European Journal of Operational Research 238 (2014) 514–526

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Stochastics and Statistics

Multi-stage stochastic fluid models for congestion control

Małgorzata M. O'Reilly *

School of Mathematics and Physics, University of Tasmania, GPO Box 252C-37, Hobart, Tasmania 7001, Australia

ARTICLE INFO

Article history:
Received 10 December 2012
Accepted 6 April 2014
Available online 21 April 2014

Keywords:
Stochastic fluid model
Markov chain
Laplace–Stieltjes transform
Matrix-analytic methods
Congestion control

ABSTRACT

We consider multi-stage stochastic fluid models (SFMs), driven by applications in telecommunications and manufacturing in which control of the behavior of the system during congestion may be required. In a two-stage SFM, the process starts from Stage 1 in level 0, and moves to Stage 2 when reaching threshold b_2 from below. Stage 1 starts again when reaching threshold $b_1 < b_2$ from above. While in a particular stage, the process evolves according to a traditional SFM with a unique set of phases, generator and fluid rates. We first consider a two-stage SFM with general, real fluid change rates. Next, we analyze a two-stage SFM with an upper boundary $B > b_2$. Finally, we discuss a generalization to multi-stage SFMs. We use matrix-analytic methods and derive efficient methodology for the analysis of this class of models.

1. Introduction

Consider an unbounded stochastic fluid model (SFM) $\{(\varphi(t), X(t)), t \geqslant 0\}$ with phase variable $\varphi(t) \in \mathcal{S} = \{1, \ldots, m\}$, level variable $X(t) \in (-\infty, +\infty)$, generator $\mathbf{T} = [T_{ij}]_{i,j\in\mathcal{S}}$, and real rates $c_i \in (-\infty, +\infty)$, for all $i \in \mathcal{S}$, defined by the following set of conditions:

- the phase process $\{\varphi(t), t \geqslant 0\}$ is an irreducible, continuous-time Markov chain (CTMC) with finite state space $\mathcal{S}$ and generator matrix $\mathbf{T}$ which records instantaneous transition rates $T_{ij} = dP(X(t) = j|X(0) = i)/dt|_{t=0}$ between the phases,
- the level $X(t)$ at time t is changing at a constant rate $c_i = dX(t)/dt$ when $\varphi(t) = i$, for some $i \in \mathcal{S}$, see Fig. 1.

The phase variable may be used to model the behaviour of the underlying environment in some system of interest, while the level variable may be used to model a continuous performance measure of the system. As a simple example, phase may record which switch is currently on in a telecommunications buffer, while level may be the amount of data currently in the buffer. Since the phase process modulates the behaviour of the level through the rates c_i, we refer to the CTMC $\{\varphi(t), t \geqslant 0\}$ as the driving process. In particular, given the initial level $X(0)$ and the full history of the process $\{\varphi(t), t \geqslant 0\}$, the evolution of the process $\{X(t), t \geqslant 0\}$ is then completely determined, since the only element of randomness is coming from the driving CTMC. For the results from the theory of the SFMs, the development of which was motivated by the applications in telecommunications, the reader is referred to Ahn and Ramaswami (2003, 2004, 2005), Asmussen (1995), Bean and O'Reilly (2013), Bean, O'Reilly, and Taylor (2005a, 2005b, 2008), da Silva Soares and Latouche (2002), O'Reilly and Palmowski (2013), Ramaswami (1997, 1999).

Now, suppose that we want to model a situation in which a congested buffer may need to operate under a different regime, with lower input rates for example. In order to achieve this, we introduce thresholds b_1 and b_2 for controlling congestion. Specifically, the process starts from level 0 in Stage 1 (original regime) and when it first hits threshold $b_2 > 0$ from below, it moves to Stage 2 (a different regime). Following this, when it hits threshold $b_1 < b_2$ from above, it moves to Stage 1 again. Of course, if the process is already in Stage 1, then no change of stage occurs upon hitting level b_1. Whenever in a particular stage, the process evolves according to the traditional SFM with a unique set of phases, generator and fluid rates. We refer to such model as a two-stage SFM and illustrate it in Fig. 2. This class of models for controlling congestion in telecommunications buffers, is motivated by and contains a model with two thresholds introduced by Malhotra, Mandjes, Scheinhardt and van den Berg in Malhotra, Mandjes, Scheinhardt, and van den Berg (2009).

Formally, we define a two-stage stochastic fluid model (SFM), denoted by $\{(\varphi(t), X(t), Z(t)), t \geqslant 0\}$, with phase variable $\varphi(t)$, level variable $X(t) \geqslant 0$, stage variable $Z(t) \in \{1, 2\}$, and thresholds b_1, b_2, with $0 < b_1 < b_2 < B$, where $B < \infty$ or $B = \infty$, as follows.

