Approximating fluid queues

Angus Lewis

August 5, 2022

Thesis submitted for the degree of
Doctor of Philosophy
in
Applied Mathematics
at The University of Adelaide
Faculty of Engineering, Computer and Mathematical Sciences
School of Mathematical Sciences

THE UNIVERSITY
of ADELAIDE

Contents

List of Tables

List of Figures

Signed Statement

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint award of this degree.

I give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

I acknowledge the support I have received for my research through the provision of an Australian Government Research Training Program Scholarship.

Signed: ..... Date: .....

Declaration

This thesis contains work taken from the publication Bean et al. (2022) which approximates the operator analytic expressions of Bean & O'Reilly (2014) via the DG method. I am a co-author on this paper. The main conceptualisation and initial writing of the manuscript was done by Vikram Sunkara, Giang Nguyen and Nigel Bean. I extended the manuscript in the following ways.

- Significant additions to Section 2 to introduce the partition induced by the approximation scheme to the theoretical operators of Bean & O'Reilly (2014) to help elucidate the connection between the theoretical operators and the approximation counterparts.
- The addition of Section 4.4 which introduces exact boundary dynamics. Prior to this boundary conditions were approximated by a small interval in the DG scheme (in this thesis, I generalised the boundary conditions further).
- The addition and proof of Corollary 4.1.
- The formal definition and derivation of the approximation to the operator $\mathbb{R}$ in Section 5.1.
- Introduction of the concept of operator kernels to the paper (used throughout Section 5.5).
- The final version of the code (Vikram did the initial version).
- Extension of the numerical experiments to consider higher-order bases, smaller cell-widths, elucidating issues around truncation errors, and computing confidence intervals.
- Addition of Appendix A on the properties of the DG approximation to the generator of the fluid queue.
- Additional details in Appendix B.

- I also made significant contributions to the writing (and rewriting) the manuscript, submission, and responding to reviewers' comments.

To clarify which parts of this thesis are related to the paper Bean et al. (2022), the following sections were taken, almost verbatim except for minor editing, notational changes so that the notation matches that of the thesis, and the addition more general boundary conditions,

- Section 2.4,
- Chapter 3, except for Sections 3.6 and 3.7,
- Appendix A,
- Appendix B.

Acknowledgements

First and foremost, to my principle supervisor, Nigel Bean. Thanks for all the support, the interest you have taken in this work, and the faith you have shown in me allowing me to essentially do whatever I want for this thesis. The effort you have put in to supervising me has been much appreciated.

To Peter Taylor, my co-supervisor, thanks for joining me on this quest. Thanks for the insightful discussions on maths, cricket and all the other important happenings of the world during our meetings. You've provided an invaluable perspective on my work and on mathematics as a science.

To Giang Nguyen. You were the one who initially inspired me to take up mathematics research which led me to where I am now. You gave me so much of your time and went well above and beyond to support me. Thank you.

To my friend Caitlin, thanks for all the support, coffee, and company, over the last five years. It has been so enjoyable. I will miss it.

To my friends Dennis, Phil and Andrew. I can't understate the value of this group (and Caitlin too). I've been lucky to be able to share ideas, get advice, and receive mentorship and support from all of you.

To my wife, Alice. Thank you for your support over the last five years. I've learnt so much about life from you.

Dedication

To my wife, Alice, for *everything*.

Abstract

A fluid-fluid queue is a stochastic fluid queue, where the driving process is a fluid queue itself. Fluid queues provide a model for a single continuous performance measure of a system in the presence of a random environment. Fluid queues have found a wide variety of applications including risk processes, telecommunications, and environmental modelling, among others. Given the success of fluid queues it is plausible that the extension to fluid-fluid queues, which enable us to track two continuous performance measures of a system, will also find success. Bean & O'Reilly (2014) provide an analysis of fluid-fluid queues and derive operator-analytic expressions for the first-return operator, and stationary distribution of a fluid-fluid queue.

This thesis provides approximations to fluid queues so that we can approximate the operators in Bean & O'Reilly (2014). It investigates three main approximation schemes; the DG scheme (Chapter 3) which is a popular finite-element scheme, the uniformisation scheme of Bean & O'Reilly (2013*a*) which approximates a fluid queue by a continuous-time Markov chain (specifically, a quasi-birth-and-death-process (QBD)), and the QBD-RAP scheme which is a generalisation of a QBD to allow matrix exponential inter event times. The QBD-RAP scheme is novel; we describe the construction of the scheme in Chapter 4, and provide an analysis to show that it is convergent in Chapters 5 and 6. We demonstrate the effectiveness of the approximation schemes in Chapter 7, focussing on problems with discontinuous solutions. In general, we find that the DG scheme performs remarkably well for smooth problems, but can produce oscillations and negative probability estimates in the presence of discontinuities, the QBD-RAP approximation performs well in the presence of discontinuities, but does not perform as well as the DG scheme for smooth problems, the uniformisation scheme produces reliable approximations in the presence of discontinuities, but converges slowest for all problems.

Chapter 1

Introduction

A fluid queue is a two-dimensional stochastic process $\{\boldsymbol{X}(t)\} = \{(X(t), \varphi(t))\}_{t\geq 0}$. The phase process, also known as the driving process, $\{\varphi(t)\}_{t\geq 0}$, is a continuous-time Markov chain (CTMC) and determines the rate at which $\{X(t)\}$ moves. The level process, $\{X(t)\}_{t\geq 0}$, is real-valued, continuous, piecewise linear and deterministic given $\{\varphi(t)\}$.

Typically, the phase, $\varphi(t)$, represents the underlying operating state or environment of a system at time t. The level variable, $X(t)$, describes some continuous variable of the system. Stochastic fluid queues have found a variety of applications such as telecommunications (see Anick et al. (1982), a canonical application in this area), power systems (Bean et al. 2010), risk processes (Badescu et al. 2005), environmental modelling (Wurm 2020), dam storage (Loynes 1962), and more. For example, Anick et al. (1982) model a *data handling switch* in which there are N information sources which asynchronously switch from ON to OFF, or OFF to ON. The phase process $\varphi(t)$ models the number of data sources which are ON at time t, and the rate at which data is received from a single source when the buffer is on is assumed (without loss of generality) to be 1. The switch stores data in a buffer and processes the data at rate c, thus the net rate of change of the buffer at time t when there are r information sources which are ON is $r - c$. The level process, $X(t)$, models the amount of data in the buffer at time t and the rate at which $X(t)$ changes when $\varphi(t) = r$ is $c_r = (r - c) \times 1(X(t) > 0)$. The model described above could be used to aid in the analysis and design of the data handling switch, for example, by determining the minimum rate r for which the system is *stable* in steady state, which practically means that the processing power of the switch is large enough that the amount of data in the buffer does not grow without bound.

The success of fluid queues largely lies in their mix of flexibility, any (finite) number of phases can be used, and their parsimony and analytic tractability which enables effective transient and stationary analysis Ahn et al. (2005), Ahn & Ramaswami (2003, 2004), Bean et al. (2005*a*,*b*, 2009*a*,*b*), da Silva Soares (2005), Latouche & Nguyen (2019), Bean et al. (2018). Fluid queues are relatively well studied. Largely, the analysis of fluid queues falls into two categories, matrix-analytic methods e.g. Ahn et al. (2005), Ahn &

Ramaswami (2003, 2004), Bean et al. (2005*a*,*b*, 2009*a*,*b*), da Silva Soares (2005), Latouche & Nguyen (2019), and differential equation-based methods Anick et al. (1982), Karandikar & Kulkarni (1995), Bean et al. (2022).

More recently, Bean & O'Reilly (2014) extend fluid queues to so-called *stochastic fluid-fluid queues*. In a fluid-fluid queue there is a second level process, $\{Y(t)\}_{t\geq 0}$ which is itself driven by a fluid queue, $\{(X(t), \varphi(t))\}_{t\geq 0}$, so $(X(t), \varphi(t))$ determines the rate at which $\{Y(t)\}$ moves. The addition of the second level process $\{Y(t)\}$ greatly increases the modelling potential. For example, Bean et al. (2010), Bean & O'Reilly (2014) describe a model for a hydroelectric power generator which can be operated 'on design' and 'off design'. When the system operates 'off design' the system is less efficient and there is more wear, but may be used to optimise overall system performance. Further, maintenance of the generator takes place periodically to improve its performance and lifespan. The system can plausibly be modelled as a fluid queue, $\{(\varphi(t), X(t))\}$, where the state space of $\{\varphi(t)\}$ consists of states representing the systems states 'start', 'stop', 'on design', 'off design', 'maintenance', and 'idle'. The level variable $\{X(t)\}$ take values in $[0, 1]$ and represents the deterioration level of the system with 0 being perfect working order and 1 meaning the system needs replacement. The second level process $\{Y(t)\}$ could be used to model the revenue of the power generator, which, logically, should depend on the state of the system, $\{\varphi(t)\}$. Moreover, the revenue could also depend on the deterioration level $\{X(t)\}$, as, for example, when the system is brand new it may operate more efficiently, and when it is near broken it may be less efficient.

The addition of the second level process $\{Y(t)\}$ greatly increases the flexibility and modelling potential. However, the added flexibility comes at a cost to the in the analysis of the model. For both fluid-fluid queue and classical fluid queues, performance measures such as the stationary distribution, and the Laplace-transforms of hitting times and return times, can be derived in terms of the infinitesimal generator of the modulating process. For fluid queues the modulating process is a CTMC for which the infinitesimal generator is a matrix. Thus, expressions for the performance measures of fluid queues are matrix expressions which are readily computable. For fluid-fluid queue the modulating processes is a fluid queue for which the infinitesimal generator is a matrix of differential operators. Thus, expressions for the performance measures of fluid-fluid queues are matrices of differential operators which, in all but the simplest of cases, are not readily computable.

The analysis of Bean & O'Reilly (2014) is in principle similar to the matrix-analytic methods of Bean et al. (2005*b*), and derives results about the second level process $\{Y(t)\}_{t\geq 0}$ in terms of the generator (a differential operator) of the fluid queue, $\{(X(t), \varphi(t))\}_{t\geq 0}$. For practical computation of the results of Bean & O'Reilly (2014), a discretisation of the infinitesimal generator of the fluid queue can be used. To this end, to date, two possible discretisations have been suggested. Taking a differential equations-based approach, Bean et al. (2022) use the discontinuous Galerkin (DG) method to discretise the operator, while Bean & O'Reilly (2013*a*) take a stochastic modelling and matrix-analytic methods

approach to approximate the fluid-queue by a quasi-birth-and-death (QBD) process. The QBD approximation of Bean & O'Reilly (2013*a*) is derived via a uniformisation argument, so we refer to it as the uniformisation scheme throughout this thesis. Both approaches are insightful and offer different tools and perspectives with which to analyse the resulting approximations. It turns out that the uniformisation scheme of Bean & O'Reilly (2013*a*) is a subclass of the former; the uniformisation scheme can be viewed as the simplest DG scheme where cells are all have equal width, the operator is projected onto a basis of piecewise constant functions, and an upwind flux is used to approximate the flow of mass between cells. This is also equivalent to a finite-volume scheme.

One drawback of the uniformisation scheme is that convergence can be relatively slow compared to higher-order DG schemes. However, in the context of approximating fluid queues, one advantage of the uniformisation scheme and, equivalently, a DG scheme with constant basis functions, is that the uniformisation scheme guarantees probabilities computed from the approximation are positive (Koltai 2011, Section 3.3). One justification for the positivity preserving property of the uniformisation scheme is from its interpretation as a stochastic process. For higher order DG schemes there is no such interpretation and positivity is not guaranteed (Koltai 2011, Section 3.3). Moreover, higher-order DG approximation schemes may produce negative and oscillatory solutions, particularly when discontinuities or steep gradients are present. This is problematic for probabilistic problems as we know that probabilities must be positive. Methods to navigate the problem of negative and oscillatory solutions have been developed, such as filtering and slope limiting (see Cockburn (1999), or Hesthaven & Warburton (2007), Section 5.6 and references therein). Slope limiting alters the discretised operator in regions where oscillations are detected and reduces the order of the approximation to linear in these regions but ensures the solution will be non-negative. Filtering is a post-hoc method which looks to recover an accurate solution, given an oscillatory approximation.

Depending on the context, filtering of the approximate solution to remove oscillations may not necessarily guarantee a strictly non-negative approximation, or may result in severe smearing of the solution at discontinuities or regions with steep gradients (Hesthaven & Warburton 2007, Section 5.6.1). Moreover, filtering requires us to make a trade-off between filtering enough of the oscillations away while maintaining sufficient accuracy – a choice which may not be obvious *a priori*. Slope limiting does guarantee positivity but reduces the approximation to linear where oscillations in the approximate solution are detected (Hesthaven & Warburton 2007, Section 5.6.1). Furthermore, limiting and filtering do not distinguish between natural oscillations, which are a fundamental feature of the solution and spurious oscillations which are caused by the approximation scheme, and they may remove both from the approximation. This can lead to an unnecessary loss of accuracy in the approximation (see (Hesthaven & Warburton 2007), Example 5.8). Other drawbacks of these methods is that they result in non-linear operators, so we cannot use matrix algebra for computations, and they incur an additional computational cost on top

of the DG scheme.

In approximating the first return operator of a fluid-fluid queue we first approximate an operator-Riccati equation by a matrix-Riccati equation by substituting in matrix approximations to the infinitesimal generator of the fluid queue which are constructed via the DG scheme (Bean et al. 2022). We then solve the matrix-Riccati equation via an iterative procedure (Bean et al. 2005*a*, 2022). Since the operators constructed with limiters are non-linear (not matrices) then we cannot apply limiters in this case. Alternatively, we could apply a post-processing based on slope limiters, or filtering. Both of these post-processing methods incur a computational cost and need to be applied on a case-by-case basis depending on the initial condition.

Furthermore, for the DG method there is a limited theoretical backing as to why we may approximate the operator-Riccati equation with the matrix-Riccati equation other than it is a sensible thing to do, and it seems to work in practice. In the case of the QBD approximation of Bean & O'Reilly (2013*a*) the approximate matrix-Riccati equation is more justified as in it can be derived by considering the first-return probabilities for a fluid queue driven by the approximating QBD where the rates of the fluid are determined by a piecewise constant approximation to $r_i(x)$ (Bean et al. 2005*b*).

Hence, for the approximation of fluid queues with application to fluid-fluid queues, on the one hand we can use the DG scheme which can produce high-order approximations when the problem is smooth, but can display spurious oscillations and result in negative probability estimates in the presence of discontinuities. While, on the other hand, we can use the uniformisation scheme which constructs in a linear operator to approximate the generator and guarantees non-negative probability estimates, but convergence is relatively slower.

Motivated by this, this thesis derives a new approximation to a fluid queue. The construction of the approximation is inspired by the observation that the Markov chain approximation of Bean & O'Reilly (2013*a*) effectively uses Erlang distributions to approximate the sojourn time of the fluid level in a given interval on the event that the phase of the fluid is constant. The sojourn time in a given interval on the event that the phase of the fluid is constant is a deterministic event, and it is known that the Erlang distribution is the least-variable Phase-type distribution, so in this sense, the best approximation to this deterministic sojourn time. Thus, it appears that the approximation of Bean & O'Reilly (2013*a*) is the best-possible Markov chain approximation. Recently, there has been much work on a class of concentrated matrix exponential distributions (Horváth, Horváth, Almousa & Telek 2020) which are postulated to be the least-variable matrix exponential distributions of a given order. Matrix exponential distributions generalise Phase-type distributions; they have the same functional form, without the restriction that the distribution has an interpretation in terms of the absorption time of a continuous-time Markov chain. A class of stochastic processes, known as quasi-birth-and-death-processes with rational-arrival-process components (QBD-RAPs) (Bean & Nielsen 2010), extends

QBDs, which have Phase-type inter-event times, to allow matrix-exponentially distributed inter-event times. Thus, by using matrix exponential distributions, we attempt to construct a QBD-RAP which better captures the dynamics of the fluid queue than the QBD approximation in Bean & O'Reilly (2013*a*), while retaining a stochastic interpretation.

As the QBD-RAP has a stochastic interpretation the approximations it produces are guaranteed to have non-negative density functions. Further, for the QBD-RAP (and uniformisation) scheme non-negativity is guaranteed for any initial condition without any further computation or post-processing. For the approximation of fluid-fluid queues, the matrix-Riccati equation we solve to approximate the first-return distribution of a fluid-fluid queue can be derived by considering the first-return probabilities for a RAP-modulated fluid queue driven by the QBD-RAP (Peralta Gutierrez 2019, Bean et al. 2021). The stochastic interpretation can also be leveraged to aid in the analysis of the approximation. Another attractive property is that the approximations to operators obtained via the QBD-RAP (and uniformisation) scheme are linear which is necessary for the application to approximating the first-return operator of fluid-fluid queues and may also provide other computational and analytical benefits.

The structure of this thesis is as follows. The next chapter, Chapter 2, is dedicated to mathematical preliminaries and introduces the main mathematical objects and tools which we will need. In particular, we go into detail describing operators arising in the analysis of fluid-fluid queues and introducing them in such a way that makes clear exactly how the approximations we develop correspond to the theoretical operators. Chapter 2 also introduces the discontinuous Galerkin (DG) method for approximating solutions to differential equations. Due to issues with oscillations in the approximations produced by the DG scheme, at the end of Chapter 2 we also describe *slope limiting* – a method which can be used to prevent oscillations and negativity. Chapter 2 also introduces the existing literature and gives context to this thesis.

Chapter 3 demonstrates a way that we can use the discontinuous Galerkin method to approximate certain operators and distributions of fluid and fluid-fluid queues. The DG method provides an effective tool for approximating performance measures of fluid-fluid queues. However, the DG method can exhibit oscillations in the presence of discontinuities and these oscillations can cause approximations to certain probabilities to be negative. This motivates the next chapter.

In the next chapter, Chapter 4, we develop a new QBD-RAP approximation scheme which, due to its interpretation as a stochastic process, ensures all estimates of probabilities will be non-negative. The chapter takes a stochastic modelling approach to developing the approximation scheme.

Once we have established the new approximation scheme, we then prove its convergence in Chapters 5 and 6. Ultimately, Chapter 5 proves that, on the event that the QBD-RAP has not-yet seen an *orbit restart epoch* (which is much like a change of level, but not always), certain Laplace transforms with respect to time of the QBD-RAP scheme

converge to corresponding Laplace transforms with respect to time of the fluid queue. This is a local convergence result as it relates only to convergence to the fluid queue in an interval.

Chapter 6 then looks to extend the local convergence results of Chapter 5 to a global convergence result on the whole domain of approximation. Chapter 6 uses more traditional Markov process arguments which rely on properties such as the strong Markov property, time-homogeneity and the law of total probability. In Chapter 6 we first consider the discrete-time process which is the process embedded in the QBD-RAP at times when the QBD-RAP sees an orbit restart epoch. We prove that the embedded process converges in distribution to a corresponding embedded process of the fluid queue. The rest of Chapter 6 proves a global convergence result. The main result is Theorem 6.12 which states that the QBD-RAP approximation scheme converges weakly (with respect to the spatial and temporal variable) to the distribution of the fluid queue.

Once we have established methods for approximating fluid queues, Chapter 7 then numerically explores some properties of the approximations. Given that the QBD-RAP scheme was developed so that it guarantees positivity of the approximation, we largely focus on problems with discontinuities.

Finally, Chapter 8 makes concluding remarks.

To help the reader understand the notation used in the DG scheme, we provide a small toy model example in Appendix A. Appendix B provides a proof that the DG scheme conserves probability. The rest of the appendices contain technical results relating to proving the convergence of the QBD-RAP approximation scheme. Appendix C proves that the *closing operators* introduced as part of the QBD-RAP scheme have the properties we claim in proving convergence in Chapter 5. Appendix C extends some results from Chapter 5 to a setting which requires slightly less computation. Lastly, Appendix D provides some algebraic results which help us to manipulate certain Laplace transforms from Chapter 5.

Chapter 2

The existing literature & mathematical preliminaries

2.1 Stochastic processes

Following Ross (1996), a stochastic process is a sequence, $\{X(t)\}_{t\in\mathcal{T}}$, of random variables (or random vectors) indexed by some index set $\mathcal{T}$. In this thesis, in the case that the index set is $\mathcal{T} = \{0, 1, 2, ...\}$ we say that the process $\{X(n)\}_{n\in\mathcal{T}}$ is a discrete-time process, and we will typically use the dummy variable n for the 'time' index. When the index set is $\mathcal{T} = [0, \infty)$ we say $\{X(t)\}_{t\in\mathcal{T}}$ is a continuous-time process, and we will typically use the dummy variable t for the 'time' index. We may omit the index set and write $\{X(t)\}$ in place of $\{X(t)\}_{t\in\mathcal{T}}$ when it is not explicitly needed, or we may write $\{X(t)\}_{t\geq t_0}$ to mean $\{X(t)\}_{t\in[t_0,\infty)}$. The *state space* of $X(t)$ is the set of possible values that $X(t)$ can take at any time $t \in \mathcal{T}$.

The *initial distribution*, μ, of a stochastic process is the distribution of $X(0)$. More generally, for a random variable, Z, we write $Z \sim \nu$ when Z has the distribution given by the probability measure ν. For the probability that $X(t)$ lies in some measurable set $E \subset \mathcal{S}$ given $X(0) \sim \mu$, we write $\mathbb{P}(X(t) \in E \mid X(0) \sim \mu)$. When μ assigns probability 1 to a single point, $x \in \mathcal{S}$, say, we write $\mathbb{P}(X(t) \in E \mid X(0) = x)$.

A stochastic process is *stationary* if, for any n, $t_0 < t_1 < ... < t_n$ and any t, $X(t_n), X(t_{n-1}), ..., X(t_0)$ and $X(t_n + t), X(t_{n-1} + t), ..., X(t_0 + t)$ have the same distribution. That is,

$$\begin{aligned}&\mathbb{P}(X(t_n) \in E_n, X(t_{n-1}) \in E_{n-1}, ..., X(t_0) \in E_0)\\ &= \mathbb{P}(X(t_n + t) \in E_n, X(t_{n-1} + t) \in E_{n-1}, ..., X(t_0 + t) \in E_0),\end{aligned} \tag{2.1}$$

for any $t, t_0, ..., t_n \in \mathcal{T}$ with $t_0 + t, ..., t_n + t \in \mathcal{T}$ and any $E_0, ..., E_n \subseteq \mathcal{S}$. A random variable, τ, is a *stopping time* for the stochastic process $\{X(t)\}_{t\in\mathcal{T}}$ if $\tau \in \sigma(X(s), s \leq t)$, where $\sigma(X(s), s \leq t)$ denotes the σ-algebra generated by $\{X(s), s \leq t\}$.

As in (Bladt & Nielsen 2017, Chapter 1.2), we call $\{X(n)\}_{n\in\{0,1,2,...\}}$ a *discrete-time Markov chain* if $\mathcal{S}$ is countable and

$$\begin{aligned}&\mathbb{P}(X(n+1)=j \mid X(n)=i, X(n-1)=i_{n-1}, ..., X(0)=i_0)\\&=\mathbb{P}(X(n+1)=j \mid X(n)=i)\end{aligned} \tag{2.2}$$

for all $n \in \{0,1,2,...\}$ and $i, i_0, ..., i_{n-1}, j \in \mathcal{S}$, and referred to (2.2) as the *Markov property*. The process $\{X(n)\}_{n\in\{0,1,2,...\}}$ is said to be time-homogeneous if $\mathbb{P}(X(n+1)=j \mid X(n)=i)$ does not depend on n. The probabilities $\mathbb{P}(X(n+1)=j \mid X(n)=i) =: p_{ij}$ are the *transition probabilities*, and $\boldsymbol{P}=[p_{ij}]_{i,j\in\mathcal{S}}$ is the *transition matrix*, of the Markov chain. The *strong Markov property* states that for each stopping time τ of the Markov chain $\{X(n)\}_{n\in\{0,1,2,...\}}$, on the event that $\{\tau<\infty\}$, then

$$\mathbb{P}(X(\tau+1)=j \mid \sigma(X(0),X(1),...,X(\tau))) = \mathbb{P}(X(\tau+1)=j \mid X(\tau)). \tag{2.3}$$

Also, as in (Bladt & Nielsen 2017, Section 1.3), we call $\{X(t)\}_{t\geq 0}$ a *continuous-time Markov chain* (CTMC) if $\mathcal{S}$ is countable and for all $t_{n+1}>t_n>t_{n-1}>...>t_0>0$ and $i, i_0, ..., i_{n-1}, j \in \mathcal{S}$,

$$\mathbb{P}(X(t_{n+1})=j \mid X(n)=i, X(n-1)=i_{n-1}, ..., X(0)=i_0) = \mathbb{P}(X(t_{n+1})=j \mid X(t_n)=i), \tag{2.4}$$

and refer to (2.4) as the *Markov property*. The process $\{X(t)\}_{t\geq 0}$ is said to be time-homogeneous if $\mathbb{P}(X(t+s)=j \mid X(s)=i)$ does not depend on s. For a time-homogeneous CTMC, the *transition function* is the matrix function

$$\boldsymbol{P}(t) = [\mathbb{P}(X(t)=j \mid X(t)=i)]_{i,j\in\mathcal{S}}.$$

The Chapman-Kolmogorov equations state that

$$\boldsymbol{P}(t+s) = \boldsymbol{P}(t)\boldsymbol{P}(s).$$

The *infinitesimal generator* matrix of a CTMC is

$$\boldsymbol{T} = [T_{ij}]_{i,j\in\mathcal{S}} = \frac{\mathrm{d}}{\mathrm{d}t}P(t),$$

and the elements T_{ij} are the *transition rates*. A CTMC is said to be *honest* if $\boldsymbol{P}(t)\mathbf{1}=\mathbf{1}$ for all $t\geq 0$ where $\mathbf{1}$ is a column vector of ones.

The *strong Markov property* states that for each stopping time τ of the Markov chain $\{X(t)\}_{t\geq 0}$, on the event that $\{\tau<\infty\}$, then

$$\mathbb{P}(X(\tau+t)=j \mid \sigma(X(s), s\leq\tau)) = \mathbb{P}(X(\tau+t)=j \mid X(\tau)). \tag{2.5}$$

Most generally, we call $\{X(t)\}_{t\geq 0}$ a *Markov process* if for all $t_n > t_{n-1} > ... > t_0 > 0$, $t_k \in \mathcal{T}$, $k = 0, ..., n+1$, and $E_0, ..., E_n \subseteq \mathcal{S}$

$$\mathbb{P}(X(t_n) \in E_n, X(t_{n-1}) \in E_{n-1}, ..., X(t_0) \in E_0) = \mathbb{P}(X(t_n) \in E_n \mid X(t_{n-1})), \tag{2.6}$$

and is time homogeneous if $\mathbb{P}(X(t+s) \in E_n \mid X(s))$ does not depend on s.

The Chapman-Kolmogorov equations state that

$$\begin{aligned}&\mathbb{P}(X(t+s) \in E \mid X(0) \sim \mu)\\ &= \int_{x\in\mathcal{S}} \mathbb{P}(X(t+s) \in E \mid X(t) = x)\mathbb{P}(X(t) \in \mathrm{d}x \mid X(0) \sim \mu).\end{aligned} \tag{2.7}$$

2.2 Some semigroup theory

The evolution of Markov processes can be described by an *operator semigroup*. Semigroups also arise in other contexts related to the analysis of fluid queues and fluid-fluid queues (see, for example, Sections 2.4 and 2.3.1). Semigroups (and therefore Markov processes) can be characterised by their *infinitesimal generator*, or generator for short. One of the aims of this thesis is to approximate the generator of a fluid queue. Here we very briefly introduce semigroups and their *infinitesimal generators*. We refer to the reader to Ethier (1986) for more details (see also Kallenberg (2021), or the more approachable course notes of Shalizi (2010)).

Operators

Let L be a *Banach space*. A Banach space is a *complete, normed* vector space. Complete means that every Cauchy sequence of points in L has a limit which also lies in L. Normed means that a norm is defined on L. Intuitively, a norm is a function which tells us the size of a vector in L, hence we can use a norm to compute a distance between two points $f, g \in L$ by computing the norm of $f - g$.

An *operator* is a mapping which takes elements of a vector space and sends them to elements of a vector space (not necessarily the same space). That is, let V and W be normed vector spaces, then $\mathbb{B} : \mathcal{D}(\mathbb{B}) \to W$ is an operator. The set $\mathcal{D}(\mathbb{B}) \subseteq V$ is the domain of the operator, and is not necessarily all of V. When specification of the domain is unnecessary we sometimes write $\mathbb{B} : V \to W$.

Let $c \in \mathbb{R}$ and $v \in V$, $w \in W$. Then $\mathbb{B}$ is a *linear* operator if $\mathbb{B}(cv) = c(\mathbb{B}v)$ and $\mathbb{B}(v+w) = \mathbb{B}v + \mathbb{B}w$. We define the norm of an operator as $||\mathbb{B}|| = \sup_{v\in V} \dfrac{||\mathbb{B}v||_W}{||v||_V}$, where $||\cdot||_V$ and $||\cdot||_W$ are the norms defined on V and W, respectively. An operator is said to be bounded if $||\mathbb{B}|| < M$ for some $M < \infty$.

Semigroups

A family of linear operators is a collection of indexed linear operators $\{\mathbb{B}(i)\}_{i\in I}$ where I is some index set. For each $i \in I$, $\mathbb{B}(i)$ is a linear operator. A family of operators $\{\mathbb{T}(t)\}_{t\geq 0}$, $\mathbb{T}(t) : L \to L$, is said to have the *semigroup* property if $\mathbb{T}(t+s) = \mathbb{T}(t)\mathbb{T}(s)$. If $\mathbb{T}(0) = I$, the identity operator, then the family $\{\mathbb{T}(t)_{t\geq 0}\}$ is known as an operator semigroup. If, for all $f \in L$, $\mathbb{T}(t)f \to f$ as $t \to 0^+$, then $\{\mathbb{T}(t)\}_{t\geq 0}$ is said to be strongly continuous. If $||\mathbb{T}(t)|| \leq 1$ for all $t \geq 0$ then $\{\mathbb{T}(t)\}_{t\geq 0}$ is known as a contraction semigroup.

For a bounded linear operator $\mathbb{B} : \mathcal{D}(\mathbb{B}) \to \mathcal{D}(\mathbb{B})$, define the operator exponential as $e^{\mathbb{B}t} := \sum_{k=0}^{\infty} \frac{1}{k!} t^k \mathbb{B}^k$ where, for $f \in \mathcal{D}(\mathbb{B})$, $\mathbb{B}^k f$ is defined recursively by $\mathbb{B}^k f = \mathbb{B}^{k-1}(\mathbb{B}f)$. It can be shown that $e^{\mathbb{B}(t+s)} = e^{\mathbb{B}t}e^{\mathbb{B}s}$, and from the definition $e^{\mathbb{B}0} = I$, which implies that $\{e^{\mathbb{B}t} : t \geq 0\}$ is a semigroup. As another example, let $\mathbb{B} = \dfrac{d}{dx} : C^\omega(\mathbb{R}) \to C^\omega(\mathbb{R})$, where $C^\omega(\mathbb{R})$ is the class of analytic functions on the whole of $\mathbb{R}$. Then $\mathbb{T}(t)f(x) = \sum_{k=0}^{\infty} \frac{1}{k!} t^k \frac{\mathrm{d}^k}{\mathrm{d}x^k} f(x) = f(x-t)$, and the sum converges since $f \in C^\omega(\mathbb{R})$; the series representation is the Taylor series of f about the point x, and we see that the operator $\mathbb{T}(t)f(x) = f(x-t)$ is the shift operator. The semigroup property follows since

$$\mathbb{T}(t)\mathbb{T}(s)f(x) = \mathbb{T}(t)f(x-s) = f(x-s-t) = \mathbb{T}(t+s)f(x).$$

Infinitesimal generators

The *infinitesimal generator* (or just *generator* for short) of a semigroup $\{\mathbb{T}(t)\}$ is the linear operator defined by

$$\mathbb{B}f = \lim_{t\to 0^+} \frac{1}{t}(\mathbb{T}(t)f - f),$$

and the domain, $\mathcal{D}(\mathbb{B})$, is the subspace of L for which this limit exists. Note that an essential part of the definition of the generator is its domain. For this reason it is common to denote the generator as the pair $(\mathbb{B}, \mathcal{D}(\mathbb{B}))$.

To determine the generator $\mathbb{B}$ we can proceed by either differentiation of the semigroup,

$$\mathbb{B}f = \lim_{t\to 0^+} \frac{1}{t}(\mathbb{T}(t)f - f),$$

or integration;

$$\int_{t=0}^{\infty} e^{-st}\mathbb{T}(t)\,\mathrm{d}t = (sI - \mathbb{B})^{-1} =: R_s.$$

The operator R_s is known as the resolvent.

One of the fundamental results of semigroup theory is the Hille-Yosida Theorem which essentially states that a semigroup is entirely characterised by its generator. This is not immediately obvious as the generator is defined on a subset of L only, whereas $\{\mathbb{T}(t)\}$ is defined on all of L.

Markov processes and semigroups

Semigroups arise naturally in Markov processes. Let $\{X(t) : t \geq 0\}$ be a Markov process with state space $\mathcal{S}$. The Markov property states that, given $X(t)$, then for $s, t \geq 0$, $X(t+s)$ is independent of $X(u)$, $0 \leq u < t$. That is,

$$P(X(t+s) \in \mathcal{E} \mid X(t), X(u), 0 \leq u < t) = P(X(t+s) \in \mathcal{E} \mid X(t)).$$

If $P(X(t+s) \in \mathcal{E} \mid X(t)) = P(X(s) \in \mathcal{E} \mid X(0))$, are invariant under translation by t, then $X(t)$ is said to be time-homogeneous.

Consider the expectation

$$\mathbb{E}[f(X(t)) \mid X(0) = x] = \int_{y \in \mathcal{S}} f(y) P(X(t) \in \mathrm{d}y \mid X(0) = x),$$

where $f < F$ is some bounded function. We can think of expectation as an operator which is acting on the function f. The expectation depends on t and x as these define the distribution with which we are taking the expectation, hence, let us write

$$\mathbb{T}(t) f(x) = \mathbb{E}[f(X(t)) \mid X(0) = x].$$

The result of the operator acting on f, $\mathbb{T}(t)f(x)$, is another function.

By the Markov property, for any $s \in [0, t]$,

$$\begin{aligned}
\mathbb{T}(t) f(x) &= \int_{y \in \mathcal{S}} f(y) P(X(t) \in \mathrm{d}y \mid X(0) = x) \\
&= \int_{y \in \mathcal{S}} \int_{z \in \mathcal{S}} f(y) P(X(t) \in \mathrm{d}y \mid X(s) = z) P(X(s) \in \mathrm{d}z \mid X(0) = x) \\
&= \int_{z \in \mathcal{S}} \mathbb{E}[f(X(t)) \mid X(s) = z] P(X(s) \in \mathrm{d}z \mid X(0) = x) \\
&= \int_{z \in \mathcal{S}} \mathbb{E}[f(X(t-s)) \mid X(0) = z] P(X(s) \in \mathrm{d}z \mid X(0) = x) \\
&= \int_{z \in \mathcal{S}} (\mathbb{T}(t-s) f(z)) P(X(s) \in \mathrm{d}z \mid X(0) = x) \\
&= \mathbb{T}(s)(\mathbb{T}(t-s) f)(x),
\end{aligned}$$

where the third equality holds by time-homogeneity. Hence, $\{\mathbb{T}(t)(\cdot)\}_{t \geq 0}$ is an operator semigroup. Moreover, since f is bounded, then $\mathbb{T}(t)f(x) = \mathbb{E}[f(X(t)) \mid X(0) = x] \leq F$, which means that $\{\mathbb{T}(t)\}$ is a contraction semigroup.

2.3 Fluid queues

An unbounded fluid queue is a two-dimensional stochastic process which we denote by $\{\ddot{\boldsymbol{X}}(t)\} = \{(\ddot{X}(t), \varphi(t))\}_{t \geq 0}$ where $\{\varphi(t)\}_{t \geq 0}$ is known as the phase or driving process,

and $\{\ddot{X}(t)\}_{t\geq 0}$ is known as the level process or buffer. The phase process $\{\varphi(t)\}_{t\geq 0}$, is an irreducible continuous-time Markov chain (CTMC) with finite state space, which we assume to be $\mathcal{S} = \{1, 2, \ldots, N\}$, and infinitesimal generator $\boldsymbol{T} = [T_{ij}]_{i,j\in\mathcal{S}}$. We assume that $\boldsymbol{T}$ is conservative and time-homogeneous. Associated with states $i \in \mathcal{S}$ are real-valued *rates* $c_i \in \mathbb{R}$ which determine the rate at which $\{X(t)\}$ moves.

Partition the state space $\mathcal{S}$ into $\mathcal{S}_+ = \{i \in \mathcal{S} \mid c_i > 0\}$, $\mathcal{S}_- = \{i \in \mathcal{S} \mid c_i < 0\}$, $\mathcal{S}_0 = \{i \in \mathcal{S} \mid c_i = 0\}$, $\mathcal{S}_{-1} = \{i \in \mathcal{S} \mid c_i \leq 0\}$, $\mathcal{S}_{K+1} = \{i \in \mathcal{S} \mid c_i \geq 0\}$*. We assume, without loss of generality, that the generator $\boldsymbol{T}$ is partitioned into sub-matrices

$$\boldsymbol{T} = \begin{bmatrix} \boldsymbol{T}_{++} & \boldsymbol{T}_{+-} & \boldsymbol{T}_{+0} \\ \boldsymbol{T}_{-+} & \boldsymbol{T}_{--} & \boldsymbol{T}_{-0} \\ \boldsymbol{T}_{0+} & \boldsymbol{T}_{0-} & \boldsymbol{T}_{00} \end{bmatrix},$$

where $\boldsymbol{T}_{mn} = [T_{ij}]_{i\in\mathcal{S}_m, j\in\mathcal{S}_n}$, $m, n \in \{+, -, 0\}$.

Let's clarify some notation. We use the notation $\boldsymbol{u} = (u_h)_{h\in\mathcal{H}}$ to denote a row-vector, $\boldsymbol{u}$, defined by its elements, u_h, indexed by $h \in \mathcal{H}$, where $\mathcal{H}$ is some index set. Similarly, $\boldsymbol{u} = (\boldsymbol{u}_h)_{h\in\mathcal{H}}$, is a row-vector defined by a collection of row-vectors $\boldsymbol{u}_h$. The notation $\boldsymbol{u}_m = (u_h)_{h\in\mathcal{H}_m}$ refers to the vector containing the subset of elements corresponding to $\mathcal{H}_m \subseteq \mathcal{H}$. When the index set is empty, the resulting vector $\boldsymbol{u}_m$ is a vector of dimension 0. In cases when there are two indices, we order the elements of the vector according to the first index, then the second; i.e. $\boldsymbol{u} = (u_g^h)_{g\in\mathcal{G},h\in\mathcal{H}} = ((u_g^h)_{g\in\mathcal{G}})_{h\in\mathcal{H}}$. Here we use the convention that for a vector $\boldsymbol{u} = (u)_{h\in\mathcal{H}}$ where the elements u do not depend on the index h and H is some index set, then we repeat u h-times; i.e. $\boldsymbol{u} = (u)_{h\in\mathcal{H}} = \underbrace{(u, \ldots, u)}_{h-\text{times}}$. The notation $\boldsymbol{U} = [u_{gh}]_{g\in\mathcal{G},h\in\mathcal{H}}$ (square brackets) is used to denote a matrix defined by its elements, or sub-blocks, u_{gh}.

Also define the diagonal matrices

$$\boldsymbol{C} = \begin{bmatrix} \boldsymbol{C}_+ & & \\ & \boldsymbol{C}_- & \\ & & \boldsymbol{0} \end{bmatrix}, \qquad \boldsymbol{C}_+ = diag(c_i, i \in \mathcal{S}_+), \qquad \boldsymbol{C}_- = diag(|c_i|, i \in \mathcal{S}_-),$$

and $\widehat{\boldsymbol{C}} = diag(c_i, i \in \mathcal{S})$, where $diag(a_i, i \in \mathcal{I})$ denotes a diagonal matrix with entries a_i down the diagonal.

The level process is given by

$$\ddot{X}(t) = \ddot{X}(0) + \int_{s=0}^{t} c_{\varphi(s)}\, \mathrm{d}s.$$

*The notation $\mathcal{S}_{-1}$ and $\mathcal{S}_{K+1}$ will make sense later when we introduce the partition into cells for the approximation schemes into K cells, plus a lower and upper boundary which we represent with indices -1 and $K+1$ respectively.

