Queueing Models and
Service Management
Vol. 1, No. 2, page 001-029, 2018

QMSM

Analysis of Fluid Flow Models

Guy Latouche[1,*] and Giang T. Nguyen[2]

[1]Département d'informatique, Université libre de Bruxelles
CP 212, Boulevard du Triomphe, 1050 Bruxelles, Belgium
[2]School of Mathematical Sciences
The University of Adelaide, SA 5005, Australia
(Received October 2017; accepted March 2018)

Abstract: Markov-modulated fluids have a long history. They form a simple class of Markov additive processes, and were initially developed in the 1950s as models for dams and reservoirs, before gaining much popularity in the 1980s as models for buffers in telecommunication systems, when they became known as fluid queues. More recent applications are in risk theory and in environmental studies. In telecommunication systems modelling, the attention focuses on determining the stationary distribution of the buffer content. Early ODE resolution techniques have progressively given way to approaches grounded in the analysis of the physical evolution of the system, and one only needs now to solve a Riccati equation in order to obtain several quantities of interest. To the early algorithms proposed in the Applied Probability literature, numerical analysts have added new algorithms, improved in terms of convergence speed, numerical accuracy, and domain of applicability. We give here a high-level presentation of the matrix-analytic approach to the analysis of fluid queues, briefly address computational issues, and conclude by indicating how this has been extended to more general processes.

Keywords: Algorithmic analysis, fluid queues, Markov modulation, Markov-regenerative process, stationary distribution, stochastic processes.

1. Introduction

Markov-modulated processes are popular because they can be used to describe the evolution of simple systems under conditions that vary in time. The fluid flow processes presented here have long found applications as models for dams and reservoirs (Loynes [42]), and for buffers in telecommunication systems, Anick *et al.* [2] being a famous early paper. It is in the latter context that the term *fluid queue* was coined. Later, the domain of applicability of fluid flows has been extended to risk theory (Avram and Usábel [5], Badescu *et al.* [6]), operations management (Bean *et al.* [12]) and others.

Fluid flows are two-dimensional processes $\{X(t), \varphi(t) : t \in \mathbb{R}^+\}$, where $\{\varphi(t)\}$ is a continuous-time, irreducible Markov chain on some finite state space $\mathcal{S} = \{1, \ldots, m\}$, and $X(t)$ takes values in $\mathbb{R}$ under the control of φ. In the simplest form,

$$X(t) = X(0) + \int_0^t c_{\varphi(s)} \mathrm{d}s, \tag{1}$$

* Corresponding author
Email : latouche@ulb.ac.be.

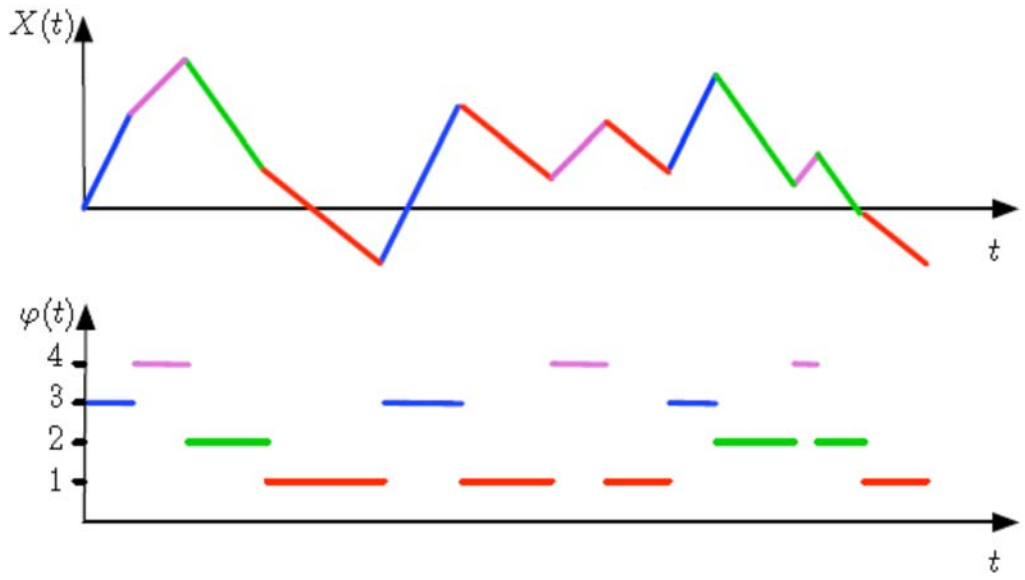

Figure 1. Sample trajectory of a fluid flow (top) and of its controlling Markov chain (bottom). The fluid rates vector is $\boldsymbol{c} = [-0.8, -1.4, 2, 1]$ so that different states correspond to different slopes.

with $\mathbf{c} = [c_i : i \in \mathcal{S}]$ being a vector of arbitrary real constants. The component X is called the level, φ is called the phase, and the level is a piecewise-linear function, with constant slope over intervals when the phase is constant. An example with four phases is shown on Figure 1.

It will be useful in the sequel to partition $\mathcal{S}$ into three subsets according to the sign of the rates c_i:

$$\mathcal{S}_+ = \{i \in \mathcal{S} : c_i > 0\}, \quad \mathcal{S}_- = \{i \in \mathcal{S} : c_i < 0\}, \quad \mathcal{S}_0 = \{i \in \mathcal{S} : c_i = 0\}. \qquad (2)$$

In many applications, X is a model for a physical quantity (water in a reservoir, packets in a buffer, energy level of a battery, etc.) and may not take negative values. In such cases, one might assume that whenever the level becomes equal to 0 and the rate is negative, the level remains equal to 0 until there is a change to a phase with positive rate. This is illustrated in Figure 2: at time θ_1 the fluid hits level 0, the phase is equal to 1 with $c_1 = -0.8$. The fluid remains equal to 0 until time δ_1 where the phase process switches to state 3, with $c_3 = 2$.

Such a mechanism is justified by the fact that the rates c_i often result from the superposition of different effects, some which remove fluid from the buffer and some which add fluid. If at some time the buffer is empty and the output rate remains greater than the input rate, fluid does not accumulate and the buffer remains empty.

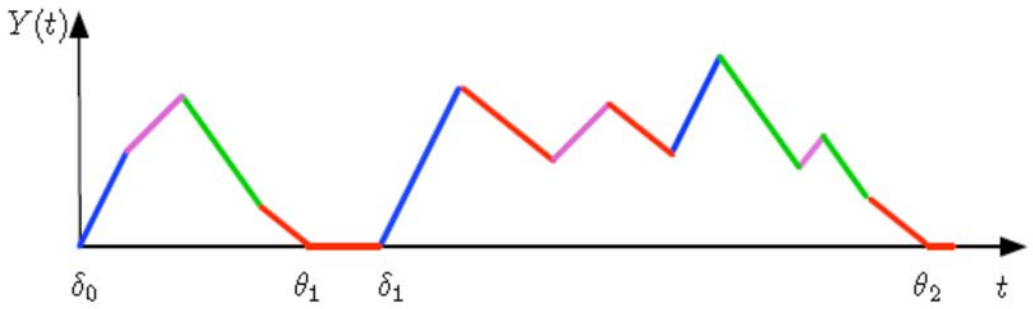

Figure 2. Sample trajectory of a regulated fluid flow.

Formally, we define the regulator $R(t) = -\min(0, \min_{0 \le s \le t} X(s))$ and the fluid queue (the regulated fluid flow) is $\{Y(t), \varphi(t)\}$, with

$$Y(t) = X(t) + R(t). \tag{3}$$

In the first part of this paper, we focus on characterising the stationary distribution of the regulated fluid queue, when it exists. Although the details of our presentation are very much inspired by Ramaswami [44] and da Silva Soares and Latouche [22], we follow a slightly different path and give explicit reference to semi-regenerative processes (Çinlar [20, Chapter 10]); this allows us to interpret in a unified manner the ad-hoc analysis of several published variants of our basic model.

Let us assume that $Y(0) = 0$ and define the sequences $\{\delta_n : n \ge 0\}$ and $\{\theta_n : n \ge 1\}$ as follows:

$$\delta_0 = \inf\{t > 0 : Y(t) > 0\},$$

$$\theta_n = \inf\{t > \delta_{n-1} : Y(t) = 0\}, \qquad \delta_n = \inf\{t > \theta_n : Y(t) > 0\}$$

(see an illustration in Figure 2). It is easily seen that $\{\theta_n : n \ge 1\}$ is a set of regenerative epochs for the process $\{Y(t), \varphi(t)\}$:

- $\{Y(\theta_n), \varphi(\theta_n)\}$ is a Markov chain on the state space $\{0\} \times \mathcal{S}_-$ as the fluid is reaching down to level zero at these epochs, in a phase of $\mathcal{S}_-$,
- the process over the interval $[\theta_n, \infty)$ is independent of the process over the interval $[0, \theta_n)$, given $\varphi(\theta_n)$, for all n, and
- the distribution of the process over the interval $[\theta_n, \infty)$, given $\varphi(\theta_n) = s$, is the same as the distribution of the process over the interval $[\theta_1, \infty)$, given $\varphi(\theta_1) = s$ for all n and $s \in \mathcal{S}_-$.

In consequence, we may immediately write that the stationary distribution $\boldsymbol{G}(x) = \left[G_i(x) : i \in \mathcal{S}\right]$, defined as[1]

[1]We use boldface letters to represent vectors, and capital letters for matrices.

$$G_j(x) = \lim_{t\to\infty} \mathrm{P}[Y(t) \le x, \varphi(t) = j], \qquad j \in \mathcal{S},$$

is given by

$$\boldsymbol{G}^{\mathrm{T}}(x) = (\boldsymbol{\rho}^{\mathrm{T}}\boldsymbol{m})^{-1}\boldsymbol{\rho}^{\mathrm{T}}M(x) \tag{4}$$

where

- $\boldsymbol{\rho}$ is the stationary probability vector of $\{(Y(\theta_n), \varphi(\theta_n))\}$,
- $M(x)$ is an $|\mathcal{S}_-|$ by $|\mathcal{S}|$ matrix, with components $M_{ij}(x)$ equal to the conditional expected sojourn time of $(Y(t), \varphi(t))$ in $[0,x]\times\{j\}$ during a regeneration interval $[\theta_n, \theta_{n+1})$, given that $\varphi(\theta_n) = i$, and
- $\boldsymbol{m} = M(\infty)\mathbf{1}$, where $\mathbf{1}$ represents a vector of ones; the components of $\boldsymbol{m}$ are the conditional expected lengths of intervals between regeneration points, given the phase at the beginning of the interval.

We give in Section 2 a few basic characteristics of the process at level 0. The vector $\boldsymbol{\rho}$ is determined in Section 3, where we define and analyse two important first passage probability matrices, and $M(x)$ is determined in Section 4 through the number of crossings of a given level during regenerative intervals. The results in these two sections are brought together in Section 5 to give the stationary distribution of the fluid queue.

The key matrices defined in Sections 3 and 4 have very distinct physical significance but they are algebraically closely related. We show this in Section 6, using results originally proved in Rogers [46].

In Section 7, we characterise the *distribution* of first passage times to a given level, and we analyse in Section 8 the first exit from an *interval*. The two sections come as complements to Section 3.

In many applications, in particular in telecommunication modeling, the buffer level is not allowed to grow without bounds. Often, the evolution of the process changes as the upper or lower boundary is reached. These, and other modifications of the basic fluid flow model, are briefly discussed in Section 10 where we show how the regenerative approach may be readily adapted to more complex assumptions.

One of the nice features of the matrix-analytic approach is that computational algorithms are easily constructed, following the development of the theoretical results. As an illustration, we give in Section 9 two of the simplest, and yet very efficient, algorithms for the numerical computation of the key matrix Ψ identified in Section 3.

Most of the results presented here have appeared earlier. For that reason, we give explanatory justifications mostly, and we refer to published sources for formal justifications. In a few cases, however, we give formal proofs: in Section 8 we give a new treatment of escape probabilities for null recurrent processes (Lemma 8.3 and Theorem 8.5), and we offer with Theorem 9.1 a novel justification for a nice computational procedure for Ψ.

