The total overflow during a busy cycle in a Markov-additive finite buffer system

Lothar Breuer

Institute of Mathematics and Statistics
University of Kent
Canterbury CT2 7NF, UK
L.Breuer@kent.ac.uk

Abstract. We consider a finite buffer system where the buffer content moves in a Markov-additive way while it is strictly between the buffer boundaries. Upon reaching the upper boundary of the buffer the content is not allowed to go higher and for every additional input into the system a penalty must be paid (to negotiate buffer overflow). At the lower boundary (empty buffer) the process terminates. For this system we determine the joint distribution of the total overflow and the last time of being at the upper boundary. The analysis is performed using exursion theory for Markov-additive processes.

1 Introduction

We shall consider a finite buffer system where the buffer content changes according to a Markov-additive process (MAP) while it is strictly between the buffer boundaries. Upon reaching the upper boundary of the buffer the content is not allowed to go higher and for every additional input into the system a penalty must be paid (to negotiate buffer overflow). At the lower boundary (empty buffer) the process terminates.

Such a system plays an important role in different areas of applied probability. It might represent a dam and its storage (or capacity) process. It is also used in insurance mathematics to model dividend payments. In queueing theory a penalty for buffer overflow is a standard consequence and termination once the buffer is empty is a natural restriction to the busy cycle. The content of the buffer is often called work load or virtual waiting time in queueing applications.

The model to be analysed in this paper is general enough to encompass a large variety of popular modelling approaches. Among them are Markovian single server queues (with BMAP input, see [17] for definition and [10] for estimation, and phase-type service time distributions, see [18] for definition, [7] for estimation, and [11] for a recent continuity result) or stochastic fluid flows with possible Brownian perturbation (see the seminal paper [3] or [1, 9] for recent related results without perturbation). An algorithmic solution for the time to buffer overflow in a Markov-additive framework is given in [6], section 6, see also [2]. An algorithmic solution for the expectation of the total overflow during a cycle is presented in [13], albeit in terms of insurance risk.

Assuming that the penalty to be paid is simply the amount of system input while the buffer is full (the total buffer overflow), we shall determine the joint distribution of this penalty and the last time of being at the upper boundary. We do not assume necessarily that the system starts with an empty buffer.

The analysis is performed mainly by matrix-analytic methods using probabilistic arguments wherever possible. This naturally results in formulas containing matrices which are to be computed via fixed point iterations. We shall present examples for the simple cases allowing explicit scalar solutions. This restriction is due to the circumstance that only for these there are solutions in the literature which can be compared with results in the present paper.

The paper is structured as follows. Section 2 contains an exact definition of the model to be analysed and the performance measures we wish to determine. Section 3 presents preparatory results from recent literature that will be needed in this paper. Section 4 finally contains the main result. Examples will be developed throughout the paper in subsequent stages.

2 The model

Let $\tilde{\mathcal{J}} = (\tilde{J}_t : t \geq 0)$ be an irreducible Markov process with finite state space $\tilde{E}$ and infinitesimal generator matrix $\tilde{Q} = (\tilde{q}_{ij})_{i,j\in\tilde{E}}$. We call $\tilde{J}_t$ the phase at time $t \geq 0$ (another common name is regime). Define the real–valued process $\tilde{\mathcal{X}} = (\tilde{X}_t : t \geq 0)$ as evolving like a Lévy process $\tilde{\mathcal{X}}^{(i)}$ with parameters $\tilde{\mu}_i$ (drift), $\tilde{\sigma}_i^2$ (variation), and $\tilde{\nu}_i$ (Lévy measure) during intervals when the phase equals $i \in \tilde{E}$. For the sake of a more concise presentation we exclude the case of $\tilde{\mu}_i = \tilde{\sigma}_i^2 = 0$, i.e. a pure jump process or the constant zero process, for any phase $i \in \tilde{E}$. Whenever $\tilde{\mathcal{J}}$ jumps from a state $i \in \tilde{E}$ to another state $j \in \tilde{E}$, this may be accompanied by a jump of $\tilde{\mathcal{X}}$ with some distribution function F_{ij}. Then the two–dimensional process $(\tilde{\mathcal{X}}, \tilde{\mathcal{J}})$ is called a Markov–additive process (or shortly MAP). In short, a MAP is a Markov-modulated Lévy process with possible jumps at phase changes. For a textbook introduction to MAPs see [4], chapter XI.

We now turn to define our model $\mathcal{B} = (B_t : t \geq 0)$ for the buffer content, where B_t shall denote the content level at time t. Let $b \geq 0$ denote the upper buffer boundary beyond which overflow occurs and penalty must be paid. As long as $0 < B_t < b$, the process $\mathcal{B}$ equals $\tilde{\mathcal{X}}$. Upon passing the upper boundary b from below, $\mathcal{B}$ does not increase above b and any additional buffer input is recorded as overflow. If a positive jump of size x occurs at time t and $b - x < B_{t-} < b$, then we agree upon the rule that the buffer content rises up to b and the overflow increases by $B_{t-} - (b - x)$. Upon passing the lower boundary 0 from above, the busy cycle concludes and we stop our examination. Thus we consider $\mathcal{B}$ as a MAP that is reflected at the upper boundary b and terminates upon passing the lower boundary 0. In exact terms,

$$B_t := \tilde{X}_t - \left(\sup_{s\leq t} \tilde{X}_s - b\right)^+$$

for all $t < \tau_B(0)$, where $(V)^+ := \max(V, 0)$ and

$$\tau_B(0) := \inf \left\{ t \geq 0 : \tilde{X}_t - \left(\sup_{s \leq t} \tilde{X}_s - b \right)^+ \leq 0 \right\} \tag{1}$$

For the sake of defining $\mathcal{B}$ at all times, say $B_t := 0$ for all $t \geq \tau_B(0)$.