Sample paths of $\{\ddot{X}(t)\}$ are continuous and piecewise linear, with $\frac{\mathrm{d}}{\mathrm{d}t}\ddot{X}(t) = c_{\varphi}(t)$, when $\ddot{X}(t)$ is differentiable. Given sample paths of $\{\varphi(t)\}$, then $\{\ddot{X}(t)\}$ is deterministic, and in this sense, $\{\varphi(t)\}$ is the only stochastic element of the fluid queue.

Often, boundary conditions are imposed. We denote a fluid queue bounded below at 0 and unbounded above by $\{\dot{\boldsymbol{X}}(t)\} = \{(\dot{X}(t), \varphi(t))\}_{t\geq 0}$, and a fluid queue bounded below at 0 and above at $b < \infty$ by $\{\boldsymbol{X}(t)\} = \{(X(t), \varphi(t))\}_{t\geq 0}$. Here, we consider a mixture of *regulated* and *reflecting* boundary conditions. Upon hitting a boundary we suppose that, with probability p_{ij}, $i, j \in \mathcal{S}$, the phase process instantaneously transitions from phase i to phase j (note that we might have $i = j$ i.e. no transition) and if $sign(c_i) = sign(c_j)$ or $sign(c_j) = 0$ then the process is absorbed in the boundary, otherwise it is reflected. At a lower boundary, if $j \in \mathcal{S}_0 \cup \mathcal{S}_-$, then $\frac{\mathrm{d}}{\mathrm{d}t}X(t) = 0$, and the phase process continues to evolve according to the sub-generator

$$\begin{bmatrix} \boldsymbol{T}_{--} & \boldsymbol{T}_{-0} \\ \boldsymbol{T}_{0-} & \boldsymbol{T}_{00} \end{bmatrix},$$

until such a time that $\{\varphi(t)\}$ transitions to a phase $k \in \mathcal{S}_+$, at which time $\{X(t)\}$ leaves the boundary. Similarly, at an upper boundary if $j \in \mathcal{S}_0 \cup \mathcal{S}_+$, then $\frac{\mathrm{d}}{\mathrm{d}t}X(t) = 0$ and the phase process continues to evolve according to the sub-generator

$$\begin{bmatrix} \boldsymbol{T}_{++} & \boldsymbol{T}_{+0} \\ \boldsymbol{T}_{0+} & \boldsymbol{T}_{00} \end{bmatrix},$$

until such a time that $\{\varphi(t)\}$ transitions to a phase $k \in \mathcal{S}_-$ at which time $\{X(t)\}$ leaves the boundary. It is without loss of generality that we assume the lower and upper boundaries (when present) are at $x = 0$ and $x = b > 0$, respectively.

In summary, the evolution of the level can be expressed as

$$\frac{\mathrm{d}}{\mathrm{d}t}X(t) = \begin{cases} c_{\varphi(t)}, & \text{if } X(t) > 0, \\ \max\{0, c_{\varphi(t)}\}, & \text{if } X(t) = 0, \\ \min\{0, c_{\varphi(t)}\}, & \text{if } X(t) = b. \end{cases}$$

Let $\boldsymbol{f}(x,t) = (f_i(x,t))_{i\in\mathcal{S}}$ be a row-vector function where $f_i(x,t)$ is the density of $\mathbb{P}(X(t) \leq x, \varphi(t) = i \mid \boldsymbol{X}(0) \sim \boldsymbol{\mu})$, assuming it exists. When a differentiable density exists, the system of partial differential equations which describes the evolution of the densities $\boldsymbol{f}(x,t)$ is

$$\frac{\partial}{\partial t}\boldsymbol{f}(x,t) = \boldsymbol{f}(x,t)\boldsymbol{T} - \frac{\partial}{\partial x}\boldsymbol{f}(x,t)\widehat{\boldsymbol{C}}, \tag{2.8}$$

on the interior $x \in (0, b)$, with appropriate boundary conditions (see Section 3.4). The initial condition is the initial distribution of the fluid queue which, when it has a density, we write as $f_i(x, 0)$. Often a differentiable density function does not exist and therefore the partial differential equation (2.8) is not well-defined. For example, for a fluid queue with no upper boundary, if the initial distribution of the fluid queue is a point mass at any point $x_0 \geq 0$ and in phase $i \in \mathcal{S}_+ \cup \mathcal{S}_0$, then a density function $f_i(x, t)$ will not exist for any finite t. Specifically, a point mass will persist along the ray $x_0 + c_i t$, $t \geq 0$. In such situations, it is the *weak solution* to (2.8) that we seek. A weak solution (ignoring boundary conditions) satisfies

$$-\int_{x=0}^{b}\int_{t=0}^{\infty} \boldsymbol{f}(x,t)\frac{\partial}{\partial t}\boldsymbol{\psi}(x,t)\,\mathrm{d}t\,\mathrm{d}x = \int_{x=0}^{b}\int_{t=0}^{\infty} \boldsymbol{f}(x,t)\boldsymbol{T}\boldsymbol{\psi}(x,t)\,\mathrm{d}t\,\mathrm{d}x + \int_{x=0}^{b}\int_{t=0}^{\infty} \boldsymbol{f}(x,t)\widehat{\boldsymbol{C}}\frac{\partial}{\partial x}\boldsymbol{\psi}(x,t)\,\mathrm{d}t\,\mathrm{d}x, \tag{2.9}$$

for every row-vector of test functions, $\boldsymbol{\psi}(x, t) = (\psi_i(x, t))_{i\in\mathcal{S}}$, which are smooth, have compact support and $\boldsymbol{\psi}(x, 0) = \boldsymbol{\psi}(0, t) = \boldsymbol{\psi}(b, t) = \mathbf{0}$ (Borthwick n.d., Chapter 10).

2.3.1 Transient analysis of fluid queues

The transient analysis of fluid queues has been relatively well studied (see, for example, Ahn & Ramaswami (2004), Bean et al. (2005*b*), da Silva Soares (2005), Bean et al. (2009*b*) which are matrix-analytics-methods-based, and also Rabehasaina & Sericola (2003) which is differential-equations-based and treats a slightly more complex model than the ones considered here, but is none-the-less relevant). Given the interest in discontinuous and point-mass initial conditions, transient analysis of fluid queues rarely relies on solving governing differential equations numerically, although it is quite possible to do so[†]. Instead, expressions for transient distributions of fluid queues are derived in terms of Laplace transforms with respect to time, and/or moments of the transient distributions are derived (Ahn & Ramaswami 2004, Bean et al. 2005*b*, 2009*b*). Moreover, these techniques often lead to quantities which have direct probabilistic interpretations which further aids in their analysis and also applications to other problems (Ahn & Ramaswami 2003, da Silva Soares 2005, Bean et al. 2018) (we also leverage these stochastic interpretations in Chapter 5 to derive expressions for certain the Laplace transforms of the fluid queue). In some contexts, the actual transient distributions are not required, and Laplace transforms or moments are all that are required. In other cases, we can invert the Laplace transforms using known methods (Abate & Whitt (2006), or in the case of the discontinuities we might

[†]We have already alluded to some difficulties with differential-equation-based approaches to problems with discontinuous solutions such as ill-defined PDEs, the need for weak solutions, and oscillatory approximations with possibly infeasible approximate solutions, for example negative probability approximations.

prefer Horváth, Horváth, Almousa & Telek (2020), which uses the same concentrated matrix exponential distribution that we do).

One such analysis of fluid queues derives the Laplace transform of the time taken for the level of an unbounded fluid queue to return to its initial level in a certain phase, given it started in a phase with positive rate (Bean et al. 2005*b*). The principles underlying the derivation of this first return operator are the same as those applied in Bean & O'Reilly (2014) to derive the first return operator for fluid-fluid queues. Moreover, certain matrices appearing in the analysis have a stochastic interpretation which we leverage to write down certain Laplace transforms in Chapter 5 in terms of these matrices. Given its relevance we briefly recount the analysis of Bean et al. (2005*b*) here.

Consider an unbounded fluid queue $\{(\ddot{X}(t), \varphi(t))\}$. Let $\zeta_X(E)$ be the random variable which is the first hitting time of $\{\ddot{X}(t)\}$ on the set E and define the matrix $\boldsymbol{\Psi}_X(s)$ with elements $[\boldsymbol{\Psi}_X(s)]_{ij}$, $i \in \mathcal{S}_+$, $j \in \mathcal{S}_-$, given by

$$[\boldsymbol{\Psi}_X(\lambda)]_{ij} = \mathbb{E}\left[e^{-\zeta_X(z)\lambda} 1(\zeta_X(z) < \infty, \varphi(\zeta_X(z) = j) \mid \ddot{X}(0) = z, \varphi(0) = i\right]. \tag{2.10}$$

$[\boldsymbol{\Psi}_X(\lambda)]_{ij}$ is the Laplace-Stieltjes transform of the time taken for the fluid queue to first return to level z and do so in phase $j \in \mathcal{S}_-$, given it started at level z in phase $i \in \mathcal{S}_+$. Define the in-out fluid level by $\beta_X(t) := \int_0^t \left|c_{\varphi(z)}\right| \mathrm{d}z$, which is the total amount of fluid to flow in to or out of the buffer $\{\ddot{X}(t)\}$ by time t, and also define $\eta_X(y) := \inf\{t > 0 : \beta_X(t) = y\}$ as the first hitting time of the in-out process $\{\beta_X(t)\}$ on level $y \geq 0$. Further, let $\boldsymbol{H}(\lambda, y)$ be the matrix with elements $h_{ij}(\lambda, y)$, $i, j \in \mathcal{S}_+ \cup \mathcal{S}_-$, given by

$$\mathbb{E}\left[e^{-\lambda\eta_X(y)} 1(\eta_X(y) < \infty, \varphi(\eta_X(y)) = j) \mid \ddot{X}(0) = 0, \varphi(0) = i\right], \tag{2.11}$$

which is the Laplace-Stieltjes transform of the time taken for y amount of fluid to flow in or out of the fluid queue and to be in phase j at this time, given the initial level of the fluid queue was 0 and the initial phase was i.

It turns out that, for fixed $\lambda \geq 0$, $\boldsymbol{H}(\lambda, y)$ is a semigroup (with variable y) (Bean et al. 2005*b*). Bean et al. (2005*b*) find the infinitesimal generator, $\boldsymbol{Q}(\lambda)$ of $\boldsymbol{H}(\lambda, y)$ and, since the generator is bounded, we can write $\boldsymbol{H}(\lambda, y) = e^{\boldsymbol{Q}(\lambda)y}$. The derivation of the generator $\boldsymbol{Q}(\lambda) = [Q_{ij}(\lambda)]_{ij \in \mathcal{S}_+ \cup \mathcal{S}_-}$ is based on a direct analysis of sample paths of the fluid queue over an infinitesimal time, u say, and by leveraging the fact that complex sample paths occur with probability $\mathcal{O}(u^2)$. As a result, there are only three types of sample paths to consider. Further, Bean et al. (2005*b*) similarly argue that the generator $\boldsymbol{Q}(\lambda)$ can be partitioned into blocks $\boldsymbol{Q}_{mn}(\lambda) = [Q_{ij}(\lambda)]_{i \in \mathcal{S}_m, j \in \mathcal{S}_n}$, $m \in \{+, -\}$, $n \in \{+, -, 0\}$ where

$$\begin{aligned}
\boldsymbol{Q}_{+0}(\lambda) &= \boldsymbol{C}_+^{-1} \boldsymbol{T}_{+0} \left[\lambda \boldsymbol{I} - \boldsymbol{T}_{00}\right]^{-1}, \\
\boldsymbol{Q}_{-0}(\lambda) &= \boldsymbol{C}_-^{-1} \boldsymbol{T}_{-0} \left[\lambda \boldsymbol{I} - \boldsymbol{T}_{00}\right]^{-1}, \\
\boldsymbol{Q}_{++}(\lambda) &= \boldsymbol{C}_+^{-1} \left(\boldsymbol{T}_{++} - \lambda \boldsymbol{I} + \boldsymbol{T}_{+0} \left[\lambda \boldsymbol{I} - \boldsymbol{T}_{00}\right]^{-1} \boldsymbol{T}_{0+}\right),
\end{aligned}$$

$$\boldsymbol{Q}_{+-}(\lambda) = \boldsymbol{C}_{+}^{-1}\left(\boldsymbol{T}_{+-} + \boldsymbol{T}_{+0}\left[\lambda\boldsymbol{I} - \boldsymbol{T}_{00}\right]^{-1}\boldsymbol{T}_{0-}\right),$$
$$\boldsymbol{Q}_{--}(\lambda) = \boldsymbol{C}_{-}^{-1}\left(\boldsymbol{T}_{--} - \lambda\boldsymbol{I} + \boldsymbol{T}_{-0}\left[\lambda\boldsymbol{I} - \boldsymbol{T}_{00}\right]^{-1}\boldsymbol{T}_{0-}\right),$$
$$\boldsymbol{Q}_{-+}(\lambda) = \boldsymbol{C}_{-}^{-1}\left(\boldsymbol{T}_{-+} + \boldsymbol{T}_{-0}\left[\lambda\boldsymbol{I} - \boldsymbol{T}_{00}\right]^{-1}\boldsymbol{T}_{0+}\right).$$

Bean et al. (2005*b*) also define the functions

$$\boldsymbol{H}^{++}(\lambda, y) = \left[h_{ij}^{++}(\lambda, y)\right]_{i\in\mathcal{S}_+, j\in\mathcal{S}_+\cup\mathcal{S}_{+0}} := e^{\boldsymbol{Q}_{++}(\lambda)y}\left[\boldsymbol{C}_{+}^{-1} \quad \boldsymbol{Q}_{+0}(\lambda)\right], \tag{2.12}$$
$$\boldsymbol{H}^{--}(\lambda, y) = \left[h_{ij}^{--}(\lambda, y)\right]_{i\in\mathcal{S}_-, j\in\mathcal{S}_-\cup\mathcal{S}_{-0}} := e^{\boldsymbol{Q}_{--}(\lambda)y}\left[\boldsymbol{C}_{-}^{-1} \quad \boldsymbol{Q}_{-0}(\lambda)\right], \tag{2.13}$$
$$\boldsymbol{H}^{+-}(\lambda, y) = \left[h_{ij}^{+-}(\lambda, y)\right]_{i\in\mathcal{S}_+, j\in\mathcal{S}_-} := e^{\boldsymbol{Q}_{++}(\lambda)y}\boldsymbol{Q}_{+-}(\lambda), \tag{2.14}$$
$$\boldsymbol{H}^{-+}(\lambda, y) = \left[h_{ij}^{-+}(\lambda, y)\right]_{i\in\mathcal{S}_-, j\in\mathcal{S}_+} := e^{\boldsymbol{Q}_{--}(\lambda)y}\boldsymbol{Q}_{-+}(\lambda), \tag{2.15}$$

for $y, \lambda \geq 0$, which have the following stochastic interpretations. The function $h_{ij}^{++}(\lambda, y)$ ($h_{ij}^{--}(\lambda, y)$) is the Laplace transform (with respect to time) of the time taken for the fluid level to shift by an amount y whilst remaining in phases in $\mathcal{S}_+ \cup \mathcal{S}_{+0}$ ($\mathcal{S}_- \cup \mathcal{S}_{-0}$), given the phase was initially $i \in \mathcal{S}_+$ ($i \in \mathcal{S}_-$). The function $h_{ij}^{+-}(\lambda, y)$ ($h_{ij}^{-+}(\lambda, y)$) is the Laplace transform (with respect to time) of the time taken for the fluid level, $\{Y(t)\}$ to shift by an amount y whilst remaining in phases in $\mathcal{S}_+ \cup \mathcal{S}_{+0}$ ($\mathcal{S}_- \cup \mathcal{S}_{-0}$), after which time the phase instantaneously changes to $j \in \mathcal{S}_-$ ($\mathcal{S}_+$), given the phase was initially $i \in \mathcal{S}_+$ ($\mathcal{S}_-$) (Bean et al. 2005*b*).

Thus, Bean et al. (2005*b*) are able to characterise, by the above matrix expressions, segments of sample paths of the fluid queue where the fluid level is non-decreasing and non-increasing. Bean et al. (2005*b*) then partition the sample paths which contribute to $\boldsymbol{\Psi}_X$ as those which either, (a) have a single transition from $\mathcal{S}_+$ to $\mathcal{S}_-$ (perhaps via $\mathcal{S}_0$), and, (b) those which have more than one transition from $\mathcal{S}_+$ to $\mathcal{S}_-$ (perhaps via $\mathcal{S}_0$). An expression for Laplace-Stieltjes transform of (a) is

$$\int_{y=0}^{\infty} e^{\boldsymbol{Q}_{++}(\lambda)y}\boldsymbol{Q}_{+-}(\lambda)e^{\boldsymbol{Q}_{--}y}\,\mathrm{d}y.$$

For the sample paths (b), there must be at least one point at which the phase process transitions from $\mathcal{S}_-$ to $\mathcal{S}_+$ (perhaps via $\mathcal{S}_0$) before the first return time. Further, one of the transitions from $\mathcal{S}_-$ to $\mathcal{S}_+$ (perhaps via $\mathcal{S}_0$) must occur lower than all others. By considering the lowest level, y, at which a phase transitions from $\mathcal{S}_-$ to $\mathcal{S}_+$ (perhaps via $\mathcal{S}_0$) occurs, Bean et al. (2005*b*) characterise the Laplace transform of the paths (b) by the expression

$$\int_{y=0}^{\infty} e^{\boldsymbol{Q}_{++}(\lambda)y}\boldsymbol{\Psi}_x(\lambda)\boldsymbol{Q}_{-+}(\lambda)\boldsymbol{\Psi}_X(\lambda)e^{\boldsymbol{Q}_{--}(\lambda)y}\,\mathrm{d}y. \tag{2.16}$$

Adding the expressions for (a) and (b) together, Bean et al. (2005*b*) state

$$\boldsymbol{\Psi}_X(\lambda) = \int_{y=0}^{\infty} e^{\boldsymbol{Q}_{++}(\lambda)y}\boldsymbol{Q}_{+-}(\lambda)e^{\boldsymbol{Q}_{--}y} + e^{\boldsymbol{Q}_{++}(\lambda)y}\boldsymbol{\Psi}_x(\lambda)\boldsymbol{Q}_{-+}(\lambda)\boldsymbol{\Psi}_X(\lambda)e^{\boldsymbol{Q}_{--}(\lambda)y}\,\mathrm{d}y, \tag{2.17}$$

which can be shown to be equivalent to (Bhatia & Rosenthal 1997, Lemma 3 and Theorem 9.2)

$$\boldsymbol{Q}_{++}(\lambda)\boldsymbol{\Psi}_X(\lambda) + \boldsymbol{\Psi}_X(\lambda)\boldsymbol{Q}_{--}(\lambda)\boldsymbol{\Psi}_X(\lambda) + \boldsymbol{Q}_{+-}(\lambda) + \boldsymbol{\Psi}_X(\lambda)\boldsymbol{Q}_{-+}(\lambda)\boldsymbol{\Psi}_X(\lambda) = \mathbf{0}. \quad (2.18)$$

The key concepts of this argument are; (1) to partition the state space of the driving process into sets on which the fluid level is increasing, decreasing, or constant; (2) to characterise the sections of sample paths of the fluid level which are non-decreasing and non-increasing as semigroups and derive their generators; (3) to partition the sample paths which comprise $\boldsymbol{\Psi}_X(\lambda)$ such that we can write down an expression for $\boldsymbol{\Psi}_X(\lambda)$ in terms of the expressions derived in (2) and $\boldsymbol{\Psi}_X(\lambda)$ itself and then to solve the resulting expression.

2.4 Fluid-fluid Queues

This subsection was largely taken from Sections 2 and 3 of Bean et al. (2022) with changes, such as notation, so that this chapter is consistent with the rest of the thesis. I am a co-author of the paper Bean et al. (2022).

A stochastic fluid-fluid queue (Bean & O'Reilly 2013*a*) is a Markov process with three elements, $\{(\ddot{Y}(t), X(t), \varphi(t))\}_{t\geq 0}$, where $\{(X(t), \varphi(t))\}_{t\geq 0}$ is a classical fluid queue[‡] and $\ddot{Y}(t)$ is the second fluid, which varies at rate $r_{\varphi(t)}(\dot{X}(t))$:

$$\ddot{Y}(t) := \ddot{Y}(0)) + \int_0^t r_{\varphi(s)}(X(s))\,\mathrm{d}s.$$

Regulated boundaries may also be included for the second fluid level. In this thesis we consider the second fluid to have a regulated boundary at $y = 0$ and unbounded above. To distinguish between unbounded and bounded processes, we use the notation $\ddot{Y}(t)$ to denote the unbounded process and $\dot{Y}(t)$ to denote the second fluid level process with a regulated lower boundary at 0.

In the following, we assume that $\dot{Y}(t) \in [0, \infty)$ and that there is a boundary at level 0 for both the first and second fluid levels and the first fluid level is also bounded above at b. Thus, the rates of change of the fluid levels can be summarised as:

$$\frac{\mathrm{d}}{\mathrm{d}t}X(t) := \max\{0, c_i\} \quad \text{if } X(t) = 0 \text{ and } \varphi(t) = i,$$
$$\frac{\mathrm{d}}{\mathrm{d}t}X(t) := \min\{0, c_i\} \quad \text{if } X(t) = b \text{ and } \varphi(t) = i,$$

[‡]We assume $X(t)$ is doubly-bounded as this results in a finite-dimensional approximation, but this is not necessary

$$\frac{\mathrm{d}}{\mathrm{d}t}\dot{Y}(t) := \max\{0, r_i(x)\} \quad \text{if } \dot{Y}(t) = 0,\ X(t) = x \text{ and } \varphi(t) = i,$$

for $i \in \mathcal{S} = \{1, ..., N\}$. Let $\boldsymbol{R}(x) := \mathrm{diag}(r_i(x))_{i\in\mathcal{S}}$ be the diagonal fluid-rate matrix of functions for $\{\dot{Y}(t)\}$.

For the remainder of this section, we summarise the findings of Bean & O'Reilly (2014) on the joint limiting distribution of $\{(\dot{Y}(t), X(t), \varphi(t))\}_{t\geq 0}$. The derivation of the limiting distribution relies on obtaining the operator Ψ which gives the distribution of the process $\{(X(t), \varphi(t))\}$ at the time when $\{\dot{Y}(t)\}$ first returns to the level 0, given $\dot{Y}(0) = 0$.

The concepts leading to the derivation of Ψ for the fluid-fluid queue are much the same as the derivation of the analogous quantity, $\boldsymbol{\Psi}_X$, for a classical fluid queue. First, we need the infinitesimal generator of the driving process, $\{(X(t), \varphi(t))\}$, which we denote by $\mathbb{B}$. Then we need to partition $\mathbb{B}$ on sets for which $\{\dot{Y}(t)\}$ is increasing, decreasing or constant – these are the sets $\mathcal{F}_i^m$, $i \in \mathcal{S}$, $m \in \{+, -, 0\}$, below. We can then derive an operator-Riccati equation for the first return operator for the fluid-fluid queue, Ψ, in terms of the partitioned generator $\mathbb{B}$, the rates $r_i(x)$. The solution to the operator-Riccati equation gives the distribution of the driving process $\{(X(t), \varphi(t))\}$ at the time when $\{\dot{Y}(t)\}$ first returns to 0.

The problem we have is to solve the operator-Riccati equation – something which is only possibly in the very simplest of cases. This is where the methods in this thesis come in. Here, we approximate $\mathbb{B}$ by a finite-dimensional matrix which we then partition according to the sets $\mathcal{F}_i^m$, $i \in \mathcal{S}$, $m \in \{+, -, 0\}$. We then substitute the resulting matrices into the operator-Riccati equation and this gives us a matrix-Riccati equation which we can then solve using known methods Bean et al. (2009*a*).

A key element of the cell-based approximation schemes is the partition of the state space of $\{(X(t), \varphi(t))\}$ into sets of the form $(\mathcal{D}_{k,i}, i)$, where $\mathcal{D}_{k,i}$ are intervals are known as cells. We must choose the cells wisely so that the partition into the sets $\mathcal{F}_i^m$, $i \in \mathcal{S}$, $m \in \{+, -, 0\}$ can be recovered from the partition into cells. This is not terribly difficult to do, but does require some notation and explanation. We now proceed to introduce the work of Bean & O'Reilly (2014) and explain how we can recover the partition into the sets $\mathcal{F}_i^m$, $i \in \mathcal{S}$, $m \in \{+, -, 0\}$ from the cells as determined by the cell-based approximations.

For each Markovian state $i \in \mathcal{S}$, we partition the state space of $X(t)$, $[0, b]$, according to the rates of change $r_i(\cdot)$ for the second fluid $\{\dot{Y}(t)\}$: $[0, b] := \mathcal{F}_i^+ \cup \mathcal{F}_i^- \cup \mathcal{F}_i^0$, where

$$\begin{aligned} \mathcal{F}_i^+ &:= \{u \in [0, b] : r_i(u) > 0\}, \\ \mathcal{F}_i^- &:= \{u \in [0, b] : r_i(u) < 0\}, \\ \mathcal{F}_i^0 &:= \{u \in [0, b] : r_i(u) = 0\}. \end{aligned} \tag{2.19}$$

For all $i \in \mathcal{S}$, the functions $r_i(\cdot)$ are assumed to be sufficiently well-behaved so that $\mathcal{F}_i^m$, $m \in \{+, -, 0\}$, is a finite union of intervals and isolated points.

We assume that the process $\{(\dot{Y}(t), X(t), \varphi(t))\}_{t\geq 0}$ is positive recurrent, in order to

guarantee the existence of the joint limiting distribution. Define limiting operators

$$\mathbb{\pi}_i(y)(\mathcal{E}) := \lim_{t\to\infty} \frac{\partial}{\partial y}\mathbb{P}\left(\dot{Y}(t) \leq y, X(t) \in \mathcal{E}, \varphi(t) = i\right), \quad y > 0, \tag{2.20}$$

$$\mathbb{p}_i(\mathcal{E}) := \lim_{t\to\infty} \mathbb{P}(\dot{Y}(t) = 0, X(t) \in \mathcal{E}, \varphi(t) = i]), \tag{2.21}$$

where $\mathcal{E} \subset [0, b]$. The keen observer will notice that (2.20) assumes a certain differentiability condition on the limiting distribution. Indeed, it is not always the case that (2.20) will be differentiable with respect to y (trivially, consider $\dot{Y}(t) = y$ for all t). Conditions on fluid-fluid queues to ensure that (2.20) will be differentiable with respect to y are not known, and here we just assume that (2.20) is differentiable.

Now, let $\mathbb{\pi}(y) = (\mathbb{\pi}_i(y))_{i\in\mathcal{S}}$ be a vector containing the joint limiting density operators and $\mathbb{p} = (\mathbb{p}_i)_{i\in\mathcal{S}}$ be a vector containing the joint limiting mass operators. The determination of $\mathbb{\pi}(y)$ involves two important matrices of operators, $\mathbb{D}$ and Ψ which we now introduce.

2.4.1 The infinitesimal generator, $\mathbb{B}$, of the driving process

Given the discussion above, if we are to replicate the arguments of Bean et al. (2005*b*) to derive the Laplace-Stieltjes transform of the first return operator of a fluid-fluid queue, we first need an expression for the generator of the driving process. We also need to partition it according to whether the second fluid is increasing, decreasing or constant.

The generator of a fluid queue is a differential operator and to enable computation of the first-return operator, approximation methods are needed. The approximation schemes which we discuss in this thesis are all cell-based methods which discretise the level of the fluid queue into cells. Thus, one complexity in approximation is to reconcile the partition according to whether the second fluid is increasing, decreasing or constant and the partition which the approximation method uses. As long as the rates of the fluid-fluid queue are not 'too-badly' behaved, this is not terribly difficult to do, but it does require some discussion and notation.

The transition semigroup of a fluid queue

Since $\{(X(t), \varphi(t))\}_{t\geq 0}$ is a Markov process, the evolution of probability can be described by a semigroup. Let $\mathcal{M}(\mathcal{S} \times [0, b])$ be the set of integrable complex-valued Borel measures on the Borel σ-algebra $\mathcal{B}_{\mathcal{S}\times[0,b]}$. For $\overline{\boldsymbol{\mu}} \in \mathcal{M}(\mathcal{S} \times [0, b])$, we can write $\overline{\boldsymbol{\mu}} = (\overline{\mu_i})_{i\in\mathcal{S}}$. The measures $\overline{\mu_i}(\cdot)$ represent an initial distribution, $\overline{\mu_i}(\cdot) = \mathbb{P}(X(0) \in \cdot, \varphi(0) = i)$. Let $\{\overline{\mathbb{V}}(t)\}_{t\geq 0}$, $\overline{\mathbb{V}}(t) : \mathcal{M}(\mathcal{S} \times [0, b] \mapsto \mathcal{M}(\mathcal{S} \times [0, b]$ be the semigroup describing the evolution of probability for $\{(X(t), \varphi(t))\}_{t\geq 0}$ structured as a matrix of operators, $\left[\overline{\mathbb{V}}(t)\right]_{ij} = \overline{\mathbb{V}}_{ij}(t)$ where,

$$\overline{\mu}_i\overline{\mathbb{V}}_{ij}(t)(\mathcal{E}) = \int_{x\in[0,b]} \mathrm{d}\overline{\mu}_i(x)\mathbb{P}(X(t) \in \mathcal{E}, \varphi(t) = j \mid X(0) = x, \varphi(0) = i).$$

Intuitively, the operator $\overline{\mathbb{V}}(t)$ maps an initial measure $\overline{\boldsymbol{\mu}}$ on $(X(0), \varphi(0))$ to the measure $\mathbb{P}(X(t) \in \mathcal{E}, \varphi(t) = j) =: \overline{\mu}_j(t)(\mathcal{E})$. The matrix of operators $\overline{\mathbb{B}} := [\overline{\mathbb{B}}_{ij}]_{i,j\in\mathcal{S}}$ is the *infinitesimal generator* of the semigroup $\{\overline{\mathbb{V}}(t)\}$ defined by

$$\overline{\mathbb{B}} = \frac{\mathrm{d}}{\mathrm{d}t}\overline{\mathbb{V}}(t)\bigg|_{t=0},$$

with domain the set of measures for which this limit exists. Specifically, the domain of $\overline{\mathbb{B}}$ is the set of measures, $\overline{\boldsymbol{\mu}} = (\overline{\mu}_i)_{i\in\mathcal{S}}$, for which each $\overline{\mu}_i$ admits an absolutely continuous density on $(0, b)$, and can have a point mass at 0 if $i \in \mathcal{S}_{-1}$ or a point mass at b if $i \in \mathcal{S}_{K+1}$; call this set of measures $\mathcal{M}_{0,b}$. The measures $\overline{\mu}_i$ cannot have a point masses at 0 if $i \notin \mathcal{S}_{-1}$, nor can they have a point masses at b if $i \notin \mathcal{S}_{K+1}$, nor can they have point masses in $(0, b)$ for any $i \in \mathcal{S}$. In the sequel we write $\overline{v}_i(x)$ as the density of $\overline{\mu}_i$, and $q_{-1,i}$ and $q_{K+1,i}$ as the point masses of $\overline{\mu}_i$ at 0 and b, respectively (if such point masses exist).

A partition with respect to rates of the second fluid

To use the operators $\{\overline{\mathbb{V}}(t)\}$ and $\overline{\mathbb{B}}$ to analyse the fluid-fluid queue, Bean & O'Reilly (2014) explicitly track when $(X(t), \varphi(t)) \in (\mathcal{F}_i^m, i)$ for $i \in \mathcal{S}$, $m \in \{+, -, 0\}$ by partitioning the operators $\overline{\mathbb{V}}(t)$ and $\overline{\mathbb{B}}$ into $\overline{\mathbb{V}}_{ij}^{mn}$ and $\overline{\mathbb{B}}_{ij}^{mn}$, for $i, j \in \mathcal{S}$, $m, n \in \{+, -, 0\}$, where

$$\overline{\mu}_i|_{\mathcal{F}_i^m}\overline{\mathbb{V}}_{ij}^{mn}(t)(\mathcal{E}) := \int_{x\in[0,b]} \mathrm{d}\overline{\mu}_i|_{\mathcal{F}_i^m}(x)\mathbb{P}(X(t) \in \mathcal{E} \cap \mathcal{F}_j^n, \varphi(t) = j \mid X(0) = x, \varphi(0) = i),$$

and $\overline{\mu}_i|_E$ is the restriction of $\overline{\mu}_i$ to E, $\overline{\mu}_i|_E(\mathcal{E}) = \overline{\mu}_i(\mathcal{E} \cap E)$. Similarly, for $\overline{\mathbb{B}}_{ij}^{mn}$, $i, j \in \mathcal{S}$, $m, n \in \{+, -, 0\}$.

A partition as dictated by the approximation schemes

We claim that numerical schemes are needed to approximate the analytic operator equations introduced in Bean & O'Reilly (2014). The schemes we choose to use here work by first partitioning the state space of the fluid level, $\{X(t)\}$, into a collection of intervals, $\mathcal{D}_k, i$ then constructing an approximation to $\overline{\mathbb{B}}$ on each interval. To help elucidate the connection between the operators $\{\overline{\mathbb{V}}(t)\}$, $\overline{\mathbb{B}}$ and their approximate counterparts we take a slightly different approach to partitioning these operators than that taken in Bean & O'Reilly (2014). Rather than partition according to the sets $\mathcal{F}_i^m$, $i \in \mathcal{S}$, $m \in \{+, -, 0\}$, we use the same partition as that in the construction of the approximation schemes. By doing so, we can directly correspond elements of the partitioned operators to their approximation counterparts. Since the partition used to construct the approximation schemes is finer, then we can reconstruct the partition in terms of the sets $\mathcal{F}_i^m$, $i \in \mathcal{S}$, $m \in \{+, -, 0\}$.

Let us first partition the space $[0, b]$ into $\mathcal{D}_{-1,i} = \mathcal{D}_{-1} = \{0\}$, $i \in \mathcal{S}_{-1}$, $\mathcal{D}_{K+1,i} = \mathcal{D}_{K+1} = \{b\}$, $i \in \mathcal{S}_{K+1}$, and non-trivial intervals $\mathcal{D}_{k,i} = [y_k, y_{k+1}) \setminus \{0\}$, $i \in \mathcal{S}_+ \cup \mathcal{S}_0$ and

$\mathcal{D}_{k,i} = (y_k, y_{k+1}] \setminus \{b\}$, $i \in \mathcal{S}_-$, with $y_0 = 0$, $y_{K+1} = b$, $y_k < y_{k+1}$, $k \in \mathcal{K}^\circ := \{0, 1, 2, ..., K\}$ and define $\mathcal{K} = \{-1, K+1\} \cup \mathcal{K}^\circ$. For $\boldsymbol{\mu} \in \mathcal{M}_{0,b}(\mathcal{S} \times [0, b])$ we write $\boldsymbol{\mu} = (\mu_{k,i})_{i \in \mathcal{S}, k \in \mathcal{K}}$, where $\mu_{k,i}(\cdot) = \mu_i(\cdot \cap \mathcal{D}_{k,i})$, $k \in \mathcal{K}$. We denote by $v_{k,i}(x)$, $x > 0$, the density associated with the measure, $\mu_{k,i}$, $k \in \mathcal{K}^\circ$. For $i, j \in \mathcal{S}$, $k, \ell \in \mathcal{K}$ define the operators

$$\mu_{k,i}\mathbb{V}_{ij}^{k\ell}(t)(\mathcal{E}) := \int_{x \in \mathcal{D}_{k,i}} \mathrm{d}\mu_{k,i}(x)\mathbb{P}(X(t) \in \mathcal{E} \cap \mathcal{D}_{\ell,i}, \varphi(t) = j \mid X(0) = x, \varphi(0) = i),$$

and the matrices of operators $\mathbb{V}^{k\ell}(t) := \left[\mathbb{V}_{ij}^{k\ell}(t)\right]_{i,j \in \mathcal{S}}$, $k, \ell \in \mathcal{K}$ and write

$$\mathbb{V}(t) = \begin{bmatrix} \mathbb{V}^{-1,-1}(t) & \mathbb{V}^{-1,0}(t) & \mathbb{V}^{-1,1}(t) & \dots & \mathbb{V}^{-1,K+1}(t) \\ \mathbb{V}^{0,-1}(t) & \mathbb{V}^{0,0}(t) & \mathbb{V}^{0,1}(t) & \dots & \mathbb{V}^{0,K+1}(t) \\ \mathbb{V}^{1,-1}(t) & \mathbb{V}^{1,0}(t) & \mathbb{V}^{1,1}(t) & \dots & \mathbb{V}^{1,K+1}(t) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbb{V}^{K+1,-1}(t) & \mathbb{V}^{K+1,0}(t) & \mathbb{V}^{K+1,1}(t) & \dots & \mathbb{V}^{K+1,K+1}(t) \end{bmatrix}.$$

Now define $\mathbb{B} = \dfrac{\mathrm{d}}{\mathrm{d}t}\mathbb{V}(t)\Big|_{t=0}$ as the infinitesimal generator of $\{\mathbb{V}(t)\}$, resulting in the tridiagonal matrix of operators

$$\mathbb{B}(t) = \begin{bmatrix} \mathbb{B}^{-1,-1}(t) & \mathbb{B}^{-1,0}(t) & & & \\ \mathbb{B}^{0,-1}(t) & \mathbb{B}^{0,0}(t) & \mathbb{B}^{0,1}(t) & & \\ & \mathbb{B}^{1,0}(t) & \mathbb{B}^{1,1}(t) & \ddots & \\ & & \ddots & \ddots & \mathbb{B}^{K,K+1} \\ & & & \mathbb{B}^{K+1,K} & \mathbb{B}^{K+1,K+1} \end{bmatrix},$$

where the blocks $\mathbb{B}^{k\ell} := \left[\mathbb{B}_{ij}^{k\ell}(t)\right]_{i,j \in \mathcal{S}}$, $k, \ell \in \mathcal{K}$.[§] The tridiagonal structure arises since, for $|k - \ell| \geq 2$ it is impossible for $\{X(t)\}$ to move from $\mathcal{D}_{k,i}$ to $\mathcal{D}_{\ell,j}$ in an infinitesimal amount of time.

By an appropriate choice of the intervals $\{\mathcal{D}_k\}$, $k \in \mathcal{K}$, the partition used in Bean & O'Reilly (2014) can be recovered. Intuitively, we must ensure that each of the boundaries

[§]We use a blackboard bold font with an overline above the character (e.g. $\overline{\mathbb{B}}$ and $\overline{\mathbb{V}}(t)$) to represent theoretical operators derived in (Bean & O'Reilly 2014) which are constructed using the partition in (2.19). The operators denoted with an overline play a minor role in the introductory sections of this thesis, but do not appear again. We use a blackboard font sans overline (e.g. $\mathbb{V}(t)$ and $\mathbb{B}$) to represent the same operators but which are constructed with the finer partition defined by $\mathcal{D}_k$, $k \in \mathcal{K}$. We use the letters $i, j \in \mathcal{S}$ to represent states of the phase process, letters $m, n, \in \{+, -, 0\}$ to refer to the partition in terms of the sets in Equations (2.19), and the letters $k, \ell \in \mathcal{K}$ to refer to the finer partition into sets $\{\mathcal{D}_k\}_k$. With a slight abuse of notation, whenever we use the dummy variables k, ℓ without qualification we imply $k, \ell \in \mathcal{K}$, the dummy variables without qualification m, n imply $m, n \in \{+, -, 0\}$ and the dummy variables i, j without qualification imply $i, j \in \mathcal{S}$. E.g. $\mathbb{B}_{ij}^{k\ell}$ means $\mathbb{B}_{ij}^{k\ell}$, $i, j \in \mathcal{S}, k, \ell \in \mathcal{K}$ and $\mathbb{B}_{ij}^{mn}$ means $\mathbb{B}_{ij}^{mn}$, $i, j \in \mathcal{S}, m, n \in \{+, -, 0\}$.

of $\mathcal{F}_i^m$, $i \in \mathcal{S}$, $m \in \{+,-,0\}$, align with a boundary of a cell $\mathcal{D}_{k,i}$. Then, each set $\mathcal{F}_i^m$, $i \in \mathcal{S}$, $m \in \{+,-,0\}$, can be written as a union of cells, $\mathcal{D}_k$, $k \in \mathcal{K}$, sans a collection of points which have measure zero for all measures in $\mathcal{M}_{0,b}$, and this collection of points is inconsequential for the purposes of the approximations presented here.

