the nth batch arrival. Note that $V_n^{(i)} = c^{(i)} W_n^{(i)}$, where $W_n^{(i)}$ is the waiting time of the nth arrival in queue i. Under the stability conditions above, the vectors $(V_t^{(1)}, \ldots, V_t^{(K)})$ and $(V_n^{(1)}, \ldots, V_n^{(K)})$ converge in distribution to the steady-state joint workload at arbitrary epochs and at arrival epochs, respectively. Owing to the PASTA property, these vectors are equal. Similarly, the vector $(W_n^{(1)}, \ldots, W_n^{(K)})$ converges in distribution to the steady-state waiting time. The Laplace–Stieltjes transform (LST) of the steady-state workload vector is given by

$$\psi(s_1, s_2, \ldots, s_K) := \mathbb{E}(\exp[-s_1 V^{(1)} - s_2 V^{(2)} - \cdots - s_K V^{(K)}]).$$

For the multidimensional ruin process defined in (1), consider a dual workload process with $V_N^{(i)}$, $i = 1, \ldots, K$, the workload seen upon arrival by the Nth customer in K initially empty queues with the time reverted arrival process (the arrival epochs are the same for all the systems):

$$\sigma_n^* = \sigma_{N-n+1}, \qquad (A_n^* = A_{N-n+1}), \qquad n = 1, \ldots, N;$$

the service time of customer n at queue i is given by $B_n^{*(i)} = B_{N-n+1}^{(i)}$, $n = 1, \ldots, N$, (time reverted service time) (cf. [1]).

The following lemma shows that the well-known duality result (cf. [1, p. 46]) between the Cramér–Lundberg model and the M/G/1 queue can be extended to the multivariate risk model and the queueing model with simultaneous arrivals. Here the connection is between the various possibilities to be ruined (i.e. we may have ruin in all books or precisely in one, at least in one, etc.). The results below are presented for the case $K = 2$, but can be directly extended to the general case.

Lemma 1. *The following identities hold.*

(a) $\{V_N^{(1)} > u_1 \wedge V_N^{(2)} > u_2\} = \{\tau^{(1)}(u_1) \leq \sigma_N \wedge \tau^{(2)}(u_2) \leq \sigma_N\}$.

(b) $\{V_N^{(1)} \leq u_1 \wedge V_N^{(2)} \leq u_2\} = \{\tau^{(1)}(u_1) > \sigma_N \wedge \tau^{(2)}(u_2) > \sigma_N\}$.

(c) $\{V_N^{(1)} > u_1 \wedge V_N^{(2)} \leq u_2\} = \{\tau^{(1)}(u_1) \leq \sigma_N \wedge \tau^{(2)}(u_2) > \sigma_N\}$.

(d) $\{V_N^{(1)} \leq u_1 \wedge V_N^{(2)} > u_2\} = \{\tau^{(1)}(u_1) > \sigma_N \wedge \tau^{(2)}(u_2) \leq \sigma_N\}$.

The above relations are pathwise identities.

Proof. The following identities hold for the cylinder sets:

$$\{V_N^{(i)} > u_i\} = \{\tau^{(i)}(u_i) \leq \sigma_N\}, \qquad i = 1, 2.$$

This follows directly from [1, p. 46] for the one-dimensional problem, and is a special case of the duality in [24].

If we intersect the above identities, we obtain (a). Identity (b) follows by intersecting their complements, and (c) and (d) by subtracting (b) and (a), respectively, from the complements of the above cylinder sets. This concludes the proof.

If we let $N \to \infty$ in Lemma 1(b), we obtain the infinite horizon joint survival probability

$$\lim_{N \to \infty} \mathbb{P}(V_N^{(1)} \leq u_1 \wedge V_N^{(2)} \leq u_2) = \mathbb{P}(\tau^{(1)}(u_1) = \infty \wedge \tau^{(2)}(u_2) = \infty). \qquad (2)$$

Denote the right-hand side by $\xi(u_1, u_2)$. This is the joint survival function for initial capital (u_1, u_2). By the PASTA property, we can replace the steady-state workload at arrival epochs with the steady-state workload at arbitrary epochs in (2).

Let

$$\xi_*(s, t) := \int e^{-sx_1 - tx_2} \xi(x_1, x_2) \, dx_1 \, dx_2$$

be the Laplace transform (LT) of the joint survival function. Via (2), this is also the LT of the cumulative distribution function (CDF) of the joint workload in steady state. By a simple integration by parts, we have the following relation with the LST of the workload:

$$\xi_*(s, t) = \frac{1}{st} \psi(s, t). \qquad (3)$$

3. The analysis of the two-dimensional problem

In this section we derive the transform of the joint steady-state workload process of the two-dimensional queueing model with simultaneous arrivals, as introduced in Section 2. We also present a probabilistic interpretation of the quantities involved in the formula of the joint workload. The results are of immediate relevance for the corresponding insurance problem, via the duality outlined in the previous section.

Before we start with the analysis, we make the following simplifying assumption.

Assumption 1. *All the premium rates in the risk model, and the corresponding service speeds in the queueing model, are* 1, *viz.,* $c^{(1)} = \cdots = c^{(K)} = 1$.

The following observation shows that this assumption is not restrictive. If we divide all terms on the right-hand side of (1) by $c^{(i)}$, we arrive at a new risk model with initial capital $u_i / c^{(i)}$, claim size $B^{(i)} / c^{(i)}$, and *unit* premium rates. Similarly, in the corresponding queueing model the service times at queue i are also divided by $c^{(i)}$ and the service speeds are equal to 1. This will not change the nth waiting time $W_n^{(i)}$ at queue i, but the workload $V_n^{(i)}$ at the nth arrival epoch is divided by $c^{(i)}$. Also, the times to ruin are preserved; hence, the identities in Lemma 1 from the previous section remain unchanged.

The LST of the joint service time/claim size vector is denoted by

$$\phi(s, t) := \mathbb{E}(e^{-sB^{(1)} - tB^{(2)}}).$$

Our key assumption is the following.

Assumption 2. *It holds that* $\mathbb{P}(B^{(1)} \geq B^{(2)}) = 1$. *In view of the above discussion, in case the speeds are* $c^{(i)}$, *our assumption would be* $\mathbb{P}(B^{(1)}/c^{(1)} \geq B^{(2)}/c^{(2)}) = 1$.

Remark 1. This model allows for a dedicated Poisson arrival stream into queue 1. Merging this separate arrival process with the simultaneous arrival process at queue 1, the distribution of $B^{(2)}$ will have an atom in 0, which is the probability that a dedicated Poisson arrival happens instead of a simultaneous arrival (see [6] for a reinsurance model with both dedicated and simultaneous arrivals).

We are interested in the joint stationary distribution of the amount of work in the two queues:

$$\psi(s, t) := \mathbb{E}(e^{-sV^{(1)}-tV^{(2)}}).$$

This can be obtained in the following way. Consider the amount of work in queue i just before the arrival of customer n. We have the following recursion for the random variables $(V_n^{(1)}, V_n^{(2)})$, $n = 1, 2, \ldots$:

$$(V_{n+1}^{(1)}, V_{n+1}^{(2)}) = (\max(V_n^{(1)} + B_n^{(1)} - A_n, 0), \max(V_n^{(2)} + B_n^{(2)} - A_n, 0)).$$

For the LST,

$$\psi_n(s, t) = \mathbb{E}(\exp[-sV_n^{(1)} - tV_n^{(2)}]), \qquad n = 1, 2, \ldots,$$

the recursion gives, after straightforward calculations,

$$\psi_{n+1}(s, t) = \frac{\lambda}{\lambda - s - t}(\phi(s, t)\psi_n(s, t) - \phi(s, \lambda - s)\psi_n(s, \lambda - s))$$

$$+ \frac{\lambda}{\lambda - s}(\phi(s, \lambda - s)\psi_n(s, \lambda - s) - \phi(\lambda, 0)\psi_n(\lambda, 0))$$

$$+ \phi(\lambda, 0)\psi_n(\lambda, 0).$$

Under the stability condition $\rho_1 < 1$, $\psi(s, t) := \lim_{n \to \infty} \psi_n(s, t)$ exists and

$$\left(1 - \frac{\lambda\phi(s, t)}{\lambda - s - t}\right)\psi(s, t) = \left(\frac{\lambda}{\lambda - s} - \frac{\lambda}{\lambda - s - t}\right)\phi(s, \lambda - s)\psi(s, \lambda - s)$$

$$+ \left(1 - \frac{\lambda}{\lambda - s}\right)\phi(\lambda, 0)\psi(\lambda, 0). \qquad (4)$$

If we let A denote a generic interarrival time, then, owing to the PASTA property,

$$\phi(\lambda, 0)\psi(\lambda, 0) = \mathbb{P}(V^{(1)} + B^{(1)} \leq A) = \mathbb{P}(V^{(1)} = 0) = 1 - \rho_1. \qquad (5)$$

This is the probability that queue 1 is empty at an arbitrary time instant.