2. Preliminaries

We partition the generator Q of the Markov process $\{\varphi(t) : t \in \mathbb{R}^+\}$ in a manner conformant to the partition (2) of $\mathcal{S}$ and write, possibly after a permutation of rows and columns,[2]

$$Q = \begin{bmatrix} Q_{++} & Q_{+-} & Q_{+0} \\ Q_{-+} & Q_{--} & Q_{-0} \\ Q_{0+} & Q_{0-} & Q_{00} \end{bmatrix}. \tag{5}$$

We assume that Q is irreducible. We also define the diagonal matrix C of fluid rates, $C = \mathrm{diag}(\boldsymbol{c})$, and we partition it as

$$C = \begin{bmatrix} C_+ & & \\ & C_- & \\ & & 0 \end{bmatrix}.$$

The fluid process $\{X(t)\}$ moves up and down in a random manner but its general direction is determined by the stationary drift $\mu = \boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{c}$, where $\boldsymbol{\alpha}$ is the stationary probability vector of Q : $\boldsymbol{\alpha}^{\mathrm{T}}Q = \mathbf{0}$, $\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{1} = 1$. If $\mu > 0$, the process eventually drifts to $+\infty$, that is, $\lim_{t\to\infty} X(t) = \infty$, if $\mu < 0$, then $\lim_{t\to\infty} X(t) = -\infty$; in both cases the process is transient. If $\mu = 0$, the process is null-recurrent and $\limsup X(t) = \infty$ while $\liminf X(t) = -\infty$ (Asmussen [4, Page 314, Proposition 2.10]).

Things are slightly different for the regulated process $\{Y(t)\}$. If $\mu < 0$, the process repeatedly alternates between intervals of time where $Y(t) > 0$ and intervals where $Y(t) = 0$. As we find in Section 6, the length of each cycle has finite expectation, the regulated process is positive recurrent, and we may determine its stationary distribution. If $\mu > 0$, then $Y(t)$ might not return to level 0 and so the process is transient. Not surprisingly, it is null-recurrent if $\mu = 0$.

We define $\overline{\varphi}_n = \varphi(\theta_n)$, $n \geq 1$. The process $\{\overline{\varphi}_n\}$ of the phases visited at epochs of regeneration is a Markov chain on the state space $\mathcal{S}_-$ by definition, and its transition matrix is H, with

$$H_{ij} = \mathrm{P}[\theta_{n+1} - \theta_n < \infty, \overline{\varphi}_{n+1} = j \mid \overline{\varphi}_n = j], \quad i, j \text{ in } \mathcal{S}_-,.$$

To determine H, we split the interval $(\theta_n, \theta_{n+1}]$ in two and we condition on the phase occupied at time δ_n. Thus, H is the product

$$H = \Phi\Psi, \tag{6}$$

where

[2]We write $\boldsymbol{v}_+$ for the subvector $[v_i : i \in \mathcal{S}_+]$ of any vector $\boldsymbol{v}$, and M_{++} for the submatrix of any matrix M at the intersection of the rows and columns in $\mathcal{S}_+$. Other sub-vectors and sub-blocks are similarly defined.

$$\Phi_{ik} = \mathrm{P}[\delta_n - \theta_n < \infty, \varphi(\delta_n) = k \mid \overline{\varphi}_n = i], \qquad i \in \mathcal{S}_-, k \in \mathcal{S}_+,$$
$$\Psi_{kj} = \mathrm{P}[\theta_{n+1} - \delta_n < \infty, \overline{\varphi}_{n+1} = j \mid \varphi(\delta_n) = k], \qquad k \in \mathcal{S}_+, j \in \mathcal{S}_-. \tag{7}$$

The matrix Φ is easily determined: $\varphi(t)$ remains in $\mathcal{S}_- \cup \mathcal{S}_0$ during the interval (θ_n, δ_n), and thus

$$\Phi = \begin{bmatrix} I & 0 \end{bmatrix} \left(- \begin{bmatrix} Q_{--} & Q_{-0} \\ Q_{0-} & Q_{00} \end{bmatrix}^{-1} \right) \begin{bmatrix} Q_{-+} \\ Q_{0+} \end{bmatrix} \tag{8}$$

(Latouche and Ramaswami [41, Section 5.5]). To determine the matrix Ψ requires more effort, and we devote Section 3 to the determination of its characteristic equation. Once Ψ is known, the vector $\boldsymbol{\rho}$ in (4) is determined by the system $\boldsymbol{\rho}^{\mathrm{T}} H = \boldsymbol{\rho}^{\mathrm{T}}$, $\boldsymbol{\rho}^{\mathrm{T}} \mathbf{1} = 1$.

In a similar manner, we decompose $M(x)$ as $M(x) = M(0) + M(0,x]$, where $M(0)$ is the expected time spent at level 0 between the epochs θ_n and δ_n, and $M(0,x]$ is the expected time spent in the semi-open interval $(0,x]$. We decompose $M(0)$ as

$$M(0) = \begin{bmatrix} M(0)_{-+} & M(0)_{--} & M(0)_{-0} \end{bmatrix},$$

and immediately note that $M(0)_{-+} = 0$ as the fluid queue does not spend any time at level 0 in a phase of $\mathcal{S}_+$. Furthermore,

$$\begin{bmatrix} M(0)_{--} & M(0)_{-0} \end{bmatrix} = \begin{bmatrix} I & 0 \end{bmatrix} \left(- \begin{bmatrix} Q_{--} & Q_{-0} \\ Q_{0-} & Q_{00} \end{bmatrix}^{-1} \right), \tag{9}$$

see [41, Section 5.5]. The second term in $M(x)$ is equal to

$$M(0,x] = \Phi \widetilde{M}(x), \tag{10}$$

where $\widetilde{M}(x)$ is the matrix of expected times spent in $(0,x]$ during the interval of time $(\delta_n, \theta_{n+1}]$. It is determined in Section 4.

3. First Passage Probabilities

We deal in this section with the fluid flow model $\{X(t), \varphi(t)\}$ without boundary. Its transition structure is independent of the level and so we shall not always pay close attention to the exact value of X, but be more interested in differences of level. For instance, the matrix Ψ defined in (7) might have been defined as

$$\Psi_{ij} = \mathrm{P}[\Theta < \infty,\ \varphi(\Theta) = j \mid \varphi(0) = i],\ i \in \mathcal{S}_+, j \in \mathcal{S}_-$$

independently of $X(0)$, where $\Theta = \inf\{t > 0 : X(t) = X(0)\}$ is the first return time to the initial level, starting in a phase of $\mathcal{S}_+$.

Furthermore, let

$$\tau_x^- = \inf\{t : X(t) < X(0) - x\}, \tag{11}$$

be the first passage time to level $X(0) - x$, for $x > 0$, and denote by $\gamma^-(x) = \varphi(\tau_x^-)$ the

value of the phase when $X(0)-x$ is reached for the first time.[3] For $\varphi(0)$ in $\mathcal{S}_-$, the process $\{\gamma^-(x): x \geq 0\}$ is a continuous-parameter Markov process on the state space $\mathcal{S}_-$, and there exists a matrix U such that

$$(e^{Ux})_{ij} = \mathrm{P}[\tau_x^- < \infty, \gamma^-(x) = j \mid \gamma^-(0) = i], \quad i, j \in \mathcal{S}_-.$$

As Q is irreducible, Ψ is strictly positive, meaning that $\Psi_{ij} > 0$ for all i in $\mathcal{S}_+$, j in $\mathcal{S}_-$. To see this, imagine a trajectory of positive probability such that, starting from $(0,i)$, the process returns at 0 for the first time in phase j, in finite time. A formal proof is in Govorun *et al.* [30, Lemma 4.3], or Guo [33, Theorem 5]. In consequence, the off-diagonal elements of the generator U are all strictly positive. We discuss at greater length the algebraic properties of U and Ψ in Section 6, but mention here the most important ones, in relation to the stationary drift μ:

- if $\mu \leq 0$, then $\Psi\mathbf{1} = \mathbf{1}$ and $U\mathbf{1} = \mathbf{0}$, that is, e^{Ux} is stochastic,
- if $\mu > 0$, then $\Psi\mathbf{1} < \mathbf{1}$ and $U\mathbf{1} < \mathbf{0}$, that is, e^{Ux} is substochastic.[4]

Finally, we define the matrix

$$T = \begin{bmatrix} Q_{++} & Q_{+-} \\ Q_{-+} & Q_{--} \end{bmatrix} + \begin{bmatrix} Q_{+0} \\ Q_{-0} \end{bmatrix} (-Q_{00})^{-1} \begin{bmatrix} Q_{0+} & Q_{0-} \end{bmatrix} \tag{12}$$

indexed by the states in $\mathcal{S}_+ \cup \mathcal{S}_-$ and we partition it as

$$T = \begin{bmatrix} T_{++} & T_{+-} \\ T_{-+} & T_{--} \end{bmatrix}.$$

This is the generator of the censored process $\{\varphi(t)\}$ observed only during the intervals of time spent in $\mathcal{S}_+ \cup \mathcal{S}_-$.

Theorem 3.1. *The matrix U is given by*

$$U = |C_-|^{-1} T_{--} + |C_-|^{-1} T_{-+} \Psi, \tag{13}$$

and Ψ, as defined in (7), is the minimal nonnegative solution of the quadratic Riccati equation

$$C_+^{-1} T_{+-} + C_+^{-1} T_{++} \Psi + \Psi |C_-|^{-1} T_{--} + \Psi |C_-|^{-1} T_{-+} \Psi = 0, \tag{14}$$

where $|C_-|$ is the matrix of absolute values of the elements of C_-.

We give here a high-level justification, a formal proof is in Ahn and Ramaswami [1]

---

[3] As the trajectory of $X(t)$ is continuous, we might have defined $\tau_x^- = \inf\{t : X(t) = X(0) - x\}$, but the strict inequality in (11) will be useful in Section 8.

[4] With $\boldsymbol{u}$ and $\boldsymbol{v}$ two vectors on the same set of indices, we write $\boldsymbol{u} < \boldsymbol{v}$ if $u_i < v_i$ for all index i.

and da Silva Soares and Latouche [24]. To determine Ψ, it is simpler to use the censored process on $\mathcal{S}_+ \cup \mathcal{S}_-$ since there is no change in the level while φ is in $\mathcal{S}_0$; that is why we use in (14) the generator T of (12) instead of the generator Q of (5).

Furthermore, instead of tracking the evolution of the phase process in *time* as one might be tempted to do, we track its evolution over changes of the *level*. The parameters $|c_i|$, for $i \in \mathcal{S}_+ \cup \mathcal{S}_-$, are conversion rates of time to fluid and their reciprocal $|c_i|^{-1}$ are conversion rates of fluid to time, so that the matrix $|C|^{-1} T$ indicates how the phase evolves as the fluid level is increasing or decreasing.

With this in mind, we write

$$\Psi = \int_0^\infty e^{C_+^{-1} T_{++} y} C_+^{-1} T_{+-} e^{Uy} \, \mathrm{d}y. \tag{15}$$

The justification goes as follows (see the illustration in Figure 3). The process starts in a phase of $\mathcal{S}_+$, and we assume without loss of generality that $X(0) = 0$. The factor $e^{C_+^{-1} T_{++} y}$ in the right-hand side of (15) is the probability that the phase remains in $\mathcal{S}_+$ until the fluid has increased up to level y. The factor $C_+^{-1} T_{+-} \mathrm{d}y$ is the probability that between the levels y and $y + \mathrm{d}y$, the phase moves to $\mathcal{S}_-$ and starts to decrease. The factor e^{Uy} is the probability that the fluid eventually goes down to level 0. Level y is reached at some unspecified moment $t(y)$, and the process moves without constraint between $t(y)$ and Θ.

We pre-multiply both sides of (15) by $C_+^{-1} T_{++}$ integrate by part, and find

$$C_+^{-1} T_{+-} + C_+^{-1} T_{++} \Psi + \Psi U = 0. \tag{16}$$

This is a nonsingular Sylvester equation (Lancaster and Tismenetsky [37]) and we may characterise Ψ as the unique solution of (16) if U is known.