Formally, to recover the partition used in (Bean & O'Reilly 2014) we choose the intervals $\mathcal{D}_{k,i}$ such that $l(\mathcal{D}_{k,i} \cap \mathcal{F}_i^m) \in \{l(\mathcal{D}_{k,i}), 0\}$ for all $i \in \mathcal{S}$, $m \in \{+,-,0\}$, $k \in \mathcal{K}$, for all $l \in \mathcal{M}_{0,b}$. That is, we choose $\mathcal{D}_{k,i}$ such that it is contained, up to sets of measure 0 with respect to measures in $\mathcal{M}_{0,b}$, within one of the sets $\mathcal{F}_i^m$ for $m \in \{+,-,0\}$ and $i \in \mathcal{S}$. We assume such a partition for the rest of the thesis. For $i \in \mathcal{S}$, $m \in \{+,-,0\}$, let $\mathcal{K}_i^m = \{k \in \mathcal{K} \mid l(\mathcal{D}_{k,i} \cap \mathcal{F}_i^m) = l(\mathcal{D}_{k,i}),\, l \in \mathcal{M}_0\}$, so that $\bigcup_{k \in \mathcal{K}_i^m} \mathcal{D}_{k,i}$ and $\mathcal{F}_i^m$ are equal up to a set of $\mathcal{M}_{0,b}$-measure 0. Define $\mathcal{K}^m = \bigcup_{i \in \mathcal{S}} \mathcal{K}_i^m$, $m \in \{+,-,0\}$.

To recover the partition defined by (2.19) we bundle together the elements of $\mathbb{V}(t)$ which correspond to $\mathcal{F}_i^m$ and $\mathcal{F}_j^n$. That is, for $m,n \in \{+,-,0\}$, define $\mathbb{V}_{ij}^{mn}(t)$ as the matrix of operators

$$\mathbb{V}_{ij}^{mn}(t) = \left[\mathbb{V}_{ij}^{k\ell}(t)\right]_{k \in \mathcal{K}_i^m, \ell \in \mathcal{K}_j^n}.$$

The same construction can be achieved with $\mathbb{B}$.

Let $\mathcal{S}_k^+ = \{i \in \mathcal{S} \mid r_i(x) > 0, \forall x \in \mathcal{D}_{k,i}\}$, $\mathcal{S}_k^0 = \{i \in \mathcal{S} \mid r_i(x) = 0, \forall x \in \mathcal{D}_{k,i}\}$, $\mathcal{S}_k^- = \{i \in \mathcal{S} \mid r_i(x) < 0, \forall x \in \mathcal{D}_{k,i}\}$, $\mathcal{S}_k^\bullet = \{i \in \mathcal{S} \mid r_i(x) \neq 0, \forall x \in \mathcal{D}_{k,i}\}$ for $k \in \mathcal{K}^\circ$ and $\mathcal{S}_k^+ = \{i \in \mathcal{S}_k \mid r_i(y_k) > 0\}$, $\mathcal{S}_k^- = \{i \in \mathcal{S}_k \mid r_i(y_k) < 0\}$, $\mathcal{S}_k^0 = \{i \in \mathcal{S}_k \mid r_i(y_k) = 0\}$, $\mathcal{S}_k^\bullet = \{i \in \mathcal{S}_k \mid r_i(y_k) \neq 0\}$ for $k \in \{-1, K+1\}$. For later reference, we need the following constructions. For $k,\ell \in \mathcal{K}$ let

$$\mathbb{B}^{k\ell} = \left[\mathbb{B}_{ij}^{k\ell}\right]_{i,j \in \mathcal{S}}, \tag{2.22}$$

for $i,j \in \mathcal{S}$ let

$$\mathbb{B}_{ij} = \left[\mathbb{B}_{ij}^{k\ell}\right]_{k,\ell \in \mathcal{K}}, \tag{2.23}$$

and for $m,n \in \{+,-,0\}$ let

$$\mathbb{B}^{mn} = \left[\left[\mathbb{B}_{ij}^{k\ell}\right]_{i \in \mathcal{S}_k^m, j \in \mathcal{S}_\ell^n}\right]_{k \in \mathcal{K}^m, \ell \in \mathcal{K}^n}, \tag{2.24}$$

$$\mathbb{B}^{kn} = \left[\left[\mathbb{B}_{ij}^{k\ell}\right]_{i \in \mathcal{S}_k^m, j \in \mathcal{S}_\ell^n}\right]_{\ell \in \mathcal{K}^n} \text{ for } k \in \mathcal{K}, \tag{2.25}$$

$$\mathbb{B}^{m\ell} = \left[\left[\mathbb{B}_{ij}^{k\ell}\right]_{i \in \mathcal{S}_k^m, j \in \mathcal{S}_\ell^n}\right]_{k \in \mathcal{K}^m} \text{ for } \ell \in \mathcal{K}. \tag{2.26}$$

We persist with the partition $\mathcal{D}_{k,i}$, $k \in \mathcal{K}$ throughout as this is consistent with the partition used in the approximation schemes considered in this thesis. Note that for all the operators defined with this partition, the partitioning used in (Bean & O'Reilly 2014) can always be recovered by the above construction.

Definition of $\mathbb{B}$

When v_k is differentiable we can write $\mu_{k,i}\mathbb{B}_{ij}^{k\ell}(\mathcal{E})$ in kernel form as

$$\int_{x\in\mathcal{D}_{k,i},y\in\mathcal{E}} \mathrm{d}\mu_{k,i}(x)\mathbb{B}_{ij}^{k\ell}(x,dy).$$

It is known that

$$\mu_{k,i}\mathbb{B}_{ij}^{kk}(\,\mathrm{d}y) := \int_{x\in\mathcal{D}_{k,i}} \mathrm{d}\mu_{k,i}(x)\mathbb{B}_{ij}^{kk}(x,\,\mathrm{d}y) = \begin{cases} v_{k,i}(y)T_{ij}\,\mathrm{d}y, & i\neq j,\\ v_{k,i}(y)T_{ii}\,\mathrm{d}y - c_i\dfrac{\mathrm{d}}{\mathrm{d}y}v_{k,i}(y)\,\mathrm{d}y, & i=j, \end{cases} \tag{2.27}$$

on the interior of $\mathcal{D}_{k,i}$, $k\in\mathcal{K}^\circ$, (Karandikar & Kulkarni 1995). Intuitively, $v_{k,i}(y)T_{ij}\,\mathrm{d}y$ represents the instantaneous rate of transition from phase i to j in the infinitesimal interval $\mathrm{d}y$, $v_{k,i}(y)T_{ii}\,\mathrm{d}y$ represents no such transition occurring, and $-c_i\dfrac{\mathrm{d}}{\mathrm{d}y}v_{k,i}(y)\,\mathrm{d}y$ represents the drift across the interval $\mathrm{d}y$ when the phase is i.

Translating the results of Bean & O'Reilly (2014) to use the partition $\{\mathcal{D}_{k,i}\}$ we may state that, for all $i,j\in\mathcal{S}$, $i\neq j$, $k\in\{1,\ldots,K-1\}$,

$$\begin{aligned}\mu_{k,i}\mathbb{B}_{ij}^{kk}(\mathcal{D}_{k,i}) &= \int_{x\in\mathcal{D}_k} v_{k,i}(x)T_{ij}\,\mathrm{d}x,\\ \mu_{k,i}\mathbb{B}_{ii}^{kk}(\mathcal{D}_{k,i}) &= \int_{x\in\mathcal{D}_k} v_{k,i}(x)T_{ii}\,\mathrm{d}x - c_i v_{k,i}(y_{k+1}^-)1(c_i>0) + c_i v_{k,i}(y_k^+)1(c_i<0),\end{aligned}$$

where $1(\cdot)$ is the indicator function and x^+ and x^- denote the right and left limits at x. Intuitively, the first expression represents the instantaneous rate of the stochastic transitions of the phase process $\{\varphi(t)\}$ from i to j while $\{X(t)\}$ remains in $\mathcal{D}_{k,i}$. The first term in the second expression represents the net rate of transition out of phase i while $\{X(t)\}$ remains in $\mathcal{D}_{k,i}$, while the second and third terms in the second expression represent the flux out of the right-hand edge of $\mathcal{D}_{k,i}$ when $c_i>0$ and the flux out of the left-hand edge of $\mathcal{D}_{k,i}$ when $c_i<0$, respectively. Essentially, this is just a rewriting of (2.27) as an integral equation and including the boundary conditions.

The results of Bean & O'Reilly (2014) also imply that,

$$\begin{aligned}\mu_{k,i}\mathbb{B}_{ii}^{k,k+1}(\mathcal{D}_{k+1}) &= c_i v_{k,i}(y_{k+1}^-)1(c_i>0),\ \text{for all } i\in\mathcal{S},\ k\in\{0,1,2,...,K-1\},\\ \mu_{k,i}\mathbb{B}_{ii}^{k,k-1}(\mathcal{D}_{k-1}) &= -c_i v_{k,i}(y_k^+)1(c_i<0),\ \text{for all } i\in\mathcal{S},\ k\in\{1,2,3,...,K\}.\end{aligned}$$

Intuitively, the first equation represents the flux from $\mathcal{D}_{k,i}$ to $\mathcal{D}_{k+1,i}$ across the shared boundary at y_{k+1} which occurs when $c_i>0$ only. The second expression represents the flux from $\mathcal{D}_{k,i}$ to $\mathcal{D}_{k-1,i}$ across the shared boundary at y_k which occurs when $c_i<0$ only.

At the boundary $x = 0$, the rate at which point masses move between phases is

$$\mu_{-1,i}\mathbb{B}_{ii}^{-1,-1} = \mu_{-1,i}(\{0\})T_{ii}, \text{ if } c_i \leq 0,$$
$$\mu_{-1,i}\mathbb{B}_{ij}^{-1,-1} = \mu_{-1,i}(\{0\})T_{ij}, \text{ if } c_i \leq 0,\ c_j \leq 0.$$

The rate at which point mass leaves the boundary is

$$\mu_{-1,i}\mathbb{B}_{ij}^{-1,0} = \mu_{-1,i}(\{0\})T_{ij}, \text{ if } c_i \leq 0,\ c_j > 0.$$

The rate at which point masses accumulate at the boundary is

$$\mu_{0,i}\mathbb{B}_{ij}^{0,-1} = -c_i p_{ij}^{-1} v_{0,i}(0^+), \text{ if } c_i < 0,\ c_j \leq 0.$$

In $\mathcal{D}_{0,i}$, if $c_i < 0$ and $c_j > 0$ the phase changes from i to j either from a stochastic jump at rate T_{ij}, or from $\{X(t)\}$ hitting the boundary in phase i and transitioning with probability p_{ij}^{-1} to phase j, hence

$$\mu_{0,i}\mathbb{B}_{ij}^{0,0}(\mathcal{D}_{0,i}) = \int_{x\in\mathcal{D}_{0,i}} v_{0,i}(x)T_{ij}\,\mathrm{d}x - c_i v_{0,i}(0^+)p_{ij}^{-1}.$$

Also in $\mathcal{D}_{0,i}$, if $i \neq j$ and $c_i \geq 0$ or $i \neq j$, $c_i < 0$ and $c_j \leq 0$, the phase changes from i to j while $\{X(t)\}$ remains in $\mathcal{D}_{0,i}$ at rate

$$\mu_{0,i}\mathbb{B}_{ij}^{0,0}(\mathcal{D}_{0,i}) = \int_{x\in\mathcal{D}_{0,i}} v_{0,i}(x)T_{ij}\,\mathrm{d}x,$$

else, $\{(X(t), \varphi(t))\}$ leaves $(\mathcal{D}_{0,i})$ at rate

$$\mu_{0,i}\mathbb{B}_{ii}^{0,0}(\mathcal{D}_{0,i}) = \int_{x\in\mathcal{D}_{0,i}} v_{0,i}(x)T_{ii}\,\mathrm{d}x - c_i v_{0,i}(y_1^-)1(c_i > 0) + c_i v_{0,i}(0^+)1(c_i < 0).$$

Similarly, at the boundary $x = b$, the rate at which point masses move between phases is

$$\mu_{K+1,i}\mathbb{B}_{ii}^{K+1,K+1} = \mu_{K+1,i}(\{0\})T_{ii}, \text{ if } c_i \geq 0,$$
$$\mu_{K+1,i}\mathbb{B}_{ij}^{K+1,K+1} = \mu_{K+1,i}(\{0\})T_{ij}, \text{ if } c_i \geq 0,\ c_j \geq 0.$$

The rate at which point mass leaves the boundary is

$$\mu_{K+1,i}\mathbb{B}_{ij}^{K+1,K} = \mu_{K+1,i}(\{0\})T_{ij}, \text{ if } c_i \geq 0,\ c_j < 0.$$

The rate at which point masses accumulate at the boundary is

$$\mu_{K,i}\mathbb{B}_{ij}^{K,K+1} = c_i p_{ij}^{K+1} v_{K,i}(0^+), \text{ if } c_i > 0,\ c_j \geq 0.$$

In $\mathcal{D}_{K,i}$, if $c_i > 0$ and $c_j < 0$ the phase changes from i to j either from a stochastic jump at rate T_{ij}, or from $\{X(t)\}$ hitting the boundary in phase i and transitioning with probability p_{ij}^{K+1} to phase j,

$$\mu_{K,i}\mathbb{B}_{ij}^{K,K}(\mathcal{D}_{K,i}) = \int_{x\in\mathcal{D}_{K,i}} v_{K,i}(x)T_{ij}\,\mathrm{d}x + c_i v_{K,i}(b^-)p_{ij}^{K+1},\ c_i > 0,\ c_j < 0.$$

If $i \neq j$ and $c_i \leq 0$ or $i \neq j$, $c_i > 0$ and $c_j \geq 0$, the phase changes from i to j within $\mathcal{D}_{K,i}$ at rate

$$\mu_{K,i}\mathbb{B}_{ij}^{K,K}(\mathcal{D}_{K,i}) = \int_{x\in\mathcal{D}_{K,i}} v_{K,i}(x)T_{ij}\,\mathrm{d}x,$$

else, $\{(X(t),\varphi(t))\}$ leaves $(\mathcal{D}_{K,i}, i)$ at rate

$$\mu_{K,i}\mathbb{B}_{ii}^{K,K}(\mathcal{D}_{K,i}) = \int_{x\in\mathcal{D}_{K,i}} v_{K,i}(x)T_{ii}\,\mathrm{d}x + c_i v_{K,i}(y_K^+)1(c_i < 0) - c_i v_{K,i}(b^-)1(c_i > 0).$$

Otherwise, $\mu_{k,i}\mathbb{B}_{ij}^{k\ell} = 0$.

Note that we have not presented $\mathbb{B}$ in its full detail here and refer the reader to (Bean & O'Reilly 2014) for the details. The main goal here is to show how $\mathbb{B}$ is used to construct the limiting distribution of the fluid-fluid queue and to illustrate the link between the operator $\mathbb{B}$ and the approximations of the same object. As we shall see later, these expressions closely resemble the approximations to the same quantities.

2.4.2 The infinitesimal generator, $\mathbb{D}$, of an in-out process

Let $\beta(t) := \int_0^t \left|r_{\varphi(z)}(X(z))\right| \mathrm{d}z$ be the total unregulated amount of fluid that has flowed into or out of the second buffer during $[0,t]$, and let $\eta(y) := \inf\{t > 0 : \beta(t) = y\}$ be the first time this accumulated in-out amount hits level y. Note that at the stopping time $\eta(y)$ it must be that $(X(\eta(y)), \varphi(\eta(y))) \in (\mathcal{F}_i^m, i)$ for some $i \in \mathcal{S}$ and $m \in \{+,-\}$, i.e. $m \neq 0$. We define the operators $\mathbb{U}_{ij}^{k\ell}(y,s) : \mathcal{M}_{0,b}(\mathcal{D}_{k,i}\cap\mathcal{F}_i^m) \mapsto \mathcal{M}_{0,b}(\mathcal{D}_{\ell,j}\cap\mathcal{F}_j^n)$, for $k \in \mathcal{K}_i^+ \cup \mathcal{K}_i^-$, $\ell \in \mathcal{K}_j^+ \cup \mathcal{K}_j^-$, and $i \in \mathcal{S}_k^\bullet$, $j \in \mathcal{S}_k^\bullet$, by

$$\begin{aligned}&\mu_{k,i}\mathbb{U}_{ij}^{k\ell}(y,s)(\mathcal{E})\\ &:= \int_{x\in\mathcal{D}_{k,i}} \mathrm{d}\mu_{k,i}(x)\mathbb{E}\left[e^{-s\eta(y)}1\left\{X(\eta(y))\in\mathcal{E}\cap\mathcal{D}_{\ell,j},\ \varphi(\eta(y)) = j\right\} \mid \varphi(0) = i, X(0) = x\right].\end{aligned}$$

Then, construct the matrix of operators

$$\mathbb{U}(y,s) := \left[\left[\mathbb{U}_{ij}^{k\ell}(y,s)\right]_{i\in\mathcal{S}_k^\bullet, j\in\mathcal{S}_\ell^\bullet}\right]_{k,\ell\in\mathcal{K}^+\cup\mathcal{K}^-}.$$

The matrix of operators $\mathbb{D}(s)$ is the infinitesimal generator of the semigroup $\{\mathbb{U}(y,s)\}_{y\geq 0}$ defined by

$$\mathbb{D}(s) = \frac{\mathrm{d}}{\mathrm{d}y}\mathbb{U}(y,s)|_{y=0},$$

whenever this limit exists.

Recalling the constructions in Equations (2.22)-(2.26) and using Lemma 4 of Bean & O'Reilly (2014) gives the following expression for $\mathbb{D}(s)$.

Lemma 2.1. *For $y \geq 0$, $s \in \mathbb{C}$ with $Re(s) \geq 0$, $i,j \in \mathcal{S}$, $k \in \mathcal{K}_i^+ \cup \mathcal{K}_i^-$, $\ell \in \mathcal{K}_j^+ \cup \mathcal{K}_j^-$,*

$$\mathbb{D}_{ij}^{k\ell}(s) = [\mathbb{R}^k(\mathbb{B}^{k\ell} - s\mathbb{I} + \mathbb{B}^{k0}(s\mathbb{I} - \mathbb{B}^{00})^{-1}\mathbb{B}^{0\ell})]_{ij},$$

where $\mathbb{I}$ is the identity operator, and $\mathbb{R}^k := diag(\mathbb{R}_i^k,\, i \in \mathcal{S})$ is a diagonal matrix of operators $\mathbb{R}_i^k$ given by

$$\mu_{k,i}\mathbb{R}_{k,i}(\mathcal{E}) := \int_{x\in\mathcal{E}\cap\mathcal{D}_{k,i}} \frac{1}{|r_i(x)|}\,\mathrm{d}\mu_{k,i}(x), \quad k \in \mathcal{K}_i^+ \cup \mathcal{K}_i^-.$$

Also, construct the matrices of operators

$$\mathbb{D}^{mn} := \left[\left[\mathbb{D}_{ij}^{k\ell}\right]_{i\in\mathcal{S}_k^m, j\in\mathcal{S}_k^n}\right]_{k\in\mathcal{K}^m,\ell\in\mathcal{K}^n}.$$

2.4.3 The first-return operator, $\Psi(s)$

We denote by $\Psi(s)$ the matrix of operators with the same dimensions as $\mathbb{D}^{+-}$, recording the Laplace-Stieltjes transforms of the time for $\{\dot{Y}(t)\}$ to return, for the first time, to the initial level of zero as introduced in Bean & O'Reilly (2014) but constructed with respect to the finer partition $\{\mathcal{D}_{k,i}\}$. Define the stopping time $\zeta_Y(E) := \inf\{t > 0 : \dot{Y}(t) \in E\}$ to be the first time $\{\dot{Y}(t)\}$ hits the set E, then each component $\Psi_{ij}^{k\ell}(s) : \mathcal{M}_{0,b}(\mathcal{D}_{k,i}) \mapsto \mathcal{M}_{0,b}(\mathcal{D}_{\ell,j}), i,j \in \mathcal{S}$, $k \in \mathcal{K}_i^+$ and $\ell \in \mathcal{K}_j^-$, is given by

$$\begin{aligned}
&\mu_{k,i}\Psi_{ij}^{k\ell}(s)(\mathcal{E}) \\
&:= \int_{x\in\mathcal{D}_{k,i}} \mathrm{d}\mu_{k,i}(x)\mathbb{E}\big[e^{-s\zeta_Y(\{0\})}1\left(X(\zeta_Y(\{0\})) \in \mathcal{E}\cap\mathcal{D}_{\ell,j},\ \varphi(\zeta_Y(\{0\})) = j\right) \mid X(0) = x, \\
&\qquad \dot{Y}(0) = 0, \varphi(0) = i\big].
\end{aligned}$$

Bean & O'Reilly (2014) Theorem 1 provides the following result which characterises $\Psi(s)$.

Theorem 2.2. *For $Re(s) \geq 0$, $\Psi(s)$ satisfies the equation:*

$$\mathbb{D}^{+-}(s) + \Psi(s)\mathbb{D}^{-+}(s)\Psi(s) + \mathbb{D}^{++}(s)\Psi(s) + \Psi(s)\mathbb{D}^{--}(s) = 0.$$

Furthermore, if s is real then $\Psi(s)$ is the minimal non-negative solution.

2.4.4 Limiting Distribution

Let $\Psi := \Psi(0)$. We define $\zeta_Y^n(\{0\}) := \inf\{t \geq \zeta_Y^{n-1}(\{0\}) : \dot{Y}(t) = 0\}$, for $n \geq 2$, to be the sequence of hitting times to level 0 of $\dot{Y}(t)$, with $\zeta_Y^1(\{0\}) := \zeta_Y(\{0\})$. Consider the discrete-time Markov process $\{X(\zeta_Y^n(\{0\})), \varphi(\zeta_Y^n(\{0\}))\}_{n\geq 1}$, and for $i \in \mathcal{S}$, $k \in \mathcal{K}_i^-$ define the measures $\mathfrak{k}_{k,i}$ given by

$$\mathfrak{k}_{k,i}(\mathcal{E}) := \lim_{n\to\infty} \mathbb{P}\left(X(\zeta_Y^n(\{0\})) \in \mathcal{E} \cap \mathcal{D}_{k,i}, \varphi(\zeta_Y^n(\{0\})) = i\right).$$

By Bean & O'Reilly (2014), the vector of measures $\mathfrak{k} := (\mathfrak{k}_{k,i})_{i\in\mathcal{S}_k^-,k\in\mathcal{K}^-}$ satisfies the set of equations

$$\begin{bmatrix}\mathfrak{k} & \mathbf{0}\end{bmatrix}\left(-\begin{bmatrix}\mathbb{B}^{--} & \mathbb{B}^{-0}\\ \mathbb{B}^{0-} & \mathbb{B}^{00}\end{bmatrix}\right)^{-1}\begin{bmatrix}\mathbb{B}^{-+}\\ \mathbb{B}^{0+}\end{bmatrix}\Psi = \mathfrak{k}, \tag{2.28}$$

$$\sum_{k\in\mathcal{K}^-}\sum_{i\in\mathcal{S}_k^-}\mathfrak{k}_{k,i}(\mathcal{F}_i^-) = 1. \tag{2.29}$$

We reproduce Theorem 2 of Bean & O'Reilly (2014) below, which gives the joint limiting distribution of $\{(\dot{Y}(t), X(t), \varphi(t))\}$. Recall that the joint limiting density operator $\mathbb{f}(y) = (\mathbb{f}_i(y))_{i\in\mathcal{S}}$ for $\{(\dot{Y}(t), X(t), \varphi(t))\}$ and the joint limiting mass operator $\mathbb{p} = (\mathbb{p}_i)_{i\in\mathcal{S}}$ are defined by (2.20) and (2.21), respectively. We can partition $\mathbb{f}$ so that

$$\begin{aligned}\mathbb{f}(y) &= \begin{bmatrix}\mathbb{f}^+(y) & \mathbb{f}^-(y) & \mathbb{f}^0(y)\end{bmatrix}\\ &= \begin{bmatrix}(\mathbb{f}_{k,i}(y))_{i\in\mathcal{S}_k^+,k\in\mathcal{K}^+} & (\mathbb{f}_{k,i}(y))_{i\in\mathcal{S}_k^-,k\in\mathcal{K}^-} & (\mathbb{f}_{k,i}(y))_{i\in\mathcal{S}_k^0,k\in\mathcal{K}^0}\end{bmatrix},\end{aligned}$$

where

$$\mathbb{f}_{k,i}(y)(\mathcal{E}) = \mathbb{f}_i(y)(\mathcal{E}\cap\mathcal{D}_{k,i}).$$

Similarly, we can write

$$\mathbb{p} = \begin{bmatrix}\mathbb{p}^- & \mathbb{p}^0\end{bmatrix} = \begin{bmatrix}(\mathbb{p}_{k,i})_{i\in\mathcal{S}_k^-,k\in\mathcal{K}^-} & (\mathbb{p}_{k,i})_{i\in\mathcal{S}_k^0,k\in\mathcal{K}^0}\end{bmatrix},$$

where $\mathbb{p}_{k,i}(\mathcal{E}) = \mathbb{p}_i(\mathcal{E}\cap\mathcal{D}_{k,i})$.

Theorem 2.3. *The operator $\mathbb{f}^m(y)$, for $m \in \{+,-,0\}$ and $y > 0$, and the probability mass $\mathbb{p}^m$, for $m \in \{-,0\}$, satisfy the set of equations:*

$$\mathbb{f}^0(y) = \begin{bmatrix}\mathbb{f}^+(y) & \mathbb{f}^-(y)\end{bmatrix}\begin{bmatrix}\mathbb{B}^{+0}\\ \mathbb{B}^{-0}\end{bmatrix}\left(-\mathbb{B}^{00}\right)^{-1}, \tag{2.30}$$

$$\begin{bmatrix}\mathbb{f}^+(y) & \mathbb{f}^-(y)\end{bmatrix} = \begin{bmatrix}\mathbb{p}^- & \mathbb{p}^0\end{bmatrix}\begin{bmatrix}\mathbb{B}^{-+}\\ \mathbb{B}^{0+}\end{bmatrix}\begin{bmatrix}e^{\mathbb{K}y} & e^{\mathbb{K}y}\Psi\end{bmatrix}\begin{bmatrix}\mathbb{R}^+ & 0\\ 0 & \mathbb{R}^-\end{bmatrix}, \tag{2.31}$$

$$\begin{bmatrix}\mathbb{p}^{-} & \mathbb{p}^{0}\end{bmatrix} = z\begin{bmatrix}\mathbb{k} & \mathbf{0}\end{bmatrix}\left(-\begin{bmatrix}\mathbb{B}^{--} & \mathbb{B}^{-0}\\ \mathbb{B}^{0-} & \mathbb{B}^{00}\end{bmatrix}\right)^{-1}, \tag{2.32}$$

$$\sum_{m\in\{+,-,0\}}\sum_{i\in\mathcal{S}}\int_{y=0}^{\infty}\mathbb{w}_i^m(y)(\mathcal{F}_i^m)\,\mathrm{d}y + \sum_{m\in\{-,0\}}\sum_{i\in\mathcal{S}}\mathbb{p}_i^m(\mathcal{F}_i^m) = 1, \tag{2.33}$$

where $\mathbb{K} := \mathbb{D}^{++}(0) + \mathbb{\Psi}\mathbb{D}^{(-+)}(0)$ *and* z *is a normalising constant.*

At this point we reiterate that Equations (2.30)-(2.33) are operator equations and are only amenable to numerical evaluation in the simplest of cases. Sources of this intractability come from, for example, the need to find the inverse operator $(-\mathbb{B}^{00})^{-1}$, and the need to find the solution, $\mathbb{\Psi}(s)$, of the operator equation in Theorem 2.2. There is also the complexity of the partition of the operators defined by the sets $\mathcal{F}_i^m$, $i \in \mathcal{S}$, $m \in \{+,-,0\}$. Therefore, there is the need for approximation schemes such as those presented in this thesis.

2.5 Quasi-birth-and-death processes with rational arrival process components

One of the approximation schemes we develop in this thesis is based on a suitably defined quasi-birth-and-death process with rational-arrival-process components (QBD-RAP). QBD-RAP processes are built from matrix-exponentially distributed inter-event times. We now introduce the class of matrix exponential distributions and recount some important properties before introducing the QBD-RAP.

2.5.1 Matrix exponential distributions

Here we recount some facts about matrix exponential distributions. See Bladt & Nielsen (2017) for a more detailed exposition. A random variable, Z, is said to have a matrix exponential distribution if it has a distribution function of the form $1 - \boldsymbol{\alpha}e^{\boldsymbol{S}x}(-\boldsymbol{S})^{-1}\boldsymbol{s}$, where $\boldsymbol{\alpha}$ is a $1 \times p$ *initial vector*, $\boldsymbol{S}$ a $p \times p$ matrix, and $\boldsymbol{s}$ a $p \times 1$ *closing vector*, and $e^{\boldsymbol{S}x} := \sum_{n=0}^{\infty}\frac{(\boldsymbol{S}x)^n}{n!}$ is the matrix exponential. The density function of Z is given by $f_Z(x) = \boldsymbol{\alpha}e^{\boldsymbol{S}x}\boldsymbol{s}$. The only restrictions on the parameters $(\boldsymbol{\alpha}, \boldsymbol{S}, \boldsymbol{s})$ are that $\boldsymbol{\alpha}e^{\boldsymbol{S}x}\boldsymbol{s}$ be a valid density function, i.e. $\boldsymbol{\alpha}e^{\boldsymbol{S}x}\boldsymbol{s} \geq 0$, for all $x \geq 0$ and $\lim_{x\to\infty} 1 - \boldsymbol{\alpha}e^{\boldsymbol{S}x}(-\boldsymbol{S})^{-1}\boldsymbol{s} = 1$. In general, in Bladt & Nielsen (2017), there is the possibility of an *atom* (a point mass) at 0, but here we do not consider this possibility. The condition that $\boldsymbol{\alpha}e^{\boldsymbol{S}x}\boldsymbol{s}$ be a valid density imposes some properties on representations $(\boldsymbol{\alpha}, \boldsymbol{S}, \boldsymbol{s})$. However, in general there is no way to determine whether, a given triplet $(\boldsymbol{\alpha}, \boldsymbol{S}, \boldsymbol{s})$ is a representation of a matrix exponential distribution, or not. Nonetheless, some properties of a triplet $(\boldsymbol{\alpha}, \boldsymbol{S}, \boldsymbol{s})$ are known, such as the following, which is used in the characterisation of QBD-RAPs.

Theorem 2.4 (Theorem 4.1.3, Bladt & Nielsen (2017)). *The density function of a matrix exponential distribution with representation* $(\boldsymbol{\alpha}, \boldsymbol{S}, \boldsymbol{s})$ *can be expressed in terms of real-valued constants as*

$$\begin{aligned}\psi(x) = &\sum_{j=1}^{m_1}\sum_{k=1}^{p_j} c_{jk}\frac{x^{k-1}}{(k-1)!}e^{\mu_j x} + \sum_{j=1}^{m+2}\sum_{k=1}^{q_j} d_{jk}\frac{x^{k-1}}{(k-1)!}e^{\eta_j x}\cos(\sigma_j x)\\ &+ \sum_{j=1}^{m_2}\sum_{k=1}^{q_j} e_{jk}\frac{x^{k-1}}{(k-1)!}e^{\eta_j x}\sin(\sigma_j x),\end{aligned} \tag{2.34}$$

for integers m_1, m_2, p_j, *and* q_j *and some real constants* c_{jk}, d_{jk}, e_{jk}, μ_j, η_j, *and* σ_j. *Here* μ_j, $j = 1, \ldots, m_1$ *are the real eigenvalues of* $\boldsymbol{S}$, *while* $\eta_j + i\sigma_j$, $\eta_j - i\sigma_j$, $j = 1, \ldots, m_2$ *denote its complex eigenvalues, which come in conjugate pairs. Thus,* $m_1 + 2m_2$ *is the total number of eigenvalues, while the dimension of the representation is given by* $p = \sum_{j=1}^{m+1} p_j + 2\sum_{j=1}^{m_2} q_j$.

Theorem 2.5 (Theorem 4.1.4, Bladt & Nielsen (2017)). *Consider the non-vanishing terms of the matrix exponential density (2.34),* i.e., *the terms for which* $c_{jk} \neq 0$, $d_{jk} \neq 0$, *or* $e_{jk} \neq 0$. *Among the corresponding eigenvalues* λ_j, *there is a real dominating eigenvalue* κ, *say. That is,* κ *is real,* $\kappa \geq Re(\lambda_j)$ *for all* j, *and the multiplicity of* κ *is at least the multiplicity of every other eigenvalue with real part* κ.

Corollary 2.6 (Corollary 4.1.5, Bladt & Nielsen (2017)). *If* $(\boldsymbol{\alpha}, \boldsymbol{S}, \boldsymbol{s})$ *is a representation for a matrix exponential distribution, then* $\boldsymbol{S}$ *has a real dominating eigenvalues.*

Theorem 2.7 (Theorem 4.1.6, Bladt & Nielsen (2017)). *Let* Z *be a matrix-exponentially distributed random variable with density (2.34). Then the dominant real eigenvalue* κ *of Theorem 2.5 is strictly negative.*

We define $dev(\boldsymbol{S})$ to be the real dominating eigenvalue of $\boldsymbol{S}$, that is $dev(\boldsymbol{S}) = \kappa$ in Theorem 2.5.

The class of matrix exponential distributions is characterised as the class of probability distributions which have a rational Laplace transform. That is, $\int_{x=0}^{\infty} e^{-\lambda x}\boldsymbol{\alpha}e^{\boldsymbol{S}x}\boldsymbol{s}\,\mathrm{d}x$ is a ratio of two polynomial functions in λ. Matrix exponential distributions are an extension of Phase-type distributions, where for the latter, $\boldsymbol{S}$ must be a sub-generator matrix of a CTMC, $\boldsymbol{s} = -\boldsymbol{S}\boldsymbol{e}$ where $\boldsymbol{e}$ is a $1 \times p$ vector of ones, and $\boldsymbol{\alpha}$ is a discrete probability distribution.

A *representation* of a matrix exponential distribution is a triplet $(\boldsymbol{\alpha}, \boldsymbol{S}, \boldsymbol{s})$, and we write $Z \sim ME(\boldsymbol{\alpha}, \boldsymbol{S}, \boldsymbol{s})$ to denote that Z has a matrix exponential distribution with this representation. The order of the representation is the dimension of the square matrix $\boldsymbol{S}$, i.e. if $\boldsymbol{S}$ is $p \times p$, then the matrix exponential distribution is said to be of order

p. Representations of matrix-exponential distributions are not unique (Bladt & Nielsen 2017). A representation is called *minimal* when $\boldsymbol{S}$ has the smallest possible dimension. Throughout this work, we assume that the representation of any matrix exponential distribution is minimal. Let $\boldsymbol{e}_i$ be a vector with a 1 in the ith position and zeros elsewhere. We assume that $\boldsymbol{s} = -\boldsymbol{S}\boldsymbol{e}$, and that $(\boldsymbol{e}_i, \boldsymbol{S}, \boldsymbol{s})$ for $i = 1, \ldots, p$ are representations of matrix exponential distributions. It is always possible to find such a representation (Bladt & Nielsen 2017, Theorem 4.5.17, Corollary 4.5.18). As such, we abbreviate our notation $Z \sim ME(\boldsymbol{\alpha}, \boldsymbol{S}, \boldsymbol{s})$ to $Z \sim ME(\boldsymbol{\alpha}, \boldsymbol{S})$. Further, given $\boldsymbol{s} = -\boldsymbol{S}\boldsymbol{e}$ then observe that $\int_{x=0}^{\infty} e^{\boldsymbol{S}x}\boldsymbol{s}\,\mathrm{d}x = (-\boldsymbol{S})^{-1}\boldsymbol{s} = \boldsymbol{e}$.

For a given $p \times p$ matrix $\boldsymbol{S}$, denote by $\mathcal{A} \subset \mathbb{R}^p$ the space of all possible vectors $\boldsymbol{a}$ such that $(\boldsymbol{a}, \boldsymbol{S})$ is a valid representation of a (possibility defective) matrix exponential distribution.

Concentrated matrix exponential distributions

Recently, the class of *concentrated matrix exponential distributions* (CMEs) has been investigated (Élteto et al. 2006, Horváth et al. 2016, Élteto et al. 2006, Horváth, Horváth & Telek 2020, Mészáros & Telek 2021). As the name suggests, concentrated matrix exponential distributions are matrix exponential distributions with a very low coefficient of variation (variance relative to the mean). As the order, p, of the representation increases, the variance of concentrated matrix exponential distributions decreases at rate approximately $\mathcal{O}(1/p^2)$. For comparison, the variance of an Erlang distribution, the most concentrated Phase-type distribution, decreases at rate $\mathcal{O}(1/p)$ as the order p increases (Aldous & Shepp 1987). In this thesis we use CMEs to approximate deterministic events. For a given order, concentrated matrix exponential distributions have a much lower coefficient of variation than any phase type of the same order, and therefore a better ability to model determinism.

The class of concentrated matrix exponential distributions (CMEs) found numerically in (Horváth, Horváth & Telek 2020) and computationally efficient expressions for the density and moments of CMEs are provided in (Horváth, Horváth & Telek 2020).

CMEs exist for all orders, however, in our computations we only use CMEs with odd order. The justification for considering representations of odd orders only is that the variance of CME distributions of orders $2p$ and $2p-1$ are relatively similar and therefore have similar abilities to represent the delta function (Horváth, Horváth & Telek 2020). Hence, if we construct an approximation with a representation of order $2p$ we expect it to perform only marginally better than an approximation constructed with representations of order $2p-1$. However, the computational cost of the latter is lower, so we opt for the order $2p-1$ representation.

From the definition of the class of CME distributions, for a given CME with odd order, p, and representation $(\boldsymbol{\alpha}, \boldsymbol{S})$, the matrix $\boldsymbol{S}$ has one real eigenvalue, and $p-1$ complex

eigenvalues and all eigenvalues have the same real part Horváth, Horváth & Telek (2020).

For CMEs considered here, the vector function $\boldsymbol{k}(t) = \dfrac{\boldsymbol{\alpha}e^{St}}{\boldsymbol{\alpha}e^{St}\boldsymbol{e}}$ is periodic with period $\rho = 2\pi/\omega$ where $\omega = min_i(|\Im(\lambda_i)|)$, λ_i are the eigenvalues of $\boldsymbol{S}$ and $\Im(z)$ is the imaginary component of a complex number z. We leverage this fact to simplify some numerical integration in Chapter 4.

2.5.2 QBD-RAPs

To define a QBD-RAP we first need a Batch (Marked) rational-arrival-process (RAP). Let $\mathscr{K} \subset \mathbb{Z}$ be a set of *marks*. Let N be a point process, and $Y_0 = 0 < Y_1 < Y_2 \cdots$ be event times of N. Let $\{N(t)\}$ be the counting process associated with N such that $N(t)$ returns the number of events by time t. Associated with the nth event is a mark M_n. For $i \in \mathscr{K}$, let N_i be simple point processes associated with events with marks of type i only, and let $\{N_i(t)\}_{t\geq 0}$ be the associated counting processes of events of mark i. Denote by $f_{N,n}(y_1, m_1, y_2, m_2, \ldots, y_n, m_n)$ the joint density, probability mass function of the first n inter-arrival times, $Y_1, Y_2 - Y_1, \ldots, Y_n - Y_{n-1}$, and the associated marks M_n. From Bean and Nielsen (Bean & Nielsen 2010, Theorem 1) we have the following.

Theorem 2.8. *A process N is a Marked RAP if there exist matrices $\boldsymbol{S}$, $\boldsymbol{D}_i$, $i \in \mathscr{K}$, and a row vector $\boldsymbol{\alpha}$ such that $dev(\boldsymbol{S}) < 0$, $dev(\boldsymbol{S}+\boldsymbol{D}) = 0$, $(\boldsymbol{S}+\boldsymbol{D})\boldsymbol{e} = 0$, $\boldsymbol{D} = \sum_{i\in\mathscr{K}} \boldsymbol{D}_i$, and*

$$f_{N,n}(y_1, m_1, y_2, m_2, \ldots, y_n, m_n) = \boldsymbol{\alpha}e^{\boldsymbol{S}y_1}\boldsymbol{D}_{m_1}e^{\boldsymbol{S}y_2}\boldsymbol{D}_{m_2}\ldots e^{\boldsymbol{S}y_n}\boldsymbol{D}_{m_n}\boldsymbol{e}. \tag{2.35}$$

Conversely, if a point process has the property (2.35) then it is a Marked RAP.

Denote such a process $N \sim BRAP(\boldsymbol{\alpha}, \boldsymbol{S}, \boldsymbol{D}_i,\ i \in \mathscr{K})$.