Let us consider the regularity domains of $\psi(s, t)$ and $\phi(s, t)$. We note that, because of the dependence $\mathbb{P}(B^{(1)} \geq B^{(2)}) = 1$, we can rewrite the transform of the joint service times as

$$\phi(s, t) = \mathbb{E}e^{-s(B^{(1)}-B^{(2)})-(s+t)B^{(2)}} =: \tilde{\phi}(s, s + t),$$

and this function is always regular in Re $s > 0$ and Re $(s + t) > 0$. If we consider $(B^{(1)}, B^{(2)})$ subject to $B^{(1)} \geq B^{(2)}$ almost surely (a.s.), $\phi(s, t)$ may not be regular beyond this domain. More precisely, if $B^{(2)}$ has a heavy-tailed distribution, this implies that $B^{(1)}$ is also heavy tailed because of the dependence structure. In this case $\phi(s, t)$ cannot be extended beyond Re $s \geq 0$ and Re $(s + t) \geq 0$. Similar considerations hold for $\psi(s, t)$ because we must also have $\mathbb{P}(V^{(1)} \geq V^{(2)}) = 1$.

It can be shown using Rouché's theorem that, for every s with Re $s > 0$, there exists a unique $t = t(s)$ with Re $t(s) > $ Re $(-s)$ that satisfies the identity $\lambda\phi(s, t) = \lambda - (s + t)$. Moreover, the function $s \to t(s)$ (which is in this case well defined) is analytic in Re $s > 0$ (for a proof of this, see [7]). Hence, the pair $(s, t(s))$ is a zero of $(1 - \lambda\phi(s, t)/(\lambda - s - t))$ in (4), which is in the regularity domain of $\psi(s, t)$. Then the right-hand side of (4) is also 0, i.e.

$$\lambda t(s)\phi(s, \lambda - s)\psi(s, \lambda - s) = -s(\lambda - t(s) - s)\phi(\lambda, 0)\psi(\lambda, 0). \qquad (6)$$

If we substitute this in (4) and use (5), we obtain

$$\psi(s, t) = (1 - \rho_1)\frac{s}{s + t - \lambda(1 - \phi(s, t))}\frac{t(s) - t}{t(s)}. \qquad (7)$$

3.1. The interpretation of the Rouché zero $t(s)$

Assume that a customer that starts a busy period $BP^{(2)}$ in queue 2 demands work x in queue 2 and work $x + y$ in queue 1. During the service time of this customer in the second queue, there are Poisson(λx) arriving customers, each generating an independent and identically distributed busy subperiod with the same distribution as $BP^{(2)}$ in queue 2. So, if we denote with U the extra work in the first queue, at the end of a busy period in the second queue, and with $U^*(s)$ its LST, we have the identity

$$U^*(s) = \int_{x=0}^{\infty} \int_{y=0}^{\infty} e^{-sy} \sum_{k=0}^{\infty} \frac{(\lambda x)^k}{k!} e^{-\lambda x} [U^*(s)]^k \, d\mathbb{P}(B^{(1)} - B^{(2)} \leq y, \ B^{(2)} \leq x).$$

The powers of $U^*(s)$ correspond to the extra work contributions at the end of the busy subperiods started during the service time of the first customer in the busy period $BP^{(2)}$. We can rewrite the above identity as

$$U^*(s) = \tilde{\phi}(s, \lambda[1 - U^*(s)]) = \phi(s, \lambda[1 - U^*(s)] - s). \tag{8}$$

Comparing this with the equation satisfied by $t(s)$, in terms of $\tilde{\phi}(s, s + t)$, we have

$$\lambda \tilde{\phi}(s, s + t(s)) = \lambda - (s + t(s)), \qquad \lambda \tilde{\phi}(s, \lambda[1 - U^*(s)]) = \lambda U^*(s).$$

We may assume without loss of generality that $\mathbb{P}(B^{(1)} > B^{(2)}) > 0$; otherwise, the two queues are a.s. identical, which is not interesting. Then it follows that the real part of $\lambda(1 - U^*(s))$ is positive, and we must have $s + t(s) = \lambda(1 - U^*(s))$ because $t(s)$ is unique in the region Re $(s + t) > 0$. We have thus proved the following result.

Proposition 1. *The relation between $t(s)$ and the transform of the extra workload in queue 1 at the end of a busy period in the shortest queue is*

$$\lambda U^*(s) = \lambda - (s + t(s)). \tag{9}$$

The transform of the joint workload in the two systems becomes

$$\psi(s, t) = (1 - \rho_1) \frac{s + t - \lambda(1 - U^*(s))}{s + t - \lambda(1 - \phi(s, t))} \frac{s}{s - \lambda(1 - U^*(s))}.$$

3.2. The workload decomposition

Based on Proposition 1, we show that the steady-state workload decomposes into an independent sum of a modified workload and an additional term, which represents the steady-state workload in a classical M/G/1 queue.

We start the joint workload process and let it run until the end of each busy period in the queue with the smallest workload. At this random time instant, we remove the extra content in queue 1, which has the largest workload of the two. Let us denote this modified joint workload process as $(\tilde{V}^{(1)}, V^{(2)})$. Then at the arrival instants of customers in the two queues, the following recurrence relation holds:

$$(\tilde{V}_{n+1}^{(1)}, V_{n+1}^{(2)}) = \begin{cases} (\tilde{V}_n^{(1)} + B_n^{(1)} - A_n, V_n^{(2)} + B_n^{(2)} - A_n) & \text{if } A_n < V_n^{(2)} + B_n^{(2)}, \\ (0, 0) & \text{if } A_n \geq V_n^{(2)} + B_n^{(2)}. \end{cases}$$

Note that, marginally, the shortest queue evolves unchanged.

If we have ergodicity then in steady state the above recurrence becomes

$$(\tilde{V}^{(1)}, V^{(2)}) \overset{\mathrm{D}}{=} \begin{cases} (\tilde{V}^{(1)} + B^{(1)} - A, V^{(2)} + B^{(2)} - A) & \text{if } A < V^{(2)} + B^{(2)}, \\ (0, 0) & \text{if } A \geq V^{(2)} + B^{(2)}. \end{cases}$$

Here and in the following '$\overset{\mathrm{D}}{=}$' denotes equality in distribution. In terms of LSTs, we obtain the following functional equation for $\tilde{\psi}(s, t) := \mathbb{E}e^{-s\tilde{V}^{(1)} - tV^{(2)}}$:

$$\left(1 - \frac{\lambda\phi(s, t)}{\lambda - s - t}\right)\tilde{\psi}(s, t) = (1 - \rho_2) - \frac{\lambda}{\lambda - s - t}\tilde{\psi}(s, \lambda - s)\phi(s, \lambda - s).$$

Here $1 - \rho_2 = \mathbb{P}(V^{(2)} = 0)$.