Let $\mathcal{S}^{(\ell)} = \{1, \ldots, m_\ell\}$ be the set of phases, $\mathbf{T}^{(\ell)}$ be the generator matrix, and $c_i^{(\ell)}$ for all $i \in \mathcal{S}^{(\ell)}$ be the rates corresponding to the (SFM of) Stage ℓ, for $\ell \in \{1, 2\}$. We say that Stage ℓ, for any

* Tel.: +61 (03) 6226 2405; fax: +61 (03) 6226 2410.
E-mail address: malgorzata.oreilly@utas.edu.au

http://dx.doi.org/10.1016/j.ejor.2014.04.010
0377-2217/

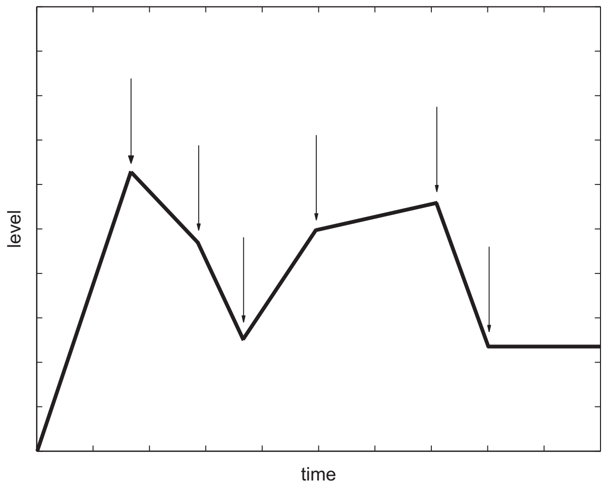

Fig. 1. SFM: The rate of change of level $X(t)$ depends on the current phase $\varphi(t)$ in the driving CTMC $\{\varphi(t), t \geqslant 0\}$. At the moment of a phase transition (indicated by the arrows) $i \to j$, the rate changes from c_i to c_j.

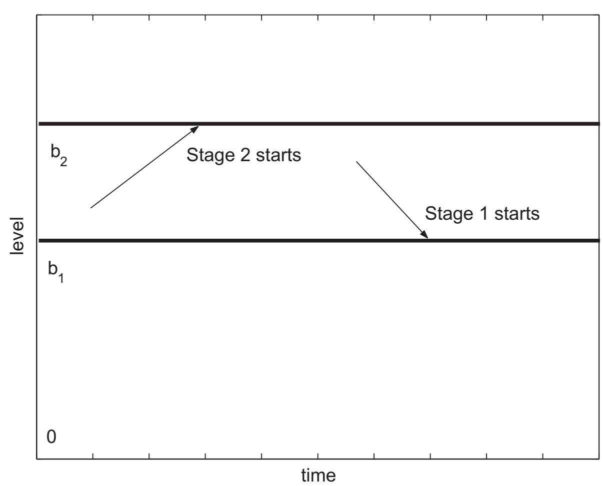

Fig. 2. Two-stage SFM: As soon the as the process hits level b_2 from below, it enters Stage 2 driven by a CTMC with state space $\mathcal{S}^{(2)}$ and generator $\mathbf{T}^{(2)}$. Following this, as soon as the process hits level b_1 from above, it enters Stage 1 driven by a CTMC with state space $\mathcal{S}^{(1)}$ and generator $\mathbf{T}^{(1)}$, and so on.

$\ell \in \{1,2\}$, is (an SFM) driven by a CTMC with state space $\mathcal{S}^{(\ell)}$ and generator $\mathbf{T}^{(\ell)}$.

- We assume $\varphi(0) \in \mathcal{S}^{(1)}, X(0) = 0$, $Z(0) = 1$. That is, the process starts from level zero in Stage 1, driven by a CTMC with state space $\mathcal{S}^{(1)}$ and generator $\mathbf{T}^{(1)}$.
- While $Z(t) = 1$, the process is in Stage 1. In such case, whenever $\varphi(t) = i$ and $X(t) > 0$, the rate of change of level $X(t)$ is given by a constant $c_i^{(1)}$, for all $i \in \mathcal{S}^{(1)}$. When $\varphi(t) = i$ and $X(t) = 0$ however, the rate of change of fluid is $\max\left\{0, c_i^{(1)}\right\}$, since zero in the lower boundary in the buffer.
- Following this, as soon as $X(t) = b_2$, the phase process moves to some phase in $\mathcal{S}^{(2)} = \{1, \ldots, m_2\}$, according to the probability matrix $\mathbf{P}^{(b_2)}$. Also, $Z(t)$ moves from 1 to 2, and we say that Stage 2 begins, driven by a CTMC with state space $\mathcal{S}^{(2)}$ and generator $\mathbf{T}^{(2)}$.
- While $Z(t) = 2$, the process is in Stage 2. In such case, whenever $\varphi(t) = i$, the rate of change of level $X(t)$ is given by a constant $c_i^{(2)}$, for all $i \in \mathcal{S}^{(2)}$.
- Following this, as soon as $X(t) = b_1$, the phase process moves to some phase in $\mathcal{S}^{(1)}$ according to the probability matrix $\mathbf{P}^{(b_1)}$. Also, $Z(t)$ moves from 2 to 1, and Stage 1 begins again.
- In the case with a finite upper boundary $B < \infty$, we assume that when $\varphi(t) = i \in \mathcal{S}^{(2)}$ and $X(t) = B$, then the rate of change of fluid is given by $\min\left\{0, c_i^{(2)}\right\}$. That is, when the buffer $X(.)$ becomes full, it can only then start to decrease (when $c_i < 0$) or remain at the same level (when $c_i \geqslant 0$).