Also, from Bean & Nielsen (2010), associated with a Marked RAP is a row-vector-valued *orbit* process, $\{\boldsymbol{A}(t)\}_{t\geq 0}$,

$$\boldsymbol{A}(t) = \frac{\boldsymbol{\alpha}\left(\prod_{i=1}^{N(t)} e^{\boldsymbol{S}(Y_i - Y_{i-1})}\boldsymbol{D}_{M_i}\right)e^{\boldsymbol{S}(t-Y_{N(t)})}}{\boldsymbol{\alpha}\left(\prod_{i=1}^{N(t)} e^{\boldsymbol{S}(Y_i - Y_{i-1})}\boldsymbol{D}_{M_i}\right)e^{\boldsymbol{S}(t-Y_{N(t)})}\boldsymbol{e}}.$$

Thus, $\{\boldsymbol{A}(t)\}$ is a piecewise-deterministic Markov process where, in between events $\{\boldsymbol{A}(t)\}$ evolves deterministically according to

$$\boldsymbol{A}(t) = \frac{\boldsymbol{A}(Y^-_{N(t)})e^{\boldsymbol{S}(t-Y_{N(t)})}}{\boldsymbol{A}(Y^-_{N(t)})e^{\boldsymbol{S}(t-Y_{N(t)})}\boldsymbol{e}},$$

where $\boldsymbol{A}(Y_{N(t)}^{-}) = \lim_{u\to 0^+} \boldsymbol{A}(Y_{N(t)} - u)$. The process $\{\boldsymbol{A}(t)\}$ can jump at event times of N (the process may not actually jump at these times, but we may still refer to it a 'jump' and typically the dynamics change at this point). At time t the intensity with which $\{\boldsymbol{A}(t)\}$ has a jump is $\boldsymbol{A}(t)\boldsymbol{D}\boldsymbol{e}$, i.e. $\mathbb{P}(N(t) = n, N(t+dt) = n+1) = \boldsymbol{A}(t)\boldsymbol{D}\boldsymbol{e}\,\mathrm{d}t$. Upon an event the event is associated with mark i with probability $\boldsymbol{A}(t)\boldsymbol{D}_i\boldsymbol{e}/\boldsymbol{A}(t)\boldsymbol{D}\boldsymbol{e}$. Upon on an event at time t with mark i, the new position of the orbit is $\boldsymbol{A}(t) = \boldsymbol{A}(t^-)\boldsymbol{D}_i/\boldsymbol{A}(t^-)\boldsymbol{D}_i\boldsymbol{e}$. It is important to note that the jumps of the orbit process are *linear* transformations of the orbit process immediately before the time of the jump.

Marked RAPs are an extension of Marked Markovian arrival processes to include matrix-exponential inter-arrival times. For Marked MAPs, the vector $\boldsymbol{A}(t)$ is a vector of posterior probabilities of a continuous-time Markov chain.

Intuitively, $\boldsymbol{A}(t)$ encodes all the information about the event times of the Marked RAP and associated marks up to time t that is needed to determine the future behaviour of the point process. Let $\mathcal{F}_t$ be the σ-algebra generated by $N(u), u \in [0,t]$. Then $N \mid \mathcal{F}_t \equiv N \mid \boldsymbol{A}(t) \sim BRAP(\boldsymbol{A}(t), \boldsymbol{S}, \boldsymbol{D}_i, i \in \mathscr{K})$. In words, the future of the point process after time t given all the information about the process up to and including time t, is distributed as a Marked RAP with initial vector $\boldsymbol{A}(t)$.

Now consider a Marked RAP, $N \sim BRAP(\boldsymbol{\alpha}, \boldsymbol{S}, \boldsymbol{D}_i,\, i \in \{-1, 0, +1\})$. The process $\{(L(t), \boldsymbol{A}(t))\}_{t\geq 0}$ formed by letting $L(t) = N_{+1}(t) - N_{-1}(t)$ is a QBD-RAP. QBDs are QBD-RAPs where the inter-event times are of Phase-type.

2.6 Laplace Transforms

In Chapters 5 and 6 we work with an object called the *Laplace transform*. For a measure μ, defined on the Borel sets of $[0,\infty)$, we define the Laplace transform of μ to be

$$\widehat{\mu}(\lambda) = \mathcal{L}(\mu)(\lambda) = \int_{t=0}^{\infty} e^{-\lambda t}\,\mathrm{d}\mu,$$

where the *region of convergence* is the set of values of $\lambda \in \mathbb{R}$ such that the integral is finite. When μ has a density, v, then the Laplace transform is

$$\widehat{\mu}(\lambda) = \widehat{v}(\lambda) = \mathcal{L}(v)(\lambda) = \int_{t=0}^{\infty} e^{-\lambda t} v(x)\,\mathrm{d}x.$$

When μ is the probability measure associated with a random variable, Z, say, then we may write

$$\widehat{\mu}(\lambda) = \mathbb{E}[e^{-\lambda Z}],$$

and the region of convergence is at least $[0,\infty)$. Further, letting E^λ be an exponentially distributed random variable with rate λ and noting that $\mathbb{P}(E^\lambda > t) = e^{-\lambda t}$, then

$$\widehat{\mu}(\lambda) = \mathbb{P}(Z < E^\lambda),$$

which gives a probabilistic interpretation of the Laplace transform.

A convenient property of the Laplace transform which we utilise is the Convolution Theorem.

Theorem 2.9 (Convolution Theorem)**.** *Let $f, g : [0, \infty) \to \mathbb{R}$ be integrable functions, then*

$$\mathcal{L}\left(\int_{u=0}^{t} f(u)g(t-u)\,\mathrm{d}u\right)(\lambda) = \mathcal{L}(f) \cdot \mathcal{L}(g).$$

The Convolution Theorem states that the Laplace transform of the convolution, given by $\int_{u=0}^{t} f(u)g(t-u)\,\mathrm{d}u$, is equal to the product of the Laplace transform of f and g.

The Laplace transform is unique in the sense that, if μ and ν are two measures on the Borel sets of $[0, \infty)$ and

$$\widehat{\mu}(\lambda) = \widehat{\nu}(\lambda)$$

for all $\lambda > a$ with $a < \infty$, then μ and ν are the same. In terms of functions, $f, g : [0, \infty) \to \mathbb{R}$, if

$$\mathcal{L}(f)(\lambda) = \mathcal{L}(g)(\lambda),$$

for all $\lambda > a$ with $a < \infty$, and f and g are continuous, then $f(t) = g(t)$ for all $t \geq 0$. Without knowing f and g are continuous, then we can only claim that $f(t) = g(t)$ for all $t \geq 0, t \notin \mathcal{N}$, where $\mathcal{N}$ is a *null set* with respect to Lebesgue measure.

2.7 Convergence theorems

We use the following convergence theorems to help us prove that the QBD-RAP scheme converges weakly to the distribution of the fluid queue. The first result we state, the Portmanteau Theorem, is a sweeping statement about convergence of measures. First, let's define the notion of weak convergence. We follow Klenke (2014).

Denote by $\mathcal{M}_f(E)$ the set of all finite measures on $(E, \mathcal{E})$, where E is a non-empty set and $\mathcal{E}$ is a σ-algebra. Further, denote by $C_b(E)$ the set of continuous bounded functions on E. Let μ, μ_1, $\mu_2, ... \in \mathcal{M}_f(E)$. We say that $\{\mu_n\}_{n\in\mathbb{N}}$ converges weakly to μ, formally $\mu_n \to \mu$ weakly as $n \to \infty$, if

$$\int f\,\mathrm{d}\mu_n \to \int f\,\mathrm{d}\mu, \text{ for all } f \in C_b(E).$$

We use part of The Portmanteau Theorem in Chapter 6. Let $\mathcal{M}_{\leq 1}(E) := \{\mu \in \mathcal{M}_f(E) \mid \mu(E) \leq 1\}$, the set of all sub-probability measures on $(E, \mathcal{E})$.

Theorem 2.10 ([Part of)**.** *The Portmanteau Theorem, Theorem 13.16 of Klenke (2014)] Let E be a metric space and let μ, μ_1, $\mu_2, ... \in \mathcal{M}_{\leq 1}(E)$. The following are equivalent.*

(i) $\mu_n \to \mu$ *weakly as* $n \to \infty$.

(ii) $\int f \, \mathrm{d}\mu_n \to \int f \, \mathrm{d}\mu$ *for all bounded, Lipschitz continuous* f.

$\vdots$

There are 8 parts to The Portmanteau Theorem in Klenke (2014), Theorem 13.16. Here we only quote to relevant parts; the other 6 parts require us to define other concepts which are not relevant to this thesis, so they are omitted. See Klenke (2014), Theorem 13.16 for details.

Another tool we can use to show convergence of measures is to show that the Laplace transforms converge, as stated in the following theorem.

Theorem 2.11 (Extended Continuity Theorem, Feller (1957), Theorem 2a)**.** *For* $p = 1, 2, ...,$ *let* U_p *be a measure with Laplace transform* ζ_p*. If* $\zeta_p(\lambda) \to \zeta(\lambda)$ *for* $\lambda > a \geq 0$*, then* ζ *is the Laplace transform of a measure* U *and* $U_p \to U$ *[weakly].*

Conversely, if $U_p \to U$ *[weakly] and the sequence* $\{\zeta_p(a)\}$ *is bounded, then* $\zeta_p(\lambda) \to \zeta(\lambda)$ *for* $\lambda > a$*.*

In Chapters 5 and 6 we use the Dominated Convergence Theorem to aid our convergence arguments. In applied probability we often want to prove convergence of certain expressions. An approach which can simplify matters is to partition the expression on certain events where we have a simpler characterisation, thereby enabling us to prove convergence on each element in the partition. The original expression can be written as an integral over the partition. To establish the convergence result we initially desired, we can use the convergence of each element of the partition and the Dominated Convergence Theorem. I have taken the following from Theorem 1.13 Stein & Shakarchi (2009). Stein & Shakarchi (2009) use the notation

$$\int f = \int f \, \mathrm{d}x = \int f \, \mathrm{d}m(x),$$

where m denotes Lebesgue measure, to denote the Lebesgue integral.

Theorem 2.12 (Dominated Convergence Theorem)**.** *Suppose* $\{f_n\}$ *is a sequence of measurable functions such that* $f_n(x) \to f(x)$ *almost everywhere with respect to* x*, as* n *tends to infinity. If* $|f_n(x)| \leq g(x)$*, where* g *is integrable, then*

$$\int |f_n - f| \to 0 \text{ as } n \to \infty,$$

and consequently

$$\int f_n \to \int f \text{ as } n \to \infty.$$

Also in Chapters 5 and 6 we want to manipulate infinite sums or integrals and rearrange the order of summation or integration. However, things can go awry when we swap the order of integration/summation if we are not careful. The next few results give some conditions under which we have equality under before and after swapping the order of summation/integration. Once again, we follow Stein & Shakarchi (2009) quoting their Theorem 2.13. If f is a function in $\mathbb{R}^d = \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$, the *slice* of f corresponding to $y \in \mathbb{R}^{d_2}$ is the function f^y of the $x \in \mathbb{R}^{d_1}$ variable, given by

$$f^y(x) = f(x, y).$$

Similarly, the slice of f for a fixed $x \in \mathbb{R}^{d_1}$ is $f_x(y) = f(x, y)$.

Theorem 2.13 (Fubini's Theorem)**.** *Suppose $f(x, y)$ is integrable on $\mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$. Then for almost every $y \in \mathbb{R}^{d_2}$:*

- *The slice f^y is integrable on $\mathbb{R}^{d_1}$.*
- *The function defined by $\int_{R^{d_1}} f^y(x)\,\mathrm{d}x$ is integrable on $\mathbb{R}^{d_2}$.*

Moreover:

(iii) $$\int_{R^{d_2}} \left(\int_{R^{d_1}} f(x, y)\,\mathrm{d}x \right) \mathrm{d}y = \int_{\mathbb{R}^d} f.$$

Stein & Shakarchi (2009) then state

> "Clearly, the [Fubini] theorem is symmetric in x and y so that we also may conclude that the slice f_x is integrable on $\mathbb{R}^{d_2}$ for almost every x. Moreover, $\int_{\mathbb{R}^{d_2}} f_x(y)\,\mathrm{d}y$ is integrable and
>
> $$\int_{R^{d_1}} \left(\int_{R^{d_2}} f(x, y)\,\mathrm{d}y \right) \mathrm{d}x = \int_{\mathbb{R}^d} f.$$
>
> In particular, Fubini's theorem states that the integral of f on $\mathbb{R}^d$ can be computed by iterating lower-dimensional integrals, and that the iterations can be taken in any order
>
> $$\int_{R^{d_2}} \left(\int_{R^{d_1}} f(x, y)\,\mathrm{d}x \right) \mathrm{d}y = \int_{R^{d_1}} \left(\int_{R^{d_2}} f(x, y)\,\mathrm{d}y \right) \mathrm{d}x = \int_{\mathbb{R}^d} f."$$

It is this last statement which is most powerful. Effectively, if either

$$\int_{R^{d_2}} \left(\int_{R^{d_1}} |f(x, y)|\,\mathrm{d}x \right) \mathrm{d}y < \infty,$$

or

$$\int_{R^{d_1}} \left(\int_{R^{d_2}} |f(x,y)| \, \mathrm{d}y \right) \mathrm{d}x < \infty,$$

then we can swap the order of integration.

A closely related theorem which is often used alongside Fubini's Theorem is Tonelli's Theorem. Define the *extended Lebesgue integral* of an extended valued (it can take the value $+\infty$) non-negative function f by

$$\int f(x) \, \mathrm{d}x = \sup_g \int g(x) \, \mathrm{d}x.$$

Theorem 2.14 (Tonelli's Theorem)**.** *Suppose $f(x,y)$ is a non-negative measurable function on $\mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$. Then for almost every $y \in \mathbb{R}^{d_2}$:*

- *The slice f^y is integrable on $\mathbb{R}^{d_1}$.*
- *The function defined by $\int_{R^{d_1}} f^y(x) \, \mathrm{d}x$ is integrable on $\mathbb{R}^{d_2}$.*

Moreover:

(iii) $\int_{R^{d_2}} \left(\int_{R^{d_1}} f(x,y) \, \mathrm{d}x \right) \mathrm{d}y = \int_{\mathbb{R}^d} f(x,y) \, \mathrm{d}x \, \mathrm{d}y$ *in the extended sense.*

Once again, we note that the theorem is symmetric in x and y, so we can establish that we may swap the order of integration provided that f is non-negative. Tonelli's Theorem allows us to exchange the order of integration for any non-negative extended real-valued function, and includes the case where the value of the integral is infinite. In contrast, Fubini's Theorem us to exchange the order of integration for any real-valued function, provided that the integral is finite.

Collectively, we refer to Theorems 2.13 and 2.14 together as the Fubini-Tonelli Theorem, but they are otherwise known collectively as just Fubini's Theorem. Often, they are used in conjunction. Since $|f|$ is a non-negative function, then we may use Tonelli's Theorem and compute (or bound) the integral $\int |f|$ via computing an iterated integral. If this is found to be finite, then Fubini's Theorem applies so f is integrable, and we may evaluate $\int f$ via an iterated integral.

In the context of probability, the function f which we are integrating is often positive, so Tonelli's Theorem is all that is required to justify a swap of iterated integrals.

2.8 Discontinuous Galerkin

The discontinuous Galerkin (DG) method is a general methodology for numerically approximating solutions to differential equations. The discontinuous Galerkin method is a finite-element method which approximates differential operators by finite dimensional

matrices. In this thesis we use the discontinuous Galerkin method to approximate the infinitesimal generator of a fluid queue. Here, we give a very brief introduction to discontinuous Galerkin methods in general and refer the reader to Chapter 3 for details on the application to fluid queues, and to Hesthaven & Warburton (2007) for more details on the method in general.

Intuitively, the DG method works by first partitioning the spatial domain into cells, on each cell a basis of (typically polynomial) functions is defined, then the differential equation is projected onto the basis of polynomials. Thus, the differential equation is discretised it two ways, by the partition of space into cells, then by the projection onto a finite dimensional basis. Linking the cells together essentially amounts to defining boundary conditions for each cell. To form the boundary condition for cells adjacent to the boundary we take the boundary condition defined by the problem we are solving and discretise it. For cells in the interior the boundary condition is defined by the solution on the adjacent cell. In the context of conservation laws (which is what we solve in this thesis), the boundary conditions on each cell are known as the *flux*, as they describe the rate, with respect to time, that the conserved quantity flows across the boundaries of the cell. The flux is discretised and is known as the *numerical flux*.

Typically, polynomial functions are used as the basis for the DG method. The main choices to make in defining the scheme are the grid which defines the cells, the order of the polynomial basis, and the choice of numerical flux.

Hesthaven & Warburton (2007) claim that the first discontinuous Galerkin finite element scheme was developed in Reed & Hill (1973) where they use the method to solve the first order differential equation

$$\sigma u + \nabla \cdot (au) = f,$$

in a 2-dimensional spatial domain, where σ is a constant, a is a piecewise constant function, and u is the unknown. The literature has since expanded and includes (but is not limited to) theory on the convergence rate of the scheme, the choice of *numerical flux* used (we will define this later), the choice of spatial discretisation used, and applications to a range of other problems including higher-order differential equations and to partial differential equations (PDEs) such as non-linear conservation law. The PDE which describes the evolution of the distribution of a fluid queue is a conservation law as probability must be conserved.

In one dimension, linear conservation laws take the form

$$\begin{aligned} &\frac{\partial u(x,t)}{\partial t} - \frac{\partial}{\partial x}(c(t)u(x,t)) = 0,\ x \in \Omega, && (2.36)\\ &u(x,t) = g(x,t), && \text{(boundary condition)}\\ &u(x,0) = f(x). && \text{(initial condition)} \end{aligned}$$

where Ω is the region of interest. Here u is the conserved quantity, which we will refer to as *mass*, but it could really be any conserved quantity.

To solve conservation laws with the DG method the spatial domain Ω is partitioned into cells (intervals) $\mathcal{D}_k = [y_k, y_{k+1}]$, $k = 0, ..., K-1$. For each cell $\mathcal{D}_k$ we choose p_k linearly independent functions $\{\phi_k^r\}_{r=1}^{p_k}$, compactly supported on $\mathcal{D}_k$ (i.e. $\phi_k^r(x) = 0$ for $x \notin \mathcal{D}_k$). Typically, we take $\phi_k^r(x)$ to be a basis of polynomial functions. The functions $\{\phi_k^r\}_{r=1}^{p_k}$ span some function space, W_k say, and further form a basis for this space. It is convenient in this work to take $\{\phi_k^r\}_{r=1}^{p_k}$ as a basis of *Lagrange interpolating polynomials* defined by the *Gauss-Lobatto quadrature points* (Hesthaven & Warburton 2007) (see also Section 2.9). For the sake of illustration, the reader may think of $\{\phi_k^r\}_{r=1}^{p_k}$ as the Lagrange polynomials. On each cell $\mathcal{D}_k$ we approximate

$$u(x,t) \approx u_k(x,t) = \sum_{r=1}^{p_k} a_k^r(t)\phi_k^r(x),$$

where $a_k^r(t)$ are yet-to-be-determined time-dependent coefficients. The approximation u_k is referred to as the *local* approximation on cell k, while the *global* approximation is given by $\sum_{k=0}^{K-1} u_k$ on the whole domain Ω. The whole approximation space is $\bigoplus_{k\in\mathbb{Z}} W_k$, the direct sum of the spaces W_k, $k \in \{0, ..., K-1\}$.

First, consider multiplying (2.36) by the row vector $\boldsymbol{\phi}_k(x) = [\phi_k^1(x), ..., \phi_k^{p_k}(x)]$ and integrating by parts over $\mathcal{D}_k$,

$$\int_{x\in\mathcal{D}_k} \frac{\partial u(x,t)}{\partial t}\boldsymbol{\phi}_k(x)\,\mathrm{d}x + \int_{x\in\mathcal{D}_k} c(t)u(x,t)\frac{\partial}{\partial x}\boldsymbol{\phi}_k(x)\,\mathrm{d}x - c(t)[u(x,t)\boldsymbol{\phi}_k(x)]_{x=y_k}^{y_{k+1}} = 0. \tag{2.37}$$

For a sufficiently large basis of functions, then (2.37) is equivalent to the PDE (2.36). The term $-c(t)[u(x,t)\boldsymbol{\phi}_k(x)]_{x=y_k}^{y_{k+1}}$ is known as the *flux* as it describes the rate at which mass flows into and out of the cell $\mathcal{D}_k$. The flux is essentially a boundary condition on the cell $\mathcal{D}_k$ and is specified by the value of the solution on the adjoining cells. When $c(t)$ is positive (negative) then mass flows into (out of) cell $\mathcal{D}_k$ at the left-hand boundary and out of (into) cell $\mathcal{D}_k$ at the right-hand boundary. Suppose that $c(t)$ is positive. When u is continuous, then $u(y_k^-, t) = u(y_k, t)$, where y^- is the left limit at y, hence the flux into cell $\mathcal{D}_k$ in Equation (2.36) is $c(t)u(y_k^-, t)\boldsymbol{\phi}_k(y_k)$. The flux out of cell $\mathcal{D}_k$ is $c(t)u(y_{k+1}^-, t)\boldsymbol{\phi}_k(y_{k+1})$. Using this, rewrite (2.37)

$$\begin{aligned}\int_{x\in\mathcal{D}_k} \frac{\partial u(x,t)}{\partial t}\boldsymbol{\phi}_k(x)\,\mathrm{d}x + \int_{x\in\mathcal{D}_k} c(t)u(x,t)\frac{\partial}{\partial x}\boldsymbol{\phi}_k(x)\,\mathrm{d}x \\ - c(t)[u(y_{k+1}^-, t)\boldsymbol{\phi}_k(y_{k+1}) - u(y_k^-, t)\boldsymbol{\phi}_k(y_k)] = 0.\end{aligned} \tag{2.38}$$

We can proceed analogously when $c(t)$ is negative.

Upon substituting the approximations u_k and u_{k-1} into (2.38)

$$\int_{x\in\mathcal{D}_k} \frac{\partial \boldsymbol{a}_k(t)}{\partial t}\boldsymbol{\phi}_k(x)'\boldsymbol{\phi}_k(x)\,\mathrm{d}x + \int_{x\in\mathcal{D}_k} c(t)\boldsymbol{a}_k(t)\boldsymbol{\phi}_k(x)'\frac{\partial}{\partial x}\boldsymbol{\phi}_k(x)\,\mathrm{d}x$$

$$
\begin{aligned}
&\quad - c(t)[\boldsymbol{a}_k(t)\boldsymbol{\phi}_k(y_{k+1}^-)'\boldsymbol{\phi}_k(y_{k+1}) - \boldsymbol{a}_{k-1}(t)\boldsymbol{\phi}_{k-1}(y_k^-)'\boldsymbol{\phi}_k(y_k)].\\
&= \frac{\mathrm{d}\boldsymbol{a}_k(t)}{\mathrm{d}t}\boldsymbol{M}_k + c(t)\boldsymbol{a}_k(t)\boldsymbol{G}_k - c(t)[\boldsymbol{a}_k(t)\boldsymbol{F}_{k,k} - \boldsymbol{a}_{k-1}(t)\boldsymbol{F}_{k-1,k}]\\
&= 0, \qquad (2.39)
\end{aligned}
$$

where

$$\boldsymbol{M}_k := \int_{x\in\mathcal{D}_k} \boldsymbol{\phi}_k(x)'\boldsymbol{\phi}_k(x)\,\mathrm{d}x, \quad \boldsymbol{G}_k := \int_{x\in\mathcal{D}_k} \boldsymbol{\phi}_k(x)'\frac{\mathrm{d}}{\mathrm{d}x}\boldsymbol{\phi}_k(x)\,\mathrm{d}x,$$

$$\boldsymbol{F}_{k,k} := \boldsymbol{\phi}_k(y_{k+1}^-)'\boldsymbol{\phi}_k(y_{k+1}), \quad \boldsymbol{F}_{k-1,k} := \boldsymbol{\phi}_{k-1}(y_k^-)'\boldsymbol{\phi}_k(y_k),$$

and are known as the *mass*, *stiffness* and *flux* matrices. Thus, the approximation has reduced the problem to a system of ordinary differential equations. One can determine the coefficients $\boldsymbol{a}_k(t)$, $k = 0, ..., K-1$, from the initial condition $\boldsymbol{a}_k(0)$, $k = 0, ..., K-1$, then integrating over time, perhaps numerically. The initial coefficients $\boldsymbol{a}_k(0)$, $k = 0, ..., K-1$, can be found by projecting the initial condition onto the basis of polynomial functions, $\{\phi_k^r\}_{k,r}$.

One subtlety is the choice of numerical flux. We arrived at the numerical flux above using a continuity argument. However, the approximate solution $\sum_{k=0}^{K-1} u_k(x,t)$ is discontinuous at the points y_k. It turns out that the choice of numerical flux is not unique and there are range of values we may choose from. The numerical flux above is known as the *upwind* flux – we use it throughout this thesis and do not explore others. Other choices which define a discontinuous Galerkin scheme are the basis functions used and the cell geometries. We refer the reader to (Hesthaven & Warburton 2007) and reference therein for more details.

Typically, DG schemes are used to solve physical problems in 2 or 3 dimensions. The cell geometries become more complex in higher dimensions and this complicates matters somewhat, but the underlying principles and intuition is the same. In this thesis we solve problems with one spatial dimension only.

Two properties make the DG scheme a natural choice to solve the problem present in this thesis. High-order accuracy enabled by the use of basis functions to model the solution on each cell, and the cell-based structure of the method with cells related to each other at cell edges only and only to neighbouring cells. Finite difference methods can achieve high-order accuracy, however they are not cell-based, which makes them not a natural choice to solve the problems in this thesis. Finite volume methods are cell-based, but do not have the same high-order accuracy as the DG method. Further, approximating flux terms with greater accuracy in the finite volume method can introduce dependencies between cells which are not neighbours and this creates difficulties for the application in this thesis.

2.8.1 Time-integration schemes

In this thesis we numerically integrate ODEs of the form

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{a}(t) = \boldsymbol{a}(t)\boldsymbol{Q} \tag{2.40}$$

where $\boldsymbol{a}(t)$ is a vector of coefficients and $\boldsymbol{Q}$ a matrix, given an initial condition $\boldsymbol{a}(0)$. To do so, we employ the strong stability preserving Runge-Kutta (SSPRK) scheme of order 4 with 5 stages (Spiteri & Ruuth 2002). SSPRK methods are a family of Runge-Kutta methods which, to quote Section 5.7 of Hesthaven & Warburton (2007), "...can be used with advantage for problems with strong shocks and discontinuities, as they guarantee that no additional oscillations are introduced as part of the time-integration process."

In our context, the SSPRK method with s stages has the form

$$\begin{aligned}
&\boldsymbol{v}^{(0)} = \boldsymbol{a}(t),\\
&\boldsymbol{v}^{(\ell)} = \sum_{k=0}^{\ell-1} \alpha_{\ell k}\boldsymbol{v}^{(k)} + h\beta_{\ell k}\boldsymbol{v}^{(k)}\boldsymbol{Q},\ \ell = 1,\ldots,s,\\
&\boldsymbol{a}(t+h) = \boldsymbol{v}^{(s)},
\end{aligned}$$

where $\alpha_{\ell k}$ and $\beta_{\ell k}$ are coefficients which define the scheme and h is the t-step size. To ensure the stability-preserving property the coefficients $\alpha_{\ell k}$ and $\beta_{\ell k}$ are chosen to be positive so that $\boldsymbol{v}^{(s)}$ is a convex combination of forward-Euler steps. Moreover, the coefficients $\alpha_{\ell k}$ and $\beta_{\ell k}$ are optimised so that the allowable t-step size is as large as possible. The maximum allowable t-step size is h_{RK}, and we require

$$h_{RK} \leq \min_{\ell k} \frac{\alpha_{\ell k}}{\beta_{\ell k}} h_E,$$

where h_E is the maximum allowable t-step size for the forward-Euler scheme.

The t-step size needs to be chosen to be sufficiently small so that the time integration is stable. Here, for an order p DG scheme we choose the time step to be less than

$$\max_{i\in\mathcal{S}} |c_i| \min_{r\in 2,\ldots,p+1} (z_r - z_{r-1}) \frac{\min_k \Delta_k}{2}$$

where $z_{r-1} \leq z_r$, $z_1 = -1$, $z_{p+1} = 1$ and $z_r, r = 2, ..., p$, are the $p-1$ zeros of the first derivative of $P_p(x)$, the order p Legendre polynomial. This follows advice from Hesthaven & Warburton (2007).

2.8.2 Slope limiters

A well-know problem with DG schemes is that, in the presence of steep gradients or discontinuities, spurious oscillations in the approximate solution can occur and positivity is

not guaranteed (see Hesthaven & Warburton (2007) Section 5.6, and Koltai (2011) Section 3.3, for example). To rectify this, in some contexts, slope limiting can be used (see Cockburn (1999), or Hesthaven & Warburton (2007), Section 5.6.2 and references therein). Slope limiting alters the DG approximation in regions where oscillations are detected to ensure non-oscillatory by reducing the order of the approximation to linear and limiting the gradient of the approximation in these regions. However, limiting does not distinguish between natural oscillations, which are a fundamental feature of the solution, and spurious oscillations, which are caused by the approximation scheme. As a result, the limiter may unnecessarily reduce the accuracy of the scheme in the presence of natural oscillations (see (Hesthaven & Warburton 2007), Example 5.8). Furthermore, slope limiting adds an extra computational cost on top of the approximation scheme and means that the resultant approximation to certain operators are no longer linear.

We now describe a simple slope limiter known as the *Generalised MUSCL* scheme (Cockburn 1999) (see also (Hesthaven & Warburton 2007), Section 5.6.1). Define the *minmod* function

$$m(a_1, a_2, a_3) = \begin{cases} s \min_{i\in\{1,2,3\}} |a_i|, & |s| = 1, \\ 0, & \text{otherwise}, \end{cases}$$

where $s = \frac{1}{3}\sum_{i=1}^{3} \text{sign}(a_i)$. When all three arguments, a_1, a_2 and a_3 have the same sign, the minmod function returns the smallest of the three arguments with the correct sign, otherwise the signs of the three arguments differ and the minmod function returns zero. In the context of DG the three arguments a_1, a_2 and a_3 are the estimates of the average gradient of the approximate solution near a given cell. When the signs of a_1, a_2 and a_3 differ, this suggests an oscillation in the approximate solution.

Now, define $\overline{u}_k$ as the average value of the approximate solution on cell k, then the slope limited solution on cell k is the linear approximant

$$\Pi_k^{lim}(u_k) = \overline{u}_k + (x - \overline{y}_k) m\left(\Pi^1_{\partial x} u_k, \frac{\overline{u}_{k+1} - \overline{u}_k}{\Delta_k}, \frac{\overline{u}_k - \overline{u}_{k-1}}{\Delta_k}\right),$$

where $\Pi^1_{\partial x} u_k$ is the slope of the linear projection of u_k on cell k and $\overline{y}_k = (y_k + y_{k+1})/2$ is the centre of the kth cell.

In non-oscillatory regions of the solution the slope limiter is unnecessary and reduces accuracy, so we only apply the limiter to cells k where oscillations are detected. To determine which cells need limiting we compute

$$\begin{aligned} v_k^L &:= \overline{u}_k - m(\overline{u}_k - u_k^L, \overline{u}_k - \overline{u}_{k-1}, \overline{u}_{k+1} - \overline{u}_k), \\ v_k^R &:= \overline{u}_k + m(\overline{u}_k - u_k^R, \overline{u}_k - \overline{u}_{k-1}, \overline{u}_{k+1} - \overline{u}_k), \end{aligned}$$

where u_k^L and u_k^R are the values of the approximate solution on cell k evaluated at the left-hand and right-hand edges of the cell. We apply the slope limiter to cell k if $v_k^L \neq u_k^L$ or $v_k^R \neq u_k^R$, otherwise the approximation on cell k is left unaltered.

2.8.3 Slope limiters and time integration

To use slope limiters within the SSPRK scheme we first project the initial condition on to W, the set of polynomial basis functions, and apply the slope limiter to the projection to determine the initial coefficients $\boldsymbol{a}(0)$. To find the coefficients for the approximate solution to the transient distribution at time t_0 we take the (limited) initial condition $\boldsymbol{a}(0)$ and evolve it forward in time until t_0 via numerical integration of the differential equation (2.40). At each stage of the time-integration we apply the slope limiter, i.e. the scheme above with a slope limiter is

$$\begin{aligned}
&\boldsymbol{v}^{(0)} = \boldsymbol{a}(t), \\
&\boldsymbol{v}^{(\ell)} = \Pi^{lim}\left(\sum_{k=0}^{\ell-1} \alpha_{\ell k}\boldsymbol{v}^{(k)} + h\beta_{\ell k}\boldsymbol{v}^{(k)}\boldsymbol{Q}\right), \ \ell = 1, \ldots, s, \\
&\boldsymbol{a}(t+h) = \boldsymbol{v}^{(s)},
\end{aligned}$$

where Π^{lim} is the slope limiter function which determines on which cells the solution needs limiting, and if so, applies the limiter Π_k^{lim}.

We refer to this (vanilla) application of a limiter to the DG scheme as the DG-lim scheme.

Consequences of slope limiting We have already mentioned that the Generalised MUSCL slope limiter reduces the order to linear in regions where oscillatory solutions are detected. When slope limiting is used in conjunction with a time-integration scheme the reduction in order may not be local in time and may persist past the time when the limiter first detected oscillations. For example, consider a problem with an initial condition which introduces oscillations into the numerical approximation. If the slope limiter is applied to the initial condition then, in the region around the oscillations, the approximation to the initial condition will be linear. Even if the slope limiter is not applied (or not needed) after this point, the initial condition has still been altered, perhaps significantly, from the original approximation and this can affect the accuracy of transient approximations. This is not to say that the slope-limited-regions of the approximation will *always* remain linear. When no oscillations are detected, the DG scheme can use the full power of the high-order polynomial basis to approximate the solution, and this is one of the advantages of generalised slope limiter described. Instead, we want to point out that once the limiter is applied at one time point, the approximation becomes linear which will affect the accuracy of subsequent computations.

2.8.4 Another slope limiting scheme

The advantage of the approach above is that, in areas of the approximation where the solution is smooth, then the high-order accuracy of the method can be retained, while maintaining positivity where necessary. However, if the problem at hand is dominated by discontinuities and/or if there are no regions of the solution of interest which are smooth, then there is no real advantage in the approach above as the limiter will reduce the approximate solution to linear in all regions of interest. Hence, we may as well just use a linear scheme and save on some computation.

Perhaps a better approach would be to use a linear scheme with a smaller cell width such that the computational work remains approximately the same. For example, for a DG scheme with $2p$ basis functions on a cell, $\mathcal{D}_k$, of width Δ, say, we construct block matrices of dimension $2p \times 2p$ (i.e. the mass, stiffness and flux matrices, $\boldsymbol{M}_k$, $\boldsymbol{G}_k$, and $\boldsymbol{F}^{k\ell}$, respectively) and compute the approximate solution using these matrices. Alternatively, consider breaking the cell, $\mathcal{D}_k$, up into p sub-cells of width Δ/p and using a DG scheme with 2 basis functions on each sub-cell (i.e. a linear basis on each cell). On each of the p sub-cells we construct block matrices of size 2×2. To approximate the solution on the original interval $\mathcal{D}_k$ we form the p approximations on each sub-cell into larger matrices which are of dimension $2p \times 2p$, and then use in the same way as before. Thus, the scheme with $2p$ basis functions on the cell $\mathcal{D}_k$, and the scheme with a 2 basis functions on each of the p sub-cell have approximately the same computational cost (the latter scheme results in a sparser matrix).

If we know *a priori* that we will apply the slope limiter to cell $\mathcal{D}_k$, then intuitively we suspect that the latter scheme may be far more accurate; it will approximate the solution on $\mathcal{D}_k$ by a piecewise linear function with p pieces, whereas the DG scheme with $2p$ basis functions on cell $\mathcal{D}_k$ will, after limiting, represent the solution as a single linear function on $\mathcal{D}_k$.

We refer to this scheme as the DG-lin-lim scheme, as it is a linear DG scheme with a limiter.

2.8.5 Briefly, on accuracy

As a rough guide on accuracy we now paraphrase some known results on the accuracy of the DG method from (Hesthaven & Warburton 2007, Sections 5.5, 5.8) for a similar kind of problem to the one under study here. For certain scalar conservation-law problems, under certain regularity conditions on the continuity and bounded-ness of the solution and its derivatives, on the grid used, on the flux function, and on the numerical flux, then one can expect accuracy proportional to Δ^p for the DG approximation of the solution, provided the DG scheme uses a polynomial basis with p functions on each cell, a regular grid of $\Delta = \max \Delta_k$ is used and a second-order SSPRK method is used to integrate over time (Hesthaven & Warburton 2007, Sections 5.5, 5.8, and references therein). Applying

this result to the DG-lin-lim scheme in Section 2.8.4, we might expect accuracy of order $(\Delta/p)^2$ under the same regularity conditions.

2.9 Sundry mathematical concepts

Polynomials

In this thesis, we use the discontinuous Galerkin method to approximate the generator, a differential operator, of a fluid queue. The discontinuous Galerkin method involves projecting the operator equation onto a basis of functions, typically polynomials. A convenient basis with which to work is the interpolating Lagrange polynomials. The order $p-1$ Lagrange polynomials are defined by a set of points, $\xi_i, i = 1, ..., p$, (the ξ_i's must be distinct), and are given by

$$\ell_i(r) = \prod_{\substack{j=1 \\ j \neq i}}^{p} \frac{r - \xi_j}{\xi_i - \xi_j},\ i = 1, ..., p.$$

A convenient property of the Lagrange polynomials is

$$l_i(\xi_j) = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{otherwise.} \end{cases}$$

Sometimes it is more convenient to work with *orthogonal* polynomials. For example, the Legendre polynomials, which can be defined recursively by

$$\begin{aligned} P_0(x) &= 1, \\ P_1(x) &= x, \\ (n+1)P_{n+1}(x) &= (2n+1)xP_n(x) - nP_{n-1}(x), \end{aligned}$$

and are orthogonal on $[-1, 1]$.

The zeros of $(1-x^2)P_n'(x)$ define the *Legendre-Gauss-Lobatto* points, which are used, among other things, for numerically approximating integrals via quadrature.

Numerical integration

To numerically approximate integrals in the discontinuous Galerkin schemes in this thesis we can use Gauss-Lobatto quadrature. Consider the integral of some function f over the interval $[-1, 1]$. Quadrature approximates the integral by evaluating the function on a set of points, x_i, $i = 1, ..., p$, and computing the weighted sum

$$\int_{-1}^{1} f(x)\,\mathrm{d}x \approx \sum_{i=1}^{p} f(x_i)w_i,$$

where w_i are weights. There are various quadrature schemes which one can use.

For the discontinuous Galerkin method, Gauss-Lobatto quadrature is convenient as in both we evaluate the function at the end points of the interval. For Gauss-Lobatto quadrature, the nodes, $x_i, i = 1, ..., p$ is the zero of $(1 - x^2)P_n'(x)$ where P_n is the nth Legendre polynomial, the weights are given by

$$\begin{aligned} w_1 &= 1, \\ w_i &= \frac{2}{n(n-1)[P_{n-1}(x_i)]^2}, \; i \neq 1, p, \\ w_p &= 1. \end{aligned}$$

Gauss-Lobatto quadrature is accurate for polynomials up to degree $2p - 3$. An integral over the interval $[a, b]$ can be approximated by

$$\int_a^b f(x)\,\mathrm{d}x \approx \frac{b-a}{2}\sum_{i=1}^{p} f\left(\frac{b-a}{2}x_i + \frac{a+b}{2}\right) w_i.$$

We also use a trapezoidal integration rule at times. Consider the integral $\int_a^b f(x)\,\mathrm{d}x$. A trapezoidal rule with p evenly spaced grid-points approximates the integral via

$$\int_a^b f(x)\,\mathrm{d}x \approx \sum_{i=2}^{p} \frac{f(x_{i-1} + f(x_i))}{2}\Delta,$$

where $x_i = a + (i-1)\Delta$, $i = 1, ..., p$, and $\Delta = \dfrac{b-a}{p-1}$.

Measuring the difference between distributions

In the numerical investigations in this thesis we evaluate approximation schemes by comparing the resulting approximations they produce with a ground truth. For comparing distributions we use L^p *norms* and the Kolmogorov-Smirnov distance. The L^p norm of a function f is given by

$$\left(\int |f(x)|^p\,\mathrm{d}x\right)^{1/p}.$$

The Kolmogorov-Smirnov distance between two distribution functions, F_1, F_2 is

$$\sup_x |F_1(x) - F_2(x)|.$$

2.10 More on relevant literature & the context of the thesis

In the process of introducing the technical concepts we have already covered some the relevant literature. Here, we provide some more context for the contributions of this thesis. The focus of this thesis is on developing numerical approximations of fluid queues such that we may approximate the operator-analytic analysis of fluid-fluid queues form Bean & O'Reilly (2014).