Now follows a similar analysis as for $\psi(s, t)$. We already know from the Rouché problem that $t(s)$ is a zero of $(1 - \lambda\phi(s, t)/(\lambda - s - t))$. We also have $\tilde{V}^{(1)} \geq V^{(2)}$ a.s. (even if we take out the extra workload at the largest queue at the end of each busy period, $\tilde{V}^{(1)}$ is still at least as large as $V^{(2)}$); therefore, $(s, t(s))$ is in the regularity domain of $\tilde{\psi}(s, t)$ and so, at the point $(s, t(s))$, the right-hand side of the above identity is equal to 0:

$$\tilde{\psi}(s, \lambda - s)\phi(s, \lambda - s) = (1 - \rho_2)\frac{\lambda - s - t(s)}{\lambda}.$$

Substituting this back into the original identity yields

$$\tilde{\psi}(s, t) = (1 - \rho_2)\frac{s + t - \lambda(1 - \phi(s, t(s)))}{s + t - \lambda(1 - \phi(s, t))}. \tag{10}$$

This is a two-dimensional Pollaczek–Khinchine type of representation. From an analytic point of view, the role of the numerator is to cancel the unique pole of the denominator in the region Re $(s + t) > 0$.

Substitute $t(s)$ from Proposition 1 and $\tilde{\psi}$ from (10) into (7):

$$\psi(s, t) = \frac{1 - \rho_1}{1 - \rho_2}\frac{s}{s - \lambda[1 - U^*(s)]}\tilde{\psi}(s, t). \tag{11}$$

We can now state the main result.

Theorem 1. (Work decomposition.) *In steady state, we have the following representation of the joint workload at the two queues as an independent sum:*

$$(V^{(1)}, V^{(2)}) \overset{\mathrm{D}}{=} (\tilde{V}^{(1)}, V^{(2)}) + (V^{(1),1}, 0).$$

Here $V^{(1),1}$ is the workload in an independent, virtual M/G/1 queue with arrival rate λ and service requirements distributed as U, the extra workload at the end of a busy period $\mathrm{BP}^{(2)}$ in the shortest queue.

Proof. It suffices to note that the factor

$$\frac{1 - \rho_1}{1 - \rho_2}\frac{s}{s - \lambda[1 - U^*(s)]} = \mathbb{E}e^{-sV^{(1),1}}$$

in (11) is the Pollaczek–Khinchine formula for the transform of the workload in the virtual M/G/1 queue with service time distribution U. This virtual queue is obtained by contracting

the busy periods in the initial shortest queue, so that an arrival in the virtual queue happens at the end of this busy period and the interarrival time is then the idle period in the initial queue, and so is exponentially distributed.

To see that indeed $(1 - \rho_1)/(1 - \rho_2)$ is the atom of $V^{(1),1}$ at 0, differentiate the identity for $U^*(s)$ in (8):

$$\mathbb{E}(U) = -\frac{\mathrm{d}}{\mathrm{d}s}\phi(s, \lambda(1 - U^*(s) - s))\bigg|_{s=0} = \mathbb{E}(B^{(1)} - B^{(2)}) + \lambda\mathbb{E}B^{(2)}\mathbb{E}(U);$$

so $1 - \lambda\mathbb{E}(U) = (1 - \rho_1)/(1 - \rho_2)$.

4. Relation with other models

In this section we point out how the results of the previous section are related to results for a risk model with proportional reinsurance, a particular tandem fluid model and with a particular priority queue. We start by showing that (7) generalizes a result obtained in [2] for the risk setting.

4.1. The case of proportional reinsurance

In [2] the joint reserve process $(R^{(1)}, R^{(2)})$ is of the form $R^{(i)}(t) = u_i + c^{(i)}t/\delta_i - S(t)$. Here $S(t)$ is a common compound Poisson input process with generic claim sizes σ, and $c^{(i)}$ are the premium rates. The claims are being divided in fixed proportions δ_i.

To bring this closer to our setting in Section 3, normalize the income rates, i.e. we consider $(R^{(1)}/p_1, R^{(2)}/p_2)$ with $p_i = c^{(i)}/\delta_i$. The assumption in [2] is that $p_1 > p_2$, which means that, in our notation, the claim sizes are $B^{(1)} := \sigma/p_1 < \sigma/p_2 =: B^{(2)}$. Note that the inequality between the $B^{(i)}$ is reversed here (which means that the role of the arguments in our transforms is interchanged, especially the Rouché zero).

Let us recall the main formula in [2, Formula (23)]:

$$\psi_{*R^{(1)}, R^{(2)}}(p, q) = \frac{\kappa_2(0+)'}{p(\kappa_1(p + q) - q(p_1 - p_2))} \frac{q + p - q^+(q(p_1 - p_2))}{q - q^+(q(p_1 - p_2))}. \qquad (12)$$

The relation between the ruin times of $(R^{(1)}, R^{(2)})$ and $(R^{(1)}/p_1, R^{(2)}/p_2)$ is

$$\tau_{R^{(1)}/p_1, R^{(2)}/p_2}(u_1, u_2) = \tau_{R^{(1)}, R^{(2)}}(p_1 u_1, p_2 u_2).$$

Hence, the relation to the LT coordinates used in (3) is $s = p_1 p$ and $t = p_2 q$. From this, the relation between the LT of the survival functions becomes, after a change of variables,

$$\psi_{*R^{(1)}/p_1, R^{(2)}/p_2}(s, t) = \frac{1}{p_1 p_2}\psi_{*R^{(1)}, R^{(2)}}(p, q). \qquad (13)$$

- $\kappa_i(\alpha)$ is the Laplace exponent of the compound Poisson process with drift p_i per unit time. This means that

$$\kappa_i(\alpha) = p_i\alpha - \lambda(1 - \mathbb{E}e^{-\alpha\sigma}).$$

Owing to the linear dependence between the $B^{(i)}$, their LST has the form $\mathbb{E}e^{-sB^{(1)} - tB^{(2)}} = \phi(s, t) =: \phi_{B^{(1)}}(s + p_1 t/p_2)$.

- $q^+(q)$ is the largest root of the equation $\kappa_1(\alpha) = q$. Then $q^+(q(p_1 - p_2))$ solves

$$p_1\alpha - \lambda(1 - \mathbb{E}e^{-\alpha p_1 B^{(1)}}) = q(p_1 - p_2).$$

Note that if we set $\alpha = p + q$, the above becomes

$$p_1 p + p_2 q - \lambda(1 - \phi_{B^{(1)}}(p_1 p + p_1 q)) = 0,$$

or, written in the (s, t) coordinates, this becomes the equation satisfied by $s(t)$ (s and t are now interchanged). Hence, the relation between the 0s in both notation is $s(t) = p_1(\alpha - q) = p_1[q^+(q(p_1 - p_2)) - q]$.

The constant $\kappa(0+)' = p_2 - \lambda\mathbb{E}B^{(2)} = p_2(1 - \rho_2)$ is the probability that the queueing system is empty in steady state (now the second queue has a higher workload).

In conclusion, (12) written via (13) and (3) in the (s, t) coordinates becomes (7), i.e.

$$\psi(t, s) = \frac{s(1 - \rho_2)}{s + t - \lambda(1 - \phi_{B^{(1)}}(s + p_1 t/p_2))} \frac{s - s(t)}{-s(t)},$$

with the arguments s and t interchanged.

4.2. Relation with work on tandem fluid queues

We now show that the workload model with ordered service times is equivalent to a particular tandem fluid queue. That is, a model of two queues in series in which the outflow from the first queue is a fluid, i.e. there is continuous outflow when the server is working (instead of customers leaving one by one). Such tandem fluid queues have been studied by various authors; see, in particular, [18]. Consider the following two-station tandem fluid network with *independent* compound Poisson input at the two stations (with arrival rate λ_i and LST of the service times $B_i^*(\cdot)$, $i = 1, 2$). Then Theorem 4.1 of [18] gives the LST of the steady-state fluid levels W_1 and W_2 in the two nodes as

$$\psi_W(\alpha_1, \alpha_2) = \mathbb{E}(e^{-\alpha_1 W_1 - \alpha_2 W_2}) = \frac{(1 - \rho_1 - \rho_2)\alpha_2}{\phi_1(\alpha_1) - \phi_1(\hat{\eta}_2(\alpha_2))} \frac{\alpha_1 - \hat{\eta}_2(\alpha_2)}{\alpha_2 - \hat{\eta}_2(\alpha_2)},$$

with

- $\rho_i = \lambda_i \mathbb{E}(B_i)$,

- $\phi_1(\alpha_1) = \alpha_1 - \eta_1(\alpha_1)$,

- $\eta_i(\alpha_i) = \lambda_i(1 - B_i^*(\alpha_i))$,

- $\hat{\eta}_2(\alpha_2)$ the solution of $\phi_1(\hat{\eta}_2(\alpha_2)) = \eta_2(\alpha_2)$.