We partition $\mathbf{P}^{(b_2)}$ according to $\mathcal{S}_1^{(1)} \times \left(\mathcal{S}_1^{(2)} \cup \mathcal{S}_2^{(2)} \cup \mathcal{S}_0^{(2)}\right)$, and $\mathbf{P}^{(b_1)}$ according to $\mathcal{S}_2^{(2)} \times \left(\mathcal{S}_1^{(1)} \cup \mathcal{S}_2^{(1)} \cup \mathcal{S}_0^{(1)}\right)$, as

$$\mathbf{P}^{(b_2)} = \left[\mathbf{P}_{11}^{(b_2)} \quad \mathbf{P}_{12}^{(b_2)} \quad \mathbf{P}_{10}^{(b_2)}\right], \; \mathbf{P}^{(b_1)} = \left[\mathbf{P}_{21}^{(b_1)} \quad \mathbf{P}_{22}^{(b_1)} \quad \mathbf{P}_{20}^{(b_1)}\right]. \tag{1}$$

In this paper, we analyze the two-stage SFM with and without an upper boundary ($B < \infty$ and $B = \infty$), respectively, and also consider a generalization to *multi-stage* SFMs.

Compared to the earlier mentioned model of Malhotra, Mandjes, Scheinhardt and van den Berg in Malhotra et al. (2009), the two-stage SFM is more general and possesses the following set of useful properties.

- We assume *any real* rates $c_i^{(\ell)}, i \in \mathcal{S}^{(\ell)}$, which allows to treat systems with general input rates.
- The change at the moment of the transition between the stages may involve not only the change in generator $\mathbf{T}^{(\ell)}$, but also *the change in the set of phases* $\mathcal{S}^{(\ell)}$, which may be useful for recording the changes in the underlying environment of a system of interest in a more practical manner.
- The change in $c_i^{(\ell)}$ at the moment of the transition between the stages allows *all possible types of changes of sign* (from $+$ or $-$ to $+, -$ and 0), which may be needed in a system with general input rates.
- We treat the model with an *upper boundary* $B < \infty$, which is a more realistic assumption for modeling real-life buffers.
- We consider a generalization to *multi-stage* SFMs, which allows to model multiple-threshold based regimes for modeling congestion.

More importantly, apart from the above listed generalizations of the model itself, we make the following contribution in terms of the methodology provided. Specifically, we would like to note that the analysis in Malhotra et al. (2009) was based on solving appropriate balance equations using a spectral expansion, and as the authors emphasized there, was difficult.

Here, we apply arguments within the theory of the matrix-analytic methods (MAMs) (Ramaswami, 1997, 1999), in particular the methods of direct analysis of SFMs developed in Bean et al. (2005b, 2009), which involve sample path decomposition and the use of the key fluid generator $\mathbf{Q}(s)$ introduced in Bean et al. (2005b). The advantages of such methodology are that the expressions are in a matrix form that do not require the use of eigenvalues, and have useful probabilistic interpretation. This makes the analysis concise and leads to efficient numerical methods based on the existing powerful algorithms (Ahn & Ramaswami, 2005; Bean et al., 2005a, 2008; Bini, Iannazzo, Latouche, & Meini, 2006; Guo, Iannazzo, & Meini, 2007).

Additionally, Malhotra et al. (2009) deals with the stationary analysis only. Here, we provide both *time-dependent and stationary* analysis of this more general model, and introduce rates $\delta_{2\to1}, \delta_{1\to2}$ and $r_{2\to1}, r_{1\to2}$ as a measure of transient and stationary tendency of switching between the two stages, respectively.

Some other related models include the multi-layer model of Bean, O'Reilly and Taylor in Bean and O'Reilly (2008), and the model of da Silva Soares and Latouche in da Silva Soares and