Fluid-fluid queues and related models Related to the analysis of fluid-fluid queues are the works of Miyazawa & Zwart (2012), Latouche et al. (2013), Bean & O'Reilly (2013*b*) and Bean et al. (2020). Miyazawa & Zwart (2012) analyse related discrete-time multidimensional Markov additive processes and derive operator-analytic expressions for the limiting distribution. Although markedly different in their approach Miyazawa & Zwart (2012), like Bean & O'Reilly (2014), is inspired by a matrix-analytic approach. The work of Latouche et al. (2013) considers a fluid-fluid queue where the driving process $\{(X(t), \varphi(t))\}$ is level-dependent, and derives computable expressions for the marginal probability distribution of the first level, and bounds for that of the second level. Their analysis is somewhat specific to the data handling model considered and only arrives at bounds for the marginal limiting distribution of the second fluid level. In contrast, the work of Bean & O'Reilly (2014) treats more general models and derives operator-analytic expressions for the joint limiting distribution. Therefore, compared to Latouche et al. (2013), the methods of this thesis therefore apply to a general class of models and enable to approximation of the joint limiting distribution. However, unlike Latouche et al. (2013), the analysis of Bean & O'Reilly (2014) does not consider a level-dependent driving process such as this, but I believe that it would be relatively straightforward to extend their results (and the results of this thesis) to the level-dependent case where the rates c_i depend on $X(t)$ in a piecewise-constant way. Bean & O'Reilly (2013*b*) analyse a model which is simple special case of the fluid-fluid queue. They consider a model where the second level process $\{\dot{Y}(t)\}$ has constant rates which do not depend on $\dot{X}(t)$ (but do depend on $\varphi(t)$), while $\{(\dot{X}(t), \varphi(t))\}$ is an *unbounded* fluid queue. They derive matrix-analytic expressions for certain Laplace transforms, with respect to the position of the first fluid $\{X(t)\}$ relative to its starting position $X(0)$, of the probability of certain transient events (such as first return and draining/filling times) of $\{Y(t)\}$. In contrast, Bean & O'Reilly (2013*b*) together with this thesis, treat a more general model and together, give computational methods for various performance measures of fluid-fluid queues. The yet-unpublished work of Bean et al. (2020) treats a similar model to that of Bean & O'Reilly (2013*b*), except that the driving process has a regulated lower boundary at 0; this significantly complicates the analysis. They derive matrix-analytic expressions for the first return operator of the second fluid given a specific exponential form of the initial distribution and a certain

boundary condition is met. Once again, Bean & O'Reilly (2013*b*) together with this thesis, treat a more general model and together, give computational methods for various performance measures of fluid-fluid queues.

Returning now to Bean & O'Reilly (2014) which introduces the fluid-fluid queue in its generality and derives operator-analytic expressions for the first-return operator of the second fluid and the joint limiting distribution. Their analysis is inspired by the analysis of classical fluid queues in Bean et al. (2005*b*) whereby matrix-analytic expressions for the fluid level are derived in terms of the generator of the driving process. The driving process of a fluid-fluid queue is a fluid queue. The analysis of Bean & O'Reilly (2014) requires specific expressions about the transient distribution of the fluid queue which cannot, in general, be obtained from the existing literature.

Transient analysis of fluid queues The analysis of fluid queues can be classified broadly into three approaches, matrix-analytic methods (for example, Ahn et al. (2005), Ahn & Ramaswami (2003, 2004), Bean et al. (2005*a*,*b*, 2009*a*,*b*), da Silva Soares (2005), Latouche & Nguyen (2019)), differential equations based approaches (such as Anick et al. (1982), Karandikar & Kulkarni (1995), Bean et al. (2022)) and analyses based on recurrence relations (for example, Sericola (1998), Sericola & Tuffin (1999), Sericola (2001)). Often, it is the limiting distribution of the fluid queue which is of interest. In the context of the analysis of fluid-fluid queue as in Bean & O'Reilly (2014), however, we require information about the transient distribution of the fluid queue. Specifically, on the event that the second fluid is non-decreasing (non-increasing), we need to know the amount of fluid to have flowed into or out of the second level process, the time (or Laplace transform of time) taken to do so, and the position of the first fluid and phase at the time when the in-out process of the second fluid reaches a given height. Moreover, we want the resulting expressions to be readily computable. None of the existing literature gives exact expressions which have all the aforementioned properties. Even if we consider a simpler model where the rates $r_i(x)$ are constant on intervals $\mathcal{D}_{k,i}$, then the existing results are not satisfactory. Perhaps the closest is the work of Bean et al. (2009*b*) who compute expressions for the Laplace transform with respect to the time take for the fluid to exit an interval, $\mathcal{D}_{k,i}$, say, on the even that fluid exits in some phase j, given it started within the interval at some point x_0 and in some phase i. These expressions could perhaps be used to analyse a fluid-fluid queue where the rates $r_i(x) = r_k$ for all $i \in \mathcal{S}$, and $x \in \mathcal{D}_{k,i}$. Even with this simplified fluid-fluid queue, it is unclear how we could use these expressions to compute the position of the second fluid.

An approximate method which is appropriate for the analysis of fluid-fluid queues has been proposed. The method of Bean & O'Reilly (2013*a*), which we refer to as the uniformisation method as it is derived via uniformisation arguments, approximates a fluid queue by a discrete-time Markov chain. Bean & O'Reilly (2013*a*) approximate the fluid queue $\{(\dot{X}(t), \varphi(t))\}$ by the quasi-birth-and-death-process $\{(L(t), \varphi(t))\}$. In both

processes, $\{\varphi(t)\}$ is the same so there is no approximation in phase process. The level process $\{L(t)\}$ approximates the level process $\{\dot{X}(t)\}$. Specifically, Bean & O'Reilly (2013*a*) derive the level process such that the event that $L(t) = k$ and $\varphi(t)$ approximates the event $\dot{X}(t) \in \mathcal{D}_{k,i}$ and $\varphi(t) = i$, where all the levels $\{\mathcal{D}_{k,i}\}_k$ have the same width, $\Delta = y_{k+1} - y_k$. By considering arbitrarily small Δ, they prove that the distribution of the quasi-birth-and-death-process $\{(L(t), \varphi(t))\}$ can be made arbitrarily close to the distribution of the fluid queue $\{(\dot{X}(t), \varphi(t))\}$. Replacing the driving process of the fluid-fluid queue, $\{(\dot{X}(t), \varphi(t))\}$, by the approximation $\{(L(t), \varphi(t))\}$ and approximating the rate functions $r_i(x)$ by constants on each interval $\mathcal{D}_k$ yields a classical fluid queue which, intuitively, approximates the fluid-fluid queue. Bean & O'Reilly (2013*a*) numerically investigate the ability of their approximation to approximate $\mathbf{\Psi}_X$. To date, the uniformisation approximation has not been applied in the context of fluid-fluid queues, and its ability to approximate transient quantities of fluid queues (other than the aforementioned investigation of $\mathbf{\Psi}_X$) has not been well studied. In Chapter 7 we investigate the numerical performance of the uniformisation approximation and show that it is effective. Of the methods considered in Chapter 7, the performance of the uniformisation approximation is the poorest and its rate of convergence is the slowest for the numerical experiments we conduct. However, due to the stochastic interpretation of the uniformisation approximation as a QBD, approximations to probabilities are guaranteed to be non-negative, and we can employ probabilistic techniques to analyse the approximation (as Bean & O'Reilly (2013*a*) did). Moreover, the simplicity of the uniformisation scheme lends itself more easily to mathematical analysis. For example, when the uniformisation scheme is used in the context of approximating a fluid-fluid queue the resultant approximation is itself a fluid queue, and this stochastic interpretation could aid in the analysis of this approximation.

It turns out that the uniformisation approximation of Bean & O'Reilly (2013*a*) is equivalent to a certain finite-element approximation which is a specific application of the *discontinuous Galerkin (DG) method*. We introduce the discontinuous Galerkin method and its application to fluid queues, and fluid-fluid queues, in Chapter 3. DG approximation schemes have found great success for approximating solutions of partial differential equations, however, to my knowledge, they have not been applied to fluid queues, or fluid-fluid queue to date (except for our paper Bean et al. (2022), of which I am a co-author and the relevant parts of that work are included in this thesis). The DG scheme has the advantage that its rate of convergence is rapid for smooth problems, however, it can produce oscillatory approximations and result in negative probabilities.

In Chapter 4 we attempt to derive a new approximation scheme which converges faster than the uniformisation scheme and still has a stochastic interpretation so that estimates of probability are guaranteed to be positive.

Chapter 3

Approximating fluid queues with the discontinuous Galerkin method

In this chapter we introduce the discontinuous Galerkin (DG) method applied to fluid queues to approximate the operator-analytic expressions for fluid-fluid queues in Bean & O'Reilly (2014).

> *Apart from Sections 3.6 and 3.7, this chapter has been taken from Section 4 and 5 of Bean et al. (2022) with only minor changes, such as notation, so that this chapter is consistent with the rest of the thesis. I am a co-author of the paper Bean et al. (2022).*

Discontinuous Galerkin (DG) methods can be used to approximate the solutions to systems of partial differential equations (PDEs). Here we describe how one can apply the DG method to approximate fluid queues and fluid-fluid queues. In finite-element methods we project the partial differential equations onto a set of piecewise polynomial functions. This projection leads to a new system of equations which represent a *weak form* of the original system of PDEs. Next, we approximate the *flux operator* which moves probability from one cell to another. This creates a system of ordinary differential equations which is known as a *semi-discrete system* (discrete in the spatial variable but continuous in time). We can solve the semi-discrete system to approximate transient distributions of the fluid queue or, as we do in Section 3.5, we can extract and manipulate an approximation to the generator of the fluid queue to approximate performance measures of fluid-fluid queues. The DG method conserves probability. Here we construct the DG approximation to the matrix of operators $\mathbb{B}$ which we use later to construct a DG approximation to $\mathbb{D}(s)$ then $\Psi(s)$, and ultimately the limiting distribution of a stochastic fluid-fluid queue.

3.1 The partial differential equation

We start by introducing the PDE from which we will extract the approximation to the generator $\mathbb{B}$.

Let $f_i(x,t)$ be the joint density of (X_t, φ_t):

$$f_i(x,t) := \frac{\partial}{\partial x}\mathbb{P}\left(X(t) \leq x, \varphi(t) = i\right), \quad 0 < x < b,\, i \in \mathcal{S},$$

which satisfies the system of partial differential equations

$$\frac{\partial}{\partial t}f_i(x,t) = \sum_{j\in\mathcal{S}} f_j(x,t)T_{ji} - c_i\frac{\partial}{\partial x}f_i(x,t), \quad 0 < x < b,\, i \in \mathcal{S},$$

subject to suitable boundary conditions. In matrix form,

$$\frac{\partial}{\partial t}\boldsymbol{f}(x,t) = \boldsymbol{f}(x,t)\boldsymbol{T} - \frac{\partial}{\partial x}\boldsymbol{f}(x,t)\widehat{\boldsymbol{C}}, \tag{3.1}$$

where $\boldsymbol{f}(x,t) = (f_i(x,t))_{i\in\mathcal{S}}$ is a row-vector.

3.2 Cells, test functions, and weak formulation

To begin with, consider an unbounded first fluid level $\{\ddot{X}_t, t \geq 0\}$, $\ddot{X}_t \in (-\infty, \infty)$. We will eventually truncate this space so that we have a finite dimensional approximation; however, this requires a discussion on boundary conditions which we save for later. Let $\mathcal{D}_{k,i} = [y_k, y_{k+1})$, $k \in \mathbb{Z}$ for $i \in \mathcal{S}_+ \cup \mathcal{S}_0$ and $\mathcal{D}_{k,i} = (y_k, y_{k+1}]$, $k \in \mathbb{Z}$ for $i \in \mathcal{S}_-$, $y_k < y_{k+1}$, partition the domain $(-\infty, \infty)$. We call the $\mathcal{D}_{k,i}$ *cells* and define $\Delta_k = y_{k+1} - y_k$.

For each k we choose p_k linearly independent functions $\{\phi_k^r\}_{r=1}^{p_k}$, compactly supported on $\mathcal{D}_{k,i}$ (i.e. $\phi_k^r(x) = 0$ for $x \notin \mathcal{D}_{k,i}$) to form a basis for the space W_k, in which we formulate the approximation. Here, as is standard in DG methods (Hesthaven & Warburton 2007), we take $\{\phi_k^r\}_{r=1}^{p_k}$ to be the space of polynomials of degree $p_k - 1$. For the sake of illustration, the reader may think of $\{\phi_k^r\}_{r=1}^{p_k}$ as the Lagrange polynomials, but any polynomial basis can be used. On each cell $\mathcal{D}_{k,i}$ we approximate

$$f_i(x,t) \approx u_{k,i}(x,t) = \sum_{r=1}^{p_k} a_{k,i}^r(t)\phi_k^r(x),$$

where $a_{k,i}^r(t)$ are yet-to-be-determined time-dependent coefficients. We refer to $u_{k,i}$ as the *local* approximation on cell k, while the *global* approximation is given by $\sum_{k\in\mathbb{Z}} u_{k,i}$ on the whole domain.

Let $\mathcal{N}_k := \{1, \dots, p_k\}$, $k \in \mathbb{Z}$. For $k \in \mathbb{Z}$, define *local* row-vectors

$$\boldsymbol{\phi}_k(x) = (\phi_k^r(x))_{r\in\mathcal{N}_k}, \quad \boldsymbol{a}_{k,i}(x) = (a_{k,i}^r(x))_{r\in\mathcal{N}_k},\ i \in \mathcal{S}.$$

Note that we will always use the letter r to index the basis function within each cell.

The DG method proceeds by first considering the *weak-formulation* of the PDE, which is constructed from the *strong-form* of the PDE, Equation (3.1). In general, to construct the weak-form we need a set of test functions, say W. Now, take the strong form of the PDE, multiply it by some test function $\psi(x) \in W$, integrate with respect to x, and apply integration by parts to get

$$\int_{x\in\mathbb{R}} \frac{\partial}{\partial t} f_j(x,t)\psi(x)\,\mathrm{d}x = \int_{x\in\mathbb{R}} \sum_{i\in\mathcal{S}} f_i(x,t)T_{ij}\psi(x)\,\mathrm{d}x + \int_{x\in\mathbb{R}} f_j(x,t)c_j\frac{\mathrm{d}}{\mathrm{d}x}\psi(x)\,\mathrm{d}x - [f_j(x,t)c_j\psi(x)]_{x=-\infty}^{x=\infty}, \tag{3.2}$$

for $j \in \mathcal{S}$. It is common to choose W such that $\psi(-\infty) = \psi(\infty) = 0$, in which case the last term on the right is zero. Requiring (3.2) to hold for every $\psi \in W$ gives the weak-formulation of the PDE. Solutions to (3.2) are known as *weak* solutions and generalise the concept of a solution of the PDE. For example, this may allow discontinuities with respect to x in the solution – something which is ill-defined for the strong form.

For the purpose of DG, we take the set of test functions to be $W = \bigoplus_{k\in\mathbb{Z}} W^k$ (the direct sum of W^k), the same as the set of basis functions of our solution space. Proceeding as described above, the weak formulation is

$$\int_{x=y_k}^{y_{k+1}} \frac{\partial}{\partial t} f_j(x,t)\phi_k^r(x)\,\mathrm{d}x = \int_{x=y_k}^{y_{k+1}} \sum_{i\in\mathcal{S}} f_i(x,t)T_{ij}\phi_k^r(x)\,\mathrm{d}x + \int_{x=y_k}^{y_{k+1}} f_j(x,t)c_j\frac{\mathrm{d}}{\mathrm{d}x}\phi_k^r(x)\,\mathrm{d}x - [f_j(x,t)c_j\phi_k^r(x)]_{x=y_k}^{x=y_{k+1}},$$

since ϕ_k^r is compactly supported on $\mathcal{D}_{k,j}$, for all $j \in \mathcal{S}$, $r \in \mathcal{N}_k$, $k \in \mathbb{Z}$. Now, note that any function $g(x)$ can be decomposed as $g(x) = g^W(x) + g^\perp(x)$ where $g^W \in W$ and $g^\perp \in W^\perp$, and $W^\perp$ is the orthogonal complement of W. Since $g^\perp$ is orthogonal to W, $\int_x g^\perp(x)\phi_k^r(x)\,\mathrm{d}x = 0$ for $r \in \mathcal{N}_k$, $k \in \mathbb{Z}$. Also, note that $\frac{\mathrm{d}}{\mathrm{d}x}\phi_k^r(x) \in W$. Using this, we can write

$$\int_{x=y_k}^{y_{k+1}} \frac{\partial}{\partial t}\left(f_j^W(x,t) + f_j^\perp(x,t)\right)\phi_k^r(x)\,\mathrm{d}x = \int_{x=y_k}^{y_{k+1}} \sum_{i\in\mathcal{S}} \left(f_i^W(x,t) + f_i^\perp(x,t)\right)T_{ij}\phi_k^r(x)\,\mathrm{d}x + \int_{x=y_k}^{y_{k+1}} \left(f_j^W(x,t) + f_j^\perp(x,t)\right)c_j\frac{\mathrm{d}}{\mathrm{d}x}\phi_k^r(x)\,\mathrm{d}x - [f_j(x,t)c_j\phi_k^r(x)]_{x=y_k}^{x=y_{k+1}},$$

which is equivalent to

$$\int_{x=y_k}^{y_{k+1}} \frac{\partial}{\partial t} f_j^W(x,t)\phi_k^r(x)\,\mathrm{d}x = \int_{x=y_k}^{y_{k+1}} \sum_{i\in\mathcal{S}} f_i^W(x,t)T_{ij}\phi_k^r(x)\,\mathrm{d}x \tag{3.3}$$
$$+ \int_{x=y_k}^{y_{k+1}} f_j^W(x,t)c_j \frac{\mathrm{d}}{\mathrm{d}x}\phi_k^r(x)\,\mathrm{d}x - [f_j(x,t)c_j\phi_k^r(x)]_{x=y_k}^{x=y_{k+1}}.$$

Now, $f_j^W(x,t) \in W$ so, on $\mathcal{D}_{k,j}$, it can be expressed as $u_{k,j}(x,t) := \boldsymbol{a}_{k,j}(t)\boldsymbol{\phi}_k(x)'$, where the prime, $'$, denotes the transpose, which we now substitute into (3.3) and repeat this for all test functions $\phi_k^r(x)$, $r = 1, ..., p_k$, to get the system of equations,

$$\int_{x=y_k}^{y_{k+1}} \frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{a}_{k,j}(t)\boldsymbol{\phi}_k(x)'\boldsymbol{\phi}_k(x)\,\mathrm{d}x = \int_{x=y_k}^{y_{k+1}} \sum_{i\in\mathcal{S}} \boldsymbol{a}_{k,i}(t)\boldsymbol{\phi}_k(x)'T_{ij}\boldsymbol{\phi}_k(x)\,\mathrm{d}x \tag{3.4}$$
$$+ \int_{x=y_k}^{y_{k+1}} \boldsymbol{a}_{k,j}(t)\boldsymbol{\phi}_k(x)'c_j\frac{\mathrm{d}}{\mathrm{d}x}\boldsymbol{\phi}_k(x)\,\mathrm{d}x - c_j[f_j(x,t)\boldsymbol{\phi}_k(x)]_{x=y_k}^{x=y_{k+1}}, \quad k \in \mathbb{Z}.$$

3.3 Mass, stiffness, and flux matrices

Recall, for $k \in \mathbb{Z}$, we defined local *mass* and *stiffness* matrices $\boldsymbol{M}_k$ and $\boldsymbol{G}_k$ by

$$\boldsymbol{M}_k := \int_{x\in\mathcal{D}_k} \boldsymbol{\phi}_k(x)'\boldsymbol{\phi}_k(x)\,\mathrm{d}x, \quad \boldsymbol{G}_k := \int_{x\in\mathcal{D}_k} \boldsymbol{\phi}_k(x)'\frac{\mathrm{d}}{\mathrm{d}x}\boldsymbol{\phi}_k(x)\,\mathrm{d}x.$$

We can write (3.4) as

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{a}_{k,j}(t)\boldsymbol{M}_k = \sum_{i\in\mathcal{S}} \boldsymbol{a}_{k,i}(t)\boldsymbol{M}_k T_{ij} + c_j\boldsymbol{a}_{k,j}(t)\boldsymbol{G}_k - c_j[f_j(x,t)\boldsymbol{\phi}_k(x)]_{x=y_k}^{x=y_{k+1}}. \tag{3.5}$$

It remains to approximate the *flux*, $f_j(x,t)$ at the cell edges y_k, $k \in \mathbb{Z}$, so that we may evaluate the terms $[f_j(x,t)\phi_k^r(x)]_{x=y_k}^{x=y_{k+1}}$, $r = 1, ..., p_k$, $k \in \mathbb{Z}$. This is the key for DG – it joins the local approximations on each cell into a global approximation on the whole domain of approximation. The flux is the instantaneous rate (with respect to time) at which density moves across the boundaries y_k, $k \in \mathbb{Z}$. There are different choices for the flux, and we refer the reader to (Cockburn 1999, Hesthaven & Warburton 2007), and references therein, for some discussion of the topic. Here, we choose the *upwind* scheme, which, as we shall see, closely resembles the flux terms from the generator $\mathbb{B}$. The approximate flux, also known as the *numerical flux*, is given by

$$f_j^*(x,t) = sign(c_j) \lim_{\varepsilon\to 0^+} u_j(x - \varepsilon c_j, t),$$

at each $x = y_k$, $k \in \mathbb{Z}$. Intuitively, the upwind flux takes the value of the density immediately on the upwind side of each y_k.

Denote by x^- and x^+ the left and right limits at x, respectively. Assume first $c_j > 0$, then

$$\begin{aligned}
-c_j[f_j(x,t)\phi_k^r(x)]_{x=y_k}^{x=y_{k+1}} &\approx -c_j[f_j^*(x,t)\phi_k^r(x)]_{x=y_k}^{x=y_{k+1}} \\
&= -c_j f_j^*(y_{k+1},t)\phi_k^r(y_{k+1}) + c_j f_j^*(y_k,t)\phi_k^r(y_k) \\
&= -c_j u_j(y_{k+1}^-,t)\phi_k^r(y_{k+1}) + c_j u_j(y_k^-,t)\phi_k^r(y_k) \\
&= -c_j u_{k,j}(y_{k+1}^-,t)\phi_k^r(y_{k+1}) + c_j u_{k-1,j}(y_k^-,t)\phi_k^r(y_k) \\
&= -c_j \boldsymbol{a}_{k,j}(t)\boldsymbol{\phi}_k(y_{k+1}^-)'\phi_k^r(y_{k+1}) + c_j \boldsymbol{a}_{k-1,j}(t)\boldsymbol{\phi}_{k-1}(y_k^-)'\phi_k^r(y_k).
\end{aligned}$$

In matrix form,

$$\begin{aligned}
-c_j[f_j(x,t)\boldsymbol{\phi}_k(x)]_{x=y_k}^{x=y_{k+1}} &\approx -c_j[f_j^*(x,t)\boldsymbol{\phi}_k(x)]_{x=y_k}^{x=y_{k+1}} \\
&= -c_j \boldsymbol{a}_{k,j}(t)\boldsymbol{\phi}_k(y_{k+1}^-)'\boldsymbol{\phi}_k(y_{k+1}) + c_j \boldsymbol{a}_{k-1,j}(t)\boldsymbol{\phi}_{k-1}(y_k^-)'\boldsymbol{\phi}_k(y_k) \\
&= c_j \boldsymbol{a}_{k,j}(t)F_j^{k,k} + c_j \boldsymbol{a}_{k-1,j}(t)F_j^{k-1,k},
\end{aligned}$$

where, for $j \in \mathcal{S}$ with $c_j > 0$, we define $\boldsymbol{F}_+^{k,k} = \boldsymbol{F}_j^{k,k} := -\boldsymbol{\phi}_k(y_{k+1}^-)'\boldsymbol{\phi}_k(y_{k+1})$, $k \in \mathbb{Z}$ and $\boldsymbol{F}_j^{k-1,k} := \boldsymbol{\phi}_{k-1}(y_k^-)'\boldsymbol{\phi}_k(y_k)$, $k \in \mathbb{Z}$.

Now proceed similarly for $c_j < 0$ to get the approximation

$$\begin{aligned}
-c_j[f_j(x,t)\boldsymbol{\phi}_k(x)]_{x=y_k}^{x=y_{k+1}} &\approx -c_j[f_j^*(x,t)\boldsymbol{\phi}_k(x)]_{x=y_k}^{x=y_{k+1}} \\
&= -c_j \boldsymbol{a}_{k+1,j}(t)\boldsymbol{\phi}_{k+1}(y_{k+1}^+)'\boldsymbol{\phi}_k(y_{k+1}) + c_j \boldsymbol{a}_{k,j}(t)\boldsymbol{\phi}_k(y_k^+)'\boldsymbol{\phi}_k(y_k) \\
&= c_j \boldsymbol{a}_{k+1,j}(t)\boldsymbol{F}_j^{k+1,k} + c_j \boldsymbol{a}_{k,j}(t)\boldsymbol{F}_j^{k,k},
\end{aligned}$$

where, for $j \in \mathcal{S}$ with $c_j < 0$, we define $\boldsymbol{F}_j^{k+1,k} := -\boldsymbol{\phi}_{k+1}(y_{k+1}^+)'\boldsymbol{\phi}_k(y_{k+1})$, $k \in \mathbb{Z}$, and $\boldsymbol{F}_-^{k,k} = \boldsymbol{F}_j^{k,k} := \boldsymbol{\phi}_k(y_k^+)'\boldsymbol{\phi}_k(y_k)$, $k \in \mathbb{Z}$.

The matrices $\boldsymbol{F}_j^{k-1,k}$, $\boldsymbol{F}_j^{k,k}$, and $\boldsymbol{F}_j^{k+1,k}$ are the local flux matrices. For convenience, we also define the matrices $\boldsymbol{F}_j^{k,k+1} = 0$ for $j \in \mathcal{S}_-$ and $\boldsymbol{F}_j^{k,k-1} = 0$ for $j \in \mathcal{S}_+$, $k \in \mathbb{Z}$.

To write this out as a *global* system, define the row-vectors

$$\boldsymbol{a}_k(t) = (\boldsymbol{a}_{k,i}(t))_{i\in\mathcal{S}}, \quad \boldsymbol{a}(t) = (\boldsymbol{a}_k(t))_{k\in\mathbb{Z}},$$

and the doubly-infinite block-tridiagonal matrix

$$\ddot{\boldsymbol{B}} = \begin{bmatrix} \ddots & \ddots & \ddots & & \\ & \ddot{\boldsymbol{B}}^{k,k-1} & \ddot{\boldsymbol{B}}^{k,k} & \ddot{\boldsymbol{B}}^{k,k+1} & \\ & & \ddots & \ddots & \ddots \end{bmatrix},$$

where, for $k \in \mathbb{Z}$,

$$\ddot{\boldsymbol{B}}^{kk} = \begin{bmatrix} T_{11}\boldsymbol{I} + c_1(\boldsymbol{F}_1^{kk} + \boldsymbol{G}_k)\boldsymbol{M}_k^{-1} & T_{12}\boldsymbol{I} & T_{1N}\boldsymbol{I} \\ T_{21}\boldsymbol{I} & & \\ \vdots & \ddots & \vdots \\ & & T_{N-1,N}\boldsymbol{I} \\ T_{N1}\boldsymbol{I} & T_{N,N-1}\boldsymbol{I} & T_{N,N}\boldsymbol{I} + c_N(\boldsymbol{F}_N^{kk} + \boldsymbol{G}_k)\boldsymbol{M}_k^{-1} \end{bmatrix},$$

$$\ddot{\boldsymbol{B}}^{k,k+1} = \begin{bmatrix} c_1\boldsymbol{F}_1^{k,k+1}\boldsymbol{M}_{k+1}^{-1} & & \\ & \ddots & \\ & & c_N\boldsymbol{F}_N^{k,k+1}\boldsymbol{M}_{k+1}^{-1} \end{bmatrix},$$

$$\ddot{\boldsymbol{B}}^{k,k-1} = \begin{bmatrix} c_1\boldsymbol{F}_1^{k,k-1}\boldsymbol{M}_{k-1}^{-1} & & \\ & \ddots & \\ & & c_N\boldsymbol{F}_N^{k,k-1}\boldsymbol{M}_{k-1}^{-1} \end{bmatrix}.$$

The global system of equations is

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{a}(t) = \boldsymbol{a}(t)\ddot{\boldsymbol{B}}. \tag{3.6}$$

3.4 Boundary conditions

To enable computation, this numerical approximation has to take place on a finite interval, which means we must consider a bounded domain and specify boundary conditions. Recall that we wish to impose lower and upper boundaries at 0 and b, respectively.

Let $[0, b]$ be the domain of the approximation, where $b < \infty$. We partition the space $[0, b]$ into $\mathcal{D}_{-1} = \{0\}$, $\mathcal{D}_{K+1} = \{b\}$, and K non-trivial intervals, $\mathcal{D}_{k,i} = [y_k, y_{k+1}) \setminus \{0\}$, $i \in \mathcal{S}_+ \cup \mathcal{S}_0, \mathcal{D}_{k,i} = (y_k, y_{k+1}] \setminus \{0\}$, $i \in \mathcal{S}_-$, $y_k < y_{k+1}$, $k \in \mathcal{K}^\circ$, $y_0 = 0$, $y_{K+1} = b$ and define $\Delta_k := y_{k+1} - y_k$.

For states with $c_i \leq 0$, there is the possibility of point mass accumulating at the boundary at 0. Denote these point masses by $q_{-1,i}(t)$ for $i \in \mathcal{S}_{-1}$. For states with $c_i > 0$ there is no possibility of a point mass at 0. Similarly, for $c_i \geq 0$ there is the possibility of a point mass at b. Denote these point masses by $q_{K+1,i}(t)$, for $i \in \mathcal{S}_{K+1}$. For states with $c_i < 0$ there is no possibility of a point mass at b. Let $\boldsymbol{q}_{-1}(t) = (q_{-1,i}(t))_{i\in\mathcal{S}_{-1}}$ and $\boldsymbol{q}_{K+1}(t) = (q_{K+1,i}(t))_{i\in\mathcal{S}_{K+1}}$ and $\boldsymbol{f}_m(x,t) = (f_i(x,t))_{i\in\mathcal{S}_m}$, $m \in \{+,-,0\}$.

Let us first consider the boundary at $X_t = 0$. The boundary conditions that describe the evolution of probability/density of a stochastic fluid queue with a boundary at 0 are;

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{q}_{-1}(t) = \boldsymbol{q}_{-1}(t)\boldsymbol{T}_{-1,-1} + \boldsymbol{f}_-(0,t)\boldsymbol{C}_-\boldsymbol{P}_{-,-1}, \tag{3.7}$$

$$\boldsymbol{f}_-(0^+,t)\boldsymbol{C}_-\boldsymbol{P}_{-,+} + \boldsymbol{q}_{-1}(t)\boldsymbol{T}_{-1,+} = \boldsymbol{f}_+(0,t)\boldsymbol{C}_+, \tag{3.8}$$

where $\boldsymbol{P}_{-,+} = \left[p_{ij}^{-1}\right]_{i\in\mathcal{S}_-,j\in\mathcal{S}_+}$, $\boldsymbol{P}_{-,-1} = \left[p_{ij}^{-1}\right]_{i\in\mathcal{S}_-,j\in\mathcal{S}_{-1}}$ and p_{ij}^{-1} is the probability of instantaneous transition of the phase process from Phase i to Phase j upon the process hitting the lower boundary. Equation (3.7) states that point mass moves between phases according to the sub-generator matrix $\boldsymbol{T}_{-1,-1}$, and that the flux of probability density into the point masses is $\boldsymbol{f}_-(0,t)\boldsymbol{P}_{-,-1}\boldsymbol{C}_{-1}$. Substituting the DG approximation for $\boldsymbol{f}_-(0,t)$ into (3.7) gives, for $j \in \mathcal{S}_{-1}$,

$$\frac{\mathrm{d}}{\mathrm{d}t}q_{-1,j}(t) = \sum_{i\in\mathcal{S}_{-1}} q_{-1,i}(t)T_{ij} - \sum_{i\in\mathcal{S}_-} \boldsymbol{a}_{0,i}(t)\boldsymbol{\phi}_0(0^+)'p_{ij}^{-1}c_i.$$

Equation (3.8) describes the reflection of process at 0, plus the flux of probability mass to density upon a transition from a phase in $\mathcal{S}_{-1}$ to a phase in $\mathcal{S}_+$. Thus, the flux into the left-hand edge of $\mathcal{D}_{0,j}$ in phase $j \in \mathcal{S}_+$ is, $\sum_{i\in\mathcal{S}_{-1}} q_{-1,i}(t)T_{ij}$. Therefore, we can now evaluate

$$\begin{aligned}
&- c_j[f_j(x,t)\boldsymbol{\phi}_0(x)]_{x=0}^{x=y_0} \\
&= -c_j f_j(y_0^-,t)\boldsymbol{\phi}_0(y_0) + c_j f_j(0^+,t)\boldsymbol{\phi}_0(0) - \sum_{i\in\mathcal{S}_-} c_i f_i(0^+,t)p_{ij}^{-1}\boldsymbol{\phi}_0(0) \\
&\approx -c_j f_j^*(y_0,t)\boldsymbol{\phi}_0(y_0) + \sum_{i\in\mathcal{S}_{-1}} q_{-1,i}(t)T_{ij}\boldsymbol{\phi}_0(0) - \sum_{i\in\mathcal{S}_-} c_i p_{ij}^{-1}\boldsymbol{a}_{0,i}(t)\boldsymbol{\phi}_0(0)'\boldsymbol{\phi}_0(0) \\
&= c_j\boldsymbol{a}_{0,j}(t)\boldsymbol{F}_j^{0,0} + \sum_{i\in\mathcal{S}_{-1}} q_{-1,i}(t)T_{ij}\boldsymbol{\phi}_0(0) - \sum_{i\in\mathcal{S}_-} c_i p_{ij}^{-1}\boldsymbol{a}_{0,i}(t)\boldsymbol{F}_j^{0,0},
\end{aligned}$$

for $j \in \mathcal{S}_+$.

Thus, the DG approximation of the flux into point masses $q_{-1,j}(t)$ is

$$-\sum_{i\in\mathcal{S}_-} \boldsymbol{a}_{0,i}(t)\boldsymbol{\phi}_0(0)'p_{ij}^{-1}c_i,\ j \in \mathcal{S}_-,$$

the rate of transition of point mass within $\boldsymbol{q}_{-1}(t)$ is $\boldsymbol{T}_{-1,-1}$, the DG approximation of the transition of point mass to density is $\sum_{i\in\mathcal{S}_{-1}} q_{-1,i}(t)T_{ij}\boldsymbol{\phi}_0(0)$, $j \in \mathcal{S}_+$, and the DG approximation to density reflected at the lower boundary is $\sum_{i\in\mathcal{S}_-} c_i p_{ij}^{-1}\boldsymbol{a}_{0,i}(t)\boldsymbol{F}_j^{0,0}$.

Similarly, for the upper boundary at b the boundary conditions are

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{q}_{K+1}(t) &= \boldsymbol{q}_{K+1}(t)T_{K+1,K+1} + \boldsymbol{f}_+(b,t)\boldsymbol{C}_+\boldsymbol{P}_{+,K+1}, \\
\boldsymbol{f}_+(b^-,t)\boldsymbol{C}_+\boldsymbol{P}_{+,-} + \boldsymbol{q}_{K+1}(t)T_{K+1-} &= \boldsymbol{f}_-(b,t)\boldsymbol{C}_-,
\end{aligned}$$

where $\boldsymbol{P}_{+,-} = \left[p_{ij}^{K+1}\right]_{i\in\mathcal{S}_+,j\in\mathcal{S}_-}$, $\boldsymbol{P}_{+,K+1} = \left[p_{ij}^{K+1}\right]_{i\in\mathcal{S}_+,j\in\mathcal{S}_{K+1}}$ and p_{ij}^{K+1} is the probability of instantaneous transition of the phase process from Phase i to Phase j upon the process

hitting the lower boundary. Using the same arguments as above,

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}q_{K+1,j}(t) &= \sum_{i\in\mathcal{S}_{K+1}} q_{K+1,i}(t)T_{ij} + \sum_{i\in\mathcal{S}_+} \boldsymbol{a}_{K,i}(t)\boldsymbol{\phi}_K(b)' p_{ij}^{K+1} c_i,\\ -c_j[f_j(x,t)\boldsymbol{\phi}_K(x)]_{x=y_K}^{x=b} &\approx c_j \boldsymbol{a}_{K,j}(t) F_j^{K,K} + \sum_{i\in\mathcal{S}_{K+1}} q_{K+1,i}(t)T_{ij}\boldsymbol{\phi}_K(b)\\ &\quad - \sum_{i\in\mathcal{S}_+} c_i p_{ij}^{K+1} \boldsymbol{a}_{0,i}(t)\boldsymbol{F}_j^{0,0}\end{aligned}$$

for $j \in \mathcal{S}_-$.

Thus, the DG approximation of the flux into the point mass $q_{K+1,j}(t)$ is

$$\sum_{i\in\mathcal{S}_+} \boldsymbol{a}_{K,i}(t)\boldsymbol{\phi}_K(0)' p_{ij}^{K+1} c_i,$$

$j \in \mathcal{S}_+$, the rate of transition of point mass within $\boldsymbol{q}_{K+1}(t)$ is $\boldsymbol{T}_{K+1,K+1}$, the DG approximation of the transition of point mass to density is $\sum_{i\in\mathcal{S}_{K+1}} q_{K+1,i}(t)T_{ij}\boldsymbol{\phi}_K(b)$, $j \in \mathcal{S}_-$ and the DG approximation to reflection of density at the upper boundary is $-\sum_{i\in\mathcal{S}_+} c_i p_{ij}^{K+1}\boldsymbol{a}_{0,i}(t)\boldsymbol{F}_j^{0,0}$.

To include the behaviour in the DG generator we truncate the doubly-infinite matrix $\ddot{\boldsymbol{B}}$ at $k = 0$ and $k = K$, then append $|\mathcal{S}_{-1}|$ rows and columns to the top and left, and $|\mathcal{S}_{K+1}|$ rows and columns to the bottom and right. These represent the point masses $\boldsymbol{q}_{-1}(t)$ and $\boldsymbol{q}_{K+1}(t)$, respectively. Given the discussion above, the truncated matrix is

$$\boldsymbol{B} = \begin{bmatrix} \boldsymbol{T}_{-1,-1} & \boldsymbol{B}^{-10} & & & & & \\ \boldsymbol{B}^{0-1} & \boldsymbol{B}^{00} & \boldsymbol{B}^{01} & & & & \\ & \boldsymbol{B}^{10} & \boldsymbol{B}^{11} & \boldsymbol{B}^{12} & & & \\ & & \ddots & \ddots & \ddots & & \\ & & \boldsymbol{B}^{K-1,K-2} & \boldsymbol{B}^{K-1,K-1} & \boldsymbol{B}^{K-1,K} & & \\ & & & \boldsymbol{B}^{K,K-1} & \boldsymbol{B}^{K,K} & \boldsymbol{B}^{K,K+1} & \\ & & & & \boldsymbol{B}^{K+1,K} & \boldsymbol{T}_{K+1,K+1} \end{bmatrix},$$

where

$$\begin{aligned}\boldsymbol{B}^{k\ell} &= \ddot{\boldsymbol{B}}^{k\ell}, \text{ for } k \in \mathcal{K}^\circ,\ \ell \in \{k-1, k, k+1\},\ k = \ell \neq 0 \text{ or } k = \ell \neq K,\\ \boldsymbol{B}^{00} &= \ddot{\boldsymbol{B}}^{00} - \left[c_i p_{ij}^{-1} 1(c_i < 0, c_j > 0)\right]_{i\in\mathcal{S},j\in\mathcal{S}} \otimes \boldsymbol{F}_-^{00},\\ \boldsymbol{B}^{KK} &= \ddot{\boldsymbol{B}}^{KK} + \left[c_i p_{ij}^{-1} 1(c_i > 0, c_j < 0)\right]_{i\in\mathcal{S},j\in\mathcal{S}} \otimes \boldsymbol{F}_+^{00},\\ \boldsymbol{B}^{-10} &:= \boldsymbol{T}_{-1+} \otimes \boldsymbol{\phi}^0(0),\\ \boldsymbol{B}^{0-1} &:= -\left[c_i p_{ij}^{-1} 1(c_i < 0)\right]_{i\in\mathcal{S},j\in\mathcal{S}_{-1}} \otimes \boldsymbol{\phi}^0(0)',\end{aligned}$$

$$\boldsymbol{B}^{K+1K} := \boldsymbol{T}_{K+1-} \otimes \boldsymbol{\phi}^K(b),$$
$$\boldsymbol{B}^{K,K+1} := \left[c_i p_{ij}^{K+1} 1(c_i > 0)\right]_{i\in\mathcal{S}, j\in\mathcal{S}_{K+1}} \otimes \boldsymbol{\phi}^K(b)',$$

and $\otimes$ is the Kronecker product.