Alternatively, the last relation can also be formulated as follows: $\hat{\eta}_2(\alpha_2)$ is the solution of

$$\lambda_1 B_1^*(\hat{\eta}_2(\alpha_2)) + \lambda_2 B_2^*(\alpha_2) = \lambda_1 + \lambda_2 - \hat{\eta}_2(\alpha_2).$$

This system is related to our model with arrival rate $\lambda = \lambda_1 + \lambda_2$ and LST of service requirements

$$\phi(s, t) = \frac{\lambda_1}{\lambda_1 + \lambda_2} B_1^*(s + t) + \frac{\lambda_2}{\lambda_1 + \lambda_2} B_2^*(s).$$
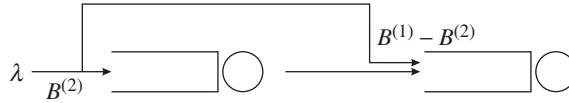
FIGURE 1: Tandem fluid queue

The corresponding notation is $B_1 \overset{\mathrm{D}}{=} B^{(2)}$ and $B_2 \overset{\mathrm{D}}{=} B^{(1)} - B^{(2)}$ (see Figure 1). Here W_1 in the tandem model corresponds to the workload in the smallest queue in our model and $W_1 + W_2$ in the tandem model corresponds to the workload in the largest queue in our model. So we have

$$
\begin{aligned}
\psi(s, t) &= \mathbb{E}(e^{-s V_1 - t V_2}) = \mathbb{E}(e^{-s(W_1 + W_2) - t W_1}) \\
&= \psi_W(s + t, s) \\
&= \frac{(1 - \rho_1 - \rho_2)s}{s + t - \lambda_1(1 - B_1^*(s + t)) - \lambda_2(1 - B_2^*(s))} \frac{s + t - \hat{\eta}_2(s)}{s - \hat{\eta}_2(s)}.
\end{aligned}
$$

Now note that

- The total traffic offered to the largest queue is $\rho_1 + \rho_2$, so indeed the factor $1 - \rho_1 - \rho_2$ in [18] corresponds to the factor $1 - \rho_1$ in (7);
- $\lambda(1 - \phi(s, t)) = \lambda_1(1 - B_1^*(s + t)) + \lambda_2(1 - B_2^*(s))$;
- $\lambda \phi(s, t(s)) = \lambda_1 B_1^*(s + t(s)) + \lambda_2 B_2^*(s) = \lambda_1 + \lambda_2 - (s + t(s))$, so indeed $\hat{\eta}_2(s)$ corresponds to our $s + t(s)$.

We conclude that (7) coincides with Theorem 4.1 of [18] in the case of independent compound Poisson input. Kella's result is more general in the sense that he has Lévy input instead of compound Poisson input. Our result is more general in the sense that we have *dependent* compound Poisson input.

4.3. Relation with work on priority queues

As already noted in Kella [18], but also in several other places in the literature, the tandem fluid network described above is also related to a priority queue with preemptive resume priorities. Hence, the same holds for our workload model. Consider the following model with two types of customer, where customers of type i arrive according to a Poisson process with rate λ_i having service times with LST $B_i^*(\cdot)$, $i = 1, 2$. Assume furthermore that customers of type 1 have preemptive resume priority over customers of type 2. If we denote by Y_1 and Y_2 the steady-state workloads in the two queues, then Y_1 and Y_2 are related to W_1 and W_2 in the tandem fluid network. The LST of the steady-state workloads in the two queues satisfies

$$
\psi_Y(s, t) = \mathbb{E}(e^{-s Y_1 - t Y_2}) = \mathbb{E}(e^{-s W_1 - t W_2}) = \mathbb{E}(e^{-s V_2 - t(V_1 - V_2)}) = \psi_V(t, s - t),
$$

where again in our model we have to take arrival rate $\lambda = \lambda_1 + \lambda_2$ and LST of service requirements

$$
\phi(s, t) = \frac{\lambda_1}{\lambda_1 + \lambda_2} B_1^*(s + t) + \frac{\lambda_2}{\lambda_1 + \lambda_2} B_2^*(s).
$$

We conclude that (7) also gives the LST of a priority queue. Again, our result is more general in the sense that we have *dependent* compound Poisson input (i.e. we can have arrivals of customers who have both low and high priority work).

5. The K-dimensional problem

In this section we consider the K-queue system with simultaneous arrivals. We give the transform for the steady-state joint workload and we show that the decomposition in Theorem 1 extends to this case if we preserve the ordering between the service requirements/claim sizes. We use an iterative argument and, for this purpose, the decomposition in Section 3 will be the starting point; the iteration step is essentially done with the help of Lemma 2 below as a work conservation identity.

We thus consider K parallel M/G/1 queues, numbered 1 to K, respectively, with simultaneous (coupled) arrivals and correlated service requirements. We use the same notation as in Section 2. The LST of the service time/claim size vector is denoted by

$$\phi(s_1, \ldots, s_K) := \mathbb{E}(\exp[-s_1 B^{(1)} - \cdots - s_K B^{(K)}]).$$

The essential assumption in the model extends Assumption 2 for the two-dimensional problem:

$$\mathbb{P}(B^{(1)} \geq B^{(2)} \geq \cdots \geq B^{(K)}) = 1.$$

Furthermore, we denote by $\rho_i := \lambda \mathbb{E} B^{(i)}$, $i = 1, \ldots, K$, the load of queue i and we assume that $\rho_1 < 1$ (hence, $\rho_i < 1$ for all i) to assure that all queues can handle the offered work.

Remark 2. As in the two-dimensional case (cf. Remark 1), this model allows for a separate Poisson arrival stream into queue 1. Merging this separate arrival process with the simultaneous arrival process, the distribution of $(B^{(2)}, \ldots, B^{(K)})$ will have an atom in $(0,\ldots,0)$, which is the probability that a dedicated Poisson arrival happens instead of a simultaneous arrival.

Similarly, the model allows for simultaneous arrivals at the first j queues only. This can be achieved by letting the distribution of $(B^{(j+1)}, \ldots, B^{(K)})$ have an atom at $(0, \ldots, 0)$.