Denote the initial buffer content by $u := B_0 \geq 0$. We may assume $u \leq b$ without loss of generality, since $b < u$ would entail an immediate pay-out of a penalty of $u - b$ and the buffer content process would continue with initial surplus b.

Let D denote the total overflow during a busy cycle. This can be defined as follows. First, let

$$\tilde{S}(t) := \left(\sup_{s \leq t} \tilde{X}_s - b \right)^+$$

denote the overflow until time $t \in [0, \tau_B(0)]$. Then $D := \tilde{S}(\tau_B(0))$ is the total overflow during a busy cycle. Define the generalised inverse of the function $\tilde{S}(t)$ by

$$\tilde{S}^{-1}(x) := \inf\{t \geq 0 : \tilde{S}(t) \geq x\}$$

for $x \geq 0$. Then $\tilde{S}^{-1}(D)$ is the last time of overflow before the end of the busy cycle.

In this paper we shall determine the joint distribution of D and $\tilde{S}^{-1}(D)$ in the form of an expression for

$$\bar{F}(x, \gamma) := \mathbb{E}\left(e^{-\gamma \tilde{S}^{-1}(x)}; D > x \right)$$

where $x \geq 0$ and $\gamma \geq 0$. Note that $\tilde{S}^{-1}(D)$ signifies the time of the last overflow and may be strictly smaller than $\tau_B(0)$, the end of the busy cycle. Further note that the process $\mathcal{B}$ and hence $\bar{F}(x, \gamma)$ is completely determined by $\tilde{\mathcal{X}}$.

Example 1 We consider the classical M/M/1 queue. Inter–arrival and service times are iid exponential with parameter $\lambda > 0$ and $\beta > 0$, respectively. The total work load at time $t \geq 0$ (including the overflow) within a busy cycle starting with a buffer content $u > 0$ is given by

$$\tilde{X}_t = u + \sum_{n=0}^{N_t} C_n - t \tag{2}$$

where $(N_t : t \geq 0)$ is a Poisson process with intensity λ and the C_n, $n \in \mathbb{N}$, are iid random variables with exponential distribution of parameter β.

The total work load process can be analysed as a MAP with exponential (and hence phase–type) positive jumps with parameter β. For this, we would need only one phase, i.e. $|\tilde{E}| = 1$. This phase governs a Lévy process with parameters $\tilde{\sigma} = 0$, $\tilde{\mu} = -1$, and $\tilde{\nu}(dx) = \lambda e^{-\beta x} \beta dx$ for all $x > 0$.

Before we can proceed by analysing our model, we first need to collect some necessary preliminary results for MAPs. This shall be the purpose of the next section.

3 Preliminaries

3.1 Markov–additive processes with phase–type jumps

In this section we introduce the restriction that all jumps have a phase-type distribution. Then we construct a new MAP $(\mathcal{X}, \mathcal{J})$ from the given MAP $(\tilde{\mathcal{X}}, \tilde{\mathcal{J}})$ without losing any information. This new MAP will have continuous paths which simplifies the one- and two-sided exit problems (cf. sections 3.2 and 3.3) considerably.

Denote the indicator function of a set A by $\mathbb{I}_A$. We assume that the Lévy measures $\tilde{\nu}_i$ have the form

$$\begin{aligned}\tilde{\nu}_i(dx) = \lambda_i^+ \mathbb{I}_{\{x>0\}}\, \alpha^{(ii)+} \exp(T^{(ii)+}x)\eta^{(ii)+}dx \\ + \lambda_i^- \mathbb{I}_{\{x<0\}}\, \alpha^{(ii)-} \exp(-T^{(ii)-}x)\eta^{(ii)-}dx \qquad (3)\end{aligned}$$

for all $i \in \tilde{E}$, where $\lambda_i^\pm \geq 0$ and $(\alpha^{(ii)\pm}, T^{(ii)\pm})$ are representations of phase–type distributions without an atom at 0. The $\eta^{(ii)\pm} := -T^{(ii)\pm}\mathbf{1}$ are called the exit vectors, where $\mathbf{1}$ denotes a column vector of appropriate dimension with all entries being 1. This means that the jump process induced by the Lévy measure ν_i is compound Poisson with jump sizes of a doubly phase–type distribution. Denote the order of $PH(\alpha^{(ii)\pm}, T^{(ii)\pm})$ by $m_{ii}^\pm$. Further write $\lambda_i := \lambda_i^+ + \lambda_i^-$.

Likewise, let p_{ij}^+ (resp. p_{ij}^-) denote the probability that a positive (resp. negative) jump is induced by a phase change from $i \in \tilde{E}$ to $j \in \tilde{E}$, and assume that these jumps have a $PH(\alpha^{(ij)\pm}, T^{(ij)\pm})$ distribution without an atom at 0. Note that $p_{ij}^+ + p_{ij}^- \leq 1$ for all $i, j \in \tilde{E}$. Let $m_{ij}^\pm$ denote the order of $PH(\alpha^{(ij)\pm}, T^{(ij)\pm})$ and define $\eta^{(ij)\pm} := -T^{(ij)\pm}\mathbf{1}$.

The class of Markov–additive processes with these assumptions of phase–type jumps is dense within the class of all MAPs, see [5], proposition 1. The main advantage of the phase–type restriction on the jump distributions is the possibility of transforming the jumps into a succession of linear pieces of exponential duration (each with slope 1 or -1) and retrieving the original process via a simple time change, see [8].