For future reference, we also define the matrices $\boldsymbol{B}_{ij}^{k\ell}$ for $k \in \{2,\dots,K-1\}$, $\ell \in \{k-1,k\}$, $i,j \in \mathcal{S}$, by

$$\boldsymbol{B}_{ij}^{kk} = \begin{cases} T_{ij}\boldsymbol{I}_{p_k} + c_i(\boldsymbol{F}_i^{kk} + \boldsymbol{G}_k)\boldsymbol{M}_k^{-1} & i = j, \\ T_{ij}\boldsymbol{I}_{p_k} & i \neq j, \end{cases}$$
$$\boldsymbol{B}_{ij}^{k\ell} = \begin{cases} c_i \boldsymbol{F}_i^{k\ell}\boldsymbol{M}_\ell^{-1} & i = j, \\ \boldsymbol{0} & i \neq j, \end{cases} \quad \ell \in \{k-1,k\}$$

and

$$\boldsymbol{B}_{ij}^{00} = \begin{cases} T_{ij}\boldsymbol{I}_{p_k} + c_i(\boldsymbol{F}_i^{00} + \boldsymbol{G}_0)\boldsymbol{M}_0^{-1} & i = j, \\ T_{ij}\boldsymbol{I}_{p_k} - 1(c_i < 0, c_j > 0) c_i p_{ij}^{-1} \boldsymbol{F}_i^{00}\boldsymbol{M}_0^{-1} & i \neq j, \end{cases}$$
$$\boldsymbol{B}_{ij}^{KK} = \begin{cases} T_{ij}\boldsymbol{I}_{p_k} + c_i(\boldsymbol{F}_i^{KK} + \boldsymbol{G}_K)\boldsymbol{M}_K^{-1} & i = j, \\ T_{ij}\boldsymbol{I}_{p_K} + 1(c_i > 0, c_j < 0) c_i p_{ij}^{K+1} \boldsymbol{F}_i^{KK}\boldsymbol{M}_K^{-1} & i \neq j. \end{cases}$$

After the addition of the boundary conditions, the system of ODEs (3.6) can now be written as

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix}\boldsymbol{q}_{-1}(t) & \boldsymbol{a}(t) & \boldsymbol{q}_{K+1}(t)\end{bmatrix} = \begin{bmatrix}\boldsymbol{q}_{-1}(t) & \boldsymbol{a}(t) & \boldsymbol{q}_{K+1}(t)\end{bmatrix}\boldsymbol{B}. \tag{3.9}$$

Approximations $\boldsymbol{B}_{ij}^{mn}$, $\boldsymbol{B}_{ij}$, and $\boldsymbol{B}^{mn}$ to $\mathbb{B}_{ij}^{mn}$, $\mathbb{B}_{ij}$, and $\mathbb{B}^{mn}$, $i,j \in \mathcal{S}$, $m,n \in \{+,-,0\}$, are constructed from the block-matrices $\boldsymbol{B}_{ij}^{k\ell}$, $i,j \in \mathcal{S}$, $k,\ell \in \mathcal{K}$, as

$$\boldsymbol{B}_{ij}^{mn} = \left[\boldsymbol{B}_{ij}^{k\ell}\right]_{k\in\mathcal{K}_i^m, \ell\in\mathcal{K}_j^n}, \quad i,j \in \mathcal{S},\ m,n \in \{+,-,0\},$$
$$\boldsymbol{B}_{ij} = \left[\boldsymbol{B}_{ij}^{k\ell}\right]_{k,\ell\in\mathcal{K}}, \ i,j \in \mathcal{S},$$
$$\boldsymbol{B}^{mn} = \left[\left[\boldsymbol{B}_{ij}^{k\ell}\right]_{i\in\mathcal{S}_k^m, j\in\mathcal{S}_\ell^n}\right]_{k\in\mathcal{K}^m,\ell\in\mathcal{K}^n}, \ m,n \in \{+,-,0\}.$$

We prove the following result in Appendix B.

Corollary 3.1. *The approximate generator $\boldsymbol{B}$ conserves probability. That is, for all $t \geq 0$,*

$$\sum_{i\in\mathcal{S}_{-1}} q_{-1,i}(t) + \sum_{i\in\mathcal{S}_{K+1}} q_{K+1,i}(t) + \sum_{i\in\mathcal{S}} \int_{x\in[0,b]} u_i(x,t)\,\mathrm{d}x$$
$$= \sum_{i\in\mathcal{S}_{-1}} q_{-1,i}(0) + \sum_{i\in\mathcal{S}_{K+1}} q_{K+1,i}(0) + \sum_{i\in\mathcal{S}} \int_{x\in[0,b]} u_i(x,0)\,\mathrm{d}x.$$

3.5 Application to a fluid-fluid queue

Given our approximation $\boldsymbol{B}$ to the generator $\mathbb{B}$ we now follow the recipe from Bean & O'Reilly (2014), replacing the actual generator $\mathbb{B}$ with its approximation $\boldsymbol{B}$, to construct approximations, $\boldsymbol{\pi}$ and $\boldsymbol{p}$, to the limiting operators, $\mathbb{\pi}$ and $\mathbb{p}$.

It may be convenient to think of our approximations in terms of approximations of kernels. Recall that the operators in (Bean & O'Reilly 2014) can be thought of in terms of kernels. That is, for some function $\boldsymbol{g} = (g_i(x))_{i\in\mathcal{S}}$, we can write $\boldsymbol{\mu}\mathbb{B}\boldsymbol{g}' = \sum_{k,\ell\in\mathcal{K}}\sum_{i,j\in\mathcal{S}}\int_{x,y} \mathrm{d}\mu_i(x)\mathbb{B}_{ij}^{k\ell}(x,\,\mathrm{d}y)g_j(y)$ where $\mathbb{B}_{ij}^{k\ell}(x,\,\mathrm{d}y)$ is the kernel of the operator $\mathbb{B}_{ij}^{k\ell}$.

Let $\boldsymbol{a}_{-1}(t) := \boldsymbol{q}_{-1}(t)$ and $\boldsymbol{a}_{K+1}(t) := \boldsymbol{q}_{K+1}(t)$, and define basis functions $\boldsymbol{\phi}_{-1}(x) = \phi_{-1}^1(x) = \delta(x)$ and $\boldsymbol{\phi}_{K+1}(x) = \phi_{K+1}^1(x) = \delta(x-b)$, where δ is the Dirac delta, $p_{-1} = p_{K+1} = 1$, and $\mathcal{N}_{-1} = \mathcal{N}_{K+1} = \{1\}$. Also define $\boldsymbol{M}_{-1} = \boldsymbol{I}_{|\mathcal{S}_{-1}|}$, $\boldsymbol{M}_{K+1} = \boldsymbol{I}_{|\mathcal{S}_{K+1}|}$, the block-diagonal matrix $\boldsymbol{M} = diag(\boldsymbol{M}_k, k\in\mathcal{K})$, and row-vectors

$$\boldsymbol{\phi}(x) = (\boldsymbol{\phi}_k(x))_{k\in\mathcal{K}}, \quad \boldsymbol{a}_i(t) = (\boldsymbol{a}_{k,i}(t))_{k\in\mathcal{K}},\ i\in\mathcal{S}.$$

To pose the approximation $\boldsymbol{B}$ in kernel form let $\boldsymbol{a}_i\boldsymbol{\phi}(x)' \in W$, $i\in\mathcal{S}$ be the initial density of the process, and $\boldsymbol{\phi}(x)\boldsymbol{b}_i' \in W$, $i\in\mathcal{S}$ be a test function. Observe that, from our DG construction earlier and the definition of $\boldsymbol{M}$,

$$\sum_{i,j\in\mathcal{S}}\int_{x,y\in[0,b]} \boldsymbol{a}_i\boldsymbol{\phi}(x)'\boldsymbol{\phi}(x)\,\mathrm{d}x\boldsymbol{M}^{-1}\boldsymbol{B}_{ij}\boldsymbol{\phi}(y)'\boldsymbol{\phi}(y)\boldsymbol{b}_j\,\mathrm{d}y = \sum_{i,j\in\mathcal{S}}\boldsymbol{a}_i\boldsymbol{B}_{ij}\boldsymbol{M}\boldsymbol{b}_j.$$

Thus, we can think of

$$\boldsymbol{\phi}(x)\boldsymbol{M}^{-1}\boldsymbol{B}_{ij}\boldsymbol{\phi}(y)'\,\mathrm{d}y,$$

as an approximation to the kernel $\mathbb{B}_{ij}(x,\,\mathrm{d}y)$. This concept can be extended to all the approximations of operators considered in this work.

3.5.1 Approximating the operator $\mathbb{R}$

Recall the operator $\mathbb{R}^k$ from Lemma 2.1. Essentially, the operator $\mathbb{R}^k$ takes an initial measure $\boldsymbol{\mu}_k$ and multiplies each element by $1/|r_i(x)|$ on cells $\mathcal{D}_k$ where $r_i(x)\neq 0$. In the context of DG the initial distribution is given by $\boldsymbol{a}_i\boldsymbol{\phi}(x)' \in W$, $i\in\mathcal{S}$. Thus, for $k\in\mathcal{K}$ such that $r_i(x)\neq 0$ on $\mathcal{D}_k$, we have

$$\boldsymbol{a}_{k,i}\boldsymbol{\phi}_k(x)'\mathbb{R}_i^k = \frac{\boldsymbol{a}_{k,i}\boldsymbol{\phi}_k(x)'}{|r_i(x)|}.$$

Decompose the right-hand side into a component which lies in W and another orthogonal to W:

$$\frac{\boldsymbol{a}_{k,i}\boldsymbol{\phi}_k(x)'}{|r_i(x)|} = \boldsymbol{\rho}_{k,i}\boldsymbol{\phi}_k(x)' + g_i^{\perp}(x),$$

where $\boldsymbol{\rho}_{k,i}\boldsymbol{\phi}_k(x)' \in W$, $g_i^{\perp} \in W^{\perp}$. Now, multiply by test functions $\{\phi_k^r(x)\}_{r=1}^{p_k}$ and integrate over $[0,b]$:

$$\begin{aligned}\boldsymbol{a}_{k,i}\int_{x\in[0,b]}\frac{\boldsymbol{\phi}_k(x)'\boldsymbol{\phi}_k(x)}{|r_i(x)|}\,\mathrm{d}x &= \boldsymbol{\rho}_{k,i}\int_{x\in[0,b]}\boldsymbol{\phi}_k(x)'\boldsymbol{\phi}_k(x)\,\mathrm{d}x + \int_{x\in[0,b]} g_i^{\perp}(x)\boldsymbol{\phi}_k(x)\,\mathrm{d}x\\ &= \boldsymbol{\rho}_{k,i}\int_{x\in[0,b]}\boldsymbol{\phi}_k(x)'\boldsymbol{\phi}_k(x)\,\mathrm{d}x = \boldsymbol{\rho}_{k,i}\boldsymbol{M}_k,\end{aligned}$$

since $g_i(x)^{\perp} \in W^{\perp}$. Define the matrix $\boldsymbol{M}_k^r := \int_{x\in[0,b]}\frac{\boldsymbol{\phi}_k(x)'\boldsymbol{\phi}_k(x)}{|r_i(x)|}\mathrm{d}x$, then $\boldsymbol{a}_{k,i}\boldsymbol{M}_k^r = \boldsymbol{\rho}_{k,i}\boldsymbol{M}_k$, which implies $\boldsymbol{\rho}_{k,i} = \boldsymbol{a}_{k,i}\boldsymbol{M}_k^r\boldsymbol{M}_k^{-1}$. Thus, we have the approximation

$$\boldsymbol{a}_{k,i}\boldsymbol{\phi}_k(x)'\mathbb{R}_i^k = \frac{\boldsymbol{a}_{k,i}\boldsymbol{\phi}_k(x)'}{|r_i(x)|} \approx \boldsymbol{a}_{k,i}\boldsymbol{M}_k^r\boldsymbol{M}_k^{-1}\boldsymbol{\phi}_k(x)'.$$

Since $\boldsymbol{a}_{k,i}$ is arbitrary, we see that we approximate $\mathbb{R}_{k,i}$ by $\boldsymbol{R}_{k,i} = \boldsymbol{M}_k^r\boldsymbol{M}_k^{-1}$, and $\mathbb{R}^k$ by $\boldsymbol{R}^k = diag(\boldsymbol{R}_{k,i}, i \in \mathcal{S}_k^{\bullet})$.

In practice, we implement a Gauss-Lobatto quadrature approximation to compute the elements of $\boldsymbol{M}_k^r$.

Remark 3.2. *We could also use interpolation to approximate* $\mathbb{R}$.

3.5.2 Approximating the operator $\mathbb{D}$ and the Riccati equation

Recalling Lemma 2.1 and replacing the operators $\mathbb{R}^k$ and $\mathbb{B}^{\ell m}$, by their approximations we have the following approximation to $\mathbb{D}^{mn}(s)$

$$\boldsymbol{D}^{mn}(s) = \left[\boldsymbol{R}^m\left(\boldsymbol{B}^{mn} - s\boldsymbol{I} + \boldsymbol{B}^{m0}\left(\boldsymbol{B}^{00} - s\boldsymbol{I}\right)^{-1}\boldsymbol{B}^{0n}\right)\right], \quad m,n \in \{+,-\}.$$

Let $\boldsymbol{\phi}_k(x)\boldsymbol{M}_k^{-1}\boldsymbol{\Psi}_{ij}^{k\ell}(s)\boldsymbol{\phi}_\ell(y)'\,\mathrm{d}y$, $i,j \in \mathcal{S}$, $k \in \mathcal{K}_i^+, \ell \in \mathcal{K}_j^-$ be a finite-dimensional approximation of the operator kernel $\Psi_{ij}^{k\ell}(s)(x,\,\mathrm{d}y)$, where $\boldsymbol{\Psi}_{ij}^{k\ell}(s)$ is a matrix of constants for a given s. Construct an approximation to $\Psi(s)(x,\,\mathrm{d}y)$ by

$$\boldsymbol{\phi}^+(x)\boldsymbol{M}_+^{-1}\boldsymbol{\Psi}(s)\boldsymbol{\phi}^-(y)'\,\mathrm{d}y = \boldsymbol{\phi}^+(x)\boldsymbol{M}_+\left[\left[\boldsymbol{\Psi}_{ij}^{k\ell}\right]_{i\in\mathcal{S}_k^+, j\in\mathcal{S}_\ell^-}\right]_{k\in\mathcal{K}^+,\ell\in\mathcal{K}^-}\boldsymbol{\phi}^-(y)'\,\mathrm{d}y$$

where $\boldsymbol{\phi}^+(x) = (\boldsymbol{\phi}_k(x))_{i\in\mathcal{S}_k^+,k\in\mathcal{K}^+}$ and $\boldsymbol{\phi}^-(y) = (\boldsymbol{\phi}_k(y))_{i\in\mathcal{S}_k^-,k\in\mathcal{K}^-}$ are row-vectors, $\boldsymbol{\Psi}(s)$ is a matrix of constants for a given s with the same size as $\boldsymbol{D}^{+-}$, and $\boldsymbol{M}_m$, $m \in \{+,-,0\}$ is a block diagonal matrix $\boldsymbol{M}_m = diag\,(\boldsymbol{M}_k, i \in \mathcal{S}_k^m, k \in \mathcal{K}^m)$, $m \in \{+,-,0\}$. Now replace the theoretical kernels in Theorem 2.2 by their DG approximations to get

$$\boldsymbol{\phi}^+(x)\boldsymbol{M}_+^{-1}\boldsymbol{D}^{+-}(s)\boldsymbol{\phi}^-(y)'\,\mathrm{d}y$$

$$
\begin{aligned}
&+ \int_{z_1,z_2} \boldsymbol{\phi}^+(x)\boldsymbol{M}_+^{-1}\boldsymbol{\Psi}(s)\boldsymbol{\phi}^-(z_1)'\boldsymbol{\phi}^-(z_1)\boldsymbol{M}_-^{-1}\boldsymbol{D}^{-+}(s)\boldsymbol{\phi}^+(z_2) \\
&\quad \times \boldsymbol{\phi}^+(z_2)\boldsymbol{M}_+^{-1}\boldsymbol{\Psi}(s)\boldsymbol{\phi}^-(y)'\,\mathrm{d}z_1\,\mathrm{d}z_2\,\mathrm{d}y \\
&+ \int_{z_1} \boldsymbol{\phi}^+(x)\boldsymbol{M}_+^{-1}\boldsymbol{D}^{++}(s)\boldsymbol{\phi}^+(z_1)'\boldsymbol{\phi}^+(z_1)\boldsymbol{M}_+^{-1}\boldsymbol{\Psi}(s)\boldsymbol{\phi}^-(y)'\,\mathrm{d}z_1\,\mathrm{d}y \\
&+ \int_{z_1} \boldsymbol{\phi}^+(x)\boldsymbol{M}_+^{-1}\boldsymbol{\Psi}(s)\boldsymbol{\phi}^-(z_1)'\boldsymbol{\phi}^-(z_1)\boldsymbol{M}_-^{-1}\boldsymbol{D}^{--}(s)\boldsymbol{\phi}^-(y)'\,\mathrm{d}z_1\,\mathrm{d}y = 0.
\end{aligned}
$$

Multiplying on the left by $\boldsymbol{\phi}^+(x)'$ and on the right by $\boldsymbol{\phi}^-(y)$, integrating over both x and y, then post-multiplying by $\boldsymbol{M}_-^{-1}$ gives the matrix Riccati equation

$$
\boldsymbol{D}^{+-}(s) + \boldsymbol{\Psi}(s)\boldsymbol{D}^{-+}(s)\boldsymbol{\Psi}(s) + \boldsymbol{D}^{++}(s)\boldsymbol{\Psi}(s) + \boldsymbol{\Psi}(s)\boldsymbol{D}^{--}(s) = \mathbf{0}. \tag{3.10}
$$

Thus, we may find $\boldsymbol{\Psi}(s)$ by solving (3.10), using one of the methods in (Bean et al. 2009*a*). Here, we use Newtons method.

Remark 3.3. *Given the stochastic interpretation of* $\Psi(0)$ *we know that* $\boldsymbol{\mu}^+\Psi(0)([0,\infty)) = 1$ *for every vector of measures* $\boldsymbol{\mu}^+$ *such that* $\boldsymbol{\mu}^+([0,\infty)\mathbf{1} = 1$*, when a fluid-fluid queue is recurrent. It appears that this result carries over to the matrix* $\boldsymbol{\Psi}(0)$ *giving the property that* $\int_{y\in[0,b]} \boldsymbol{\Psi}(0)\boldsymbol{\phi}^-(y)'\,\mathrm{d}y = \mathbf{1}$*. However, we have only observed this numerically and have no proof of this property.*

3.5.3 Putting it all together: constructing an approximation to the limiting distribution

We find an approximation to the limiting distribution by replacing the theoretical operators in Theorem 2.3 with their approximations. Table 3.1 defines the notation we use for the DG approximations to limiting operators.

With the notation in Table 3.1 define row-vectors

$$
\begin{aligned}
\boldsymbol{\xi}_k &:= (\boldsymbol{\xi}_{k,i})_{i\in\mathcal{S}_k^-}, \quad k \in \mathcal{K}_i^-, \\
\boldsymbol{\xi} &:= (\boldsymbol{\xi}_k)_{k\in\mathcal{K}^-}, \\
\boldsymbol{p}_k^m &:= (\boldsymbol{p}_{k,i})_{i\in\mathcal{S}_k^m}, \quad k \in \mathcal{K}^m,\, m \in \{-,0\}, \\
\boldsymbol{p}^m &:= (\boldsymbol{p}_k^m)_{k\in\mathcal{K}^m}, \quad m \in \{-,0\}, \\
\boldsymbol{p} &:= (\boldsymbol{p}^m)_{m\in\{-,0\}}, \\
\boldsymbol{\pi}_m^k(y) &:= (\boldsymbol{\pi}_{k,i}(y))_{i\in\mathcal{S}_k^m}, \quad k \in \mathcal{K},\, m \in \{+,-,0\}, \\
\boldsymbol{\pi}^m(y) &:= (\boldsymbol{\pi}_m^k(y))_{k\in\mathcal{K}^m}, \quad m \in \{+,-,0\}, \\
\boldsymbol{\pi}(y) &:= (\boldsymbol{\pi}^m(y))_{m\in\{+,-,0\}}.
\end{aligned}
$$

Exact	Operator indices	Approximation notation	Approximations
$\mathfrak{k}_{k,i}$	$i \in \mathcal{S}_k^-$, $k \in \mathcal{K}^-$	$\boldsymbol{\xi}_{k,i} := (\xi_{k,i}^r)_{r\in\mathcal{N}_k}$	$\mathfrak{k}_{k,i}(\,\mathrm{d}x) \approx \boldsymbol{\xi}_{k,i}\boldsymbol{\phi}^k(x)'\,\mathrm{d}x$,
$\mathbb{p}_{k,i}$	$i \in \mathcal{S}_k^- \cup \mathcal{S}_k^0$, $k \in \bigcup_{m\in\{-,0\}} \mathcal{K}_m$	$\boldsymbol{p}_{k,i} := (p_{k,i}^r)_{r\in\mathcal{N}_k}$	$\mathbb{p}_{k,i}(\,\mathrm{d}x) \approx \boldsymbol{p}_{k,i}\boldsymbol{\phi}_k(x)'\,\mathrm{d}x$
$\mathbb{m}_{k,i}(y)$	$i \in \mathcal{S}$, $k \in \mathcal{K}$	$\boldsymbol{\pi}_{k,i}(y) := (\pi_{k,i}^r(y))_{r\in\mathcal{N}_k}$	$\mathbb{m}_{k,i}(y)(\,\mathrm{d}x) \approx \boldsymbol{\pi}_{k,i}(y)\boldsymbol{\phi}_k(x)'\,\mathrm{d}x$

Table 3.1: Notation for the approximation of the limiting operators of a fluid-fluid queue. The first column contains the operators which we are approximating, the second column contains indices for which the operators are defined, the third column defines the notation we use for the coefficients of the approximation, and the last column defines how the coefficients and basis functions are used to approximate the operators.

Proceeding similarly to the derivation of the Riccati equation (3.10), we can argue that the coefficients $\boldsymbol{\xi}$ are the solution to the matrix system

$$\begin{aligned}\begin{bmatrix}\boldsymbol{\xi} & \mathbf{0}\end{bmatrix}\left(-\begin{bmatrix}\boldsymbol{B}^{--} & \boldsymbol{B}^{-0}\\ \boldsymbol{B}^{0-} & \boldsymbol{B}^{00}\end{bmatrix}\right)^{-1}\begin{bmatrix}\boldsymbol{B}^{-+}\\ \boldsymbol{B}^{0+}\end{bmatrix}\boldsymbol{\Psi}(0) &= \boldsymbol{\xi},\\ \int_{x\in[0,b]}\boldsymbol{\xi}\begin{bmatrix}\boldsymbol{\phi}^-(x)'\\ \boldsymbol{\phi}^0(x)'\end{bmatrix}\mathrm{d}x\mathbf{1} &= 1.\end{aligned}$$

Essentially, we replace the theoretical operators in (2.28) and (2.29) with their DG counterparts.

Similarly, the coefficients $\boldsymbol{p}$ are given by

$$\begin{bmatrix}\boldsymbol{p}^- & \boldsymbol{p}^0\end{bmatrix} = z\begin{bmatrix}\boldsymbol{\xi} & \mathbf{0}\end{bmatrix}\left(-\begin{bmatrix}\boldsymbol{B}^{--} & \boldsymbol{B}^{-0}\\ \boldsymbol{B}^{0-} & \boldsymbol{B}^{00}\end{bmatrix}\right)^{-1}, \tag{3.11}$$

where z is a normalising constant. The coefficients $\boldsymbol{\pi}(y)$ are given by

$$\boldsymbol{\pi}^0(y) = \begin{bmatrix}\boldsymbol{\pi}^+(y) & \boldsymbol{\pi}^-(y)\end{bmatrix}\begin{bmatrix}\boldsymbol{B}^{+0}\\ \boldsymbol{B}^{-0}\end{bmatrix}\left(-\boldsymbol{B}^{00}\right)^{-1}, \tag{3.12}$$

$$\begin{bmatrix}\boldsymbol{\pi}^+(y) & \boldsymbol{\pi}^-(y)\end{bmatrix} = \begin{bmatrix}\boldsymbol{p}^- & \boldsymbol{p}^0\end{bmatrix}\begin{bmatrix}\boldsymbol{B}^{-+}\\ \boldsymbol{B}^{0+}\end{bmatrix}\begin{bmatrix}e^{\boldsymbol{K}y} & e^{\boldsymbol{K}y}\boldsymbol{\Psi}(0)\end{bmatrix}\begin{bmatrix}\boldsymbol{R}^+ & 0\\ 0 & \boldsymbol{R}^-\end{bmatrix}, \tag{3.13}$$

$$\sum_{i\in\mathcal{S}}\sum_{k\in\mathcal{K}}\int_{y=0}^{\infty}\int_{x\in[0,b]}\boldsymbol{\pi}_{k,i}(y)\boldsymbol{\phi}_k(x)'\,\mathrm{d}x\,\mathrm{d}y \tag{3.14}$$

$$+\sum_{i\in\mathcal{S}}\sum_{\ell\in\{-,0\}}\sum_{k\in\mathcal{K}_i^\ell}\int_{x\in[0,b]}\boldsymbol{p}_{k,i}\boldsymbol{\phi}_k(x)'\,\mathrm{d}x = 1, \tag{3.15}$$

where $\boldsymbol{K} := \boldsymbol{D}^{++}(0) + \boldsymbol{\Psi}(0)\boldsymbol{D}^{(-+)}(0)$, and z is a normalising constant.

To assist the reader in understanding these constructions and the notation we provide an explicitly worked toy-example in Appendix A.

3.6 Other problems we can solve with cell-based approximation schemes

The utility of the DG scheme (or any of the approximation schemes considered in this thesis, for that matter) for fluid queues is not limited to approximating the first return and limiting operators of fluid-fluid queues. For example, we can use a DG scheme to approximate the transient distribution of the fluid queue at time t, which requires us to find the coefficients $\begin{bmatrix}\boldsymbol{q}_{-1}(t) & \boldsymbol{a}(t) & \boldsymbol{q}_{K+1}(t)\end{bmatrix}$, given an initial condition, typically by numerically integrating (3.9). An approximation to the limiting distribution of the fluid queue can be found by solving

$$\begin{aligned}\boldsymbol{b}\boldsymbol{B} &= \boldsymbol{0},\\ \boldsymbol{b}\boldsymbol{1} &= \boldsymbol{1},\end{aligned}$$

for the coefficients $\boldsymbol{b}$. We can also approximate first hitting times. For example, given an initial condition on cell k, we first find the initial coefficients $\boldsymbol{a}_k(0)$, then approximate the probability that the level process first hits $\{y_k, y_{k+1}\}$ after time t_0 by finding $\boldsymbol{a}_k(0)e^{B^{kk}t_0}\int_{y_k}^{y_{k+1}}\boldsymbol{\phi}_k(x)\,\mathrm{d}x$ where the coefficients $\boldsymbol{a}_k(t_0)$ can be found by integrating the differential equation

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{a}_k(t) = \boldsymbol{a}_k(t)\boldsymbol{B}^{kk},$$

over time.

Indeed, the DG scheme has the potential to answer many mathematical questions for fluid queues. However, one reason DG schemes may not have found such widespread use in the context of fluid queues is the difficulties surrounding discontinuous solutions, which we elaborate on in Section 2.8.2 and demonstrate in Chapter 7 (see also Section 5.6 Hesthaven & Warburton (2007)), which we often want to consider. Moreover, there are sometimes other techniques which can give exact solutions, for example exact expressions for transient quantities are derived in Rabehasaina & Sericola (2003), Ahn & Ramaswami (2004), Bean et al. (2005*b*, 2009*b*), and for the limiting distribution are derived in Ahn et al. (2005), da Silva Soares (2005), Latouche et al. (2013), Sonenberg (2017).

Nonetheless, the DG scheme can be used to answer a very broad range of questions about fluid queues (such as transient distributions, hitting times on complex geometries, and of course analysis of fluid-fluid queues) which have no exact, readily computable solution. Moreover, we will numerically investigate the performance of the DG scheme

(and other numerical schemes which we will develop) in Chapter 7 which require numerical integration with respect to time. Hence, we now briefly introduce the numerical time integration scheme which we use in this thesis, as well as some issues pertaining to oscillatory approximations.

3.7 Remarks on slope limiters, linear operators and application to fluid-fluid queues

The process of slope limiting, as described in Section 2.8.2, essentially amounts to a non-linear post-processing of the approximate solution after applying the operator $\boldsymbol{B}$. Thus, the slope-limited scheme arrives at an approximation which is the action of a non-linear operator applied to the initial condition. The fact that the approximate operator is non-linear makes solving the Riccati Equation (3.10) difficult/impossible, whereas there are efficient techniques to solve matrix-Riccati equations Bean et al. (2009*a*). Thus, for the application to fluid-fluid queues slope limiting is not a viable option. We could, however, proceed with a vanilla DG scheme (with no limiter) and, when we apply the operator $\boldsymbol{\Psi}$ to some initial condition and realise an approximation to the distribution of the fluid queue when $\{Y(t)\}$ first return to its initial level, we could then post-process the solution with a limiter or filter if oscillations are present in the approximation. The application of the limiter/filter is dependent on the initial condition, so the post-processing needs to be done for every initial condition under consideration. If there are many initial distributions to consider, this could result in an increased computational cost. We do not investigate this post-processing approach in this thesis.

On a related note, the advantage of a linear approximation can also be useful if we want to approximate transient distributions of fluid queues for many initial conditions. For example, consider solving (2.40) numerically with a forward-Euler scheme to integrate over time. To advance the coefficients from time t to time $t + h$ we compute

$$\boldsymbol{a}(t+h) = \boldsymbol{a}(t) + h\boldsymbol{a}(t)\boldsymbol{Q} = \boldsymbol{a}(t)\left[\boldsymbol{I} + h\boldsymbol{Q}\right].$$

Thus, the coefficients at time $t_0 = nh$ are given by

$$\boldsymbol{a}(nh) = \boldsymbol{a}(0)\left[\boldsymbol{I} + h\boldsymbol{Q}\right]^n,$$

which is a linear transformation of the coefficients $\boldsymbol{a}(0)$. Thus, if we determine the matrix $[\boldsymbol{I} + h\boldsymbol{Q}]^n$, then we may apply it to any initial condition to approximate the solution at time $t_0 = nh$. In the presence of a limiter, this is not possible as the coefficients at time $t_0 = nh$ are not a linear transformation of the initial coefficients.

Another point to note is that limiters incur a computational cost. There is a computational cost in reducing the solution to linear when oscillations are detected, and also in determining whether there are oscillations or not. Thus, even if the solution needs

no limiting, there is an added computational cost in determining this is the case. Given the limiter is applied for each stage and each time-step of the time integration scheme in Section 2.8.3, then this computational cost may not be insignificant.

One advantage of the QBD-RAP (and uniformisation) scheme we develop in this thesis is that the resulting operator is positivity-preserving and linear.

Chapter 4

A stochastic modelling approach to approximating fluid queues

One of the main issues with the DG scheme we introduced in the previous chapter is that discontinuities in the problem can introduce spurious oscillations in the approximate solution, making the solution take infeasible values, such as giving negative probabilities. In this chapter we develop a new discretisation of a fluid queue using a *quasi-birth-and-death process with rational arrival process components* (QBD-RAP) which, due to its stochastic interpretation, ensures that approximations obey the rules of probability. The discretisation is inspired by the QBD approximation of Bean & O'Reilly (2013*a*), which also ensures approximations obey the rules of probability due to its stochastic interpretation. However, the convergence of the approximation scheme can be slow. Thus, here we attempt to construct a new stochastic approximation of a fluid queue which converges faster than Bean & O'Reilly (2013*a*). In Chapters 5 and 6 we prove the scheme converges weakly to the distribution of the fluid queue.

To improve upon the discretisation of Bean & O'Reilly (2013*a*), we argue that the shift from QBD to QBD-RAP is necessary. We may view the QBD construction of Bean & O'Reilly (2013*a*) as using Erlang distributions to approximate certain deterministic events of the fluid queue. Specifically, the deterministic time that the fluid queue spends in an interval on the event that the phase (of the fluid queue) is constant is approximated by Erlang distributions. It is well known that the Erlang distribution is the least variable distribution of phase type (has the smallest coefficient of variation) of a given order (Aldous & Shepp 1987). In this sense, an Erlang distributed random variable is the best phase type approximation to a deterministic event, and therefore we argue that the QBD construction of Bean & O'Reilly (2013*a*) is also the best possible (or at least close to). Thus, if we are to improve upon Bean & O'Reilly (2013*a*) we must look past QBD processes.

Concentrated matrix exponential distributions are matrix exponential distributions with very low variance (relative to the mean) for a given order. As the order, p, of the

representation increases the variance of the concentrated matrix exponential distributions decreases at rate approximately $\mathcal{O}(1/p^2)$ Horváth, Horváth & Telek (2020). The variance of the most-concentrated Phase-type distribution, the Erlang distribution, decreases at rate $\mathcal{O}(1/p)$ as the order p increases Aldous & Shepp (1987). Thus, the variance of concentrated matrix exponential distributions decreases at a faster rate than the variance of Phase-type distributions as order increases, and we claim that this means concentrated matrix-exponential distributions have a superior ability to model determinism than any Phase-type distribution.

To improve upon the QBD of Bean & O'Reilly (2013*a*) this chapter constructs a QBD-RAP from concentrated matrix exponential. We show in Chapters 5 and 6 that, by using a matrix exponential distribution with sufficiently small variance in the construction of the QBD-RAP, the approximation can be made arbitrarily accurate. Due to its stochastic interpretation, the QBD-RAP discretisation ensures solutions are positive, even when discontinuities are present. Unlike slope limiters and post-hoc filtering, for the QBD-RAP the positivity preserving nature is a property of the discretised operator and the discretised operator is linear.

The structure of this chapter is as follows. Section 4.1 motivates the idea using the QBD approximation of Bean & O'Reilly (2014). Sections 4.2 and 4.3 describe the modelling of certain events of the fluid queue with matrix exponential distributions to ultimately construct the behaviour of the QBD-RAP on the event that the QBD-RAP remains in a given level. Section 4.4 describes the dynamics of the construction and also constructs the level process of the QBD-RAP. Section 4.5 deals with modelling boundary behaviour. Section 4.6 describes how to model initial conditions of the fluid queue. Section 4.7 introduces the concept of a *closing operator* which maps the state space of the QBD-RAP to an estimate of the density of the fluid queue.

4.1 Inspiration and motivation

The inspiration for the approach in the chapter stems from the QBD discretisation of Bean & O'Reilly (2013*a*), which they derive via uniformisation arguments. Recall that a QBD is a two-dimensional CTMC, $\{(L(t), \phi(t))\}_{t\geq 0}$, where $\{L(t)\}$ is the discrete level process, and $\{\varphi(t)\}$ is the phase process. In Bean & O'Reilly (2013*a*) the process $\{\varphi(t)\}$ of the fluid queue and the QBD approximation are the same, thus they are able to capture the dynamics of the phase process exactly. To approximate the level process of the fluid queue, Bean & O'Reilly (2013*a*) effectively discretise the state space of the fluid level into small intervals of width Δ. Their approximation supposes that when $L(t) = k$, $\varphi(t) = i$, then $X(t) \approx k\Delta$, $\varphi(t) = i$.

The dynamics of their approximation are as follows. When in level $k > 1$ and phase i, the approximation sees events at rate $|T_{ii}| + |c_i|\Delta$. Upon an event, with probability $T_{ij}/(|T_{ii}| + |c_i|\Delta)$ a change of phase from i to j occurs, and the approximation remains

in level k; with probability $|c_i|\Delta/(|T_{ii}| + |c_i|\Delta)$ a change of level, to $k+1$ if $c_i > 0$, or $k-1$ if $c_i < 0$, occurs. The generator of their QBD approximation (for an unbounded fluid queue) is

$$\boldsymbol{B} = \begin{bmatrix} \ddots & \ddots & \ddots & \\ \boldsymbol{B}_{-1}(\Delta) & \boldsymbol{B}_0(\Delta) & \boldsymbol{B}_{+1}(\Delta) & \\ & \boldsymbol{B}_{-1}(\Delta) & \boldsymbol{B}_0(\Delta) & \boldsymbol{B}_{+1}(\Delta) \\ & \ddots & \ddots & \ddots \end{bmatrix}$$

$$\boldsymbol{B}_0 = \boldsymbol{T} - \boldsymbol{C}\Delta, \qquad \boldsymbol{B}_{-1} = \begin{bmatrix} \mathbf{0} & & \\ & \boldsymbol{C}_-\Delta & \\ & & \mathbf{0} \end{bmatrix}, \qquad \boldsymbol{B}_{+1} = \begin{bmatrix} \boldsymbol{C}_+\Delta & & \\ & \mathbf{0} & \\ & & \mathbf{0} \end{bmatrix}. \tag{4.1}$$

Bean & O'Reilly (2013*a*) show that when we take $\Delta \to 0$ then the approximation converges weakly to the fluid queue. For a given discretisation level, Δ, when the phase of the QBD is i and on the event of no change of phase, the sojourn time in a given level has an exponential distribution with rate $|c_i|\Delta$. However, the corresponding sojourn time of the fluid queue is deterministic: given $X(t) = x \in \mathcal{D}_k$, then $\{X(t)\}$ will leave $\mathcal{D}_k$ in exactly $((k+1)\Delta - x)/|c_i|$ units of time if $c_i > 0$, and $(x - k\Delta)/|c_i|$ units of time if $c_i < 0$, on the event that the phase does not change before this time. At a first glance it may seem that we should be able to do better than Bean & O'Reilly (2013*a*), which approximates a deterministic event with an exponential random variable. However, on closer inspection it seems that this is the best we can do if we want to keep the interpretation of the approximation as a QBD. We now elaborate on this point.

In an attempt to extend the QBD model of Bean & O'Reilly (2013*a*) to model this determinism more accurately, consider using Erlang distributions of order 2 and mean Δ, rather than exponential distributions with mean Δ, to approximate the sojourn time of the fluid level in the intervals $\mathcal{D}_k$. Skipping over the details about what happens when there is a change of phase, we might arrive at the QBD with generator

$$\boldsymbol{B} = \begin{bmatrix} \ddots & \ddots & \ddots & \\ \boldsymbol{B}_{-1}(\Delta) & \boldsymbol{B}_0(\Delta) & \boldsymbol{B}_{+1}(\Delta) & \\ & \boldsymbol{B}_{-1}(\Delta) & \boldsymbol{B}_0(\Delta) & \boldsymbol{B}_{+1}(\Delta) \\ & \ddots & \ddots & \ddots \end{bmatrix}$$

$$\boldsymbol{B}_0 = \boldsymbol{T} \otimes I_2 + \begin{bmatrix} \begin{bmatrix} -\Delta/2 & \Delta/2 \\ 0 & -\Delta/2 \end{bmatrix} \otimes \boldsymbol{C}_+ & & \\ & \begin{bmatrix} -\Delta/2 & 0 \\ \Delta/2 & -\Delta/2 \end{bmatrix} \otimes \boldsymbol{C}_- & \\ & & \mathbf{0} \end{bmatrix}, \tag{4.2}$$

$$\boldsymbol{B}_{-1} = \begin{bmatrix} \mathbf{0} & & \\ & \boldsymbol{C}_- \otimes \begin{bmatrix} 0 & \Delta/2 \\ 0 & 0 \end{bmatrix} & \\ & & \mathbf{0} \end{bmatrix}, \quad \boldsymbol{B}_{+1} = \begin{bmatrix} \boldsymbol{C}_+ \otimes \begin{bmatrix} 0 & 0 \\ \Delta/2 & 0 \end{bmatrix} & & \\ & \mathbf{0} & \\ & & \mathbf{0} \end{bmatrix}. \quad (4.3)$$

But this is the same discretisation we would get if we were to use intervals of width $\Delta/2$ in the QBD approximation of Bean & O'Reilly (2013*a*), with the rows/columns reordered. The same arguments show that the QBD approximation with intervals of width Δ/p is equivalent constructing a QBD from Erlang distributions of order p to approximate the behaviour of the fluid queue over the intervals $\mathcal{D}_k$. Since the Erlang distribution is the least variable distribution of phase type, it appears the best we can do is the QBD approximation of Bean & O'Reilly (2013*a*). This motivates the extension to QBD-RAPs, whereby we can find more concentrated distributions than any Phase-type and retain enough stochastic interpretation and matrix-analytic tools.