5.1. The LST of $(V^{(1)}, \ldots, V^{(K)})$

The K-dimensional Lindley recursion holds for the random variables $(V_n^{(1)}, \ldots, V_n^{(K)})$:

$$(V_{n+1}^{(1)}, \ldots, V_{n+1}^{(K)}) = (\max(V_n^{(1)} + B_n^{(1)} - A_n, 0), \ldots, \max(V_n^{(K)} + B_n^{(K)} - A_n, 0)).$$

For $\psi_n(s_1, \ldots, s_K) := \mathbb{E}(\exp[-s_1 V_n^{(1)} - \cdots - s_K V_n^{(K)}]), n \geq 1$, the Lindley recursion gives, after straightforward calculations,

$$
\begin{aligned}
\psi_{n+1}&(s_1, \ldots, s_K) \\
&= \sum_{j=1}^{K} \frac{\lambda}{\lambda - \sum_{i=1}^{j} s_i} [\phi^{(j)}(s_1, \ldots, s_j) \psi_n^{(j)}(s_1, \ldots, s_j) \\
&\qquad\qquad - \phi^{(j-1)}(s_1, \ldots, s_{j-1}) \psi_n^{(j-1)}(s_1, \ldots, s_{j-1})] + \phi^{(0)} \psi_n^{(0)}, \quad (14)
\end{aligned}
$$

where we have used the following notation for simplicity: $\psi_n^{(K)}(s_1, \ldots, s_K) := \psi_n(s_1, \ldots, s_K)$ with $\psi_n^{(0)} := \psi_n(\lambda, 0, \ldots, 0)$, and

$$\psi_n^{(j)}(s_1, \ldots, s_j) := \psi_n\left(s_1, \ldots, s_j, \lambda - \sum_{i=1}^{j} s_i, \underbrace{0, \ldots, 0}_{K-j-1 \text{ arguments}}\right) \quad \text{for } 1 \leq j \leq K-1.$$

We define $\phi^{(j)}(s_1, \ldots, s_j)$ analogously for $j = 0, \ldots, K$. By taking $n \to \infty$ in (14), we obtain, for $\psi(s_1, \ldots, s_K) := \lim_{n \to \infty} \psi_n(s_1, \ldots, s_K)$,

$$
\left(1 - \frac{\lambda \phi(s_1, \ldots, s_K)}{\lambda - \sum_{i=1}^{K} s_i}\right) \psi(s_1, \ldots, s_K)
$$
$$
= \sum_{j=0}^{K-1} \left(\frac{\lambda}{\lambda - \sum_{i=1}^{j} s_i} - \frac{\lambda}{\lambda - \sum_{i=1}^{j+1} s_i}\right) \phi^{(j)}(s_1, \ldots, s_j) \psi^{(j)}(s_1, \ldots, s_j), \qquad (15)
$$

with $\psi^{(j)} := \lim_{n \to \infty} \psi_n^{(j)}$ and $\phi^{(0)} \psi^{(0)} = \mathbb{P}(V^{(1)} + B^{(1)} \leq A) = 1 - \rho_1$.

Formula (15) has a simple recursive structure, and we can rewrite it as

$$
\left(1 - \frac{\lambda \phi(s_1, \ldots, s_K)}{\lambda - \sum_{i=1}^{K} s_i}\right) \psi(s_1, \ldots, s_K)
$$
$$
= \left(\frac{\lambda}{\lambda - \sum_{i=1}^{K-1} s_i} - \frac{\lambda}{\lambda - \sum_{i=1}^{K} s_i}\right) \phi^{(K-1)}(s_1, \ldots, s_{K-1}) \psi^{(K-1)}(s_1, \ldots, s_{K-1})
$$
$$
+ \left(1 - \frac{\lambda \phi(s_1, \ldots, s_{K-1}, 0)}{\lambda - \sum_{i=1}^{K-1} s_i}\right) \psi(s_1, \ldots, s_{K-1}, 0). \qquad (16)
$$

Denote by

$$
C_j := \left(1 - \frac{\lambda \phi(s_1, \ldots, s_j, 0, \ldots, 0)}{\lambda - \sum_{i=1}^{j} s_i}\right) \psi(s_1, \ldots, s_j, 0, \ldots, 0),
$$

and note that $\psi(s_1, \ldots, s_j, 0, \ldots, 0)$ is the transform of the workload in the j-dimensional system obtained by ignoring the last $K - j$ queues, $j = 1, \ldots, K$.

Proposition 2. *The LST of the steady-state workload in the $K \geq 3$ systems is given by*

$$
\psi(s_1, \ldots, s_K) = \frac{(1 - \rho_K)(S_K - s_K)}{\sum_{i=1}^{K} s_i - \lambda(1 - \phi(s_1, \ldots, s_K))} \prod_{j=2}^{K-1} \frac{1 - \rho_j}{1 - \rho_{j+1}} \frac{S_j - s_j}{S_{j+1}} \frac{1 - \rho_1}{1 - \rho_2} \frac{s_1}{S_2}, \qquad (17)
$$

with $S_j = S_j(s_1, \ldots, s_{j-1})$ the unique solution of the equation

$$
\lambda \phi(s_1, \ldots, s_j, 0, \ldots, 0) = \lambda - \sum_{i=1}^{j} s_i,
$$

with $\operatorname{Re}(s_1 + \cdots + s_{j-1} + S_j(s_1, \ldots, s_{j-1})) > 0$ *for all* $j = 2, \ldots, K$.

Proof. The key remark is that s_K is not among the arguments of the functions $\psi^{(j)}$ that appear on the right-hand side of (15) or (16). Similarly as in Section 3, Rouché's theorem applied to $s = s_1 + \cdots + s_{K-1}$ and $t = s_K$ yields the existence of a unique solution $S_K = S_K(s_1, \ldots, s_{K-1})$ of the equation

$$
\lambda \phi(s_1, \ldots, s_K) = \lambda - \sum_{i=1}^{K} s_i,
$$

such that $S_K(s_1, \ldots, s_{K-1}) + \sum_{i=1}^{K-1} s_i$ has positive real part. Hence, the hypersurface given by $S_K = S_K(s_1, \ldots, s_{K-1})$ is contained in the regularity domain of $\psi(s_1, \ldots, s_K)$, and then the right-hand side of (16) must be 0. This gives the following relation for $\psi^{(K-1)}(s_1, \ldots, s_{K-1})$:

$$(\phi^{(K-1)}\psi^{(K-1)})(s_1, \ldots, s_{K-1}) = \frac{(\lambda - \sum_{i=1}^{K-1} s_i)\phi(s_1, \ldots, s_{K-1}, S_K)}{S_K} C_{K-1}.$$

By substituting this into (16), we obtain the recursion

$$C_K = \frac{\lambda - \sum_{i=1}^{K-1} s_i}{\lambda - \sum_{i=1}^{K} s_i} \frac{S_K - s_K}{S_K} C_{K-1}$$

with initial condition

$$C_2 = -(1 - \rho_1) \frac{s_1}{\lambda - s_1 - s_2} \frac{S_2 - s_2}{S_2},$$

which follows from (7). From this, we obtain (17), after rearranging the factors. The proof is complete.

5.2. Interpretation of the Rouché zero

It is worthwhile to change the coordinates: $(s_1, s_2, \ldots, s_K) \to (s_1, s_2, \ldots, s_{K-1}, \sum_{i=1}^{K} s_i)$. We can rewrite

$$\phi(s_1, \ldots, s_K) = \mathbb{E} \exp\left[-s_1(B^{(1)} - B^{(K)}) - \cdots - s_{K-1}(B^{(K-1)} - B^{(K)}) - \left(\sum_{i=1}^{K} s_i\right) B^{(K)}\right].$$

Let us denote it by $\tilde{\phi}(s_1, \ldots, s_{K-1}, \sum_{i=1}^{K} s_i)$. This is the transform of the extra service time (relative to the shortest queue) in the first $K - 1$ queues, together with the service time in the shortest queue. It turns out there is a connection between $s_K(s_1, \ldots, s_{K-1})$ and the joint extra work in systems 1 to $K - 1$ at the end of a busy period in system K. Let us denote this extra work by $(U_1, U_2, \ldots, U_{K-1})$, with LST $U_K^*(s_1, \ldots, s_{K-1})$, and let $F(x_1, x_2, \ldots, x_K)$ be the CDF of $(B^{(1)} - B^{(K)}, \ldots, B^{(K-1)} - B^{(K)}, B^{(K)})$. Then, by a similar argument as that leading to (8), $U_K^*(s_1, \ldots, s_{K-1})$ satisfies the identity

$$
\begin{aligned}
U_K^*(s_1, \ldots, s_{K-1}) &= \int \exp\left[-\sum_{i=1}^{K-1} s_i x_i\right] \sum_{n=0}^{\infty} \frac{(\lambda x_K)^n}{n!} e^{-\lambda x_K} [U_K^*(s_1, \ldots, s_{K-1})]^n \\
&\quad \times F(dx_1 \cdots dx_K) \\
&= \tilde{\phi}(s_1, \ldots, s_{K-1}, \lambda[1 - U_K^*(s_1, \ldots, s_{K-1})]).
\end{aligned}
\tag{18}
$$

Comparing this with the identity for the Rouché zero,

$$\lambda - (s_1 + \cdots + s_{K-1} + S_K) = \lambda \tilde{\phi}(s_1, \ldots, s_{K-1}, s_1 + \cdots + s_{K-1} + S_K),$$

gives a relation analogous to (9), i.e.

$$\lambda U_K^*(s_1, \ldots, s_{K-1}) = \lambda - (s_1 + \cdots + s_{K-1} + S_K), \tag{19}$$

which follows because the Rouché zero is unique.