This is done in the following way. Without the jumps, the Lévy process $\tilde{\mathcal{X}}^{(i)}$ during a phase $i \in \tilde{E}$ is either a linear drift (i.e. $\tilde{\sigma}_i = 0$) or a Brownian motion (with parameters $\tilde{\sigma}_i > 0$ and $\tilde{\mu}_i \in \mathbb{R}$). Considering this MAP (without the jumps) we can partition its phase space $\tilde{E}$ into the subspaces E_p (for positive drifts), E_σ (for Brownian motions), and E_n (for negative drifts). We thus define

$$\begin{aligned}& E_p := \{i \in \tilde{E} : \tilde{\mu}_i > 0, \tilde{\sigma}_i = 0\}, \quad E_n := \{i \in \tilde{E} : \tilde{\mu}_i < 0, \tilde{\sigma}_i = 0\} \qquad (4) \\ \text{and} \quad & E_\sigma := \{i \in \tilde{E} : \tilde{\sigma}_i > 0\}\end{aligned}$$

Note that $\tilde{E} = E_p \cup E_\sigma \cup E_n$, since we have excluded the case of $\tilde{\mu}_i = \tilde{\sigma}_i^2 = 0$ for any phase $i \in \tilde{E}$. Then we introduce two new phase spaces

$$E_\pm := \{(i, j, k, \pm) : i, j \in E_p \cup E_\sigma \cup E_n, 1 \leq k \leq m_{ij}^\pm\} \qquad (5)$$

to model the jumps. Define now the enlarged phase space $E = E_+ \cup \tilde{E} \cup E_-$. We define the modified MAP $(\mathcal{X}, \mathcal{J})$ over the phase space E as follows. Set the parameters $(\mu_i, \sigma_i^2, \nu_i)$ for $i \in E$ as

$$(\mu_i, \sigma_i^2, \nu_i) := \begin{cases} (\pm 1, 0, \mathbf{0}), & i \in E_\pm \\ (\tilde{\mu}_i, \tilde{\sigma}_i, \mathbf{0}), & i \in \tilde{E} = E_p \cup E_\sigma \cup E_n \end{cases} \tag{6}$$

This leads to the cumulant functions

$$\psi_i(\alpha) = \begin{cases} \pm\alpha, & i \in E_\pm \\ \mu_i \alpha, & i \in E_p \cup E_n \\ \frac{1}{2}\sigma_i^2 \alpha^2 + \mu_i \alpha, & i \in E_\sigma \end{cases} \tag{7}$$

We shall order the new phase space $E = E_+ \cup E_p \cup E_\sigma \cup E_n \cup E_-$ such that $i_+ < i_p < i_\sigma < i_n < i_-$ for phases $i_* \in E_*$. Let $E_c := E_p \cup E_\sigma \cup E_n$ denote the subspace of E that contains all phases under which the real time movements are continuous. The modified phase process $\mathcal{J}$ is determined by its generator $Q = (q_{ij})_{i,j \in E}$. For this the construction above yields

$$q_{ih} = \begin{cases} \tilde{q}_{ii} - \lambda_i, & h = i \in E_c \\ \tilde{q}_{ih} \cdot (1 - p_{ih}^+ - p_{ih}^-), & h \in E_c, h \neq i \\ \lambda_i^\pm \alpha_k^{(ii)\pm}, & h = (i, i, k, \pm) \\ \tilde{q}_{ij} \cdot p_{ij}^\pm \cdot \alpha_k^{(ij)\pm}, & h = (i, j, k, \pm) \end{cases} \tag{8}$$

for $i \in E_c$ as well as

$$q_{(i,j,k,\pm),(i,j,l,\pm)} = T_{kl}^{(ij)\pm} \qquad \text{and} \qquad q_{(i,j,k,\pm),j} = \eta_k^{(ij)\pm} \tag{9}$$

for $i, j \in E_c$ and $1 \leq k, l \leq m_{ij}^\pm$. For later use we define $q_i := -q_{ii}$ for all $i \in E$.

The original level process $\tilde{\mathcal{X}}$ is retrieved via the time change

$$c(t) := \int_0^t \mathbb{I}_{\{J_s \in E_c\}}\, ds \quad \text{and} \quad \tilde{X}_{c(t)} = X_t \tag{10}$$

for all $t \geq 0$. Likewise, we obtain

$$S(t) := \left(\sup_{s \leq t} X_s - b \right)^+ = \left(\sup_{s \leq c(t)} \tilde{X}_s - b \right)^+ = \tilde{S}(c(t))$$

and for the generalised inverse

$$\tilde{S}^{-1}(x) = c(S^{-1}(x))$$

The inverses of the cumulant functions ψ_i can be given explicitly as

$$\phi_i(\beta) = \begin{cases} \pm\beta, & i \in E_\pm \\ \frac{\beta}{\mu_i}, & i \in E_p \cup E_n \\ \frac{1}{\sigma_i}\sqrt{2\beta + \frac{\mu_i^2}{\sigma_i^2}} - \frac{\mu_i}{\sigma_i^2}, & i \in E_\sigma \end{cases} \tag{11}$$

We shall, however, use them only for the so–called ascending phases $i \in E_a := E_+ \cup E_p \cup E_\sigma$.

Example 2 Continuing example 1, we obtain the MAP $(\mathcal{X}, \mathcal{J})$ as follows. In equation (3) we have $\lambda_2^+ = \lambda$ and $\lambda_2^- = 0$. Further $m_{22}^+ = 1$ since the positive jumps have an exponential distribution. Hence the enlarged phase space is given by $E_+ = \{1\}$, $E_n = \{2\}$, and $E_p = E_\sigma = E_- = \emptyset$. The parameters are given by $\sigma_1 = \sigma_2 = 0$, $\mu_1 = 1$, $\mu_2 = -1$, $\nu_1 = \nu_2 = \mathbf{0}$, and

$$Q = \begin{pmatrix} -\beta & \beta \\ \lambda & -\lambda \end{pmatrix}$$

3.2 First passage times

Of central use in the present paper will be the recent derivation of the Laplace transforms for the first passage times of MAPs as given in [12]. We call the phases $i \in E_d := E_n \cup E_-$ descending. Define the first passage times