To prove (13), we write

$$\begin{aligned} U_{ij} h &= \mathrm{P}[\gamma^-(u+h) = j \,|\, \gamma^-(u) = i] + o(h), \\ &= (|C_-|^{-1} T_{--})_{ij} h + (|C_-|^{-1} T_{-+} \Psi)_{ij} h + o(h) \end{aligned} \tag{17}$$

for i, j in $\mathcal{S}_-$. The first term is the probability that φ changes from i to j, the second

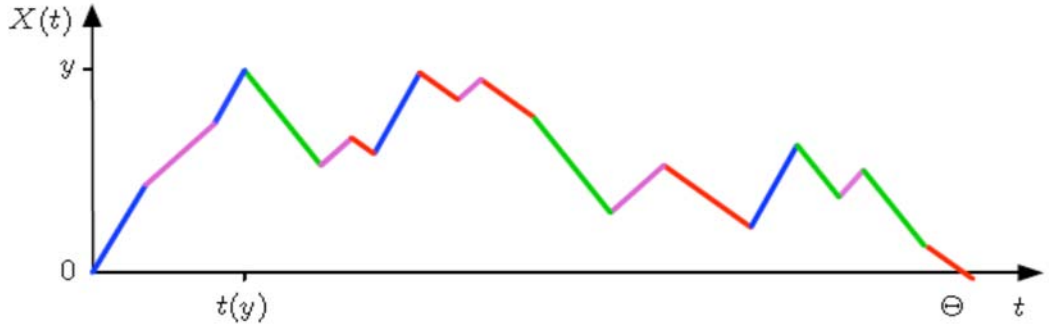

Figure 3. The fluid increases and reaches y at some unspecified time $t(y)$. Between $t(y)$ and Θ the fluid moves without constraint.

term is the probability that φ changes from i to a phase in $\mathcal{S}_+$ at some level u and is in phase j when the process later returns to level u. We illustrate this in Figure 4, where we plot the trajectory of the process from level y to level 0, and draw with solid lines the part that corresponds to $\gamma^-(x)$. At level y_a there is a simple change from phase 2 to phase 1; at level y_b there is a change from phase 2 to phase 4 in $\mathcal{S}_+$ and upon return to level y_b the process is in phase 1.

The Riccati equation (14) is obtained by replacing in (16) U by its expression in (13). Over the years, very efficient algorithms have been developed to solve (14), we describe two of these in Section 9.

Two other matrices, $\widehat{U}$ and $\widehat{\Psi}$, may be defined at this stage. Let $\tau_x^+ = \inf\{t : X(t) > X(0) + x\}$ be the first passage time to level $X(0)+x$, for $x > 0$, and denote by $\gamma^+(x) = \varphi(\tau_x^+)$ the value of the phase when level $X(0)+x$ is reached for the first time. The matrix $\widehat{U}$ is the generator of $\{\gamma^+(x)\}$ and $\widehat{\Psi}$ is the matrix of first passage probability back to the initial level, given that the phase at time 0 is in $\mathcal{S}_-$:

$$\widehat{\Psi}_{ij} = \mathrm{P}[\Theta < \infty, \varphi(\Theta) = j | \varphi(0) = i], \qquad i \in \mathcal{S}_-, j \in \mathcal{S}_+.$$

We easily adapt the argument in Theorem 3.1 to prove the following.

Corollary 3.2. *The matrix $\widehat{U}$ is given by*

$$\widehat{U} = C_+^{-1} T_{++} + C_+^{-1} T_{+-} \widehat{\Psi} \tag{18}$$

and $\widehat{\Psi}$ is the minimal nonnegative solution of the equation

$$|C_-|^{-1} T_{-+} + |C_-|^{-1} T_{--} \widehat{\Psi} + \widehat{\Psi} C_+^{-1} T_{++} + \widehat{\Psi} C_+^{-1} T_{+-} \widehat{\Psi} = 0. \tag{19}$$

Furthermore, $\widehat{\Psi}\mathbf{1} < \mathbf{1}$ and $\widehat{U}\mathbf{1} < \mathbf{0}$ if $\mu < 0$, while $\widehat{\Psi}\mathbf{1} = \mathbf{1}$ and $\widehat{U}\mathbf{1} = \mathbf{0}$ if $\mu \geq 0$.

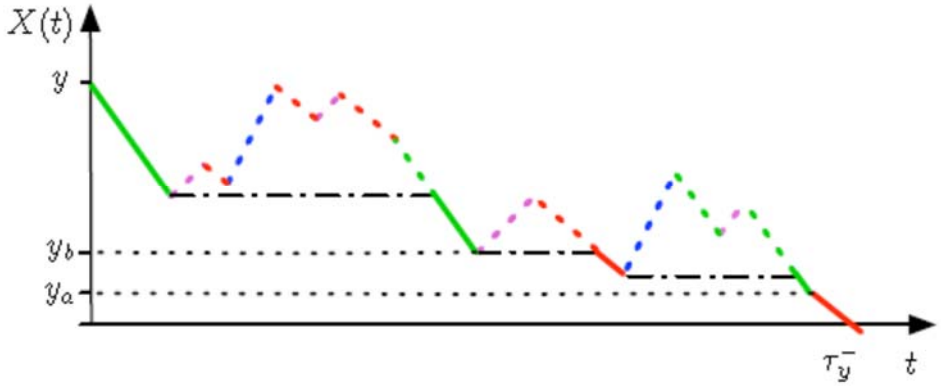

Figure 4. The sample path of $X(t)$ from y to 0 is reproduced from Figure 3, the path of $\gamma^-(x)$ corresponds to the solid downward segments, y_a and y_b are two of the levels where the value of $\gamma^-(x)$ changes.

4. Number of Crossings

As a preliminary step to determining the expected sojourn time $\widetilde{M}(x)$ in $(0,x]$ during an interval $(\delta_n, \theta_{n+1}]$, we analyse the number of times the fluid crosses a given level during a regenerative interval. We define $N_{ij}(x)$ to be the expected number of times $(X,\varphi) = (X(0)+x, j)$ during the interval $(0,\Theta]$, given that $\varphi(0) = i$, for i and j in $\mathcal{S}_+$.

Theorem 4.1. *The matrix $N(x)$ is given by $N(x) = e^{Kx}$, where*

$$K = C_+^{-1} T_{++} + \Psi \mid C_- \mid^{-1} T_{-+} \tag{20}$$

is a matrix indexed by $\mathcal{S}_+$.

The formal proof is given in Ramaswami [44] and proceeds as follows.[5] Assume without loss of generality that $X(0) = 0$, take x and $y > 0$, and count the expected number of visits to $(x+y, j)$, starting from $(0,i)$, before the first return to level 0. We group the visits to $(x+y, j)$ into subintervals between successive up-crossings of level x, and write that

$$N_{ij}(x+y) = \sum_{k \in \mathcal{S}_+} N_{ik}(x) N_{kj}(x, x+y), \qquad i, j \in \mathcal{S}_+,$$

Where $N_{kj}(x, x+y)$ is the expected number of visits to $(x+y, j)$ between two successive visits to level x, starting from (x,k). Remember that it is the *distance* y between the target $x+y$ and the starting level x that matters, not the specific location of the latter. Thus, $N(x, x+y) = N(0, y)$ and so the equation above is also written as $N(x+y) = N(x)N(y)$. From the semi-group property, we conclude that there exists a matrix K such that $N(x) = e^{Kx}$.

Next, we approximate $(e^{Kh})_{ij}$ for h small and $i \neq j$ as

$$\begin{aligned}
(e^{Kh})_{ij} &= K_{ij} h + o(h) \\
&= c_i^{-1} Q_{ij} h + \sum_{k \in \mathcal{S}_0} c_i^{-1} Q_{ik} ((-Q_{00})^{-1} Q_{0+})_{kj} h \\
&\quad + (1 + c_i^{-1} Q_{ii} h) \sum_{\ell \in \mathcal{S}_-} \Psi_{i\ell} \mid c_\ell \mid^{-1} Q_{\ell j} h \\
&\quad + (1 + c_i^{-1} Q_{ii} h) \sum_{\ell \in \mathcal{S}_-} \Psi_{i\ell} \sum_{k \in \mathcal{S}_0} \mid c_\ell \mid^{-1} Q_{\ell k} ((-Q_{00})^{-1} Q_{0+})_{kj} h + o(h).
\end{aligned}$$

The first two terms are about the process being in phase j at the first crossing of level h;

- the first is the probability that during the interval of time $c_i^{-1} h$ the phase process changes from i to j and is still in j when the fluid crosses level h;
- the second term is the probability that during that interval of time, the phase switches

[5] The argument is similar to the one used in Latouche and Ramaswami [41, Theorem 6.2.7] in the context of Quasi-Birth-and-Death processes.

to k in $\mathcal{S}_0$ at some unspecified level below h, remains at that level until there is a jump to j, and is still in j when the fluid crosses level h.

The next two terms cover the circumstances where the process does not leave phase i before crossing level h, returns to level h in some phase ℓ in $\mathcal{S}_-$, switches to phase j during the interval of time $|c_\ell|^{-1}h$ and eventually crosses level h in phase j. The $o(h)$ term captures the negligible probabilities of crossing level h more than once. Simple manipulations give us

$$(e^{Kh})_{ij} = (c_i^{-1}T_{ij} + \sum_{\ell\in\mathcal{S}_-} \Psi_{i\ell}|c_\ell|^{-1}T_{\ell j})h + o(h), \qquad i \neq j \in \mathcal{S}_+.$$

A similar argument holds for $i = j$ and so is (20) justified.

We partition the matrix $\widetilde{M}(x)$ in three blocks, corresponding to the three subsets $\mathcal{S}_+$, $\mathcal{S}_-$ and $\mathcal{S}_0$ of phases, and we deal with each one separately.

Theorem 4.2. *The matrix of expected sojourn times in* $(0,x]$ *during a regenerative interval is*

$$\widetilde{M}(x) = \begin{bmatrix}\widetilde{M}_+(x) & \widetilde{M}_-(x) & \widetilde{M}_0(x)\end{bmatrix}$$

with

$$\widetilde{M}_+(x) = \mathcal{F}(K,x)C_+^{-1}, \tag{21}$$

$$\widetilde{M}_-(x) = \mathcal{F}(K,x)\Psi|C_-|^{-1}, \tag{22}$$

$$\widetilde{M}_0(x) = \mathcal{F}(K,x)\begin{bmatrix}C_+^{-1} & \Psi|C_-|^{-1}\end{bmatrix}\begin{bmatrix}Q_{+0}\\ Q_{-0}\end{bmatrix}(-Q_{00})^{-1}. \tag{23}$$

where $\mathcal{F}(K,x) = \int_0^x e^{Ku}\,\mathrm{d}u$.

Proof. A formal proof of (21, 22) is given in Latouche and Nguyen [38], we give here a heuristic argument, and we give a justification for (23).

For i, j in $\mathcal{S}_+$, we argue that

$$\widetilde{M}_{ij}(x) = \int_0^x N_{ij}(u)c_j^{-1}\,\mathrm{d}u.$$

To see this, we interpret $c_j^{-1}\mathrm{d}u$ as the expected time spent by X in $(u,u+\mathrm{d}u)$ each time there is a visit to (u,j), and multiply this by the expected number $N_{ij}(u)$ of such visits, so that $N_{ij}(u)c_j^{-1}\mathrm{d}u$ is the total *time* spent by the process. There only remains for us to integrate over the interval $(0,x)$ and use Theorem 4.1.

Next, we define the matrix $N'(x)$ of expected number of down-crossings of level x in a phase of $\mathcal{S}_-$. As $X(0) = 0 < x$, of necessity each down-crossing of level x is preceded by an up-crossing of that same level. So, if we condition on the phase at the up-crossing, we find that

$$N'_{ij}(x) = \sum_{k\in\mathcal{S}_+} N_{ik}(x)\Psi_{kj}, \qquad i \in \mathcal{S}_+,\ j \in \mathcal{S}_-$$

and (22) is proved by repeating the argument of (21).