Let us fix our attention on the case $K = 3$ for the moment. Then identity (17) becomes

$$\psi(s_1, s_2, s_3) = \frac{(1 - \rho_3)(S_3 - s_3)}{s_1 + s_2 + s_3 - \lambda[1 - \phi(s_1, s_2, s_3)]} \frac{1 - \rho_2}{1 - \rho_3} \frac{S_2 - s_2}{S_3} \frac{1 - \rho_1}{1 - \rho_2} \frac{s_1}{S_2}. \tag{20}$$
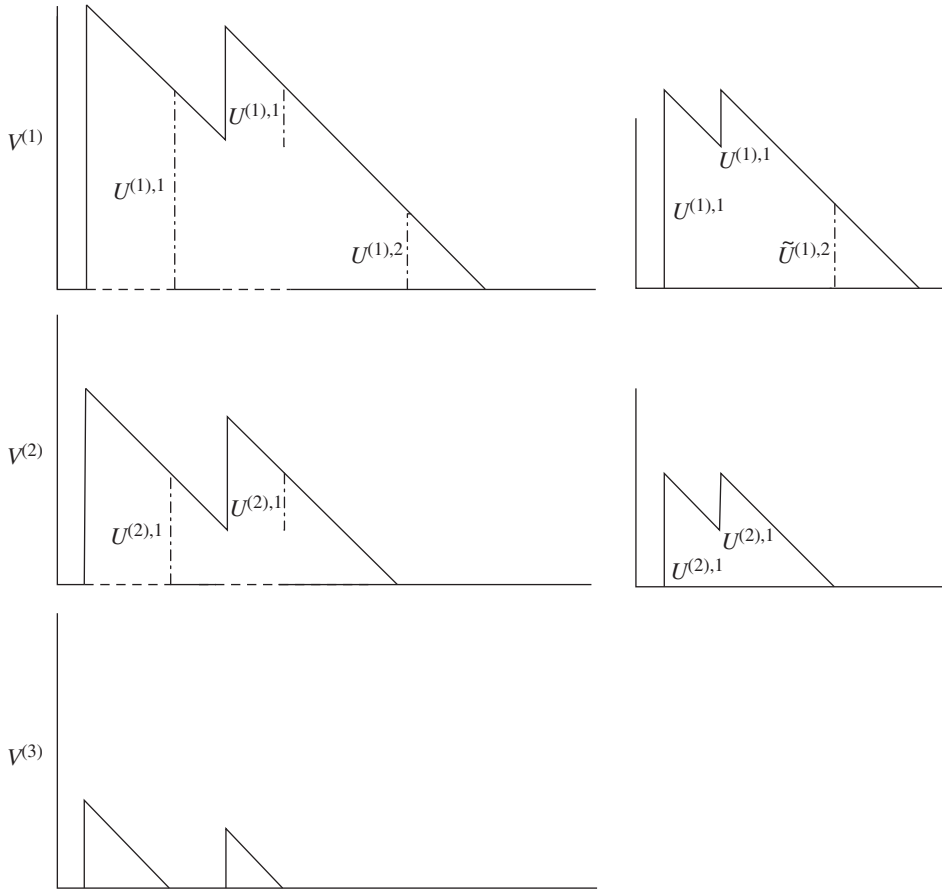
FIGURE 2: Work in the original system (*left*) and in the virtual system (*right*).

5.3. Work conservation

We would like to give a probabilistic interpretation of (20). In order to achieve this, we start by considering the joint extra work in queues 1 and 2 at the end of a busy period in queue 3. This has LST $U_3^*(s_1, s_2)$ as input in a two-dimensional system with simultaneous Poisson arrivals, which is obtained by contracting the busy cycles in queue 3. We call this the two-dimensional virtual system. Note that the interarrival times in the virtual system are precisely the idle periods in queue 3.

For this construction, the key observation is that the steady-state extra work in virtual queue 1 at the end of the busy period in virtual queue 2 is the same as the extra work in initial queue 1 at the end of the busy period in original queue 2. In analytic form, let $\tilde{U}_2^*(s_1)$ be the LST of the extra work in the virtual system and let $U_2^*(s_1)$ be the LST of the extra work in the original system, see Figure 2.

Lemma 2. *It holds that*

$$\tilde{U}_2^*(s_1) = U_2^*(s_1).$$

Proof. We begin by noting that the extra work $(U^{(1),1}, U^{(2),1})$ in the first two queues at the end of a busy period in queue 3 satisfies the inequality $U^{(1),1} \geq U^{(2),1}$ a.s. Since this is the

input in the virtual system, from Proposition 1, $\tilde{U}_2^*(s_1)$ satisfies identity (8) with $U_3^*(s_1, s_2)$ instead of $\phi(s_1, s_2)$:

$$U_3^*(s_1, \lambda[1 - \tilde{U}_2^*(s_1)] - s_1) = \tilde{U}_2^*(s_1). \tag{21}$$

At the same time, via (18), $U_3^*(s_1, s_2)$ satisfies

$$\phi(s_1, s_2, \lambda(1 - U_3^*(s_1, s_2)) - s_1 - s_2) = U_3^*(s_1, s_2).$$

If we substitute this fixed point identity into (21), we have

$$\phi(s_1, \lambda(1 - \tilde{U}_2^*(s_1)) - s_1, 0) = \tilde{U}_2^*(s_1).$$

On the other hand, this is also identity (8) satisfied by $U_2^*(s_1)$ in the two-dimensional system obtained by ignoring the last queue. Hence, from the uniqueness of Rouché's zero, $\tilde{U}_2^*(s_1) = U_2^*(s_1)$ (See Figure 2). This completes the proof.

We can rewrite (17) using (19):

$$\psi(s_1, s_2, s_3) = (1 - \rho_3) \frac{s_1 + s_2 + s_3 - \lambda(1 - U_3^*(s_1, s_2))}{s_1 + s_2 + s_3 - \lambda(1 - \phi(s_1, s_2, s_3))}$$
$$\times \frac{1 - \rho_2}{1 - \rho_3} \frac{s_1 + s_2 - \lambda(1 - U_2^*(s_1))}{s_1 + s_2 - \lambda(1 - U_3^*(s_1, s_2))} \frac{1 - \rho_1}{1 - \rho_2} \frac{s_1}{s_1 - \lambda(1 - \tilde{U}_2^*(s_1))}. \tag{22}$$

Note that the atom $(1 - \rho_1)/(1 - \rho_2)$ above is the conditional probability that queue 1 is empty, given that queue 2 is empty; and, similarly, for $(1 - \rho_2)/(1 - \rho_3)$. In addition, the last factor in (22) is the Pollaczek–Khinchine representation for an M/G/1 queue with service times having LST $\tilde{U}_2^*(s_1)$. Now we are ready to give the main result of this section.

Theorem 2. *In steady state, the joint workload distribution decomposes as an independent sum:*

$$(V^{(1)}, V^{(2)}, V^{(3)}) \overset{\mathrm{D}}{=} (\tilde{V}^{(1),1}, \tilde{V}^{(2),1}, V^{(3)}) + (\tilde{V}^{(1),2}, V^{(2),2}, 0) + (V^{(1),3}, 0, 0).$$

The first term in the sum represents the steady-state distribution of the modified joint workload process obtained by removing the extra work in the first two queues at the end of a busy period in the third queue. The second term is the workload in the first two queues obtained by removing the extra work in the first queue at the end of a busy cycle in the second queue. Finally, the third term represents the workload in the virtual M/G/1 queue with input distributed as the extra work in queue 1, at the end of a busy period in queue 2.