$$\tilde{\tau}(x) := \inf\{t \geq 0 : \tilde{X}_t > x\} \qquad \text{and} \qquad \tau(x) := \inf\{t \geq 0 : X_t > x\}$$

for all $x \geq 0$ and assume that $\tilde{X}_0 = X_0 = 0$. Note that $\tilde{\tau}(x)$ is the first passage time over the level x for the original MAP $\tilde{\mathcal{X}}$, meaning that we do not count the time spent in jump phases $i \in E_\pm$. This means that

$$\tilde{\tau}(x) = c(\tau(x)) = \int_0^{\tau(x)} \mathbb{I}_{\{J_s \in E_c\}} ds$$

according to (10). In particular, we may compute expectations over $\tilde{\tau}(x)$ using the distribution of the modified MAP $(\mathcal{X}, \mathcal{J})$ only and without needing to recur to the original MAP $(\tilde{\mathcal{X}}, \tilde{\mathcal{J}})$. For $\gamma \geq 0$ denote

$$\mathbb{E}_{ij}(e^{-\gamma\tilde{\tau}(x)}) := \mathbb{E}(e^{-\gamma\tilde{\tau}(x)}; J_{\tau(x)} = j | J_0 = i, X_0 = 0)$$

for all $i, j \in E$. Let $\mathbb{E}(e^{-\gamma\tilde{\tau}(x)})$ denote the matrix with these entries and write

$$\mathbb{E}(e^{-\gamma\tilde{\tau}(x)}) = \begin{pmatrix} \mathbb{E}_{(a,a)}(e^{-\gamma\tilde{\tau}(x)}) & \mathbb{E}_{(a,d)}(e^{-\gamma\tilde{\tau}(x)}) \\ \mathbb{E}_{(d,a)}(e^{-\gamma\tilde{\tau}(x)}) & \mathbb{E}_{(d,d)}(e^{-\gamma\tilde{\tau}(x)}) \end{pmatrix}$$

in obvious block notation with respect to the subspaces $E_a = E_+ \cup E_p \cup E_\sigma$ (ascending phases) and $E_d = E_n \cup E_-$ (descending phases).

Since a first passage to a level above cannot occur in a descending phase, we obtain first $\mathbb{P}(J_{\tau(x)} = j) = 0$ for all $j \in E_d$ and thus

$$\mathbb{E}_{(d,d)}(e^{-\gamma\tilde{\tau}(x)}) = \mathbb{E}_{(a,d)}(e^{-\gamma\tilde{\tau}(x)}) = \mathbf{0}$$

where $\mathbf{0}$ denotes a zero matrix of suitable dimension. Equation (6) in [12] states that

$$\mathbb{E}_{(d,a)}(e^{-\gamma\tilde{\tau}(x)}) = A(\gamma)e^{U(\gamma)x}$$

and

$$\mathbb{E}_{(a,a)}(e^{-\gamma\tilde{\tau}(x)}) = e^{U(\gamma)x}$$

for some sub–generator matrix $U(\gamma)$ of dimension $E_a \times E_a$ and a sub–transition matrix $A(\gamma)$ of dimension $E_d \times E_a$. Altogether we can write

$$\mathbb{E}(e^{-\gamma\tilde{\tau}(x)}) = \begin{pmatrix} I_a \\ A(\gamma) \end{pmatrix} \begin{pmatrix} e^{U(\gamma)x} & \mathbf{0} \end{pmatrix} \tag{12}$$

where I_a denotes the identity matrix of dimension $E_a \times E_a$.

Write $\Delta_q := diag(q_i)_{i\in E}$ and let $P = \Delta_q^{-1}Q + I$ denote the transition matrix of phase changes. Note that $p_{ii} = 0$ for all $i \in E$.

In order to shorten notation, we shall write $A = A(\gamma)$ and $U = U(\gamma)$ unless we wish to stress dependence on γ. According to theorem 3 in [12], A and U satisfy the following equations:

$$\begin{aligned}
e_h'U &= \sum_{l=1}^{m_{ij}^+} T_{kl}^{(ij)+} e_{(i,j,l,+)}' + \eta_k^{(ij)+} e_j' \begin{pmatrix} I_a \\ A \end{pmatrix} && \text{for } h = (i,j,k,+) \in E_+, \\
e_i'U &= \phi_i(q_i) \sum_{j\in E} p_{ij}\ e_j' \begin{pmatrix} I_a \\ A \end{pmatrix} L_i(-U) - \phi_i(q_i+\gamma)e_i' && \text{for } i \in E_p \cup E_\sigma, \\
e_i'A &= \sum_{j\in E, j\neq i} q_{ij} e_j' \begin{pmatrix} I_a \\ A \end{pmatrix} \left((q_i+\gamma)I_a - \psi_i(-U)\right)^{-1} && \text{for } i \in E_n, \text{ and} \\
e_i'A &= \sum_{j\in E, j\neq i} q_{ij} e_j' \begin{pmatrix} I_a \\ A \end{pmatrix} \left(q_i I_a - \psi_i(-U)\right)^{-1} && \text{for } i \in E_-.
\end{aligned}$$

where e_i', e_j' and e_h' denote canonical row base vectors with suitable dimension. For the MAP $(\mathcal{X}, \mathcal{J})$ with continuous level process, the matrix function

$$L_i(-U) = \frac{q_i}{\phi_i(q_i)} \cdot (\phi_i(q_i+\gamma)I_a + U) \cdot ((q_i+\gamma)I_a - \psi_i(-U))^{-1}$$

can be simplified considerably. For $i \in E_\sigma$, the same arguments as in [12], example 2, lead to

$$L_i(-U) = \phi_i^*(q_i) \cdot (\phi_i^*(q_i+\gamma)I_a - U)^{-1} \tag{13}$$

with

$$\phi_i^*(\beta) = \frac{1}{\sigma_i}\sqrt{2\beta + \frac{\mu_i^2}{\sigma_i^2}} + \frac{\mu_i}{\sigma_i^2} \tag{14}$$

Furthermore, $L_i(-U) = I_a$ for $i \in E_p$ (see example 3 in [12]), while according to (7) $\psi_i(-U) = -\mu_i U$ for $i \in E_n$, and $\psi_i(-U) = U$ for $i \in E_-$. Hence the equations above involve rather simple expressions only.