Finally, let us use the expression *excursion to* $\mathcal{S}_0$ for intervals of time spent in $(0,x)\times\mathcal{S}_0$ and separated by visits to $(0,x)\times\mathcal{S}_+\cup\mathcal{S}_-$. Equation (23) expresses $\widetilde{M}_0(x)$ as the product of the expected number of excursions by the time spent in individual phases during excursion. Indeed, Q_{+0} and Q_{-0} are the matrices of transition rates from $\mathcal{S}_+$ and $\mathcal{S}_-$ to $\mathcal{S}_0$, and $\mathcal{F}(K,x)[C_+^{-1} \quad \Psi|C_-|^{-1}]$ is the matrix of expected time spent in the states of $(0,x)\times\mathcal{S}_+\cup\mathcal{S}_-$. Thus, the component $N''_{ik}(x)$ of the product

$$N''(x)=\mathcal{F}(K,x)\begin{bmatrix} C_+^{-1} & \Psi|C_-|^{-1}\end{bmatrix}\begin{bmatrix} Q_{+0}\\ Q_{-0}\end{bmatrix}$$

is the expected number of excursions that start in phase k. As $(-Q_{00}^{-1})_{kj}$ is the expected time spent in j during an excursion which starts in k, this proves (23).

As we shall see in Section 6, all the eigenvalues of K are in $\mathbb{C}_{\le 0}$, that is, they have a negative real part. If $\mu<0$, the real parts are all strictly negative, otherwise K has one eigenvalue equal to 0, the others being in $\mathbb{C}_{<0}$. With this, we may express the integral $\int_0^x e^{Ku}\mathrm{d}u$ as follows.

- If $\mu<0$, then

$$\mathcal{F}(K,x)=(-K^{-1})(I-e^{Kx}), \tag{24}$$

- if $\mu\ge 0$, then

$$\mathcal{F}(K,x)=(-K^{\#})(I-e^{Kx})+x\boldsymbol{v}\boldsymbol{u}^{\mathrm{T}},$$

where $\boldsymbol{v}$ and $\boldsymbol{u}$ are the normalised right- and left-eigenvectors of K for the eigenvalue 0, and $K^{\#}$ is the group inverse[6] of K.

In the same manner as we defined $\widehat{U}$ and $\widehat{\Psi}$, we may define $N^*(x)$ as the number of crossings of level $X(0)-x$ in a phase of $\mathcal{S}_-$, starting from a phase in $\mathcal{S}_-$, before the first return to level $X(0)$, and we have the following corollary to Theorem 4.1:

Corollary 4.3. *The matrix* $N^*(x)$ *is given by* $N^*(x)=e^{\widehat{K}x}$, *where*

$$\widehat{K}=|C_-|^{-1}T_{--}+\widehat{\Psi}C_+^{-1}T_{+-} \tag{25}$$

is indexed by $\mathcal{S}_-$. *Furthermore, the eigenvalues of* $\widehat{K}$ *are in* $\mathbb{C}_{<0}$, *with the exception of one eigenvalue equal to 0 if* $\mu\le 0$.

5. Stationary Distribution

We collect the results obtained in the preceding sections and we express as follows the stationary distribution of the fluid queue, when it exists.

[6]The group inverse $\boldsymbol{K}^{\#}$ of $\boldsymbol{K}$ is the unique matrix such that $\boldsymbol{K}^{\#}\boldsymbol{K}=\boldsymbol{I}-\boldsymbol{v}\boldsymbol{u}^{\mathrm{T}},\boldsymbol{K}^{\#}\boldsymbol{v}=\boldsymbol{0}$. See Campbell and Meyer [19, Chapter 7].

Theorem 5.1. *If* $\mu < 0$, *the regulated process* $(Y(t), \varphi(t))$ *has a stationary distribution, given by*

$$\boldsymbol{G}^{\mathrm{T}}(x) = c(\begin{bmatrix} 0 & \boldsymbol{p}_-^{\mathrm{T}} & \boldsymbol{p}_0^{\mathrm{T}} \end{bmatrix} + (\boldsymbol{p}_-^{\mathrm{T}} Q_{-+} + \boldsymbol{p}_0^{\mathrm{T}} Q_{0+}) \widetilde{M}(x)),$$

where

$$\begin{bmatrix} \boldsymbol{p}_-^{\mathrm{T}} & \boldsymbol{p}_0^{\mathrm{T}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\rho}^{\mathrm{T}} & 0 \end{bmatrix} \left(- \begin{bmatrix} Q_{--} & Q_{-0} \\ Q_{0-} & Q_{00} \end{bmatrix}^{-1} \right),$$

$$\widetilde{M}(x) = (-K)^{-1}(I - e^{Kx}) \begin{bmatrix} C_+^{-1} & \Psi |C_-|^{-1} & \Gamma \end{bmatrix}, \tag{26}$$

$$\Gamma = (C_+^{-1} Q_{+0} + \Psi |C_-|^{-1} Q_{-0})(-Q_{00})^{-1},$$

and c *is the normalising constant, with*

$$c^{-1} = \begin{bmatrix} \boldsymbol{p}_-^{\mathrm{T}} & \boldsymbol{p}_0^{\mathrm{T}} \end{bmatrix} (\mathbf{1} + \begin{bmatrix} Q_{-+} \\ Q_{0+} \end{bmatrix} (-K)^{-1} (C_+^{-1}\mathbf{1} + \Psi |C_-|^{-1}\mathbf{1} + \Gamma \mathbf{1})).$$

Proof. The expression (26) for $\widetilde{M}(x)$ results from Theorem 4.2 and Equation (24). Furthermore, the vector $\boldsymbol{m}$ of conditional expected duration of a regeneration interval, given the initial phase, is given by

$$\begin{aligned} \boldsymbol{m} &= \lim_{x \to \infty} M(x)\mathbf{1} \\ &= M(0)\mathbf{1} + \lim_{x \to \infty} \Phi \widetilde{M}(x)\mathbf{1} \qquad \text{by (9, 10),} \\ &= M(0)\mathbf{1} + \Phi(-K)^{-1}(C_+^{-1}\mathbf{1} + \Psi |C_-|^{-1}\mathbf{1} + \Gamma \mathbf{1}) \end{aligned} \tag{27}$$

by (26). The normalising constant c equals $(\boldsymbol{\rho}^{\mathrm{T}} \boldsymbol{m})^{-1}$ by (4); simple calculations complete the proof.

The stationary distribution appears under various forms in the literature (Asmussen [3], Govorun *et al.* [30], Ramaswami [44], Rogers [46]); the matrix $\widetilde{M}(x)$ is a common feature in papers that rely on matrix-analytic methods, the vector of probability mass at zero and the left-factor of $\widetilde{M}(x)$ are given very different expressions, depending on the specific approach followed by the authors.

6. Wiener-Hopf Factorisation

The matrices U, Ψ and K defined in Sections 3 and 4 are related in many ways. Obviously, U and K are determined by Ψ through (13) and (20), respectively, but Ψ may be seen as a function of U by (16) and we might combine (14) and (20) to write

$$C_+^{-1} T_{+-} + K\Psi + \Psi \, | \, C_- \, |^{-1} \, T_{--} = 0,$$

from which we conclude that Ψ is a function of K. There exist also numerous relations between the 3-tuples (U, Ψ, K) and $(\widehat{U}, \widehat{\Psi}, \widehat{K})$, as we briefly discuss below.

The starting point is the Wiener-Hopf factorisation

$$C^{-1}T\begin{bmatrix} I & \Psi \\ \widehat{\Psi} & I \end{bmatrix} = \begin{bmatrix} I & \Psi \\ \widehat{\Psi} & I \end{bmatrix}\begin{bmatrix} \widehat{U} & 0 \\ 0 & -U \end{bmatrix} \tag{28}$$

proved in Rogers [46], with

$$C = \begin{bmatrix} C_+ & \\ & C_- \end{bmatrix}.$$

Equation (28) may be proved by direct verification, starting from the Riccati equations (14) and (19), and the expressions (13) and (18) for U and $\widehat{U}$. The equation below may also be proved by direct verification:

$$\begin{bmatrix} I & -\Psi \\ -\widehat{\Psi} & I \end{bmatrix} C^{-1}T = \begin{bmatrix} K & 0 \\ 0 & -\widehat{K} \end{bmatrix}\begin{bmatrix} I & -\Psi \\ -\widehat{\Psi} & I \end{bmatrix}, \tag{29}$$

An immediate consequence is that U and $\widehat{U}$ on the one hand, K and $\widehat{K}$ on the other hand, share the eigenvalues of $C^{-1}T$. We denote by m_+ and m_- respectively the number of phases in $\mathcal{S}_+$ and $\mathcal{S}_-$, and we denote by $\{\lambda_i : 1 \le i \le m_+ + m_-\}$ the eigenvalues of $C^{-1}T$, labeled in increasing values of their real parts.

Theorem 6.1. *The eigenvalues* λ_{m_+} *and* $\lambda_{m_+ +1}$ *are real, and are distinct from the others:*

$$\mathrm{Re}(\lambda_1) \le \ldots \le \mathrm{Re}(\lambda_{m_+ -1}) < \lambda_{m_+} \le \lambda_{m_+ +1} < \mathrm{Re}(\lambda_{m_+ +2}) \le \ldots \le \mathrm{Re}(\lambda_m). \tag{30}$$

Furthermore, λ_i, $1 \le i \le m_+$ *are the eigenvalues of* K *and also the eigenvalues of* $\widehat{U}$, *while* $\lambda_{m_+ +1}, \ldots \lambda_{m_+ + m_-}$ *are those of* $-U$ *and of* $-\widehat{K}$. *Finally,*

- *if* $\mu < 0$, *then* $\lambda_{m_+} < 0 = \lambda_{m_+ +1}$,
- *if* $\mu = 0$, *then* $\lambda_{m_+} = 0 = \lambda_{m_+ +1}$,
- *if* $\mu > 0$, *then* $\lambda_{m_+} = 0 < \lambda_{m_+ +1}$.

Details are given in Govorun *et al.* [30]. The main properties are summarised in Table 1.

Table 1. Crossing probabilities, and maximal eigenvalues of U, K, $\widehat{U}$, and $\widehat{K}$ in relation to the asymptotic drift μ.

drift μ	< 0	0	> 0
$\Psi\mathbf{1}$	**1**	**1**	$< \mathbf{1}$
$\widehat{\Psi}\mathbf{1}$	$< \mathbf{1}$	**1**	**1**
U, $\widehat{K}$	0	0	< 0
$\widehat{U}$, K	< 0	0	0

Actually, K and $\widehat{U}$ are similar matrices, and $\widehat{K}$ is similar to U. This is easy to show if the stationary drift is different from 0, for we find from (28, 29) after some simple algebraic manipulation that

$$K(I - \Psi\widehat{\Psi}) = (I - \Psi\widehat{\Psi})\widehat{U} \quad \text{and} \quad \widehat{K}(I - \widehat{\Psi}\Psi) = (I - \widehat{\Psi}\Psi)U. \tag{31}$$

If $\mu \neq 0$, then both $\Psi\widehat{\Psi}$ and $\widehat{\Psi}\Psi$ are strictly sub-stochastic matrices, $I - \Psi\widehat{\Psi}$ and $I - \widehat{\Psi}\Psi$ are non-singular, and we may write

$$\widehat{U} = (I - \Psi\widehat{\Psi})^{-1} K (I - \Psi\widehat{\Psi}),$$

$$\widehat{K} = (I - \widehat{\Psi}\Psi) U (I - \widehat{\Psi}\Psi)^{-1},$$

which shows that K is similar to $\widehat{U}$ and $\widehat{K}$ is similar to U.

If $\mu = 0$, however, $\Psi\widehat{\Psi}\mathbf{1} = \mathbf{1}$ and $\widehat{\Psi}\Psi\mathbf{1} = \mathbf{1}$, $I - \Psi\widehat{\Psi}$ and $I - \widehat{\Psi}\Psi$ are both singular, and the argument above fails. Instead, one must develop the Jordan chain argument from the proof of Govorun *et al.* [30, Lemma 4.6].