Proof. Consider the modified work process that evolves in steady state as

$$(\tilde{V}^{(1),1}, \tilde{V}^{(2),1}, V^{(3)}) \overset{\mathrm{D}}{=} (\tilde{V}^{(1),1} + B^{(1)} - A, \tilde{V}^{(2),1} + B^{(2)} - A, V^{(3)} + B^{(3)} - A)$$

if $A < V^{(3)} + B^{(3)}$, and $(\tilde{V}^{(1),1}, \tilde{V}^{(2),1}, V^{(3)}) = (0, 0, 0)$ otherwise.

By similar computations as those leading to (10), we obtain

$$\tilde{\psi}(s_1, s_2, s_3) = (1 - \rho_3) \frac{s_1 + s_2 + s_3 - \lambda(1 - U_3^*(s_1, s_2))}{s_1 + s_2 + s_3 - \lambda(1 - \phi(s_1, s_2, s_3))}.$$

This is the first factor in (22). For the second factor, consider the following modified virtual workload process that evolves in steady state as

$$(\tilde{V}^{(1),2}, V^{(2),2}, 0)$$
$$\stackrel{\text{D}}{=} \begin{cases} (\tilde{V}^{(1),2} + U^{(1),1} - A, V^{(2),2} + U^{(2),1} - A, 0) & \text{if } A < V^{(2),2} + U^{(2),1}, \\ (0, 0, 0) & \text{if } A \geq V^{(2),2} + U^{(2),1}, \end{cases}$$

with $(U^{(1),1}, U^{(2),1})$ the extra work vector in the first two queues at the end of a busy period in queue 3. Here we remove the excess workload in virtual queue 1 at the end of the busy period in virtual queue 2, which, by Lemma 2, is the same as in the original system. In terms of LSTs, this becomes

$$\tilde{\psi}_1(s_1, s_2) = \frac{1 - \rho_1}{1 - \rho_2} \frac{s_1 + s_2 - \lambda(1 - U_2^*(s_1))}{s_1 + s_2 - \lambda(1 - U_3^*(s_1, s_2))}.$$

Finally, the third factor in (22) is the Pollaczek–Khinchine representation of the steady-state workload in the M/G/1 queue with service time distributed as the extra work in queue 1 at the end of a busy period in queue 2. This completes the proof.

These considerations can be iterated now for the general K-dimensional system.

Corollary 1. *The steady-state joint workload in the K systems decomposes into the independent sum*

$$(V^{(1)}, \ldots, V^{(K)}) \stackrel{\text{D}}{=} (\tilde{V}^{(1),1}, \ldots, \tilde{V}^{(K-1),1}, V^{(K)}) + (\tilde{V}^{(1),2}, \ldots, \tilde{V}^{(K-2),2}, V^{(K-1),2}, 0)$$
$$+ \cdots + (\tilde{V}^{(1),K-1}, V^{(2),K-1}, 0, \ldots, 0) + (V^{(1),K}, 0, \ldots, 0),$$

where the jth term in the sum satisfies the identity in distribution (j = 2, \ldots, K):

$$(\tilde{V}^{(1),j}, \tilde{V}^{(2),j}, \ldots, \tilde{V}^{(K-j),j}, V^{(K-j+1),j}, 0, \ldots, 0)$$
$$\stackrel{\text{D}}{=} (\tilde{V}^{(1),j} + U^{(1),j-1} - A, \tilde{V}^{(2),j} + U^{(2),j-1} - A, \ldots,$$
$$V^{(K-j+1),j} + U^{(K-j+1),j-1} - A, 0, \ldots, 0) \quad \text{if } A \leq V^{(K-j+1),j} + U^{(K-j+1),j-1},$$

and $(0, \ldots, 0)$ otherwise. Here $U^{(i),j}$ is the extra workload in queue i at the end of a busy period in queue $(K - j + 1)$ for $i > K - j + 1$.

6. The general two-dimensional workload/reinsurance problem

In this section we consider the general two-dimensional workload problem: pairs of customers arrive simultaneously at two parallel queues Q_1 and Q_2 according to a Poisson(λ) process, the nth pair requiring service times $(B_n^{(1)}, B_n^{(2)})$ with LST $\phi(s, t)$. We are interested in the steady-state workload vector $(V^{(1)}, V^{(2)})$ with LST $\psi(s, t)$. By the duality that is exposed in Section 2, $\psi(s, t)$ also is the LT (with respect to u_1 and u_2) of the probability that both portfolios of an insurance company with simultaneous claims $(B_n^{(1)}, B_n^{(2)})$, with initial capital u_1 and u_2, will survive.

In Section 3 we determined $\psi(s, t)$ for the special case that $\mathbb{P}(B^{(1)} \geq B^{(2)}) = 1$. We now show how the general case—$B_n^{(1)}$ and $B_n^{(2)}$ having an arbitrary joint distribution—has been solved in the literature (with the solution of that special case emerging as a degenerate solution). Baccelli [4], De Klein [12], and Cohen [10] treated the two-dimensional workload problem with simultaneous arrivals in increasing generality. The starting point in those three studies is the following functional equation for $\psi(s, t)$, which is derived by studying the two-dimensional

Markovian workload process during an infinitesimal amount of time Δt:

$$K(s, t)\psi(s, t) = t\psi_1(s) + s\psi_2(t), \qquad \operatorname{Re} s, t \geq 0. \tag{23}$$

Here the so-called *kernel* $K(s, t)$ is given by

$$K(s, t) := s + t - \lambda(1 - \phi(s, t)),$$

and

$$\psi_1(s) := \mathbb{E}[\mathrm{e}^{-sV_1}(V_2 = 0)], \qquad \psi_2(t) := \mathbb{E}[\mathrm{e}^{-tV_2}(V_1 = 0)],$$

with $(\cdot)$ denoting an indicator function.

Remark 3. In the special case of Section 3, with $\mathbb{P}(B^{(1)} \geq B^{(2)}) = 1$, we have $\psi_2(t) \equiv \mathbb{P}(V_1 = 0)$, because V_2 cannot be positive when $V_1 = 0$. It then remains to find $\psi_1(s)$. This is done by observing (see Section 3) that, for all s with $\operatorname{Re} s > 0$, there is a unique zero $t(s)$ of the kernel, with $\operatorname{Re} t(s) > \operatorname{Re}(-s)$. This immediately yields $\psi_1(s) = -s\mathbb{P}(V_1 = 0)/t(s)$, which is readily seen to be in agreement with (7).

Equation (4), which was obtained by studying the workloads at arrival epochs (i.e. the waiting times; by the PASTA property, they have the same distribution as the steady-state workloads), looks slightly different from (23), but, using (6), it is readily seen that they are equivalent.

Globally speaking, the essential steps in [4], [10], and [12] are the following.

Step 1: *find a suitable set of zeroes* $(\hat{s}, \hat{t})$, *with* $\operatorname{Re} \hat{s} \geq 0$, $\operatorname{Re} \hat{t} \geq 0$, *of the kernel* $K(s, t)$, i.e. $K(\hat{s}, \hat{t}) = 0$. Because $\psi(s, t)$ is regular for all (s, t) with $\operatorname{Re} s, t \geq 0$, we must have, for all these zeroes,

$$\hat{t}\psi_1(\hat{s}) = -\hat{s}\psi_2(\hat{t}).$$

It is further observed that $\psi_1(s)$ is regular for $\operatorname{Re} s > 0$ and continuous for $\operatorname{Re} s \geq 0$, and that $\psi_2(t)$ is regular for $\operatorname{Re} t > 0$ and continuous for $\operatorname{Re} t \geq 0$.