Considering (11), the matrices $A(\gamma)$ and $U(\gamma)$ can be determined by successive approximation as the limit of the sequence $((A_n, U_n) : n \geq 0)$ with initial values $A_0 := \mathbf{0}$, $U_0 := -diag(\phi_i(q_i+\gamma)1_{i\in E_\sigma\cup E_p} + \phi_i(q_i)1_{i\in E_+})_{i\in E_a}$, and the following iteration:

$$\begin{aligned}
e_h' U_{n+1} &= \sum_{l=1}^{m_{ij}^+} T_{kl}^{(ij)+} e_{(i,j,l,+)}' + \eta_k^{(ij)+} e_j' \begin{pmatrix} I_a \\ A_n \end{pmatrix} \qquad \text{for } h=(i,j,k,+)\in E_+, \\
e_i' U_{n+1} &= -\frac{q_i+\gamma}{\mu_i} e_i' + \frac{1}{\mu_i} \sum_{j\in E, j\neq i} q_{ij}\ e_j' \begin{pmatrix} I_a \\ A_n \end{pmatrix} \qquad \text{for } i\in E_p, \\
e_i' A_{n+1} &= \sum_{j\in E, j\neq i} q_{ij} e_j' \begin{pmatrix} I_a \\ A_n \end{pmatrix} ((q_i+\gamma)I + \mu_i U_n)^{-1} \qquad \text{for } i\in E_n, \\
e_i' A_{n+1} &= \sum_{j\in E, j\neq i} q_{ij} e_j' \begin{pmatrix} I_a \\ A_n \end{pmatrix} (q_i I - U_n)^{-1} \qquad \text{for } i\in E_-, \text{ and} \\
e_i' U_{n+1} &= \frac{2}{\sigma_i^2} \sum_{j\in E, j\neq i} q_{ij}\ e_j' \begin{pmatrix} I_a \\ A_n \end{pmatrix} (\phi_i^*(q_i+\gamma)I - U_n)^{-1} - \phi_i(q_i+\gamma) e_i'
\end{aligned}$$

for $i \in E_\sigma$. For the last equality the relation $\phi_i(q_i)\phi_i^*(q_i) = 2q_i/\sigma_i^2$ has been used. Note that the only difference between the iterations for E_n and E_- is the missing γ in the last factor for E_-, reflecting that we do not discount the time for phases $i \in E_-$ as they are jump phases in real time.

Example 3 Continuing example 2, first note that phase 1 represents the upwards jumps and we will not discount the time during sojourns in it. As shown in [12], example 5, the Laplace transform of the first passage time over a level $x > 0$, is given by

$$\mathbb{E}(e^{-\gamma\tilde{\tau}(x)}) = Ae^{Ux} \quad \text{where} \quad A = \frac{\beta - R}{\beta}, \quad U = -R$$

and

$$-R = \frac{1}{2}\left(\lambda + \gamma - \beta - \sqrt{(\beta-\gamma-\lambda)^2 + 4\beta\gamma}\right)$$

This coincides with equation (4.24) in [15], noting that γ is denoted as δ there and $c = 1$ in our case.

3.3 The two-sided exit problem

Define the stopping times $\tau(0,b) := \inf\{t \geq 0 : X_t < 0 \quad \text{or} \quad X_t > b\}$ and

$$\tilde{\tau}(0,b) := \int_0^{\tau(0,b)} \mathbb{I}_{\{J_s\in E_c\}} ds = \inf\{t \geq 0 : \tilde{X}_t < 0 \quad \text{or} \quad \tilde{X}_t > b\} \qquad (15)$$

which are the exit times of $\mathcal{X}$ and $\tilde{\mathcal{X}}$ from the interval $[0, b]$, respectively. For the main result we need an expression for

$$\Psi_{ij}^+(b|x) := \mathbb{E}\left(e^{-\gamma\tilde{\tau}(0,b)}; X_{\tau(0,b)} = b, J_{\tau(0,b)} = j | J_0 = i, X_0 = x\right)$$

where $x \in [0, b]$ and $i, j \in E$. Clearly $\Psi^+_{ij}(b|x) = 0$ for $j \in E_d$ since an exit over the upper boundary can occur only in an ascending phase. Define the matrix $\Psi^+(b|x) := (\Psi^+_{ij}(b|x))_{i\in E, j\in E_a}$. A formula for $\Psi^+(b|x)$ has been derived in [16]. In order to state it we need some additional notation.

Let $(\mathcal{X}^+, \mathcal{J})$ denote the MAP as constructed in section 3.1 and define the process $\mathcal{X}^- = (X^-_t : t \geq 0)$ by $X^-_t := -X^+_t$ for all $t > 0$ given that $X^+_0 = X^-_0 = 0$. Thus $(\mathcal{X}^-, \mathcal{J})$ is the negative of $(\mathcal{X}^+, \mathcal{J})$. The two processes have the same generator matrix Q for $\mathcal{J}$, but the cumulant functions of the Lévy process governed by phase $i \in E$ are different and relate as $\psi^-_i(\alpha) = \psi^+_i(-\alpha)$. Denoting variation and drift parameters for $\mathcal{X}^\pm$ by $\sigma^\pm_i$ and $\mu^\pm_i$, respectively, this means $\sigma^+_i = \sigma^-_i$ and $\mu^-_i = -\mu^+_i$ for all $i \in E$. This of course implies that phases $i \in E_+ \cup E_p$ (resp. $i \in E_- \cup E_n$) are descending (resp. ascending) phases for $\mathcal{X}^-$.