As a matter of fact, we often find that $\mu = 0$ is a case that presents additional difficulties. This will be seen in Theorem 8.5 about escape probabilities — that is but one example. Moreover, the convergence of computational algorithms is much slower (Guo [34]).

7. First Passage Times

The matrix H analysed in Section 2 gives us the joint probability that a regeneration interval is finite, and the phase at the end of the interval. Here, we are interested in the *distribution* of the length of the regeneration interval, that is, the distribution functions

$$H_{ij}(t) = \mathrm{P}[\theta_{n+1} - \theta_n \leq t, \overline{\varphi}_{n+1} = j \mid \overline{\varphi}_n = i],$$

for i, j in $\mathcal{S}_-$. Using the same partition that gave us (6), we write $H(t)$ as the convolution product $H(t) = \Phi * \Psi(t)$, with

$$\Phi_{ik}(t) = \mathrm{P}[\delta_n - \theta_n \leq t, \varphi(\delta_n) = k \mid \overline{\varphi}_n = i], \ i \in \mathcal{S}_-, \qquad k \in \mathcal{S}_+,$$

$$\Psi_{kj}(t) = \mathrm{P}[\theta_{n+1} - \delta_n \leq t, \overline{\varphi}_{n+1} = j \mid \varphi(\delta_n) = k], \ k \in \mathcal{S}_+, \ j \in \mathcal{S}_-.$$

As usual, it is easier to characterise the regenerative intervals through their Laplace-Stieltjes transforms

$$\widehat{H}_{ij}(s) = \int_0^\infty e^{-st} \, \mathrm{d}H_{ij}(t), \tag{32}$$

and to write $\widehat{H}(s) = \widehat{\Phi}(s)\widehat{\Psi}(s)$ where $\widehat{\Phi}(s)$ and $\widehat{\Psi}(s)$ are the matrices of LS transforms of $\Phi(t)$ and $\Psi(t)$, respectively. It is well-known that LS transforms may be interpreted in probabilistic terms through the introduction of an exponential random variable V with

parameter s, independent of the fluid flow process: we rewrite (32) as

$$\widehat{H}_{ij}(s) = \mathrm{P}[\theta_{n+1} - \theta_n \le V, \overline{\varphi}_{n+1} = j \mid \overline{\varphi}_n = i],$$

and similarly

$$\widehat{\Phi}_{ik}(s) = \mathrm{P}[\delta_n - \theta_n \le V, \varphi(\delta_n) = k \mid \overline{\varphi}_n = i],$$

$$\widehat{\Psi}_{kj}(s) = \mathrm{P}[\theta_{n+1} - \delta_n \le V, \overline{\varphi}_{n+1} = j \mid \varphi(\delta_n) = k].$$

At this point, it is easy to verify that $\widehat{\Phi}(s)$ and $\widehat{\Psi}(s)$ are given by slight modifications of (8) and (14):

$$\widehat{\Phi}(s) = \begin{bmatrix} I & 0 \end{bmatrix} \left(- \begin{bmatrix} Q_{--} - sI & Q_{-0} \\ Q_{0-} & Q_{00} - sI \end{bmatrix}^{-1} \right) \begin{bmatrix} Q_{-+} \\ Q_{0+} \end{bmatrix}, \tag{33}$$

and $\widehat{\Psi}(s)$ is the minimal nonnegative solution of

$$C_+^{-1} T_{+-}(s) + C_+^{-1} T_{++}(s) \widehat{\Psi}(s) + \widehat{\Psi}(s) |C_-|^{-1} T_{--}(s) + \widehat{\Psi}(s) |C_-|^{-1} T_{-+}(s) \widehat{\Psi}(s) = 0, \tag{34}$$

where

$$T(s) = \begin{bmatrix} Q_{++} - sI & Q_{+-} \\ Q_{-+} & Q_{--} - sI \end{bmatrix} + \begin{bmatrix} Q_{+0} \\ Q_{-0} \end{bmatrix} (sI - Q_{00})^{-1} \begin{bmatrix} Q_{0+} & Q_{0-} \end{bmatrix}.$$

The Riccati equation may be solved for any given s by the same algorithms as discussed in Section 9, and this makes it feasible to compute the distributions themselves by numerical inversion procedures.

Moments of first passage times are obtained by taking the derivatives of $\widehat{H}(s) = \widehat{\Phi}(s)\widehat{\Psi}(s)$ and evaluating it for $s = 0$. Derivatives of $\widehat{\Phi}(s)$ are easily obtained from (33) but those of $\widehat{\Psi}(s)$ are more involved, as shown in Bean *et al.* [14], and are expressed as solutions of nonsingular Sylvester equations. The *first moment* may be obtained in a more straightforward manner, as we show in (27).

8. Escape from an Interval

Section 3 is about the distribution of the phase upon the first passage of $X(t)$ to a given level. Here, we deal with the first passage to the boundary of a finite interval: assuming that $X(0) = 0$, we look for the distribution of the phase when the level escapes for the first time from the interval $(-a, b)$, with a and $b \ge 0$.

We define as follows the matrices $A^{(a,b)}$ and $B^{(a,b)}$ indexed by $\mathcal{S} \times \mathcal{S}_-$ and $\mathcal{S} \times \mathcal{S}_+$, respectively:

$$A_{ij}^{(a,b)} = \mathrm{P}[\tau_a^- < \tau_b^+, \varphi(\tau_a^-) = j \mid X(0) = 0, \varphi(0) = i], \quad i \in \mathcal{S}, j \in \mathcal{S}_-,$$

$$B_{ij}^{(a,b)} = \mathrm{P}[\tau_b^+ < \tau_a^-, \varphi(\tau_b^+) = j \mid X(0) = 0, \varphi(0) = i], \quad i \in \mathcal{S}, j \in \mathcal{S}_+.$$

The matrices are partitioned into the usual subblocks:

$$\begin{bmatrix} B^{(a,b)} & A^{(a,b)} \end{bmatrix} = \begin{bmatrix} B_{++}^{(a,b)} & A_{+-}^{(a,b)} \\ B_{-+}^{(a,b)} & A_{--}^{(a,b)} \\ B_{0+}^{(a,b)} & A_{0-}^{(a,b)} \end{bmatrix}.$$

If $\varphi(0)$ is in $\mathcal{S}_0$, the process remains at level 0 for a while, before jumping to a phase in either $\mathcal{S}_+$ or $\mathcal{S}_-$. We condition on the first phase visited either in $\mathcal{S}_+$ or $\mathcal{S}_-$ and we find that

$$\begin{bmatrix} B_{0+}^{(a,b)} & A_{0-}^{(a,b)} \end{bmatrix} = (-Q_{00})^{-1} \begin{bmatrix} Q_{0+} & Q_{0-} \end{bmatrix} \begin{bmatrix} B_{++}^{(a,b)} & A_{+-}^{(a,b)} \\ B_{-+}^{(a,b)} & A_{--}^{(a,b)} \end{bmatrix}. \tag{35}$$

Remark 8.1. *We see here why we defined* τ_x^- *in Section 3 as* $\tau_x^- = \inf\{t : X(t) < X(0) - x\}$. *If we had defined it as* $\tau_x^- = \inf\{t : X(t) = X(0) - x\}$, *then (35) would not have held for* $a = 0$. *Similarly, it would not have held for* $b = 0$ *if we had defined* $\tau_x^+ = \inf\{t : X(t) \geq X(0) + x\}$.

We need the following lemma to determine the remaining entries of $B^{(a,b)}$ and $A^{(a,b)}$.

Lemma 8.2. *For* $\varphi(0)$ *in* $\mathcal{S}_+$ *or* $\mathcal{S}_-$, *the escape probability matrices are solutions of the linear system*

$$\begin{bmatrix} B_{++}^{(a,b)} & A_{+-}^{(a,b)} \\ B_{-+}^{(a,b)} & A_{--}^{(a,b)} \end{bmatrix} (I + \mathcal{P}) = \mathcal{U}, \tag{36}$$

where

$$\mathcal{P} = \begin{bmatrix} 0 & \Psi e^{U(a+b)} \\ \widehat{\Psi} e^{\widehat{U}(a+b)} & 0 \end{bmatrix},$$

$$\mathcal{U} = \begin{bmatrix} e^{\widehat{U}b} & \Psi e^{Ua} \\ \widehat{\Psi} e^{\widehat{U}b} & e^{Ua} \end{bmatrix}.$$

Proof. Assume $\varphi(0)$ is in $\mathcal{S}_+$. We have

$$e^{\widehat{U}b} = B_{++}^{(a,b)} + A_{+-}^{(a,b)} \widehat{\Psi} e^{\widehat{U}(a+b)}. \tag{37}$$

Indeed, the left-hand side gives the distribution of the phase when the process has moved up from level 0 to level b, it is decomposed in the right-hand side as the sum of the probability that the process reaches b without going down to $-a$ and the probability that it goes down to $-a$ first, then returns to level $-a$ from below, and eventually goes up by $a+b$ units, from $-a$ to b. Similarly,

$$\Psi e^{Ua} = A_{+-}^{(a,b)} + B_{++}^{(a,b)} \Psi e^{U(a+b)}. \tag{38}$$

Equations (37, 38) form the first row of the system (36). The argument for the second row is similar.

If $\mu \neq 0$, then (36) is nonsingular and it has a unique solution. The reason is that either $\Psi e^{U(a+b)}$ or $\widehat{\Psi} e^{\widehat{U}(a+b)}$ is substochastic, and so the series $\sum_{\nu \geq 0} (-1)^{\nu} \mathcal{P}^{\nu}$ is converging to $(I + \mathcal{P})^{-1}$ (see da Silva Soares and Latouche [24] for details). If $\mu = 0$, then both $\Psi e^{U(a+b)}$ and $\widehat{\Psi} e^{\widehat{U}(a+b)}$ are stochastic matrices, $I + \mathcal{P}$ is singular and we need to add one equation.

Equation (39) is one such choice, as we prove in Theorem 8.5. This equation is identical to the one given for Markov-modulated *Brownian motion* in Ivanovs [36] in a comment after Theorem 3.1, referring to a result obtained in D'Auria et al. [27, Section 7] by a spectral decomposition argument. The proof given here is based on the analysis of the stochastic process itself, it is new and for that reason we give all technical details.