Recall from Section 2.5.2 that a QBD-RAP is a two-dimensional process $\{(L(t), \boldsymbol{A}(t))\}$ where $\{L(t)\}$ is the level and $\boldsymbol{A}(t)$ is the orbit process. In this chapter we choose to use an equivalent, but slightly different, representation of the QBD-RAP, we write $\{(L(t), \boldsymbol{A}(t), \phi(t))\}_{t\geq 0}$, where $\{\phi(t)\}$ is the phase process. We are able to write our QBD-RAP in this way due to the Markovian nature of the phase process. The advantage of this representation is that it elucidates the connection between the phase processes of the QBD-RAP and the fluid queue. In some cases we can show even show that the phase processes of the QBD-RAP and fluid queue have the same distribution.

The construction of the approximating QBD-RAP is developed intuitively from a stochastic modelling perspective. The key observation is that, on the event that $\{\varphi(t)\}$ is constant, then $\{X(t)\}$ moves deterministically. Upon discretising the state space of the level process into intervals of width Δ, on the event that $\{\varphi(t)\}$ remains in the same phase then the distribution of time it takes for $\{X(t)\}$ to leave a given interval is deterministic, given $X(t)$. We start by modelling this deterministic behaviour approximately by concentrated matrix exponential distributions (Élteto et al. 2006, Horváth et al. 2016, Élteto et al. 2006, Horváth, Horváth & Telek 2020, Mészáros & Telek 2021).

Next, we argue how the orbit process $\{\boldsymbol{A}(t)\}$ must evolve upon a change of phase, and then upon a change of level. It turns out that the value of orbit process $\boldsymbol{A}(t)$ can be used to approximate how far $X(t)$ is from the left of a given interval when the phase is in $\mathcal{S}_+$, or how far from the right of the interval $X(t)$ is when the phase is in $\mathcal{S}_-$. Thus, we must determine, on the event of a transition from $\mathcal{S}_+$ to $\mathcal{S}_-$ (perhaps via $\mathcal{S}_0$) or $\mathcal{S}_-$ to $\mathcal{S}_+$ (perhaps via $\mathcal{S}_0$), how must $\boldsymbol{A}(t)$ jump to retain this information about where $X(t)$ is within a given interval. We also would like a way to use the orbit position, $\boldsymbol{A}(t)$, to obtain an approximation for the density of $X(t)$ within a given level. Answers to both problems can be derived from the *residual time* of a matrix exponential distribution. Let us now formalise these concepts some more.

Finally, we construct a level process for the QBD-RAP and map it to a discretisation of

the level process of the fluid queue. This is similar to Bean & O'Reilly (2013*a*) where the level process of their QBD corresponds a discretisation of the fluid level, however, unlike Bean & O'Reilly (2013*a*), however, here we do not take $\Delta \to 0$. Rather, we suppose that the variance of the concentrated matrix exponential distributions we use to approximate deterministic events gets small.

4.2 Time to exit an interval

For simplicity, first consider a fluid queue with $\mathcal{S}_0 = \emptyset$. Consider partitioning the state space of the level of a fluid queue, $[0, b]$ into $K + 1$ intervals of width $\Delta = b/(K + 1)$, specifically $\mathcal{D}_{k,i} = [k\Delta, (k + 1)\Delta)$ if $i \in \mathcal{S}_+$ and $\mathcal{D}_{k,i} = (k\Delta, (k + 1)\Delta]$ if $i \in \mathcal{S}_-$, $k \in \{0, 1, ..., K\} =: \mathcal{K}^\circ$. The distinction between $\mathcal{D}_{k,i}$ for $i \in \mathcal{S}_+$ and $\mathcal{S}_-$ is a technical one and will be discussed later. For now, one reason we might want to specify the discretisation in this way, is that it ensures the stochastic process

$$\left\{ \sum_{k \in \mathcal{K}} \sum_{i \in \mathcal{S}} k 1(X(t) \in \mathcal{D}_{k,i}) \right\}_{t \geq 0} \tag{4.4}$$

is right-continuous with left limits (cádlág). The expression (4.4) is the discrete process of the fluid-queue which the level process of the QBD-RAP will approximate. Let $y_k = k\Delta$, $k = 0, \ldots, K + 1$ and $\mathcal{D}_k = [y_k, y_{k+1}]$. We now look to model the behaviour of $\{X(t)\}$ on an interval $\mathcal{D}_k$ over the time $[t, t + u]$ on certain events of the phase process over this time.

4.2.1 Modelling the residual time to exit on no change of phase

First consider no change of phase in the time interval $[t, t + u]$ and suppose that $\varphi(t) = i \in \mathcal{S}_+$ and $X(t) = x \in \mathcal{D}_{k,i}$. On the event that there is no change of phase $\varphi(t + s) = i$ and $X(t + s) = x + c_i s$, $s \in [0, u]$. Thus, provided that u is sufficiently large, $u > (y_{k+1} - x)/c_i$, the level process of the fluid queue, $\{X(t)\}$, leaves the interval $\mathcal{D}_{k,i}$ at exactly time $t + (y_{k+1} - x)/c_i$. Similarly, for $\varphi(t) = i \in \mathcal{S}_-$, $\{X(t)\}$ leaves the interval $\mathcal{D}_{k,i}$ at exactly time $t + (x - y_k)/|c_i|$ on the event that there is no change of phase on $[t, t + u]$ for $u > (x - y_k)/|c_i|$.

To approximate the deterministic behaviour above, consider a matrix exponential distribution $Z \sim ME(\boldsymbol{\alpha}, \boldsymbol{S}, \boldsymbol{s})$ with mean Δ and low variance. As we shall see later, the assumption that Z has low variance will allow us to claim that Z will approximate deterministic behaviour. Let $Z_i \sim ME(\boldsymbol{\alpha}, \boldsymbol{S}_i, \boldsymbol{s}_i)$ where $\boldsymbol{S}_i = |c_i|\boldsymbol{S}$, $\boldsymbol{s}_i = -\boldsymbol{S}_i\boldsymbol{e}$, $i \in \mathcal{S}$. The random variable Z_i has mean $\Delta/|c_i|$.

Suppose first that $X(t) = y_k$ (we will build up to $X(t) = x_0 \in \mathcal{D}_k$ eventually, but for now assume $X(t) = y_k$) and $\varphi(t) = i \in \mathcal{S}_+$. In this case, $\{X(t)\}$ will leave $\mathcal{D}_{k,i}$

in exactly $\Delta/|c_i|$ units of time on the event that the phase process remains in i for at least $\Delta/|c_i|$ amount of time. A sensible approximation of the time until this deterministic event is Z_i. Thus, consider that at time t the value of the orbit process is $\boldsymbol{A}(t) = \boldsymbol{\alpha}$ and on the event that the phase does not change by time t the orbit evolves according to $\boldsymbol{A}(t+s) = \boldsymbol{\alpha}e^{\boldsymbol{S}_i s}/\boldsymbol{\alpha}(t)e^{\boldsymbol{S}_i s}\boldsymbol{e}$, $s \in [0,u]$ until there is an event. With this choice, the distribution of time until the next event of the QBD-RAP, on the event that the phase process remains constant will be the distribution of Z_i, which will be concentrated around $\Delta/|c_i|$. To see this, observe that

$$\begin{aligned}\mathbb{P}(Z_i \in ((\Delta-\varepsilon)/|c_i|, (\Delta+\varepsilon)/|c_i|)) &= \mathbb{P}(Z \in (\Delta-\varepsilon, \Delta+\varepsilon)) \\ &\geq 1 - \frac{\mathrm{Var}(Z)}{\varepsilon^2},\end{aligned}$$

since, by Chebyshev's inequality, $\mathbb{P}(Z \in (\Delta-\varepsilon, \Delta+\varepsilon)) \geq 1 - \dfrac{\mathrm{Var}(Z)}{\varepsilon^2}$. Choosing $\varepsilon = \mathrm{Var}(Z)^{1/3}$, then

$$\mathbb{P}(R_i(u) \in ((\Delta-\varepsilon)/|c_i| - u, (\Delta+\varepsilon)/|c_i|)) \geq 1 - \mathrm{Var}(Z)^{1/3} \approx 1,$$

when $\mathrm{Var}(Z)$ is small. Hence, when the variance of Z (equivalently Z_i) is low, then with high probability the time until the next event of the QBD-RAP, on the event that phase process remains constant will be approximately $\Delta/|c_i|$.

After $u \in [0, \Delta/|c_i|)$ amount of time has elapsed, on the event that the phase has not changed, $\varphi(t+s) = i$, for all $s \in [0,u)$, then $X(t+u) = y_k + c_i u \in \mathcal{D}_{k,i}$. If the phase remains i for a further $\Delta/|c_i| - u$ amount of time, then $\{X(t)\}$ will leave $\mathcal{D}_{k,i}$ in exactly $\Delta/|c_i| - u$ amount of time.

At time $t+u$, on the event that there is no change of phase, then the orbit position of the QBD-RAP will be $\boldsymbol{\alpha}e^{\boldsymbol{S}_i u}/\boldsymbol{\alpha}(t)e^{\boldsymbol{S}_i u}\boldsymbol{e}$ provided there has been no event of the QBD-RAP either, and the time until the next event of the QBD-RAP will be $R_i(u)$, where $R_i(u)$ is the residual time, $R_i(u) = (Z_i - u)1\{Z_i - u > 0\}$, $i \in \mathcal{S}$. Given $Z_i > u$, the distribution of the residual time, $R_i(u)$, has density

$$f_{R_i(u)}(r) = \frac{\boldsymbol{\alpha}e^{\boldsymbol{S}_i(u+r)}\boldsymbol{s}_i}{\boldsymbol{\alpha}e^{\boldsymbol{S}_i u}\boldsymbol{e}} = \boldsymbol{A}(t+u)e^{\boldsymbol{S}_i r}\boldsymbol{s}_i.$$

Here, the event $Z_i > u$ approximates the event $X(t+u) \in \mathcal{D}_{k,i}$, and we want the residual time, $R_i(u)$, to approximate the time until $\{X(t)\}$ leaves $\mathcal{D}_{k,i}$. That is, we want $R_i(u)$ to approximate a deterministic random variable at $\Delta/|c_i| - u$. To this end, we observe that for $\varepsilon > 0$ and $u < (\Delta-\varepsilon)/|c_i|$, then

$$\begin{aligned}&\mathbb{P}(R_i(u) \in ((\Delta-\varepsilon)/|c_i| - u, (\Delta+\varepsilon)/|c_i| - u)) \\ &= \mathbb{P}(Z_i \in ((\Delta-\varepsilon)/|c_i|, (\Delta+\varepsilon)/|c_i|))\end{aligned}$$

$$= \mathbb{P}(Z \in (\Delta - \varepsilon, \Delta + \varepsilon))$$
$$\geq 1 - \frac{\mathrm{Var}(Z)}{\varepsilon^2}.$$

Choosing $\varepsilon = \mathrm{Var}(Z)^{1/3}$, then

$$\mathbb{P}(R_i(u) \in ((\Delta - \varepsilon)/|c_i| - u, (\Delta + \varepsilon)/|c_i|)) \geq 1 - \mathrm{Var}(Z)^{1/3} \approx 1,$$

when $\mathrm{Var}(Z)$ is small. That is, when the variance of Z (equivalently Z_i) is low, the residual time $R_i(u)$ will be concentrated around $\Delta/|c_i| - u$, as required. Figure 4.1 gives an example of a density function for a Z_i with mean $\Delta = 1$ and $c_i = 1$, as well as the density function of $R_i(0.3)$ conditional on $Z_i > 0.3$ and $R_i(0.6)$ conditional on $Z_i > 0.6$, for comparison. Observe how the density of the residual life $R_i(0.3)$ conditional on $Z_i > 0.3$ is concentrated around $\Delta - 0.3 = 0.7$ and, similarly the density of the residual life $R_i(0.6)$ conditional in $Z_i > 0.6$ is concentrated around $\Delta - 0.6 = 0.4$.

For notational convenience, define the row-vector-valued function $\boldsymbol{k}(t) : \mathbb{R} \to \mathcal{A}$,

$$\boldsymbol{k}(t) = \frac{\boldsymbol{\alpha} e^{\boldsymbol{S}t}}{\boldsymbol{\alpha} e^{\boldsymbol{S}t} \boldsymbol{e}}, \tag{4.5}$$

and the operator $\boldsymbol{K}(t) : \mathcal{A} \to \mathbb{R}$, $t \geq 0$,

$$\boldsymbol{a}\boldsymbol{K}(t) = \frac{\boldsymbol{a} e^{\boldsymbol{S}t}}{\boldsymbol{a} e^{\boldsymbol{S}t} \boldsymbol{e}}, \tag{4.6}$$

for any $a \in \mathcal{A}$.

We interpret the position of the orbit $\boldsymbol{k}(|c_i|u)$ as corresponding to the fact that $X(t+u)$ is $|c_i|u$ units from the left-hand boundary of the interval $\mathcal{D}_{k,i}$ when $i \in \mathcal{S}_+$. This gives a heuristic argument as to how we can model the sojourn times in a given interval $\mathcal{D}_{k,i}$ on the event that the phase does not change. We can apply analogous arguments to heuristically develop a model for the sojourn time of the fluid queue in an interval, $\mathcal{D}_{k,i}$, $i \in \mathcal{S}_-$, on the event that the phase does not change. For $i \in \mathcal{S}_-$, the orbit position $\boldsymbol{A}(t + u) = \boldsymbol{k}(|c_i|u)$ is interpreted as corresponding to $X(t)$ being $|c_i|u$ units from the right-hand boundary of the interval $\mathcal{D}_{k,i}$, y_{k+1}.

In summary, when $\varphi(t) = i \in \mathcal{S}_+$ and $X(t) = y_k$, the orbit position should be $\boldsymbol{A}(t) = \boldsymbol{\alpha}$ and, on the event of no change of phase, should evolve according to $\boldsymbol{A}(t+u) = \boldsymbol{k}(|c_i|u)$ until an event occurs. At time $u < \Delta/|c_i|$ the time until the next event of the QBD-RAP, on the event of no change of phase, is $R_i(u)$ which is concentrated around $\Delta/|c_i| - u$. Accordingly, the orbit position $\boldsymbol{A}(t + u) = \boldsymbol{k}(|c_i|u)$ corresponds to fluid level $X(t) = y_k + c_i u$.

Similarly, when $\varphi(t) = i \in \mathcal{S}_-$ and $X(t) = y_{k+1}$, the orbit position should be $\boldsymbol{A}(t) = \boldsymbol{\alpha}$ and, on the event of no change of phase, should evolve according to $\boldsymbol{A}(t+u) = \boldsymbol{k}(|c_i|u)$ until an event occurs. At time $u < \Delta/|c_i|$ the time until the next event of the QBD-RAP, on the

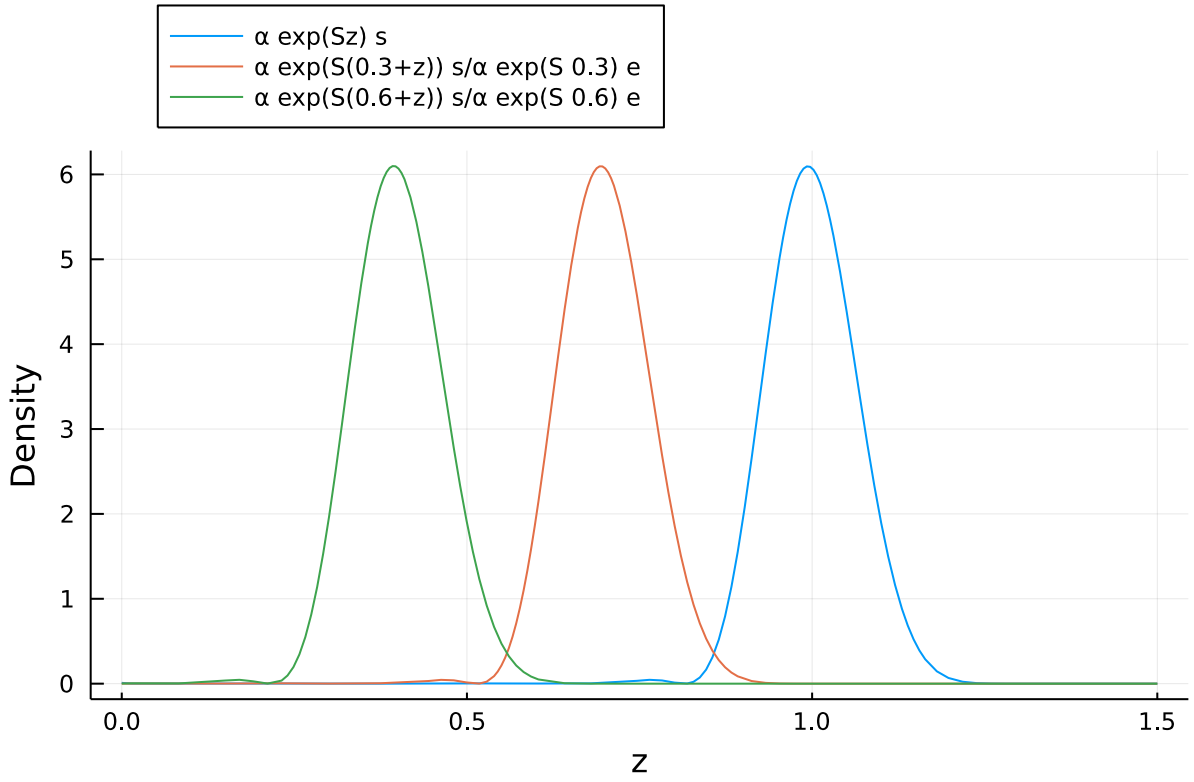

Figure 4.1: The density function for a *concentrated matrix exponential of order 21* from (Horváth, Horváth & Telek 2020) (blue) and corresponding density functions of the residual lives, $R_{0.3}$ (red), $R_{0.6}$ (blue). Observe how the density function of the Z_i (blue) approximates a point mass at $\Delta = 1$, while the density functions of $R_{0.3}$ (red) and $R_{0.6}$ (blue) approximate point masses at 0.7 and 0.4, respectively.

event of no change of phase, is $R_i(u)$ which is concentrated around $\Delta/|c_i|-u$. Accordingly, the orbit position $\boldsymbol{A}(t+u) = \boldsymbol{k}(|c_i|u)$ corresponds to fluid level $X(t) = y_{k+1} + c_i u$.

Thus, we have developed the dynamics of the QBD-RAP on the event that the phase is constant. We now develop the dynamics of the QBD-RAP on the event of a change of phase.

4.2.2 The residual time on a change of phase from $i \in \mathcal{S}_+$ to $j \in \mathcal{S}_+$

Let E be the event that, at time t the orbit of the QBD-RAP is $\boldsymbol{A}(t) = \boldsymbol{\alpha}$, the phase process is $\varphi(t) = i \in \mathcal{S}_+$, and there are no events of the QBD-RAP before time $t+u$. On the event E the orbit position at time $t+u$ will be

$$\boldsymbol{A}(t+u) = \boldsymbol{k}(|c_i|u) = \frac{\boldsymbol{\alpha} e^{\boldsymbol{S}_i u}}{\boldsymbol{\alpha} e^{\boldsymbol{S}_i u}\boldsymbol{e}}. \tag{4.7}$$

Correspondingly, on the event that at time t the level process of the fluid queue is $X(t) = y_k$, the phase is $\varphi(t) = i$ and there are no change of phase by time $t+u$, then $X(t+u) = y_k + c_i u$. That is, $X(t+u)$ is $c_i u$ units from the left-hand edge of $\mathcal{D}_k$.

If, at time $t+u$ there is a change of phase from $i \in \mathcal{S}_+$ to $j \in \mathcal{S}_+$, then we need to map the orbit position $\boldsymbol{A}(t+u^-) = \boldsymbol{k}(|c_i|u)$ to an orbit position which corresponds to the being $c_i u$ units from the left-hand edge of $\mathcal{D}_k$ in phase j. Call this mapping $\boldsymbol{D}(i,j)(\cdot)$. As noted in Section 2.5.2, the mapping must be linear, $\boldsymbol{D}(i,j)(\boldsymbol{a}) = \boldsymbol{a}\boldsymbol{D}(i,j)$, $\boldsymbol{a} \in \mathcal{A}$, for some matrix $\boldsymbol{D}(i,j)$. The matrix $\boldsymbol{D}(i,j)$ is a modelling choice.

So that the process is a QBD-RAP, the matrix $\boldsymbol{D}(i,j)$ must have the property that, for any $\boldsymbol{a} \in \mathcal{A}$, $(\boldsymbol{a}\boldsymbol{D}(i,j), \boldsymbol{S})$ is a representation of a matrix exponential distribution. We would also like the matrix $\boldsymbol{D}(i,j)$ to have the property $\boldsymbol{D}(i,j)\boldsymbol{e} = \boldsymbol{e}$ as this will mean that the rate at which a change of phase from i to j of the QBD-RAP occurs proportional to

$$\boldsymbol{A}(t+u)\boldsymbol{D}(i,j)\boldsymbol{e} = 1,$$

which is constant and therefore the distribution of time until a change from phase i to j is exponential. We can use this to show, for certain models, the distribution of the phase process of the fluid queue and the distribution of the phase process of the QBD-RAP share the same distribution.

An orbit position which corresponds to the level process of the QBD-RAP being $c_i u$ units to the right of y_k in phase $j \in \mathcal{S}_+$ is

$$\boldsymbol{A}(t+u) = \frac{\boldsymbol{\alpha} e^{\boldsymbol{S}_j(|c_i|u/|c_j|)}}{\boldsymbol{\alpha} e^{\boldsymbol{S}_j(|c_i|u/|c_j|)u}\boldsymbol{e}}.$$

To see this, consider the event that at time t the orbit process of the QBD-RAP is $\boldsymbol{A}(t) = \boldsymbol{\alpha}$, the phase process is $\varphi(t) = j \in \mathcal{S}_+$, and there are no events of the QBD-RAP

before time $t + |c_i|u/|c_j|$. On this event, the orbit position at time $t + u$ will be

$$\boldsymbol{A}(t+u) = \frac{\boldsymbol{\alpha} e^{\boldsymbol{S}_j(|c_i|u)/|c_j|}}{\boldsymbol{\alpha} e^{\boldsymbol{S}_j(|c_i|u/|c_j|)u}\boldsymbol{e}}.$$

Correspondingly, on the event that at time t the level process of the fluid queue is $X(t) = y_k$, the phase is $\varphi(t) = j$ and there are no change of phase by time $t + |c_i|u/|c_j|$, then $X(t+u) = y_k + c_j|c_i|u/|c_j| = y_k + c_i u$. That is, $X(t+u)$ is $c_i u$ units from the left-hand edge of $\mathcal{D}_k$.

Now, observe

$$\frac{\boldsymbol{\alpha} e^{\boldsymbol{S}_j(|c_i|u/|c_j|)}}{\boldsymbol{\alpha} e^{\boldsymbol{S}_j(|c_i|u/|c_j|)u}\boldsymbol{e}} = \frac{\boldsymbol{\alpha} e^{\boldsymbol{S}_i u}}{\boldsymbol{\alpha} e^{\boldsymbol{S}_i u}\boldsymbol{e}} = \boldsymbol{k}(|c_i|u),$$

which is exactly (4.7), the orbit position on the event E. Hence, it is sensible to choose $\boldsymbol{D}(i,j) = \boldsymbol{I}$.

Moreover, the residual time $R_j(|c_i|u/|c_j|)$ has density

$$f_{R_j(|c_i|u/|c_j|)}(r) = \frac{\boldsymbol{\alpha} e^{\boldsymbol{S}_i u}}{\boldsymbol{\alpha} e^{\boldsymbol{S}_i u}\boldsymbol{e}} e^{\boldsymbol{S}_j r}\boldsymbol{s}_j,$$

and

$$\begin{aligned}
&\mathbb{P}(R_j(|c_i|u/|c_j|) \in ((\Delta - |c_i|u - \varepsilon)/|c_j|, (\Delta - |c_i|u + \varepsilon)/|c_j|)) \\
&= \mathbb{P}(Z_j \in ((\Delta - \varepsilon)/|c_j|, (\Delta + \varepsilon)/|c_j|)) && (4.8)\\
&= \mathbb{P}(Z \in (\Delta - \varepsilon, \Delta + \varepsilon)) && (4.9)\\
&\geq 1 - \mathrm{Var}(Z)^{1/3} \approx 1. && (4.10)
\end{aligned}$$

Hence, when the variance of Z is low, the residual time, $R_j(|c_i|u/|c_j|)$, is concentrated around $(\Delta - |c_i|u)/|c_j|$, as required.

Analogous arguments suggest that the same applies for changes of phase from $i \in \mathcal{S}_-$ to $j \in \mathcal{S}_-$.

4.2.3 The residual time on a change of phase from $i \in \mathcal{S}_+$ to $j \in \mathcal{S}_-$

Now suppose $X(t) = y_k$, $\varphi(t) = i \in \mathcal{S}_+$, and the phase remains in state i until there is a change of phase from $i \in \mathcal{S}_+$ to $j \in \mathcal{S}_-$ at time $t + u$, $u \in [0, \Delta/|c_i|)$. As before, we need to find a matrix $\boldsymbol{D}(i,j)$ to map the orbit position on the event E when there is a change of phase from $i \in \mathcal{S}_+$ to $j \in \mathcal{S}_-$ at time $t + u$, from

$$\boldsymbol{A}(t + u^-) = \boldsymbol{k}(|c_i|u^-),$$

to

$$\boldsymbol{A}(t+u) = \frac{\boldsymbol{k}(u^-)\boldsymbol{D}(i,j)}{\boldsymbol{k}(u^-)\boldsymbol{D}(i,j)\boldsymbol{e}} = \frac{\boldsymbol{\alpha}e^{\boldsymbol{S}_i u}\boldsymbol{D}(i,j)}{\boldsymbol{\alpha}e^{\boldsymbol{S}_i u}\boldsymbol{D}(i,j)\boldsymbol{e}}$$

in such a way that the orbit position after the jump corresponds, in some way, to the fluid level being a distance of $|c_i|u$ from y_k in phase $j \in \mathcal{S}_-$.

Again, the matrix $\boldsymbol{D}(i,j)$ is a modelling choice. We first discuss how we might choose the matrix $\boldsymbol{D}(i,j)$ for when the matrix exponential Z is a phase-type distribution.

The Phase-type case If $Z \sim ME(\boldsymbol{\alpha}, \boldsymbol{S})$ is chosen to be a phase-type distribution then Z has the interpretation as the distribution of time until absorption of a finite-state continuous-time Markov chain with transient states $\{1, \ldots, p\}$ and a single absorbing state. The sub-generator matrix describing the dynamics of the Markov chain on transient states is $\boldsymbol{S}$, and $\boldsymbol{\alpha}$ is an initial probability distribution over the transient states. Let $\{J(t)\}$ be the Markov chain associated with the phase-type distribution.

In the discussions above, we have relied on the relationship between the event E and the orbit position $\boldsymbol{k}(|c_i|u)$. The relationship allows us to associate this orbit position with the level of the fluid queue. For phase-type distributions the vector $\boldsymbol{k}(|c_i|u)$ is the vector of posterior probabilities $[\mathbb{P}(J(u) = k \mid E)]_{k\in\{1,\ldots,p\}}$. We can use this vector of posterior probabilities in the same way as we do for matrix exponential distributions and QBD-RAPs. However, the orbit interpretation forgoes the presence of the phase process $\{J(t)\}$. Here, we will use the phase process $\{J(t)\}$ to derive a choice of $\boldsymbol{D}(i,j)$.

Recall the position of the orbit process can be used to determine the residual time. This idea can be replicated when we are given the phase of the phase-type rather than the value of the orbit process. Given the phase of the phase-type distribution is $k \in \{1, ..., p\}$, the distribution of the residual time is

$$\mathbb{P}(R_i \leq r \mid \text{phase} = k) = 1 - \boldsymbol{e}_k e^{\boldsymbol{S}_i r}\boldsymbol{e}.$$

Notice that this is independent of the time since the phase-type distribution was initialised. We call the time since the phase-type distribution was initialised the *age*. As noted by Hautphenne et al. (2017), the distribution of the age, given the phase of the phase-type is $k \in \{1, ..., p\}$, depends on the sampling scheme which determines the observation time. Here, on the event that at time t the phase of the QBD-RAP is i, a change of phase occurs after an exponential amount of time with rate $-T_{ii}$. Proposition 4.1, Hautphenne et al. (2017) states that the distribution function of the age, given the phase of the phase-type is k and the process is observed after an exponential time with rate $-T_{ii}$ is

$$\mathbb{P}(\text{age} \leq u \mid \text{phase} = k) = 1 - \frac{\boldsymbol{\alpha}e^{(\boldsymbol{S}_i + T_{ii}\boldsymbol{I})u}(-(\boldsymbol{S}_i + T_{ii}\boldsymbol{I}))^{-1}\boldsymbol{e}_k}{\boldsymbol{\alpha}(-(\boldsymbol{S}_i + T_{ii}\boldsymbol{I}))^{-1}\boldsymbol{e}_k}.$$

Let $\widehat{\boldsymbol{S}}_i(T_{ii}) = diag(\boldsymbol{\nu})^{-1}\boldsymbol{S}_i' diag(\boldsymbol{\nu})$, where $\boldsymbol{\nu} = \boldsymbol{\alpha}(-(\boldsymbol{S}_i + T_{ii}\boldsymbol{I}))^{-1}/(\boldsymbol{\alpha}(-(\boldsymbol{S}_i + T_{ii}\boldsymbol{I}))^{-1}\boldsymbol{e})$. Algebraic manipulations show

$$1 - \frac{\boldsymbol{\alpha}e^{(\boldsymbol{S}_i+T_{ii}\boldsymbol{I})u}(-(\boldsymbol{S}_i + T_{ii}\boldsymbol{I}))^{-1}\boldsymbol{e}_k}{\boldsymbol{\alpha}(-(\boldsymbol{S}_i + T_{ii}\boldsymbol{I}))^{-1}\boldsymbol{e}_k} = 1 - \boldsymbol{e}_k' e^{(\widehat{\boldsymbol{S}}_i(T_{ii})+T_{ii}\boldsymbol{I})u}\boldsymbol{e}, \tag{4.11}$$

which is of Phase-type.

Recall that in the matrix exponential case we associated the value of the orbit process $\boldsymbol{k}(|c_i|u)$ with being $|c_i|u$ units to the right of y_k. In contrast, for the phase-type case if we use the phase process $\{J(t)\}$ to inform our modelling of the fluid queue (as opposed to the value of the orbit process), then we associate phase $k \in \{1, ..., p\}$ of $\{J(t)\}$ with being a *random* distance to the right of y_k where this distance has distribution

$$1 - \boldsymbol{e}_k' e^{(\widehat{\boldsymbol{S}}_i(T_{ii})+T_{ii}\boldsymbol{I})u}\boldsymbol{e}. \tag{4.12}$$

Therefore, on the event E, upon a change from of phase of the QBD-RAP from $i \in \mathcal{S}_+$ to $j \in \mathcal{S}_-$ at time $t + u$, it seems reasonable to want the distribution of time until the next event of the QBD-RAP to be

$$1 - \boldsymbol{e}_k' e^{(\widehat{\boldsymbol{S}}_i(T_{ii})+T_{ii}\boldsymbol{I})|c_j|u/|c_i|+T_{jj}Iu}\boldsymbol{e}. \tag{4.13}$$

The factor $|c_j|u/|c_i|$ arises as a conversion between the speed at which the fluid level moves in phase j compared to phase i. The rate T_{jj} in the exponent is the rate at which the next change of phase occurs.

While this does achieve what we want, it is not quite satisfactory for the purpose of our approximation scheme due to dependence on the sample path of $\{\varphi(t)\}$. Specifically, the evolution of the QBD-RAP from time $t+u$ until the next event depends on the phase immediately before the change of phase at time $t + u$, $\varphi(t + u^-) = i$. This increases the size of the approximating QBD-RAP as we need a separate model for each $\varphi(t+u^-) \in \mathcal{S}_+$. Furthermore, we have not yet considered how to model any further changes from $\mathcal{S}_-$ to $\mathcal{S}_+$ or beyond, which further complicates matters.

A solution is to suppose that, rather than observing the phase-type random variable at an exponential time with rate $-T_{ii}$, we instead observe the process uniformly randomly on the lifetime of length Z_i. Let $\boldsymbol{\Pi} = diag\,(\boldsymbol{\pi})$, where $\left[\pi_k\right]_{k\in\{1,...,p\}} =: \boldsymbol{\pi} = \boldsymbol{\alpha}(-\boldsymbol{S})^{-1}/m$ and $m = \boldsymbol{\alpha}(-\boldsymbol{S})^{-1}\boldsymbol{e}$. There is a time-reverse representation of a Phase-type distribution given by $(\widetilde{\boldsymbol{\alpha}}, \widetilde{\boldsymbol{S}}, \widetilde{\boldsymbol{s}})$, where $\widetilde{\boldsymbol{\alpha}} = m\boldsymbol{s}'\boldsymbol{\Pi}$, $\widetilde{\boldsymbol{S}} = \boldsymbol{\Pi}^{-1}\boldsymbol{S}'\boldsymbol{\Pi}$ and $\widetilde{\boldsymbol{s}} = \boldsymbol{\Pi}^{-1}\boldsymbol{\alpha}'/m$ (Asmussen 2008, Page 91). In the case where the phase is observed randomly on the lifetime of Z_i, the distribution function of the age, given the phase is k is (Hautphenne et al. 2017, Lemma 3.1)

$$\mathbb{P}(\text{age} \leq u \mid \text{phase} = k) = 1 - \frac{\boldsymbol{\alpha}e^{\boldsymbol{S}_i u}(-\boldsymbol{S}_i)^{-1}\boldsymbol{e}_k}{\boldsymbol{\alpha}(-\boldsymbol{S}_i)^{-1}\boldsymbol{e}_k} = 1 - \boldsymbol{e}_k e^{\widetilde{\boldsymbol{S}}_i u}\boldsymbol{e}. \tag{4.14}$$

With this interpretation, we may associate phase $k \in \{1, ..., p\}$ of $\{J(t)\}$ with being a *random* distance, with distribution (4.14), to the right of y_k.

Therefore, on the event E and on a change of phase from $i \in \mathcal{S}_+$ to $j \in \mathcal{S}_-$ at time $t+u$, a reasonable model for the time until the next event of the QBD-RAP has distribution

$$1 - \boldsymbol{e}_k e^{\widetilde{\boldsymbol{S}}_i u|c_j|/|c_i| + T_{jj}\boldsymbol{I}u}\boldsymbol{e} = 1 - \boldsymbol{e}_k e^{\widetilde{\boldsymbol{S}}_j u + T_{jj}\boldsymbol{I}u}\boldsymbol{e}. \tag{4.15}$$

This suggests that, at a jump from $\mathcal{S}_+$ to $\mathcal{S}_-$, the state of $\{J(t)\}$ does not change, but begins to evolve according to the time-reverse generator at an appropriate speed $\widetilde{\boldsymbol{S}}_j$. Since the time-reverse of $\widetilde{\boldsymbol{S}}$ is $\boldsymbol{S}$, then upon a jump back to $\mathcal{S}_+$ from $\mathcal{S}_-$, the phase of the phase-type random variable remains k but begins to evolve according to $\boldsymbol{S}$. This suggests that we use the representation $(\boldsymbol{\alpha}, \boldsymbol{S})$ when in phases in $\mathcal{S}_+$, and use the time-reverse representation $(\widetilde{\boldsymbol{\alpha}}, \widetilde{\boldsymbol{S}})$ when in phases in $\mathcal{S}_-$. The matrices $\boldsymbol{D}(i,j) = \boldsymbol{I}$ for all $i, j \in \mathcal{S}$.

With this construction, and choosing $Z \sim Erlang(p, \Delta/p)$, we recover the discretisation of Bean & O'Reilly (2013*a*) with discretisation parameter Δ/p.

The matrix exponential case For matrix exponential distributions we cannot rely on the phase process $\{J(t)\}$ as we did in the phase-type case (because $\{J(t)\}$ does not exist for matrix exponential distributions). Recall that we want to find a matrix $\boldsymbol{D}(i,j)$ to map the orbit position on the event E when there is a change of phase from $i \in \mathcal{S}_+$ to $j \in \mathcal{S}_-$ at time $t+u$, from $\boldsymbol{A}(t+u^-) = \boldsymbol{k}(|c_i|u)$, to

$$\boldsymbol{A}(t+u) = \frac{\boldsymbol{k}(|c_i|u)\boldsymbol{D}(i,j)}{\boldsymbol{k}(|c_i|u)\boldsymbol{D}(i,j)\boldsymbol{e}} = \frac{\boldsymbol{\alpha}e^{\boldsymbol{S}_i u}\boldsymbol{D}(i,j)}{\boldsymbol{\alpha}e^{\boldsymbol{S}_i u}\boldsymbol{D}(i,j)\boldsymbol{e}}$$

in such a way that the orbit position after the jump corresponds, in some way, to the fluid level being a distance of $|c_i|u$ from y_k in phase j.

Given our interpretation of the orbit position, $\boldsymbol{k}(x)$, a solution would be to find a linear map which takes $\boldsymbol{k}(x)$ and maps it directly to

$$\boldsymbol{k}(\Delta - x) = \frac{\boldsymbol{\alpha}e^{\boldsymbol{S}(\Delta - x)}}{\boldsymbol{\alpha}e^{\boldsymbol{S}(\Delta - x)}\boldsymbol{e}}.$$

However, we have been unsuccessful in finding such a mapping with the main hurdle being that the map must be linear. Instead, we approximate as follows.

Recall that we use $R_i(u)$ to approximate the distance of the fluid level to the left of y_{k+1}. Suppose we are given $R_i(u) = r$ which corresponds, approximately, to the fluid level being $|c_i|r$ units to the left of y_{k+1}. The position of the orbit process which corresponds to a distance of $|c_i|r$ to the left of y_{k+1} in phase j is $\boldsymbol{k}(|c_i|r)$. Hence, on $R_i(u) = r$, the event E and on the event that a change of phase from $i \in \mathcal{S}_+$ to $j \in \mathcal{S}_-$ occurs at time $t+u$, the orbit position should jump from $\boldsymbol{k}(|c_i|u)$ to $\boldsymbol{k}(|c_i|r)$. However, at time $t+u$,

$R_i(u)$ is a random variable about the future of the process and therefore not known, so instead, we take the expected initial vector

$$\mathbb{E}\left[\boldsymbol{k}(|c_i|R_i(u))\right] = \int_{r=0}^{\infty} \frac{\boldsymbol{\alpha}e^{\boldsymbol{S}_i u}}{\boldsymbol{\alpha}e^{\boldsymbol{S}_i u}\boldsymbol{e}} e^{\boldsymbol{S}_i r}\boldsymbol{s}_i \boldsymbol{k}(|c_i|r)\,\mathrm{d}r = \frac{\boldsymbol{\alpha}e^{\boldsymbol{S}_i u}}{\boldsymbol{\alpha}e^{\boldsymbol{S}_i u}\boldsymbol{e}} \int_{r=0}^{\infty} e^{\boldsymbol{S}_i r}\boldsymbol{s}_i \frac{\boldsymbol{\alpha}e^{\boldsymbol{S}_i r}}{\boldsymbol{\alpha}e^{\boldsymbol{S}_i r}\boldsymbol{e}}\,\mathrm{d}r.$$

After a change of variables $x = |c_i|r$ we get

$$\frac{\boldsymbol{\alpha}e^{\boldsymbol{S}_i u}}{\boldsymbol{\alpha}e^{\boldsymbol{S}_i u}\boldsymbol{e}} \int_{x=0}^{\infty} e^{\boldsymbol{S}x}|c_i|\boldsymbol{s}\frac{\boldsymbol{\alpha}e^{\boldsymbol{S}x}}{\boldsymbol{\alpha}e^{\boldsymbol{S}x}\boldsymbol{e}}\,\mathrm{d}x/|c_i| = \boldsymbol{A}(t+u^-)\int_{x=0}^{\infty} e^{\boldsymbol{S}x}\boldsymbol{s}\frac{\boldsymbol{\alpha}e^{\boldsymbol{S}x}}{\boldsymbol{\alpha}e^{\boldsymbol{S}x}\boldsymbol{e}}\,\mathrm{d}x,$$

since at time $t+u^-$, the orbit position is $\boldsymbol{A}(t+u^-) = \boldsymbol{\alpha}e^{\boldsymbol{S}_i u}/\boldsymbol{\alpha}e^{\boldsymbol{S}_i u}\boldsymbol{e}$. Thus, we find

$$\boldsymbol{D}(i,j) = \int_{x=0}^{\infty} e^{\boldsymbol{S}x}\boldsymbol{s}\boldsymbol{k}(x)\,\mathrm{d}x =: \boldsymbol{D}.$$

Observe that $\boldsymbol{D}\boldsymbol{e} = \int_{x=0}^{\infty} e^{\boldsymbol{S}x}\boldsymbol{s}\boldsymbol{k}(x)\,\mathrm{d}x\boldsymbol{e} = \int_{x=0}^{\infty} e^{\boldsymbol{S}x}\boldsymbol{s}\,\mathrm{d}x = \boldsymbol{e}$, since $\boldsymbol{k}(x)\boldsymbol{e} = 1$ for all $x \geq 0$. Further, since $\mathcal{A}$ is closed and convex (Bladt & Nielsen 2017), then $(\boldsymbol{a}\boldsymbol{D}, \boldsymbol{S}, \boldsymbol{s})$ is a representation of a matrix exponential distribution for any $\boldsymbol{a} \in \mathcal{A}$.