Let $A^\pm(\gamma)$ and $U^\pm(\gamma)$ denote the matrices that determine the first passage times in (12). We shall write $A^\pm = A^\pm(\gamma)$ and $U^\pm = U^\pm(\gamma)$ except in cases when we wish to underline the dependence on γ. Define the matrices

$$C^+ := \begin{pmatrix} \mathbf{0} & I_{E_\sigma} \\ A^+ & \end{pmatrix} \qquad \text{and} \qquad C^- := \begin{pmatrix} A^- & \\ I_{E_\sigma} & \mathbf{0} \end{pmatrix}$$

of dimensions $(E_\sigma \cup E_d) \times E_a$ and $E_a \times (E_\sigma \cup E_d)$, respectively. Further define

$$W^+ := \begin{pmatrix} I_{E_a} \\ A^+ \end{pmatrix} \qquad \text{and} \qquad W^- := \begin{pmatrix} A^- \\ I_{E_\sigma \cup E_d} \end{pmatrix}$$

which are matrices of dimensions $E \times E_a$ and $E \times (E_\sigma \cup E_d)$. Finally, let $Z^\pm := C^\pm e^{U^\pm \cdot b}$. Then equation (23) in [16] states that

$$\Psi^+(b|x) = \left(W^+ e^{U^+ \cdot (b-x)} - W^- e^{U^- \cdot x} Z^+\right) \cdot \left(I - Z^- Z^+\right)^{-1} \tag{16}$$

for $0 \leq x \leq b$. Note that this expression depends on a choice of $\gamma \geq 0$.

Remark 1 Noting that $\left(I - Z^- Z^+\right)^{-1} = \sum_{n=0}^\infty \left(Z^- Z^+\right)^n$ and $Z^- Z^+$ represents a crossing of the interval $[0, b]$ from b to 0 and back, this formula has a clear probabilistic interpretation. The term $W^+ e^{U^+ \cdot (b-x)}$ simply yields the event that $\mathcal{X}$ exits from b. The correction term $W^- e^{U^- \cdot x} Z^+$ refers to the event that $\mathcal{X}$ descends below 0 before exiting from b. Multiplication by $\left(I - Z^- Z^+\right)^{-1}$ yields all possible combinations with any number of subsequent (down and up) crossings over the complete interval $[0, b]$.

Remark 2 Since $Z^+ = C^+ e^{U^+ \cdot b}$ we can write $\Psi^+(b|x)$ in the form

$$\Psi^+(b|x) = \left(W^+ e^{-U^+ \cdot x} - W^- e^{U^- \cdot x} C^+\right) \left(e^{-U^+ \cdot b} - C^- e^{U^- \cdot b} C^+\right)^{-1}$$

This comes closer to the usual expression of the exit time distribution in terms of scale functions. For instance, let $\mathcal{X}$ be a Brownian motion with variation $\sigma > 0$ and drift $\mu \in \mathbb{R}$. We then obtain

$$U^\pm = \frac{\pm\mu - \sqrt{\mu^2 + 2\gamma\sigma^2}}{\sigma^2}$$

Denote $-r := U^+$ and $s := U^-$. Then

$$\Psi^+(b|x) = \frac{e^{rx} - e^{sx}}{e^{rb} - e^{sb}}$$

cf. [14], (2.12 - 2.15), where the γ-scale function is given as $g(x) = e^{rx} - e^{sx}$.

Example 4 Another example is the M/M/1 queue during a busy cycle, which is the negative of the classical compound Poisson model with exponential jumps used in insurance mathematics. This continues example 3. We obtain

$$U^\pm = \frac{1}{2}\left(\pm(\lambda + \gamma - \beta) - \sqrt{(\beta - \gamma - \lambda)^2 + 4\beta\gamma}\right)$$

Denote $R := -U^+$ and $\rho := -U^-$ and compare this to [15], equations (3.12) and (4.24), with $\delta = \gamma$ and $c = 1$. Section 3.2 further yields $A^- = \beta/(\beta + \rho)$ and $A^+ = (\beta - R)/\beta$. Thus, starting with buffer content x in the descending phase, we obtain

$$\begin{aligned}
\Psi^+(b|x) &= \left(A^+ e^{-U^+ \cdot x} - e^{U^- \cdot x} A^+\right)\left(e^{-U^+ \cdot b} - A^- e^{U^- \cdot b} A^+\right)^{-1} \\
&= \left(\frac{\beta - R}{\beta} e^{Rx} - e^{-\rho x}\frac{\beta - R}{\beta}\right) \Big/ \left(e^{Rb} - \frac{\beta}{\beta + \rho} e^{-\rho b}\frac{\beta - R}{\beta}\right) \\
&= (\beta + \rho) \cdot \frac{\beta - R}{\beta} \cdot \frac{e^{Rx} - e^{-\rho x}}{(\beta + \rho)e^{Rb} - (\beta - R)e^{-\rho b}}
\end{aligned}$$

This is the Laplace transform of the time to buffer overflow within a busy cycle.

4 Main result

Starting with an initial buffer content $u < b$ or with $u = b$ but in a descending phase, there is a positive probability of no overflow at all before the buffer empties. Let α denote the initial phase distribution of $(\mathcal{X}, \mathcal{J})$, i.e. $\alpha_i = \mathbb{P}(J_0 = i)$ for all $i \in E$. Then equation (16) yields, with $\gamma := 0$,

$$\mathbb{P}(D = 0) = \begin{cases} 1 - \alpha\Psi^+(b|u)\mathbf{1}, & u < b \\ 1 - \alpha \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & I_{E_d} \end{pmatrix} \Psi^+(b|b)\mathbf{1}, & u = b \end{cases}$$

where I_{E_d} denotes the identity matrix of dimension $|E_d|$. Clearly the event $D = 0$ means that $\mathcal{X}$ exits the interval $[0, b]$ at the lower boundary first. We further observe that an overflow can occur only in ascending phases, i.e. on the time set $\{t \geq 0 : J_t \in E_a\}$.