Lemma 8.3. *If* $\mu = 0$*, then the escape probability matrix is such that*

$$\begin{bmatrix} B_{++}^{(a,b)} & A_{+-}^{(a,b)} \\ B_{-+}^{(a,b)} & A_{--}^{(a,b)} \end{bmatrix} \boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{h}_{+} \\ \boldsymbol{h}_{-} \end{bmatrix}, \tag{39}$$

where $\boldsymbol{h} = -Q^{\#} \boldsymbol{c}$ *and*

$$\boldsymbol{\beta} = \begin{bmatrix} b\mathbf{1}_{+} + \boldsymbol{h}_{+} \\ -a\mathbf{1}_{-} + \boldsymbol{h}_{-} \end{bmatrix}. \tag{40}$$

Proof. We know from Coolen-Schrijner and van Doorn [21, Section 3] that the deviation matrix of Q is equal to the group inverse of $-Q$, that is, $\int_0^{\infty} (e^{Qu} - \mathbf{1} \cdot \boldsymbol{\alpha}^{\mathrm{T}}) \, \mathrm{d}u = -Q^{\#}$. Therefore,

$$\begin{aligned} \boldsymbol{h} &= \int_0^{\infty} (e^{Qu} - \mathbf{1} \cdot \boldsymbol{\alpha}^{\mathrm{T}}) \, \mathrm{d}u \, \boldsymbol{c} \\ &= \lim_{t \to \infty} \int_0^{t} (e^{Qu} - \mathbf{1} \cdot \boldsymbol{\alpha}^{\mathrm{T}}) \, \mathrm{d}u \, \boldsymbol{c} \\ &= \lim_{t \to \infty} \int_0^{t} (e^{Qu} \boldsymbol{c} - \mathbf{1} \cdot \boldsymbol{\alpha}^{\mathrm{T}} \boldsymbol{c}) \, \mathrm{d}u \\ &= \lim_{t \to \infty} \int_0^{t} e^{Qu} \boldsymbol{c} \, \mathrm{d}u \end{aligned} \tag{41}$$

as $\boldsymbol{\alpha}^{\mathrm{T}} \boldsymbol{c} = \mu$ and $\mu = 0$ by assumption. Furthermore, $(e^{Qu})_{ij}$ is the probability $\mathrm{P}[\varphi(u) = j \,|\, \varphi(0) = i]$ and so (41) may be interpreted by (1) as

$$h_i = \lim_{t \to \infty} \mathrm{E}[X_t \,|\, X(0) = 0, \varphi(0) = i],$$

and we write $\boldsymbol{h} = \mathrm{E}_0[X_{\infty} | \boldsymbol{\varphi}(0)]$ for short. By conditioning on the first time the process escapes from $(-a, b)$, it is straightforward to verify that

$$\begin{aligned} \boldsymbol{h} &= \mathrm{E}_0[X_{\infty} \mathbb{1}\{\tau_b^{+} < \tau_a^{-}\} | \boldsymbol{\varphi}(0)] + \mathrm{E}_0[X_{\infty} \mathbb{1}\{\tau_a^{-} < \tau_b^{+}\} | \boldsymbol{\varphi}(0)] \\ &= B^{(a,b)}(b\mathbf{1}_{+} + \boldsymbol{h}_{+}) + A^{(a,b)}(-a\mathbf{1}_{-} + \boldsymbol{h}_{-}) \end{aligned}$$

and this concludes the proof.

Remark 8.4. *The vector* $\boldsymbol{h}$ *has a number of interesting properties that we need later. We observe that* $\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{h} = 0$ *since* $\boldsymbol{\alpha}^{\mathrm{T}}Q^{\#} = \mathbf{0}$. *Furthermore,*

$$\boldsymbol{h}_{+} = \Psi \boldsymbol{h}_{-} \quad \text{if } \mu \le 0, \tag{42}$$

$$\boldsymbol{h}_{-} = \widehat{\Psi}\boldsymbol{h}_{+} \quad \text{if } \mu \ge 0. \tag{43}$$

To justify the first equation, we use the interpretation given to $\boldsymbol{h}$, *and we use the fact that if* $\mu \le 0$, *starting from a phase in* $\mathcal{S}_{+}$, *the process returns to 0 in finite time with probability 1. The justification of (43) is similar.*

In summary, the distribution of the phase upon escaping from the interval $(-a,b)$ is given in the next theorem.

Theorem 8.5. *The distribution of the phase at first escape from* $(-a,b)$ *is given by*

$$\begin{bmatrix} B_{0+}^{(a,b)} & A_{0-}^{(a,b)} \end{bmatrix} = (-Q_{00})^{-1} \begin{bmatrix} Q_{0+} & Q_{0-} \end{bmatrix} \begin{bmatrix} B_{++}^{(a,b)} & A_{+-}^{(a,b)} \\ B_{-+}^{(a,b)} & A_{--}^{(a,b)} \end{bmatrix}$$

if $\varphi(0) \in \mathcal{S}_0$. *If* $\mu \neq 0$ *and* $\varphi(0)$ *is in* $\mathcal{S}_{+}$ *or* $\mathcal{S}_{-}$, *then*

$$\begin{bmatrix} B_{++}^{(a,b)} & A_{+-}^{(a,b)} \\ B_{-+}^{(a,b)} & A_{--}^{(a,b)} \end{bmatrix} = \mathcal{U}(I+\mathcal{P})^{-1}, \tag{44}$$

where $\mathcal{U}$ *and* $\mathcal{P}$ *are defined in Lemma 8.2.*

If $\mu \neq 0$ *and* $\varphi(0)$ *is in* $\mathcal{S}_{+}$ *or* $\mathcal{S}_{-}$, *then*

$$\begin{bmatrix} B_{++}^{(a,b)} & A_{+-}^{(a,b)} \\ B_{-+}^{(a,b)} & A_{--}^{(a,b)} \end{bmatrix} = \mathcal{U}(I+\mathcal{P})^{\#} + \omega \cdot \eta^{T}, \tag{45}$$

where $\boldsymbol{\eta}$ *is the left eigenvector of* $I+\mathcal{P}$ *for the eigenvalue 0,*

$$\boldsymbol{w} = (\boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{\beta})^{-1}(\boldsymbol{h} - \mathcal{U}(I+\mathcal{P})^{\#}\boldsymbol{\beta}),$$

and $\boldsymbol{\beta}$ *is defined in (40).*

Proof. The first two statements have been justified before, we include them for completeness. If $\mu = 0$, the matrix $\mathcal{P}$ is stochastic and irreducible. As

$$I - \mathcal{P} = \begin{bmatrix} I & -\Psi e^{U(a+b)} \\ -\widehat{\Psi} e^{\widehat{U}(a+b)} & I \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & -I \end{bmatrix} (I+\mathcal{P}) \begin{bmatrix} I & 0 \\ 0 & -I \end{bmatrix},$$

the matrices $I-\mathcal{P}$ and $I+\mathcal{P}$ are similar and $I+\mathcal{P}$ has a unique eigenvalue equal to 0. The corresponding left eigenvector $\boldsymbol{\eta}$ is such that

$$\begin{aligned}\boldsymbol{\eta}_-^{\mathrm{T}} &= -\boldsymbol{\eta}_+^{\mathrm{T}}\Psi e^{U(a+b)},\\ \boldsymbol{\eta}_+^{\mathrm{T}} &= \boldsymbol{\eta}_+^{\mathrm{T}}\Psi e^{U(a+b)}\widehat{\Psi}e^{\widehat{U}(a+b)},\end{aligned} \tag{46}$$

and we may choose $\boldsymbol{\eta}_+^{\mathrm{T}} > \mathbf{0}$. Thus, the system (36) has the solution

$$\begin{bmatrix} B_{++}^{(a,b)} & A_{+-}^{(a,b)} \\ B_{-+}^{(a,b)} & A_{--}^{(a,b)} \end{bmatrix} = \mathcal{U}(I+\mathcal{P})^{\#} + \boldsymbol{w}\cdot\boldsymbol{\eta}^{\mathrm{T}} \tag{47}$$

for some vector $\boldsymbol{w}$. We post-multiply (47) by $\boldsymbol{\beta}$, the left-hand side is equal to $\boldsymbol{h}$ by Lemma 8.3 and we obtain

$$\boldsymbol{w} = (\boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{\beta})^{-1}(\boldsymbol{h} - \mathcal{U}(I+\mathcal{P})^{\#}\boldsymbol{\beta}),$$

provided that $\eta^{\mathrm{T}}\beta \neq 0$. This is equivalent to showing that β is, indeed, linearly independent of the columns of $I+\mathcal{P}$, which in turn implies that (36, 39) is a non-singular system when $\mu = 0$.

Now,

$$\begin{aligned}\boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{\beta} &= b\boldsymbol{\eta}_+^{\mathrm{T}}\mathbf{1}_+ + \boldsymbol{\eta}_+^{\mathrm{T}}\boldsymbol{h}_+ - a\boldsymbol{\eta}_-^{\mathrm{T}}\mathbf{1}_- + \boldsymbol{\eta}_-^{\mathrm{T}}\boldsymbol{h}_-\\ &= b\boldsymbol{\eta}_+^{\mathrm{T}}\mathbf{1}_+ + a\boldsymbol{\eta}_+^{\mathrm{T}}\mathbf{1}_+ + \boldsymbol{\eta}_+^{\mathrm{T}}(\boldsymbol{h}_+ - \Psi e^{U(a+b)}\boldsymbol{h}_-)\end{aligned} \tag{48}$$

by (46). The vector $\boldsymbol{h}_-$ is indexed by phases in $\mathcal{S}_-$ and we write, for short,

$$\begin{aligned}\boldsymbol{h}_- &= \lim_{t\to\infty} \mathrm{E}_0[X(t)|\varphi(0)\in\mathcal{S}_-]\\ &= \lim_{t\to\infty}(\mathrm{E}_0[X(t)\mathbb{1}\{\tau_0^+ < t\}|\varphi(0)\in\mathcal{S}_-] + \mathrm{E}_0[X(t)\mathbb{1}\{\tau_0^+ \geq t\}|\varphi(0)\in\mathcal{S}_-])\\ &= \lim_{t\to\infty}(\mathrm{P}[\tau_0^+ < t, \varphi(\tau_0^+)|\varphi(0)\in\mathcal{S}_-]\,\mathrm{E}_0[X(t)|\tau_0^+ < t, X(\tau_0^+), \varphi(\tau_0^+)]\\ &\quad + \mathrm{P}[\tau_0^+ \geq t|\varphi(0)\in\mathcal{S}_-]\,\mathrm{E}_0[X(t)|\tau_0^+ \geq t, \varphi(0)\in\mathcal{S}_-])\\ &= \widehat{\Psi}\boldsymbol{h}_+ + \lim_{t\to\infty}\mathrm{P}[\tau_0^+ \geq t|\varphi(0)\in\mathcal{S}_-]\,\mathrm{E}_0[X(t)|\tau_0^+ \geq t, \varphi(0)\in\mathcal{S}_-])\end{aligned}$$

as $\mathrm{X}(\tau_0^+) = 0$ and $\varphi(\tau_0^+)\in\mathcal{S}_+$,

$$\leq \widehat{\Psi}\boldsymbol{h}_+$$

since $X(t) < 0$ for $t < \tau_0^+$. The inequality is strict for at least one component of $\boldsymbol{h}_-$. Otherwise, by Remark 8.4, we would get $\boldsymbol{h}_- = \widehat{\Psi}\boldsymbol{h}_+ = \widehat{\Psi}\Psi\boldsymbol{h}_-$, from which we would successively conclude that $\boldsymbol{h}_- = c\mathbf{1}$ for some scalar c, that $\boldsymbol{h}_+ = c\mathbf{1}$, and that $\boldsymbol{h} = c\mathbf{1}$, which would be in contradiction with $\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{h} = 0$, by Remark 8.4. Consequently, (48) becomes

$$\boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{\beta} > (a+b)\boldsymbol{\eta}_+^{\mathrm{T}}\mathbf{1}_+ + \boldsymbol{\eta}_+^{\mathrm{T}}(I - \Psi e^{U(a+b)}\widehat{\Psi})\boldsymbol{h}_+. \tag{49}$$

By an argument similar to the one held above, we find that

$$\begin{aligned}\boldsymbol{h}_+ &= \lim_{t\to\infty} \mathrm{E}_0[X(t)-(a+b)|\varphi(0)\in\mathcal{S}_+]+(a+b)\mathbf{1}\\ &= \lim_{t\to\infty}(\mathrm{P}_0[\tau^+_{a+b}<t,\varphi(\tau^+_{a+b})|\varphi(0)\in\mathcal{S}_+]\mathrm{E}_0[X(t)-(a+b)|\tau^+_{a+b}<t,\varphi(\tau^+_{a+b})]\\ &\quad +(a+b)\mathbf{1}\\ &\quad +\mathrm{P}_0[\tau^+_{a+b}\geq t|\varphi(0)\in\mathcal{S}_+]\mathrm{E}_0[(X(t)-(a+b)|\tau^+_{a+b}\geq t,\varphi(0)\in\mathcal{S}_-])\\ &\leq e^{\widehat{U}(a+b)}\boldsymbol{h}_+ +(a+b)\mathbf{1}\end{aligned}$$

as $X(\tau^+_{a+b})-(a+b)=0$, $\varphi(\tau^+_{a+b})\in\mathcal{S}_+$, and $X(t)-(a+b)<0$ for $t<\tau^+_{a+b}$. With this, (49) becomes

$$\begin{aligned}\boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{\beta} &> (a+b)\boldsymbol{\eta}_+^{\mathrm{T}}\mathbf{1}_+ +\boldsymbol{\eta}_+^{\mathrm{T}}(I-\Psi e^{U(a+b)}\widehat{\Psi}e^{\widehat{U}(a+b)})\boldsymbol{h}_+ -(a+b)\boldsymbol{\eta}_+^{\mathrm{T}}\mathbf{1}_+\\ &= 0.\end{aligned}$$

This completes the proof.