Step 2: *formulate a boundary-value problem for* $\psi_1(s)$ *and* $\psi_2(t)$. There are various types of boundary-value problem, such as the Riemann and the Wiener–Hopf boundary-value problems. Typically, they ask to determine two functions $P_1(\cdot)$ and $P_2(\cdot)$, which satisfy a relation on a particular boundary B, while $P_1(\cdot)$ is regular in the interior B^+ and $P_2(\cdot)$ is regular in the exterior B^-. Here B could be the unit circle (Riemann boundary-value problem), or the imaginary axis (Wiener–Hopf boundary-value problem; B^+ now is the left-half plane). We refer the reader to [17] and [20] for excellent expositions of such boundary-value problems and their variants, such as the boundary-value problem with a shift. The latter occurs in the approach of De Klein [12]; see below.

Step 3: *solve the boundary-value problem for* $\psi_1(\cdot)$ *and* $\psi_2(\cdot)$ *with boundary* B. If B is a smooth closed contour that is not a circle, the use of a conformal mapping from B to the unit circle C is required to arrive at a Riemann boundary-value problem for the unit circle, the solution of which can be found in [17] and [20]. Thus, we obtain $\psi_1(s)$ and $\psi_2(t)$ inside certain regions; subsequently, one may use analytic continuation to find them in $\operatorname{Re} s, t \geq 0$. Finally, $\psi(s, t)$ follows from (23).

Remark 4. Application of the boundary value method in queueing theory was pioneered by Fayolle and Iasnogorodski [13]. They used this method to analyze the joint queue length process in two coupled processors, viz. two M/M/1 queues which operate at unit speeds when the other queue is not empty, but at different speeds when the other queue is empty. The method was subsequently developed in [11] for a large class of two-dimensional random walks; various queueing applications were also discussed in [11]. See [9] for a survey of the method in queueing

theory, and see [10] and [14], for two monographs which have further developed the theory of two-dimensional random walks. Part IV of [10] explores the analysis of N-dimensional random walks with $N > 2$. Results for $N > 2$ are very limited, and it seems fair to conclude that the boundary value method is, apart from a few special cases, restricted to two-dimensional random walks.

Remark 5. We strongly believe that the boundary value method also has a large potential in the analysis of two-dimensional risk models. Owing to the duality between the reinsurance model and the two-queue model with simultaneous arrivals, the publications [4], [10], and [12], are of immediate relevance to the reinsurance problem. These publications seem unknown in the insurance community (see, e.g. Chan *et al.* [8], who posed the two-dimensional risk problem and stopped at (23)—where [4], [10], and [12], begin). They have remained largely unnoticed even in the queueing community, perhaps because of their complexity and because [4] and [12] did not appear in the open literature.

The approaches in [4], [10], and [12] are successively exposed at some length in the full, preprint version of this paper [7].

Remark 6. It should be observed that Baccelli [4], Cohen [10], and De Klein [12] all also solve the more complicated *transient* problem of determining the joint time-dependent distribution of the two workloads.

7. Conclusions and future work

We have studied a multivariate queueing system, which is shown to correspond to a dual risk process with multiple lines of insurance that receive coupled claims. We find the LST of the multivariate workload distribution in the case in which the service requirements are ordered with probability 1. Duality then yields the Laplace transform of the survival probabilities. For general service requirement (respectively claim size) vectors the workload (respectively ruin) problem can be solved in the two-dimensional case, by solving a Riemann boundary-value problem. For dimension $K > 2$, the problem seems analytically intractable in its full generality. That raises the need for approximations and asymptotics. It would in particular be interesting to obtain explicit multidimensional tail asymptotics of workloads and ruin probabilities, both for light-tailed and heavy-tailed service requirements (or claim sizes). Even for $K = 2$ queues, this is already quite challenging. Moreover, a wide range of different cases must be studied, giving rise to quite different techniques and results. Therefore, we intend to devote a separate study to tail asymptotics.

Acknowledgements

References

[1] Asmussen, S. and Albrecher, H. (2010). *Ruin Probabilities*, 2nd edn. World Scientific, Hackensack, NJ.

[2] Avram, F., Palmowski, Z. and Pistorius, M. (2008). A two-dimensional ruin problem on the positive quadrant. *Insurance Math. Econom.* **42,** 227–234.

[3] AVRAM, F., PALMOWSKI, Z. AND PISTORIUS, M. R. (2008). Exit problem of a two-dimensional risk process from the quadrant: exact and asymptotic results. *Ann. Appl. Prob.* **18,** 2421–2449.

[4] BACCELLI, F. (1985). Two parallel queues created by arrivals with two demands: The M/G/2 symmetrical case. Res. Rep. 426, INRIA-Rocquencourt.

[5] BACCELLI, F., MAKOWSKI, A. M. AND SHWARTZ, A. (1989). The fork-join queue and related systems with synchronization constraints: stochastic ordering and computable bounds. *Adv. Appl. Prob.* **21,** 629–660.

[6] BADESCU, A. L., CHEUNG, E. C. K. AND RABEHASAINA, L. (2011). A two-dimensional risk model with proportional reinsurance. *J. Appl. Prob.* **48,** 749–765.

[7] BADILA, E. S., BOXMA, O. J., RESING, J. A. C. AND WINANDS, E. M. M. (2012). Queues and risk models with simultaneous arrivals. Preprint. Available at http://arxiv.org/abs/1211.2193.

[8] CHAN, W.-S., YANG, H. AND ZHANG, L. (2003). Some results on ruin probabilities in a two-dimensional risk model. *Insurance Math. Econom.* **32,** 345–358.

[9] COHEN, J. W. (1988). Boundary value problems in queueing theory. *Queueing Systems* **3,** 97–128.

[10] COHEN, J. W. (1992). *Analysis of Random Walks*. IOS Press, Amsterdam.

[11] COHEN, J. W. AND BOXMA, O. J. (1983). *Boundary Value Problems in Queueing System Analysis*. North-Holland, Amsterdam.

[12] DE KLEIN, S. J. (1988). *Fredholm integral equations in queueing analysis*. Doctoral Thesis, University of Utrecht.

[13] FAYOLLE, G. AND IASNOGORODSKI, R. (1979). Two coupled processors: the reduction to a Riemann–Hilbert problem. *Z. Wahrscheinlichkeitsth.* **47,** 325–351.

[14] FAYOLLE, G., IASNOGORODSKI, R. AND MALYSHEV, V. (1999). *Random Walks in the Quarter-Plane*. Springer, Berlin.

[15] FLATTO, L. AND HAHN, S. (1984). Two parallel queues created by arrivals with two demands. I. *SIAM J. Appl. Math.* **44,** 1041–1053. (Erratum: **45** (1985), 168.)

[16] FROSTIG, E. (2004). Upper bounds on the expected time to ruin and on the expected recovery time. *Adv. Appl. Prob.* **36,** 377–397.

[17] GAKHOV, F. D. (1990). *Boundary Value Problems*. Dover, New York.

[18] KELLA, O. (1993). Parallel and tandem fluid networks with dependent Lévy inputs. *Ann. Appl. Prob.* **3,** 682–695.

[19] LÖPKER, A. AND PERRY, D. (2010). The idle period of the finite $G/M/1$ queue with an interpretation in risk theory. *Queueing Systems* **64,** 395–407.

[20] MUSKHELISHVILI, N. I. (2008). *Singular Integral Equations*, 2nd edn. Dover, Mineola, NY.

[21] NELSON, R. AND TANTAWI, A. N. (1987). *Approximating task response times in fork/join queues*. IBM Res. Rep. RC13012.

[22] NELSON, R. AND TANTAWI, A. N. (1988). Approximate analysis of fork/join synchronization in parallel queues. *IEEE Trans. Comput.* **37,** 739–743.

[23] ROLSKI, T., SCHMIDLI, H., SCHMIDT, V. AND TEUGELS, J. (1999). *Stochastic Processes for Insurance and Finance*. John Wiley, Chichester.

[24] SIEGMUND, D. (1976). The equivalence of absorbing and reflecting barrier problems for stochastically monotone Markov Processes. *Ann. Prob.* **4,** 914–924.

[25] WRIGHT, P. E. (1992). Two parallel processors with coupled inputs. *Adv. Appl. Prob.* **24,** 986–1007.