We wish to derive an expression for the function

$$\bar{F}(x, \gamma) := \mathbb{E}\left(e^{-\gamma \tilde{S}^{-1}(x)}; D > x\right)$$

where $x, \gamma \geq 0$. The strong Markov property and the fact that an exit from $[0, b]$ at the upper boundary can occur only in an ascending phase yield together

$$\bar{F}(x, \gamma) = \Psi^{+}(b|u)\, \mathbb{E}\left(e^{-\gamma \tilde{S}^{-1}(x)}; D > x | X_0 = b\right)$$

where the last factor (written as an expectation) is an $E_a \times E_a$ matrix with entries

$$\mathbb{E}\left(e^{-\gamma \tilde{S}^{-1}(x)}; D > x, J_{S^{-1}(x)} = j | X_0 = b, J_0 = i\right)$$

for $i, j \in E_a$. This observation may be compared with equation (2.16) in [14]. Thus it suffices to determine the matrix-valued function

$$M(x, \gamma) := \mathbb{E}\left(e^{-\gamma \tilde{S}^{-1}(x)}; D > x | X_0 = b\right)$$

This is the content of our main result.

Theorem 1 *The distribution of the total overflow above the level b, given that $X_0 = b$ and $J_0 \in E_a$, is matrix-exponential. Specifically,*

$$M(x, \gamma) = e^{G(b) \cdot x}$$

for $\gamma, x \geq 0$, where

$$G(b) = \left(U^{+} e^{-U^{+} b} + C^{-} e^{U^{-} b} U^{-} C^{+}\right) \left(e^{-U^{+} b} - C^{-} e^{U^{-} b} C^{+}\right)^{-1}$$

Proof: We employ the following approximation. Assume that the penalty for an overflow is paid out in small batches of sizes $\varepsilon > 0$ rather than continuously. More exactly, we define a process $(\mathcal{X}^{\varepsilon}, \mathcal{J}^{\varepsilon})$ as follows. The phase process $\mathcal{J}^{\varepsilon}$ shall equal $\mathcal{J}$ almost surely. The level process $\mathcal{X}^{\varepsilon}$ behaves like $\mathcal{X}$ in the interval $[0, b]$ but may go above the level b. Whenever $\mathcal{X}^{\varepsilon}$ reaches the level $b + \varepsilon$, we pay a penalty of amount ε whereupon $\mathcal{X}^{\varepsilon}$ jumps back to the level b. The phase process $\mathcal{J}^{\varepsilon}$ remains unchanged by this jump. The original process $(\mathcal{X}, \mathcal{J})$ is obtained if we let ε tend to 0.

Let D^{ε} denote the total penalty obtained for $(\mathcal{X}^{\varepsilon}, \mathcal{J}^{\varepsilon})$. Then D^{ε} has a matrix-geometric distribution, i.e.

$$M^{\varepsilon}(n, \gamma) := \mathbb{E}\left(e^{-\gamma T_n(\varepsilon)}; D^{\varepsilon} \geq n \cdot \varepsilon | X_0^{\varepsilon} = b\right) = \left(\Psi^{+}_{(a,a)}(b + \varepsilon | b)\right)^{n}$$

for $n \in \mathbb{N}$ and $\gamma \geq 0$, where

$$\Psi^{+}_{(a,a)}(b + \varepsilon | b) = \left(e^{U^{+} \varepsilon} - C^{-} e^{U^{-} b} C^{+} e^{U^{+} \cdot (b+\varepsilon)}\right) \times \left(I - C^{-} e^{U^{-} \cdot (b+\varepsilon)} C^{+} e^{U^{+} \cdot (b+\varepsilon)}\right)^{-1}$$

according to (16), and $T_n(\varepsilon)$ denotes the time of the nth payment of an ε-penalty.

Now letting ε tend to 0 we obtain that $M(x,\gamma)$ has a matrix-exponential distribution with parameter

$$\begin{aligned}G(b) &= \lim_{\varepsilon\downarrow 0}\frac{1}{\varepsilon}\left(\Psi^+_{(a,a)}(b+\varepsilon|b)-I\right)\\ &= \lim_{\varepsilon\downarrow 0}\frac{1}{\varepsilon}\left(e^{U^+\varepsilon}-I+C^-e^{U^-b}\left(e^{U^-\varepsilon}-I\right)C^+e^{U^+\cdot(b+\varepsilon)}\right)\\ &\quad\times\left(I-C^-e^{U^-\cdot(b+\varepsilon)}C^+e^{U^+\cdot(b+\varepsilon)}\right)^{-1}\\ &= \left(U^++C^-e^{U^-b}U^-C^+e^{U^+b}\right)\left(I-C^-e^{U^-b}C^+e^{U^+b}\right)^{-1}\end{aligned}$$

which is equivalent to the statement.
□

Remark 3 Defining an analogue of the γ-scale function by

$$W(x) := e^{-U^+x}-C^-e^{U^-x}C^+$$

for $x>0$, we see first that $G(b)=-W'(b)[W(b)]^{-1}$ where $W'(b)$ denotes the derivative of the function $W(x)$ at b. Setting $\gamma=0$, the mean total overflow during a busy cycle can be computed as

$$\begin{aligned}V(b|u) &:= \mathbb{E}\left(D|X_0=u\right)=\Psi^+(b|u)\,\mathbb{E}\left(D|X_0=b\right)\\ &= \Psi^+(b|u)\int_0^\infty \mathbb{P}(D>x|X_0=b)\,dx\\ &= \Psi^+(b|u)\int_0^\infty M(x,0)\,dx=\Psi^+(b|u)\,[-G(b)]^{-1}\\ &= \left(W^+e^{-U^+u}-W^-e^{U^-u}C^+\right)\left(-U^+e^{-U^+b}+C^-e^{U^-b}\left(-U^-\right)C^+\right)^{-1}\end{aligned}$$