In the same manner as in Section 7, we may determine the LS transform of the random variable $\min(\tau^-_a,\tau^+_b)$, details are in Bean *et al.* [13].

9. Numerical Procedures

It should be clear by this point that the numerical evaluation of many quantities of interest is dependent on being able to compute the matrices Ψ and $\widehat{\Psi}$. If we replace U in (15) by the right-hand side of (13) and write

$$\Psi=\int_0^\infty e^{C_+^{-1}T_{++}y}C_+^{-1}T_{+-}e^{(|C_-|^{-1}T_{--}+|C_-|^{-1}T_{-+}\Psi)y}\,\mathrm{d}y,$$

then an obvious approach to compute Ψ is to proceed by successive substitution: define iteratively

$$\Psi_{n+1}=\int_0^\infty e^{C_+^{-1}T_{++}y}C_+^{-1}T_{+-}e^{(|C_-|^{-1}T_{--}+|C_-|^{-1}T_{-+}\Psi_n)y}\,\mathrm{d}y. \tag{50}$$

for $n\geq 0$, starting from $\Psi_0=0$. The resulting sequence is monotonically convergent to Ψ as we show in Theorem 9.1. The proof is new, and we give it in detail. It is based on the evolution of a stack σ associated to the fluid queue.

At the epochs when the phase process enters $\mathcal{S}_+$ after a sojourn in $\mathcal{S}_-$, we put the level at the top of the stack; the value recorded on top of the stack is removed when the fluid decreases to that level. Formally, we define the sequence $\{s_k : 0\leq k\leq L\}$ of epochs where the stack increases during the interval $[0,\Theta]$:

$$\begin{aligned}&s_0=0,\\ &s_k=\inf\{t>f_k:\varphi(t)\in\mathcal{S}_+\}\quad\text{with}\quad f_k=\inf\{t>s_{k-1}:\varphi(t)\in\mathcal{S}_-\},\end{aligned}$$

and $L=\sup\{k:s_k<\Theta\}$. On the sample path of Figure 5, $L=6$ and we have marked s_0 to s_6.

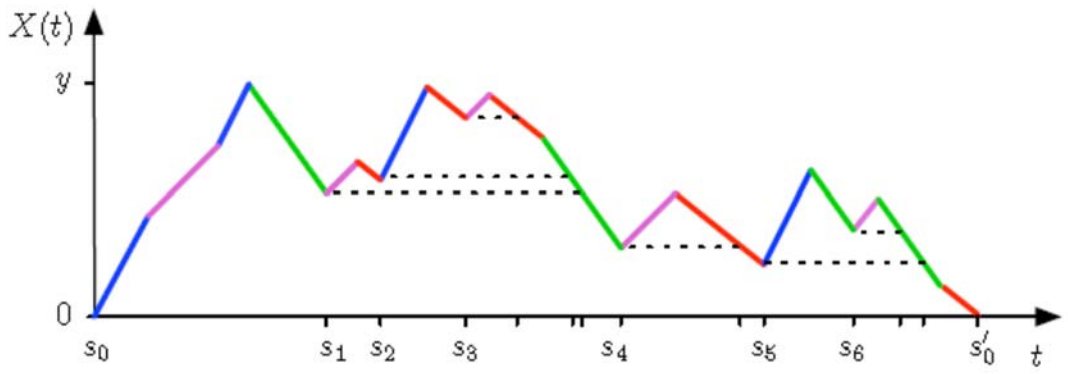

Figure 5. The fluid level $X(s_i)$ is pushed on the stack at time s_i, $i = 0,1\ldots$. The epochs s_1' to s_6' of removal of the top are marked but not labeled, they are ordered as $s_3' < s_2' < s_1' < s_4' < s_6' < s_5' < s_0'$. The maximum size of the stack is 4, reached at time s_3.

Next, we define $\{s_k' : 0 \le k \le L\}$:

$$s_k' = \inf\{t > s_k : X(t) = X(s_k)\}.$$

On Figure 5, the epochs s_1' to s_6' are marked, but not labeled so as not to clutter the graph.

At time s_k, the size $|\sigma|$ increases by one and we record $X(s_k)$ on top of σ, at time s_k' we remove the top of σ and $|\sigma|$ decreases by one. Note that $s_0' = \Theta$, and that the stack becomes empty for the first time.

Theorem 9.1. *The sequences* Ψ_n, $n \ge 0$ *defined by*

$$(\Psi_n)_{ij} = \mathrm{P}[\Theta < \infty, \varphi(\Theta) = j, \max_{0 \le t \le \Theta} |\sigma(t)| \le n \,|\, \varphi(0) = i], \tag{51}$$

for i *in* $\mathcal{S}_+$, j *in* $\mathcal{S}_-$, *satisfies (50). Furthermore,* Ψ_n *is the unique solution of the linear equation*

$$C_+^{-1}T_{+-} + C_+^{-1}T_{++}\Psi_n + \Psi_n U_{n-1} = 0, \tag{52}$$

where

$$U_{n-1} = |C_-|^{-1} T_{--} + |C_-|^{-1} T_{-+}\Psi_{n-1} \tag{53}$$

for $n \ge 1$.

The sequence $\{\Psi_n\}$ *converges monotonically to* Ψ *and* $\{U_n\}$ *converges to* U, *as* $n \to \infty$.

Proof. It is obvious that the sequence defined by (51) is monotone and converges to Ψ as there are fewer constraints on the trajectories for increasing n until there is none in the limit.

Next, we show that the return probabilities defined in (51) are solutions of (50). For $n = 1$, we have

$$\Psi_1 = \int_0^\infty e^{C_+^{-1}T_{++}y} C_+^{-1}T_{+-}\, e^{|C_-|^{-1}T_{--}y}\, \mathrm{d}y.$$

This means that the fluid queue spends some time in $\mathcal{S}_+$ and grows up to some level y, then switches to $\mathcal{S}_-$ and never returns to $\mathcal{S}_+$ until hitting level 0. We push the value 0 on σ at time 0, remove it at time Θ and so $|\sigma| = 1$ over the whole interval. This shows that Ψ_1 is a solution of (50) with $\Psi_0 = 0$.

For the general case, we illustrate on Figure 6 the physical meaning of the right-hand side of (50): the fluid process grows up to some level y, then goes down to 0 with occasional episodes of growth; during those episodes, the trajectories followed by the process are constrained by the definition of Ψ_n.

Such episodes, if any, occur during intervals (a_i, b_i), $1 \le i \le L'$, with

$$a_1 = s_1,$$
$$b_i = \inf\{t > a_i : X(t) = X(a_i)\}$$
$$a_{i+1} = \inf\{t > b_i : \varphi(t) \in \mathcal{S}_+\},$$

$L' = \sup\{i : a_i < \Theta\}$. During the intervals $(s_0, a_1 \cup_{1 \le i \le L'-1} (b_i, a_{i+1})) \cup (b_{L'}, s_0')$, σ contains only the value 0 which was pushed at time 0, and $|\sigma| = 1$. During the intervals (a_i, b_i), the stack may increase at most by n units. Thus, $|\sigma| \le n+1$ over the whole interval $[0, \Theta]$ and the right-hand side of (50) is equal to Ψ_{n+1}.

With U_n defined in (53), (50) becomes

$$\Psi_{n+1} = \int_0^\infty e^{C_+^{-1} T_{++} y} C_+^{-1} T_{+-} e^{U_n y} \, \mathrm{d}y. \tag{54}$$

and (52) follows in the same manner as we proved (16). The coefficients $C_+^{-1} T_{++}$ and U_n are both defective generators, all of their eigenvalues are in $\mathbb{C}_{<0}$, and so the system (52) has a unique solution Ψ_n. Therefore, the sequence $\{U_n, \Psi_n\}$ is well-defined and the theorem follows.

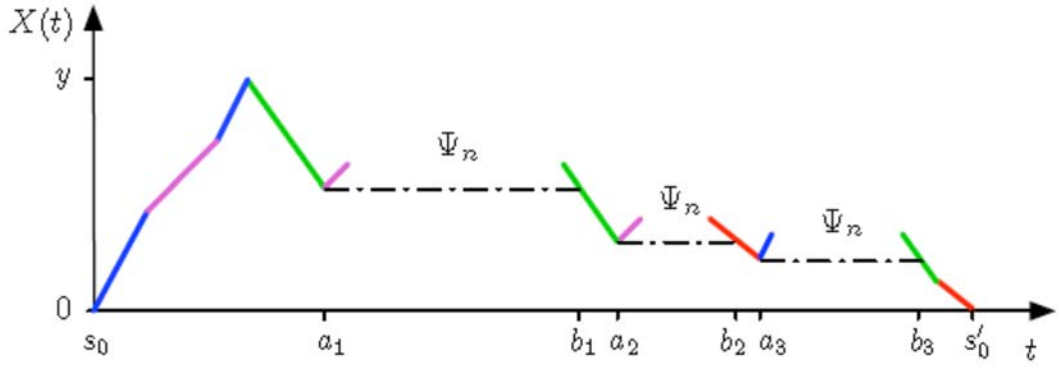

Figure 6. Approximation Ψ_{n+1} by functional iteration: the size of the stack increases at most by n over the intervals $[a_1, b_1]$, $[a_2, b_2]$ and $[a_3, b_3]$ and is at most $n+1$ over the whole interval $[s_0, s_0']$.

The algorithm defined by (52, 53) is easily implemented and is the most efficient among several *linearly* convergent algorithms, as shown in Bean *et al.* [8]. Several other procedures have been proposed in Guo [32] and Bini *et al.* [16, 17]. A special mention should be made of the Newton method as it is easily implemented and much faster than functional iteration: the sequence $\{\Psi_n^{(N)}\}$ defined by

$$C_+^{-1}T_{+-} + (C_+^{-1}T_{++} + \Psi_n^{(N)} \mid C_- \mid^{-1} T_{-+})\Psi_{n+1}^{(N)}$$
$$+\Psi_{n+1}^{(N)}(\mid C_- \mid^{-1} T_{--} + \mid C_- \mid^{-1} T_{-+}\Psi_n^{(N)}) = \Psi_n^{(N)} \mid C_- \mid^{-1} T_{-+}\Psi_n^{(N)}$$

for $n \geq 0$, with $\Psi_0^{(N)} = 0$, is well-defined and converges *quadratically* if $\mu \neq 0$.

The most efficient algorithms today form the family of doubling algorithms, which solve simultaneously for Ψ and for $\widehat{\Psi}$; they include the structure-preserving doubling algorithm (SDA, Guo *et al.* [35]), SDA shrink-and-shift (Bini *et al.* [15]), componentwise-accurate doubling algorithms (Nguyen and Poloni [43]), and alternating-directional doubling algorithm (Wang *et al.* [48]). These algorithms are quadratically convergent if $\mu \neq 0$ and, furthermore, each iteration is faster due to fewer computations.

10. Extensions

After the publication of Ramaswami's seminal paper [44], the basic fluid queue model defined through Equations (1) and (3) has been extended in many ways. We cite some of these, without getting into details, the list is far from exhaustive.

Finite buffers. In some applications, $Y(t)$ represents the content of a finite buffer, so that the level may only take values in some interval $0 \leq Y(t) \leq B < \infty$ (see da Silva Soares and Latouche [23, 24] and Bean *et al.* [13]). In that case, it is natural to choose as regeneration points the epochs of return to the upper boundary in addition to the returns to level 0, and the transition matrix H between regeneration points makes use of first passage probabilities from boundary to boundary. These may be obtained from Lemma 8.2 with $a = 0$ and $b = B$, if the process starts from the lower boundary, or $a = B$ and $b = 0$, if the process starts from the upper boundary.

The final expression for the stationary density is a mixture of two matrix exponentials, e^{Kx} and $e^{\widehat{K}(B-x)}$. Details may be found in da Silva Soares [25] and in Latouche and Nguyen [38].

Level feedback. In our presentation so far, the level is driven by the phase, subject to the boundary constrains that $Y(t) \geq 0$, or $0 \leq Y(t) \leq B$, while the phase evolves independently of the level. In many cases, the level has a direct influence on the evolution of the phase; for instance, one might reduce the flow into the buffer as the level gets nears the upper boundary, so as to avoid spillage.

In Bean *et al.* [9] and da Silva Soares [25] (and other references cited there), one defines

a number of threshold values $0 \le b_1 < b_2 < \cdots < b_N \le \infty$ such that the parameters C and Q of the phase process change upon the crossing of a threshold. Here, regenerations occur when the fluid reaches any of the threshold, and the analysis of such systems may be based on a systematic extension of the results for the system with a finite buffer.

Fluid with jumps. In Remiche [45] $Y(t)$ represents the supply of tokens in a leaky bucket system. It increases linearly in time and drops by a positive amount each time a file is transmitted. In Bean *et al.* [12], $Y(t)$ represents the amount of wear of a power generator and it may jump from B (indicating that the generator is unusable) to 0 (indicating that it has been replaced by a new equipment). As discussed in Badescu *et al.* [7] and Stanford *et al.* [47], risk processes may be analysed as fluid queues with jumps.

If the jumps have a phase-type distribution, then the analysis of the process requires very little adaptation from the material presented in the present paper.

Markov modulated Brownian motion. The definition of these processes is very similar to that of fluid flows. The difference is that the fluid evolves like a Brownian motion with parameters (drift and variance) which depend on $\varphi(t)$. Recent references are d'Auria *et al.* [26, 28], Ivanovs [36], Gribaudo *et al.* [31] where the authors focused on obtaining time-dependent distributions and first hitting times using different approaches: stochastic ODE resolution, spectral decomposition and martingale theory. Breuer [18] determined the occupation time of the process in an interval before a one- or two-sided exit. Latouche and Nguyen [38, 39] are two recent papers that follow a regenerative approach similar to the one developed here.

Two-dimensional fluid. A few authors have considered systems where the component X is two-dimensional: Bean and O'Reilly [10, 11], Foss and Miyazawa [29] and Latouche *et al.* [40] among others. The area of two- or higher-dimensional fluids is still wide open, with many exciting unanswered questions.

Acknowledgements

The authors would like to acknowledge the support of ACEMS (Australian Research Council Centre of Excellence for Mathematical and Statistical Frontiers).

References

[1] Ahn, S., & Ramaswami, V. (2003). Fluid flow models and queues: A connection by stochastic coupling. *Stochastic Models*, 19, 325–348.

[2] Anick, D., Mitra, D., & Sondhi, M. M. (1982). Stochastic theory of a data-handling system with multiple sources. *The Bell System Technical Journal*, 61, 1871–1894.

[3] Asmussen, S. (1995). Stationary distributions for fluid flow models with or without Brownian noise. *Communications in Statistics, Part C: Stochastic Models*, 11, 21–49.

[4] Asmussen, S. (2003). *Applied Probability and Queues*. Springer, New York, NY, Second edition.

[5] Avram, F., & Usábel, M. (2004). Ruin probabilities and deficit for the renewal risk model with phase-type interarrival times. *Astin Bulletin*, 34, 315–332.

[6] Badescu, A., Breuer, L., da Silva Soares, A., Latouche, G., Remiche, M. A., & Stanford, D. (2005). Risk processes analyzed as fluid queues. *Scandinavian Actuarial Journal*, 105, 127–141.

[7] Badescu, A., Breuer, L., Drekic, S., Latouche, G., & Stanford, D. (2005). The surplus prior to ruin and the deficit at ruin for a correlated risk process. *Scandinavian Actuarial Journal*, 433–445, 2005.

[8] Bean, N. G., O'Reilly, M. M., & Taylor, P. G. (2005). Algorithms for return probabilities for stochastic fluid flows. *Stochastic Models*, 21, 149–184.

[9] Bean, N. G., & O'Reilly, M. M. (2008). Performance measures of a multi-layer Markovian fluid model. *Annals of Operations Research.*, 160, 99–120.

[10] Bean, N. G., & O'Reilly, M. M. (2013). A stochastic two-dimensional fluid model. *Stochastic Models*, 29, 31–63.

[11] Bean, N. G., & O'Reilly, M. M. (2014). The stochastic fluid-fluid model: A stochastic fluid model driven by an uncountable-state process, which is a stochastic fluid model itself. *Stochastic Process and their Applications*, 124, 1741–1772.

[12] Bean, N. G., O'Reilly, M. M., & Sargison, J. E. (2010). A stochastic flow model of the operation and maintenance of power generation systems. *IEEE Transactions on Power Systems*, 25, 1361–1374.

[13] Bean, N. G., O'Reilly, M. M., & Taylor, P. G. (2009). Hitting probabilities and hitting times for stochastic fluid flows. the bounded model. *Probability in the Engineering and Informational Sciences*, 23, 121–147.

[14] Bean, N. G., O'Reilly, M. M., & Taylor, P. G. (2005). Hitting probabilities and hitting times for stochastic fluid flows. *Stochastic Process and their Applications*, 115, 1530–1556.

[15] Bini, D. A., Meini, B., & Poloni, F. (2010). Transforming algebraic Riccati equations into unilateral quadratic matrix equations. *Numerische Mathematik*,116, 553-578.

[16] Bini, D. A., Iannazzo, B., Latouche, G., & Meini, B. (2006). On the solution of Riccati equations arising in fluid queues. *Linear Algebra and its Applications*, 413, 474–494.

[17] Bini, D. A., Iannazzo, B., & Meini, B. (2006). *Numerical Solution of Algebraic Riccati Equations*. Number 9 in Fundamentals of Algorithms. SIAM, Philadelphia PA, 2012.

[18] Breuer, L. (2012). Occupation times for Markov-modulated Brownian motion. *Journal of Applied Probability*, 49, 549–565.

[19] Campbel, S. L., & Meyer, C. D. (1991). *Generalized Inverses of Linear Transformations*. Dover Publications, New York. Republication.

[20] Çinlar, E. (1975). *Introduction to Stochastic Processes*. Prentice–Hall, Englewood Cliffs, NJ.

[21] Coolen-Schrijner P. & van Doorn, E. A. (2002). The deviation matrix of a continuous time Markov chain. *Probability in the Engineering and Informational Sciences*, 16, 351–366.

[22] da Silva Soares, A., & Latouche, G. (2002). Further results on the similarity between fluid queues and QBDs. In Latouche, G. and Taylor, P. editors, *Proceedings of the 4th International Conference on Matrix-Analytic Methods*, 89–106. World Scientific, River Edge, NJ, 2002.

[23] da Silva Soares, A., & Latouche, G. (2005). A matrix-analytic approach to fluid queues with feedback control. *International Journal of Simulation: Systems, Science & Technology*, 6, 4–12.

[24] da Silva Soares, A., & Latouche, G. (2006). Matrix-analytic methods for fluid queues with finite buffers. *Performance Evaluation*, 63, 295–314.

[25] da Silva Soares, A., & Latouche, G. (2009). Fluid queues with level dependent evolution. *European Journal of Operational Research*, 196, 1041–1048.

[26] D'Auria, B., Ivanovs, J., Kella, O., & Mandjes, M. (2010). First passage of a Markov additive process and generalized Jordan chains. *Journal of Applied Probability*, 47, 1048–1057.

[27] D'Auria, B., Ivanovs, J., Kella, O., & Mandjes, M. (2010). First passage process of a Markov additive process with applications to reflection problems. *ArXiv e-prints*, arXiv:1006.2965v1 [math.PR].

[28] D'Auria, B., Ivanovs, J., Kella, O., & Mandjes, M. (2012). Two-sided reflection of Markov-modulated Brownian motion. *Stochastic Models*, 28, 316–332.

[29] Foss S., & Miyazawa, M. (2014). Two-node fluid network with a heavy-tailed random input: the strong stability case. *Journal of Applied Probability*, 51, 249–265.

[30] Govorun, M., Latouche, G., & Remiche, M. A. (2013). Stability for fluid queues: characteristic inequalities. *Stochastic Models*, 29, 64–88.

[31] Gribaudo, M., Manini, D., Sericola, B., & Telek, M. (2008). Second order fluid models with general boundary behaviour. *Annals of Operations Research*, 160, 69–82.

[32] Guo, C.-H. (2001). Nonsymmetric algebraic Riccati equations and Wiener-Hopf factorization for *M*-matrices. *SIAM Journal on Matrix Analysis and Applications*, 23, 225–242.

[33] Guo, C.-H. (2002). A note on the minimal nonnegative solution of a nonsymmetric algebraic Riccati equation. *Linear Algebra and its Applications*, 357, 299–302.

[34] Guo, C.-H. (2006). Efficient methods for solving a nonsymmetric algebraic Riccati equation arising in stochastic fluid models. *Journal of Computational and Applied Mathematics*, 192, 353–373.

[35] Guo, X.-X., Lin, W.-W., & Xu, S.-F. (2006). A structure-preserving doubling algorithm for nonsymmetric algebraic Riccati equation. *Numerische Mathematik*, 103, 93–412.

[36] Ivanovs, J. (2010). Markov-modulated Brownian motion with two reflecting barriers. *Journal of Applied Probability*, 47, 1034–1047.

[37] Lancaster, P., & Tismenetsky, M. (1985). *The Theory of Matrices*. Academic Press, New York, NY.

[38] Latouche, G., & Nguyen, G. T. (2016). Feedback control: Markov-modulated Brownian motion with instantaneous change of phase. *Performance Evaluation*, 106, 30–49.

[39] Latouche, G., & Nguyen, G. T. (2017). Slowing time: Markov-modulated Brownian motion with a sticky boundary. *Stochastic Models*, 33, 297–321.

[40] Latouche, G., Nguyen, G. T., & Palmowski, Z. (2013). Two-dimensional fluid queues with temporary assistance. In G. Latouche, V. Ramaswami, J. Sethuraman, K. Sigman, M. S. Squillante, and D. D. Yao, editors, *Matrix-Analytic Methods in Stochastic Models*, volume 27 of *Springer Proceedings in Mathematics & Statistics*, Chapter 9, 187–207. Springer, New York, NY.

[41] Latouche, G., & Ramaswami, V. (2011). *Introduction to Matrix Analytic Methods in Stochastic Modeling*. ASA-SIAM Series on Statistics and Applied Probability. SIAM, Philadelphia PA, 1999. Second printing 2011.

[42] Loynes, R. M. (1962). A continuous-time treatment of certain queues and infinite dams. *Journal of the Australian Mathematical Society*, 2, 484–498.

[43] Nguyen, G. T., & Poloni, F. (2015).Componentwise accurate fluid queue computations using doubling algorithms. *Numerische Mathematik*, 130, 763–792.

[44] Ramaswami, V. (1999). Matrix analytic methods for stochastic fluid flows. In Smith, D.and Hey, P.editors, *Teletraffic Engineering in a Competitive World (Proceedings of the 16th International Teletraffic Congress)*, 1019–1030. Elsevier Science B.V., Edinburgh, UK.

[45] Remiche, M. A. (2005). Compliance of the token bucket model for Markovian type traffic. *Stochastic Models*, 21, 615–630.

[46] Rogers, L. C. G. (1994). Fluid models in queueing theory and Wiener-Hopf factorization of Markov chains. *Annals of Applied Probabability*, 4, 390–413.

[47] Stanford, D., Avram,F., Badescu, A., Breuer, L., da Silva Soares, A., & Latouche, G. (2005). Phase-type approximations to finite-time ruin probabilities in the Sparre Andersen and stationary renewal risk models. *ASTIN Bulletin*, 35, 131–144.

[48] Wang,W.-G., Wang, W.-C., & Li, R.-C. (2012). Alternating-directional doubling algorithm for M-matrix algebraic Riccati equations. *SIAM Journal on Matrix Analysis and Applications*, 33, 170-194.