Example 5 We continue the example in remark 2 of a Brownian motion fluid flow. Since there is only one phase, we get $W^+=W^-=C^+=C^-=1$ and hence

$$V(b|u)=\frac{e^{ru}-e^{su}}{re^{rb}-se^{sb}}$$

which is equation (2.11) in [14]. Note that for $\gamma=0$ we obtain

$$(s,r)=\begin{cases}\left(-2\frac{\mu}{\sigma^2},0\right), & \mu>0\\ \left(0,-2\frac{\mu}{\sigma^2}\right), & \mu<0\end{cases}$$

This implies

$$\mathbb{E}(D)=\begin{cases}\dfrac{\sigma^2}{2\mu}\left(e^{2\mu b/\sigma^2}-e^{2\mu(b-u)/\sigma^2}\right), & \mu>0\\ -\dfrac{\sigma^2}{2\mu}\left(e^{2\mu(b-u)/\sigma^2}-e^{2\mu b/\sigma^2}\right), & \mu<0\end{cases}$$

cf. equation (2.22) in [14] for the case $\mu>0$.

Example 6 Another example is the M/M/1 queue. Starting in the descending phase with initial buffer content $X_0 = u$, we obtain for the mean discounted overflow during a busy cycle

$$\begin{aligned} V(b|u) &= \left(A^{+}e^{-U^{+}u} - e^{U^{-}u}A^{+}\right)\left(-U^{+}e^{-U^{+}b} + A^{-}e^{U^{-}b}\left(-U^{-}\right)A^{+}\right)^{-1} \\ &= \frac{\frac{\beta-R}{\beta}e^{Ru} - e^{-\rho u}\frac{\beta-R}{\beta}}{Re^{Rb} + \frac{\beta}{\beta+\rho}e^{-\rho b}\rho\frac{\beta-R}{\beta}} \\ &= \frac{\beta - R}{\beta}(\beta+\rho)\frac{e^{Ru} - e^{-\rho u}}{(\beta+\rho)R \cdot e^{Rb} + (\beta - R)\rho \cdot e^{-\rho b}} \end{aligned}$$

Note that this is different from formula (7.8) in [15], as the M/M/1 queue is the negative of the compound Poisson model in risk theory. Setting $\gamma = 0$ and assuming the stability condition $\beta > \lambda$ holds, we obtain

$$R = \beta - \lambda \qquad \text{and} \qquad \rho = 0$$

This yields

$$\mathbb{E}(D) = \frac{\lambda}{\beta \cdot (\beta - \lambda)}\left(e^{(\lambda-\beta)\cdot(b-u)} - e^{(\lambda-\beta)\cdot b}\right)$$

for the mean total overflow during a busy cycle.

References

1. Ahn, S., Badescu, A., Ramaswami, V.: Time dependent analysis of finite buffer fluid flows and risk models with a dividend barrier. Queueing Syst. 55(4), 207–222 (2007)
2. Asmussen, S., Jobmann, M., Schwefel, H.P.: Exact buffer overflow calculations for queues via martingales. Queueing Syst. 42, 63–90 (2002)
3. Asmussen, S.: Stationary distributions for fluid flow models with or without Brownian motion. Stochastic Models 11, 1–20 (1995)
4. Asmussen, S.: Applied Probability and Queues. New York etc.: Springer (2003)
5. Asmussen, S., Avram, F., Pistorius, M.: Russian and American put options under exponential phase-type Lévy models. Stochastic Processes and their Applications 109, 79–111 (2004)
6. Asmussen, S., Kella, O.: A multi-dimensional martingale for Markov additive processes and its applications. Adv. Appl. Prob. 32, 376–393 (2000)
7. Asmussen, S., Nerman, O., Olsson, M.: Fitting phase-type distributions via the EM algorithm. Scand. J. Stat. 23(4), 419–441 (1996)
8. Badescu, A., Breuer, L., da Silva Soares, A., Latouche, G., Remiche, M.A., Stanford, D.: Risk processes analyzed as fluid queues. Scandinavian Actuarial Journal pp. 127–141 (2005)
9. Bean, N., O'Reilly, M., Taylor, P.G.: Hitting probabilities and hitting times for stochastic fluid flows. Probab. Eng. Inf. Sci. 23, 121–147 (2009)
10. Breuer, L.: An EM Algorithm for Batch Markovian Arrival Processes and its Comparison to a Simpler Estimation Procedure. Annals of Operations Research 112, 123–138 (2002)

11. Breuer, L.: Continuity of the M/G/c queue. Queueing Syst. 58(4), 321–331 (2008)
12. Breuer, L.: First passage times for Markov–additive processes with positive jumps of phase–type. J. Appl. Prob. 45(3), 779–799 (2008)
13. Frostig, E.: On risk model with dividends payments perturbed by a Brownian motion - an algorithmic approach. Astin Bulletin 38(1), 183–206 (2008)
14. Gerber, H., Shiu, E.: Optimal Dividends: Analysis with Brownian motion. North American Actuarial Journal 8(1), 1–20 (2004)
15. Gerber, H., Shiu, E.: On the time value of ruin. North American Actuarial Journal 2(1), 48–78 (1998)
16. Jiang, Z., Pistorius, M.: On perpetual American put valuation and first-passage in a regime-switching model with jumps. Finance Stoch. 12, 331–355 (2008)
17. Lucantoni, D.M.: New results on the single server queue with a batch Markovian arrival process. Commun. Stat., Stochastic Models 7(1), 1–46 (1991)
18. Neuts, M.F.: Matrix–Geometric Solutions in Stochastic Models. Johns Hopkins University Press, Baltimore (1981)