We pose the choice of the matrix $\boldsymbol{D}$ as a modelling choice. Other choices are possible, for example, $\boldsymbol{D} = \int_{x=0}^{\Delta} e^{\boldsymbol{S}x}\boldsymbol{s}\boldsymbol{k}(x)\,\mathrm{d}x + \int_{x=\Delta}^{\infty} e^{\boldsymbol{S}x}\boldsymbol{s}\boldsymbol{\alpha}\,\mathrm{d}x$, or $\boldsymbol{D} = \int_{x=0}^{\Delta} \boldsymbol{v}(x)\boldsymbol{k}(x)\,\mathrm{d}x$, where $\boldsymbol{v}(x)$ is a *closing operator* as introduced later in Section 4.7, are other possible choices. It may also be possible to construct other matrices $\boldsymbol{D}$, perhaps via geometric arguments. We do not investigate other choices of $\boldsymbol{D}$ in this thesis.

Computing $\boldsymbol{D}$ In practice, we use the class of concentrated matrix exponential distributions (CMEs) found numerically in (Horváth, Horváth & Telek 2020). Recall that we take the index p to be the order of the representation of the CME and consider p odd. Also recall that for a given CME with odd order, p, and representation $(\boldsymbol{\alpha}, \boldsymbol{S})$, the matrix $\boldsymbol{S}$ has one real eigenvalue, and $p-1$ complex eigenvalues and all eigenvalues have the same real part. Further, the vector function $\boldsymbol{k}(t)$ is periodic with period $\rho = 2\pi/\omega$ where $\omega = min_i(|\Im(\lambda_i)|)$, λ_i are the eigenvalues of $\boldsymbol{S}$ and $\Im(z)$ is the imaginary component of a complex number z.

We numerically evaluate the matrix $\boldsymbol{D}$ where

$$\boldsymbol{D} = \int_{t=0}^{\infty} e^{\boldsymbol{S}t}\boldsymbol{s}\cdot\boldsymbol{k}(t)\,\mathrm{d}t$$

using a trapezoidal rule as follows. Let $\boldsymbol{f}(t) = e^{\boldsymbol{S}t}\boldsymbol{s}$. Then $\boldsymbol{f}(t)e^{-\lambda t}$, where $\lambda = \Re(\lambda_i)$ is the real part of the eigenvalues of $\boldsymbol{S}$, (they all share the same real part), is also periodic with the period ρ. Hence, we can simplify the integral to a finite one;

$$\boldsymbol{D} = \int_{t=0}^{\infty} \boldsymbol{f}(t)\cdot\boldsymbol{k}(t)\,\mathrm{d}t$$

$$= \sum_{k=0}^{\infty} \int_{k\rho}^{(k+1)\rho} e^{\lambda t} e^{-\lambda t} \boldsymbol{f}(t) \cdot \boldsymbol{k}(t) \, \mathrm{d}t$$
$$= \sum_{k=0}^{\infty} \int_{0}^{\rho} e^{\lambda(k\rho+t)} e^{-\lambda(k\rho+t)} \boldsymbol{f}(k\rho+t) \cdot \boldsymbol{k}(k\rho+t) \, \mathrm{d}t. \tag{4.16}$$

By periodicity, then $e^{-\lambda(k\rho+t)} \boldsymbol{f}(k\rho+t) \cdot \boldsymbol{k}(k\rho+t) = e^{-\lambda t} \boldsymbol{f}(t) \cdot \boldsymbol{k}(t)$, hence (4.16) is equal to

$$\sum_{k=0}^{\infty} (e^{\lambda\rho})^k \int_0^{\rho} e^{\lambda t} e^{-\lambda t} \boldsymbol{f}(t) \cdot \boldsymbol{k}(t) \, \mathrm{d}t = \frac{1}{1 - e^{\lambda\rho}} \int_0^{\rho} \boldsymbol{f}(t) \cdot \boldsymbol{k}(t) \, \mathrm{d}t, \tag{4.17}$$

where the sum converges as it is a geometric series and $\lambda < 0$, $\rho > 0$.

To approximate (4.17) numerically, we first partition $[0, \rho)$ into N equal-width intervals $[t_n, t_{n+1})$, where $t_n = (n-1)\rho/N$, $n = 1, 2, ..., N+1$. On $[t_n, t_{n+1})$ we approximate the orbit $\boldsymbol{k}(t)$ by a constant $\boldsymbol{k}(t) \approx \boldsymbol{k}_n := \frac{1}{2}(\boldsymbol{k}(t_n) + \boldsymbol{k}(t_{n+1}))$, $t \in [t_n, t_{n+1})$. Substituting this approximation into the expression for $\boldsymbol{D}$ gives

$$\boldsymbol{D} \approx \frac{1}{1 - e^{\lambda\rho}} \sum_{n=1}^{N} \int_{t_n}^{t_{n+1}} \boldsymbol{f}(t) \cdot \boldsymbol{k}_n \, \mathrm{d}t$$
$$= \frac{1}{1 - e^{\lambda\rho}} \sum_{n=1}^{N} \left[e^{\boldsymbol{S}t_{n+1}} - e^{\boldsymbol{S}t_n} \right] \boldsymbol{e} \cdot \boldsymbol{k}_n.$$

This approximation preserves the property that $\boldsymbol{D}\boldsymbol{e} = \boldsymbol{e}$.

Computationally efficient expressions for $e^{\boldsymbol{S}t}\boldsymbol{e}$ and $\boldsymbol{k}(t)$ are provided in (Horváth, Horváth & Telek 2020).

4.2.4 Upon exiting $\mathcal{D}_{k,i}$ the interval

Suppose that upon exiting $\mathcal{D}_{k,i}$ at time t the phase is $\varphi(t) = i \in \mathcal{S}_+$. At this time $X(t) = y_{k+1}$ which is the left-hand endpoint of $\mathcal{D}_{k+1,i}$. Hence, we restart the model of the sojourn time with the initial condition $\boldsymbol{A}(t) = \boldsymbol{\alpha}$. Similarly, upon exiting $\mathcal{D}_{k,i}$ at time t in phase $\varphi(t) = i \in \mathcal{S}_-$, then $X(t) = y_k$, which is the right-hand endpoint of $\mathcal{D}_{k,i}$, and so we restart the model of the sojourn time with the initial condition $\boldsymbol{A}(t) = \boldsymbol{\alpha}$.

4.3 The association of $j \in \mathcal{S}_0$ with $\mathcal{S}_+$ or $\mathcal{S}_-$

So far we have assumed $\mathcal{S}_0 = \emptyset$. We now remove this assumption and describe how we can include phases in $\mathcal{S}_0$ into the approximation. To do so, we can associate phases $j \in \mathcal{S}_0$ with either $\mathcal{S}_+$ or $\mathcal{S}_-$.

Let $X(0) = y_\ell$ and consider the event where $\{\varphi(t)\}$ transitions from $j_0 \to j_1 \to j_2$ where $j_0 \in \mathcal{S}_+$, $j_1 \in \mathcal{S}_0$ and $j_2 \in \mathcal{S}_-$, before there is a change of level, i.e.

$$\varphi(t) = \begin{cases} j_0 & t \in [0, t_1)\,, \\ j_1 & t \in [t_1, t_2)\,, \\ j_2 & t \in [t_2, t_3)\,, \end{cases}$$

and $X(t) \in \mathcal{D}_\ell$, $t \in [0, t_3)$. On approximating this event, the initial orbit position is $\boldsymbol{A}(0) = \boldsymbol{\alpha}$. The corresponding sample path of the orbit process on $t \in [0, t_3)$, is

$$\boldsymbol{A}(t) = \begin{cases} \dfrac{\boldsymbol{\alpha} e^{(\boldsymbol{S}_{j_0} + T_{j_0 j_0}\boldsymbol{I})t}}{\boldsymbol{\alpha} e^{(\boldsymbol{S}_{j_0} + T_{j_0 j_0}\boldsymbol{I})t}\boldsymbol{e}} & t \in [0, t_1)\,, \\ \dfrac{\boldsymbol{\alpha} e^{(\boldsymbol{S}_{j_0} + T_{j_0 j_0}\boldsymbol{I})t_1}\boldsymbol{D}(j_0, j_1) e^{T_{j_1 j_1}(t - t_1)}}{\boldsymbol{\alpha} e^{(\boldsymbol{S}_{j_0} + T_{j_0 j_0}\boldsymbol{I})t_1}\boldsymbol{D}(j_0, j_1) e^{T_{j_1 j_1}(t - t_1)}\boldsymbol{e}} & t \in [t_1, t_2)\,, \\ \dfrac{\boldsymbol{\alpha} e^{(\boldsymbol{S}_{j_0} + T_{j_0 j_0}\boldsymbol{I})t_1}\boldsymbol{D}(j_0, j_1) e^{T_{j_1 j_1}(t_2 - t_1)}\boldsymbol{D}(j_1, j_2) e^{(\boldsymbol{S}_{j_2} + T_{j_2 j_2}\boldsymbol{I})(t - t_2)}}{\boldsymbol{\alpha} e^{(\boldsymbol{S}_{j_0} + T_{j_0 j_0}\boldsymbol{I})t_1}\boldsymbol{D}(j_0, j_1) e^{T_{j_1 j_1}(t_2 - t_1)}\boldsymbol{D}(j_1, j_2) e^{(\boldsymbol{S}_{j_2} + T_{j_2 j_2}\boldsymbol{I})(t - t_2)}\boldsymbol{e}} & t \in [t_2, t_3)\,, \end{cases}$$

$$= \begin{cases} \dfrac{\boldsymbol{\alpha} e^{\boldsymbol{S}_{j_0} t}}{\boldsymbol{\alpha} e^{\boldsymbol{S}_{j_0} t}\boldsymbol{e}} & t \in [0, t_1)\,, \\ \dfrac{\boldsymbol{\alpha} e^{\boldsymbol{S}_{j_0} t_1}\boldsymbol{D}(j_0, j_1)}{\boldsymbol{\alpha} e^{\boldsymbol{S}_{j_0} t_1}\boldsymbol{D}(j_0, j_1)\boldsymbol{e}} & t \in [t_1, t_2)\,, \\ \dfrac{\boldsymbol{\alpha} e^{\boldsymbol{S}_{j_0} t_1}\boldsymbol{D}(j_0, j_1)\boldsymbol{D}(j_1, j_2) e^{\boldsymbol{S}_{j_2}(t - t_2)}}{\boldsymbol{\alpha} e^{\boldsymbol{S}_{j_0} t_1}\boldsymbol{D}(j_0, j_1)\boldsymbol{D}(j_1, j_2) e^{\boldsymbol{S}_{j_2}(t - t_2)}\boldsymbol{e}} & t \in [t_2, t_3)\,, \end{cases}$$

for some matrices $\boldsymbol{D}(j_0, j_1)$ and $\boldsymbol{D}(j_2, j_1)$. We want to find sensible choices for the matrices $\boldsymbol{D}(j_0, j_1)$ and $\boldsymbol{D}(j_2, j_1)$. Notice that $\{\boldsymbol{A}(t)\}$ is constant on $t \in [t_1, t_2)$. For $t \in [t_2, t_3)$ The matrix product $\boldsymbol{D}(j_0, j_1)\boldsymbol{D}(j_1, j_2)$ is there to capture the change in direction due to the change from $\mathcal{S}_+$ to $\mathcal{S}_-$. Hence, $\boldsymbol{D}(j_0, j_1)\boldsymbol{D}(j_1, j_2)$ should be equal to $\boldsymbol{D}$. These types of sample paths are the reason we need to associate states $j_1 \in \mathcal{S}_0$ with either $\mathcal{S}_+$ or $\mathcal{S}_-$.

Associating j_1 with $\mathcal{S}_+$, amounts to choosing $\boldsymbol{D}(j_0, j_1) = \boldsymbol{I}$ and $\boldsymbol{D}(j_1, j_2) = \boldsymbol{D}$; associating j_1 with $\mathcal{S}_-$, amounts to choosing $\boldsymbol{D}(j_0, j_1) = \boldsymbol{D}$ and $\boldsymbol{D}(j_1, j_2) = \boldsymbol{I}$.

There are some consequences of this choice. Let $k_2 \in \mathcal{S}_+$. Consider an event where the phase process of the fluid queue transitions from $j_0 \to j_1 \to k_2$ and there is no change of level. If j_1 is associated with $\mathcal{S}_+$, then $\boldsymbol{D}(j_0, j_1) = \boldsymbol{D}(j_1, k_2) = \boldsymbol{I}$ and the corresponding

orbit process, given $\boldsymbol{A}(0) = \boldsymbol{\alpha}$, is

$$\boldsymbol{A}(t) = \begin{cases} \dfrac{\boldsymbol{\alpha} e^{\boldsymbol{S}_{j_0} t}}{\boldsymbol{\alpha} e^{\boldsymbol{S}_{j_0} t} \boldsymbol{e}} & t \in [0, t_1)\,, \\ \dfrac{\boldsymbol{\alpha} e^{\boldsymbol{S}_{j_0} t_1}}{\boldsymbol{\alpha} e^{\boldsymbol{S}_{j_0} t_1} \boldsymbol{e}} & t \in [t_1, t_2)\,, \\ \dfrac{\boldsymbol{\alpha} e^{\boldsymbol{S}_{j_0} t_1} e^{\boldsymbol{S}_{k_2}(t-t_2)}}{\boldsymbol{\alpha} e^{\boldsymbol{S}_{j_0} t_1} e^{\boldsymbol{S}_{k_2}(t-t_2)} \boldsymbol{e}} & t \in [t_2, t_3)\,. \end{cases}$$

Notice that there is no matrix $\boldsymbol{D}$ in this expression.

Compare this to if j_1 is associated with $\mathcal{S}_-$. In this case $\boldsymbol{D}(j_0, j_1) = \boldsymbol{D}(j_1, k_2) = \boldsymbol{D}$ and the corresponding orbit process of the approximation, given $\boldsymbol{A}(0) = \boldsymbol{\alpha}$ and there are no change of level in $[t_1, t_3)$, is

$$\boldsymbol{A}(t) = \begin{cases} \dfrac{\boldsymbol{\alpha} e^{\boldsymbol{S}_{j_0} t}}{\boldsymbol{\alpha} e^{\boldsymbol{S}_{j_0} t} \boldsymbol{e}} & t \in [0, t_1)\,, \\ \dfrac{\boldsymbol{\alpha} e^{\boldsymbol{S}_{j_0} t_1} \boldsymbol{D}}{\boldsymbol{\alpha} e^{\boldsymbol{S}_{j_0} t_1} \boldsymbol{D} \boldsymbol{e}} & t \in [t_1, t_2)\,, \\ \dfrac{\boldsymbol{\alpha} e^{\boldsymbol{S}_{j_0} t_1} \boldsymbol{D} \boldsymbol{D} e^{\boldsymbol{S}_{k_2}(t-t_2)}}{\boldsymbol{\alpha} e^{\boldsymbol{S}_{j_0} t_1} \boldsymbol{D} \boldsymbol{D} e^{\boldsymbol{S}_{k_2}(t-t_2)} \boldsymbol{e}} & t \in [t_2, t_3)\,. \end{cases}$$

Ideally $\boldsymbol{D}^2 = \boldsymbol{I}$, however this is not the case here. Recall that a jump according to $\boldsymbol{D}$ corresponds to approximating the residual life by an expectation. With this interpretation as an approximation, it suggests that we might want to minimise the number of jumps according to $\boldsymbol{D}$ which occur. Therefore, for $j_1 \in \mathcal{S}_0$, if transitions $\mathcal{S}_+ \to j_1 \to \mathcal{S}_+$ occur with high probability compared to transition $\mathcal{S}_- \to j_1 \to \mathcal{S}_-$, then this suggests we might want to associate j_1 with $\mathcal{S}_+$. Which association is chosen will depend on the parameters of the fluid queue and on which aspects of the model we wish to approximate.

Augmented state space schemes

Another way to approach the problem is to augment the state space of the phase process by duplicating $\mathcal{S}_0$ and associating one copy of $\mathcal{S}_0$ with $\mathcal{S}_+$ and one copy of $\mathcal{S}_0$ with $\mathcal{S}_-$. Let $\{\varphi^*(t)\}$ be the augmented CTMC with state space $\mathcal{S}^*$ and generator $\boldsymbol{T}^*$. Let $\mathcal{S}_+$ and $\mathcal{S}_-$ be as before and $\mathcal{S}_{m0} = \{(m, i) \mid i \in \mathcal{S}_m\}$, $m \in \{+, -\}$, then $\mathcal{S}^* = \mathcal{S}_+ \cup \mathcal{S}_- \cup \mathcal{S}_{+0} \cup \mathcal{S}_{-0}$. The generator of $\varphi^*(t)$ can be written as

$$\boldsymbol{T}^* = \begin{bmatrix} \boldsymbol{T}_{++} & \boldsymbol{T}_{+0} & \boldsymbol{T}_{+-} & 0 \\ \boldsymbol{T}_{0+} & \boldsymbol{T}_{00} & \boldsymbol{T}_{0-} & 0 \\ \boldsymbol{T}_{-+} & 0 & \boldsymbol{T}_{--} & \boldsymbol{T}_{-0} \\ \boldsymbol{T}_{0+} & 0 & \boldsymbol{T}_{0-} & \boldsymbol{T}_{00} \end{bmatrix}.$$

Also define a fluid level $\{X^*(t)\}$ using $\{\varphi^*(t)\}$, with rates $c_i^* = c_i$ for $i \in \mathcal{S}_+ \cup \mathcal{S}_-$ and $c^*_{(m,i)} = 0$ for $(m,i) \in \mathcal{S}_{+0} \cup \mathcal{S}_{-0}$. The process $\{\varphi(t)\}$ is embedded within $\{\varphi^*(t)\}$ and is recovered by marginalising over $\mathcal{S}_{+0}$ and $\mathcal{S}_{-0}$. On the event $X^*(0) = X(0)$, the fluid levels $X^*(t)$ and $X(t)$ match exactly. Hence, by approximating $\{(X^*(t), \varphi^*(t))\}$, we can recover an approximation to $\{(X(t), \varphi(t))\}$. This construction removes the problem of having to choose how to associate states $j \in \mathcal{S}_0$ with either $\mathcal{S}_+$ or $\mathcal{S}_-$

The generator for the QBD-RAP approximation to the augmented fluid process (ignoring boundaries) is

$$\boldsymbol{B} = \begin{bmatrix} \ddots & \ddots & \ddots & & & \\ & \boldsymbol{B}_{-1} & \boldsymbol{B}_0 & \boldsymbol{B}_{+1} & & \\ & & \boldsymbol{B}_{-1} & \boldsymbol{B}_0 & \boldsymbol{B}_{+1} & \\ & & & \ddots & \ddots & \ddots \end{bmatrix},$$

where,

$$\boldsymbol{B}_0 = \begin{bmatrix} \boldsymbol{C}_+ \otimes \boldsymbol{S} + \boldsymbol{T}_{++} \otimes \boldsymbol{I} & \boldsymbol{T}_{+0} \otimes \boldsymbol{I} & \boldsymbol{T}_{+-} \otimes \boldsymbol{D} & \boldsymbol{0} \\ \boldsymbol{T}_{0+} \otimes \boldsymbol{I} & \boldsymbol{T}_{00} \otimes \boldsymbol{I} & \boldsymbol{T}_{0-} \otimes \boldsymbol{D} & \boldsymbol{0} \\ \boldsymbol{T}_{-+} \otimes \boldsymbol{D} & \boldsymbol{0} & \boldsymbol{C}_- \otimes \boldsymbol{S} + \boldsymbol{T}_{--} \otimes \boldsymbol{I} & \boldsymbol{T}_{-0} \otimes \boldsymbol{I} \\ \boldsymbol{T}_{0+} \otimes \boldsymbol{D} & \boldsymbol{0} & \boldsymbol{T}_{0-} \otimes \boldsymbol{I} & \boldsymbol{T}_{00} \otimes \boldsymbol{I} \end{bmatrix},$$

$$\boldsymbol{B}_{-1} = \begin{bmatrix} \boldsymbol{0} & & & \\ & \boldsymbol{0} & & \\ & & \boldsymbol{C}_- \otimes (\boldsymbol{s}\boldsymbol{\alpha}) & \\ & & & \boldsymbol{0} \end{bmatrix}, \qquad \boldsymbol{B}_{+1} = \begin{bmatrix} \boldsymbol{C}_+ \otimes (\boldsymbol{s}\boldsymbol{\alpha}) & & & \\ & \boldsymbol{0} & & \\ & & \boldsymbol{0} & \\ & & & \boldsymbol{0} \end{bmatrix}.$$

With this construction, jumps according to the matrix $\boldsymbol{D}$ occur only on transitions from $\mathcal{S}_m \to \mathcal{S}_n$, or $\mathcal{S}_m \to \mathcal{S}_{0m} \to \mathcal{S}_n$, $m, n \in \{+.-\}$, $m \neq n$.

4.4 The dynamics of the QBD-RAP approximation

We now have all the elements we need to describe the dynamics of the QBD-RAP approximation. Since the phase dynamics are a CTMC, we choose to use an alternate notation to the standard presented in Section 2.5.2. Let $\{\boldsymbol{Y}(t)\}_{t\geq 0} = \{(L(t), \boldsymbol{A}(t), \phi(t))\}_{t\geq 0}$ be the QBD-RAP approximation of a fluid queue, where $\{L(t)\}$ is the level, $\{\boldsymbol{A}(t)\}$ is the orbit process and $\{\phi(t)\}$ is the phase process. We use this representation to make explicit how the approximation captures the phase dynamics of the fluid queue.

We will show later that $\phi(t)$ captures the phase dynamics of the fluid queue exactly, provided the phase process is independent of the fluid level $\{X(t)\}$. The level $L(t) = \ell$ approximates which band $\mathcal{D}_\ell$ that $X(t)$ is in at time t, and the orbit $\boldsymbol{A}(t)$ can be use to obtain an approximation of where $X(t)$ is within the interval $\mathcal{D}_\ell$.

We now proceed to describe the evolution of the orbit and phase processes, before introducing the level variable later.

The dynamics of the orbit process and phase On $\phi(t) = i$, between event epochs, the process $\{\boldsymbol{A}(t)\}$ evolves deterministically according to the differential equation

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{A}(t) = \boldsymbol{A}(t)(\boldsymbol{S}_i - \boldsymbol{A}(t)(\boldsymbol{S}_i + T_{ii}\boldsymbol{I}_p)\boldsymbol{e})\boldsymbol{I}_p. \tag{4.18}$$

Let $\boldsymbol{a} \in \mathcal{A}$ be an arbitrary vector in $\mathcal{A}$. On the event that no jumps or changes of phase occur before time $t + u$, $\boldsymbol{A}(u) = \boldsymbol{a} \in \mathcal{A}$ and $\phi(u) = i$, the solution to (4.18) states that $\boldsymbol{A}(t + u)$ evolves deterministically according to

$$\boldsymbol{A}(t+u) = \frac{\boldsymbol{a}e^{(\boldsymbol{S}_i + T_{ii}\boldsymbol{I}_p)t}}{\boldsymbol{a}e^{(\boldsymbol{S}_i + T_{ii}\boldsymbol{I}_p)t}\boldsymbol{e}} = \frac{\boldsymbol{a}e^{\boldsymbol{S}_i t}}{\boldsymbol{a}e^{\boldsymbol{S}_i t}\boldsymbol{e}}.$$

At time t an event occurs at rate

$$\boldsymbol{A}(t)(\boldsymbol{S}_i - T_{ii}\boldsymbol{I}_p)\boldsymbol{e} = \boldsymbol{A}(t)\boldsymbol{s}_i - T_{ii}.$$

More precisely, an event corresponding to a change in phase for $\phi(t)$ occurs at rate

$$-\boldsymbol{A}(t)T_{ii}\boldsymbol{e} = -T_{ii}$$

and an event corresponding to a *change of level* occurs at rate

$$-\boldsymbol{A}(t)\boldsymbol{S}_i\boldsymbol{e} = \boldsymbol{A}(t)\boldsymbol{s}_i.$$

Later, we will make clear why we say that the latter event corresponds to a change of level. Upon an event occurring at time t, with probability $-T_{ii}/(-T_{ii} + \boldsymbol{A}(t^-)\boldsymbol{s}_i)$ the event corresponds to a change of phase and with probability $\boldsymbol{A}(t^-)\boldsymbol{s}_i/(-T_{ii} + \boldsymbol{A}(t^-)\boldsymbol{s}_i)$ the event corresponds a change of level.

Upon an event corresponding to a change of level occurring at time t the process $\{\boldsymbol{A}(t)\}$ jumps to $\boldsymbol{A}(t) = \boldsymbol{\alpha}$.

Upon an event corresponding to a change of phase from i to $j \neq i$ occurring at time t, there are two possibilities; either $sign(c_i) = sign(c_j)$, or $sign(c_i) \neq sign(c_j)$. As discussed earlier, for states $i \in \mathcal{S}_0$ we must specify some association with either $\mathcal{S}_+$ or $\mathcal{S}_-$. Here we choose the augmented state space approach and duplicate $\mathcal{S}_0$ and associate one copy with $i \in \mathcal{S}_+$ and one copy with $\mathcal{S}_-$; call these $\mathcal{S}_{+0}$ and $\mathcal{S}_{-0}$, respectively. We take $sign(c_i) = +$ for $i \in \mathcal{S}_{+0}$ and $sign(c_i) = -$ for $i \in \mathcal{S}_{-0}$.

Upon an event corresponding to a change of phase from i to $j \neq i$ occurring at time t, in the case $sign(c_i) = sign(c_j)$, at the time of the event $\boldsymbol{A}(t)$ is unchanged but immediately begins to evolve according to (4.18) with i replaced by j, while $\{\phi(t)\}$ jumps from $\phi(t^-) = i$ to $\phi(t) = j$.

Upon an event corresponding to a change of phase from i to $j \neq i$ occurring at time t, in the case $sign(c_i) \neq sign(c_j)$, the process $\{\boldsymbol{A}(t)\}$ jumps to

$$\boldsymbol{A}(t) = \frac{\boldsymbol{A}(t^-)T_{ij}\boldsymbol{D}}{\boldsymbol{A}(t^-)T_{ij}\boldsymbol{D}\boldsymbol{e}} = \frac{\boldsymbol{A}(t^-)T_{ij}\boldsymbol{D}}{T_{ij}} = \boldsymbol{A}(t^-)\boldsymbol{D}$$

and then immediately proceeds to evolve according to (4.18) with i replaced by j, and $\{\phi(t)\}$ jumps from $\phi(t^-) = i$ to $\phi(t) = j$.

The two scenarios, $sign(c_i) \neq sign(c_j)$ and $sign(c_i) = sign(c_j)$, can be written succinctly by stating that, at the time of an event corresponding to a change of phase from i to j, $\{\boldsymbol{A}(t)\}$ jumps to $\boldsymbol{A}(t)\boldsymbol{D}^{1(sign(c_i)\neq sign(c_j))}$ and begins to evolve according to (4.18) with i replaced by j, meanwhile $\{\phi(t)\}$ jumps from $\phi(t^-) = i$ to $\phi(t) = j$.

The following result states that $\{\phi(t)\}$ has the same distribution as $\{\varphi(t)\}$ when the fluid queue is unbounded, the latter being the phase process of the fluid queue.

Theorem 4.1. *Let Θ_i be the time at which the first jump of the phase process of the unbounded QBD-RAP, $\{\phi(t)\}$, occurs, given $\phi(0) = i$. For any initial orbit $\boldsymbol{a} \in \mathcal{A}$, then Θ_i has an exponential distribution with rate parameter $|T_{ii}|$. Furthermore, given $\phi(t)$ leaves state i, it jumps to state j with probability $T_{ij}/|T_{ii}|$. Hence, $\phi(t)$ and $\varphi(t)$ have the same probability law.*

Proof. Let $\{\tau_n\}_{n\geq 0}$ with $\tau_0 = 0$ and τ_n the time of the n-th change of level of the QBD-RAP. Consider partitioning $\{\Theta_i > t\}$ with respect to $\{\tau_{n-1} < t \leq \tau_n\}$, $n = 1, 2, \ldots$. For $n = 1$ we can write

$$\mathbb{P}(\Theta_i > t, \tau_0 < t \leq \tau_1 \mid \boldsymbol{A}(0) = \boldsymbol{a}) = \boldsymbol{a}e^{(\boldsymbol{S}_i + T_{ii}\boldsymbol{I}_p)t}\boldsymbol{e}$$

and since $T_{ii}\boldsymbol{I}_p$ commutes with $\boldsymbol{S}_i$ and $e^{T_{ii}t}$ is a scalar, then this is equal to

$$\boldsymbol{a}e^{\boldsymbol{S}_i t}e^{T_{ii}\boldsymbol{I}_p t}\boldsymbol{e} = \boldsymbol{a}e^{\boldsymbol{S}_i t}\boldsymbol{e}e^{T_{ii}t} = \mathbb{P}(\tau_0 < t \leq \tau_1)e^{T_{ii}t}.$$

For $n > 1$, by partitioning on the times of the first $n-1$ level changes, $\tau_1, \ldots, \tau_n$, we get

$$\begin{aligned}
&\mathbb{P}(\Theta_i > t, \tau_{n-1} < t \leq \tau_n \mid \boldsymbol{A}(0) = \boldsymbol{a}) \\
&= \int_{t_1=0}^{t}\int_{t_2=t_1}^{t}\cdots\int_{t_{n-1}=t_{n-2}}^{t} \mathbb{P}(\Theta_i > t, t \leq \tau_n, \tau_{n-1} \in \mathrm{d}t_{n-1}, \ldots, \tau_1 \in \mathrm{d}t_1 \mid \boldsymbol{A}(0) = \boldsymbol{a}) \\
&= \int_{t_1=0}^{t}\int_{t_2=t_1}^{t}\cdots\int_{t_{n-1}=t_{n-2}}^{t} \boldsymbol{a}e^{(\boldsymbol{S}_i + T_{ii}\boldsymbol{I}_p)t_1}\boldsymbol{s}_i \left(\prod_{k=2}^{n-1}\boldsymbol{\alpha}e^{(\boldsymbol{S}_i + T_{ii}\boldsymbol{I}_p)(t_k - t_{k-1})}\boldsymbol{s}_i\right) \\
&\quad \times \boldsymbol{\alpha}e^{(\boldsymbol{S}_i + T_{ii}\boldsymbol{I}_p)(t - t_{n-1})}\boldsymbol{e}\,\mathrm{d}t_{n-1}\,\mathrm{d}t_{n-2}\ldots\,\mathrm{d}t_1.
\end{aligned}$$

Since $T_{ii}\boldsymbol{I}_p$ commutes with $\boldsymbol{S}_i$, $e^{T_{ii}t_k}$, $k = 1, ..., n-1$ are scalars, and since $t_1 + (t_2 - t_1) + \ldots + (t_{n-1} - t_{n-2}) + (t - t_{n-1}) = t$, then this is equal to

$$\int_{t_1=0}^{t}\int_{t_2=t_1}^{t}\cdots\int_{t_{n-1}=t_{n-2}}^{t} \boldsymbol{a}e^{\boldsymbol{S}_i t_1}\boldsymbol{s}_i \left(\prod_{k=2}^{n-1}\boldsymbol{\alpha}e^{\boldsymbol{S}_i(t_k - t_{k-1})}\boldsymbol{s}_i\right)\boldsymbol{\alpha}e^{\boldsymbol{S}_i(t - t_{n-1})}\boldsymbol{e} \times e^{T_{ii}t}\,\mathrm{d}t_{n-1}\ldots\,\mathrm{d}t_1$$

$= \mathbb{P}(\tau_{n-1} < t \leq \tau_n) e^{T_{ii}t}.$

Hence, by the law of total probability,

$$\begin{aligned}\mathbb{P}(\Theta_i > t) &= \sum_{n=1}^{\infty} \mathbb{P}(\Theta_i > t, \tau_{n-1} < t \leq \tau_n) \\ &= \sum_{n=1}^{\infty} \mathbb{P}(\tau_{n-1} < t \leq \tau_n) e^{T_{ii}t} \\ &= e^{T_{ii}t},\end{aligned}$$

and therefore Θ_i has an exponential distribution with rate $|T_{ii}|$.

Upon leaving state i at time t, $\phi(t)$ transitions to state j with probability

$$\frac{\left(\dfrac{\boldsymbol{A}(t)\boldsymbol{D}^{1(sign(c_i)\neq sign(c_j))}T_{ij}\boldsymbol{e}}{\displaystyle\sum_{j\in\mathcal{S}}\boldsymbol{A}(t)\boldsymbol{D}^{1(sign(c_i)\neq sign(c_j))}T_{ij}\boldsymbol{e} + \boldsymbol{A}(t)\boldsymbol{s}_i}\right)}{\left(\dfrac{\displaystyle\sum_{j\in\mathcal{S}}\boldsymbol{A}(t)\boldsymbol{D}^{1(sign(c_i)\neq sign(c_j))}T_{ij}\boldsymbol{e}}{\displaystyle\sum_{j\in\mathcal{S}}\boldsymbol{A}(t)\boldsymbol{D}^{1(sign(c_i)\neq sign(c_j))}T_{ij}\boldsymbol{e} + \boldsymbol{A}(t)\boldsymbol{s}_i}\right)} = \frac{\boldsymbol{A}(t)\boldsymbol{e}T_{ij}}{\displaystyle\sum_{j\in\mathcal{S}}\boldsymbol{A}(t)\boldsymbol{e}T_{ij}} = \frac{T_{ij}}{-T_{ii}}.$$

Therefore, the process $\{\phi(t)\}$ has the same probability law as $\{\varphi(t)\}$. □

Remark 4.2. *The same result can be shown for a regulated boundary. For boundary conditions which interact with the phase dynamics, such as a reflecting boundary, the result does not hold. The cause is the fact that the phase dynamics are level dependent – we may see a forced change of phase upon a boundary being hit – and the QBD-RAP can only approximate the level process of the fluid queue. However, until a boundary is hit (by either the fluid queue or QBD-RAP) then the phase processes match. We show later that, in the limit as the variance of the matrix exponential distribution used in the construction of the QBD-RAP goes to zero, then the dynamics of the level process of the fluid queue, $X(t)$, are captured by the QBD-RAP, and boundary behaviour which interacts with the phase dynamics can be captured too.*

Since the phase processes $\{\phi(t)\}$ and $\{\varphi(t)\}$ have the same law when boundaries are not present, henceforth, we shall assume $\{\phi(t)\}$ and $\{\varphi(t)\}$ are coupled when possible (they share the same sample path). Specifically, in Section 5.2, we will analyse the QBD-RAP on the event that it remains in the same level, ℓ, say, and we compare this to the fluid queue on the event that the level remains in the band $\mathcal{D}_\ell$. No boundary behaviour is involved in this calculation, so we treat $\{\phi(t)\}$ and $\{\varphi(t)\}$ as coupled and use the latter notation. When we must distinguish the two processes, we use $\phi(t)$ for the phase of the QBD-RAP and $\varphi(t)$ for the phase of the fluid.

The level process To event epochs of $\{(\boldsymbol{A}(t), \varphi(t))\}_{t\geq 0}$ we associate marks $\{-1, 0, +1\}$ in the following way.

- To events epochs corresponding to a change of phase of $\varphi(t)$ we associate the mark 0.
- To event epochs at time t which correspond to a change in level and for which $\varphi(t^-) = i \in \mathcal{S}_-$ we associate the mark -1.
- To event epochs at time t which correspond to a change in level and for which $\varphi(t^-) = i \in \mathcal{S}_+$ we associate the mark $+1$.

Now define $N_+(t)$ ($N_-(t)$) as the simple point process which counts the number of event epochs with marks $+1$ (-1) which have occurred in the time up to and including time t. The level process of the QBD-RAP is given by $L(t) = N_+(t) - N_-(t)$. The process $\{(L(t), \boldsymbol{A}(t), \varphi(t))\}_{t\geq 0}$ forms a QBD with RAP components.

One way to specify the QBD with RAP components, $\{(L(t), \boldsymbol{A}(t), \varphi(t))\}$, is to describe its generator:

$$\boldsymbol{B} = \begin{bmatrix} \ddots & \ddots & \ddots & & \\ & \boldsymbol{B}_{-1} & \boldsymbol{B}_0 & \boldsymbol{B}_{+1} & \\ & & \boldsymbol{B}_{-1} & \boldsymbol{B}_0 & \boldsymbol{B}_{+1} \\ & & & \ddots & \ddots & \ddots \end{bmatrix},$$

where,

$$\boldsymbol{B}_0 = \begin{bmatrix} \boldsymbol{B}_{++} & \boldsymbol{B}_{+-} \\ \boldsymbol{B}_{-+} & \boldsymbol{B}_{--} \end{bmatrix}$$

$$\boldsymbol{B}_{-1} = \begin{bmatrix} \mathbf{0} & & & \\ & \mathbf{0} & & \\ & & \boldsymbol{C}_- \otimes (\boldsymbol{s}\boldsymbol{\alpha}) & \\ & & & \mathbf{0} \end{bmatrix}, \qquad \boldsymbol{B}_{+1} = \begin{bmatrix} \boldsymbol{C}_+ \otimes (\boldsymbol{s}\boldsymbol{\alpha}) & & & \\ & \mathbf{0} & & \\ & & \mathbf{0} & \\ & & & \mathbf{0} \end{bmatrix},$$

and

$$\boldsymbol{B}_{++} = \begin{bmatrix} \boldsymbol{C}_+ \otimes \boldsymbol{S} + \boldsymbol{T}_{++} \otimes \boldsymbol{I} & \boldsymbol{T}_{+0} \otimes \boldsymbol{I} \\ \boldsymbol{T}_{0+} \otimes \boldsymbol{I} & \boldsymbol{T}_{00} \otimes \boldsymbol{I} \end{bmatrix}, \boldsymbol{B}_{+-} = \begin{bmatrix} \boldsymbol{T}_{+-} \otimes \boldsymbol{D} & \mathbf{0} \\ \boldsymbol{T}_{0-} \otimes \boldsymbol{D} & \mathbf{0} \end{bmatrix},$$
$$\boldsymbol{B}_{-+} = \begin{bmatrix} \boldsymbol{T}_{-+} \otimes \boldsymbol{D} & \mathbf{0} \\ \boldsymbol{T}_{0+} \otimes \boldsymbol{D} & \mathbf{0} \end{bmatrix}, \boldsymbol{B}_{--} = \begin{bmatrix} \boldsymbol{C}_- \otimes \boldsymbol{S} - \boldsymbol{T}_{--} \otimes \boldsymbol{I} & \boldsymbol{T}_{-0} \otimes \boldsymbol{I} \\ \boldsymbol{T}_{0-} \otimes \boldsymbol{I} & \boldsymbol{T}_{00} \otimes \boldsymbol{I} \end{bmatrix}.$$

4.5 Boundary conditions

In the study and practical application of fluid queues, regulated, reflecting, sticky, or a mixture of these boundary conditions may be imposed. We now present the intuition as to how one may include such boundary conditions in the QBD-RAP approximation scheme.

Without loss of generality, assume that there is a boundary for the fluid level at $y_0 = 0$. This boundary can only be hit from above in phases $i \in \mathcal{S}_-$. Suppose that, upon hitting the boundary in phase $i \in \mathcal{S}_-$, the phase jumps from i to $j \in \mathcal{S}$ with probability p_{ij}^{-1}. If $j \in \mathcal{S}_- \cup \mathcal{S}_0$ the process remains at the boundary and the phase process evolves among states in $\mathcal{S}_- \cup \mathcal{S}_0$ until the first transition to a state in $\mathcal{S}_+$, at which point the level $\{X(t)\}$ immediately leaves the boundary – we will call this a *sticky* boundary. If $j \in \mathcal{S}_+$ the process immediately leaves the boundary – we will call this a reflecting boundary. We collect the probabilities p_{ij}^{-1} into the matrices $\boldsymbol{P}_{-+}^{-1} = [p_{ij}^{-1}]_{i\in\mathcal{S}_-,j\in\mathcal{S}_+}$, $\boldsymbol{P}_{-0}^{-1} = [p_{ij}^{-1}]_{i\in\mathcal{S}_-,j\in\mathcal{S}_0}$, and $\boldsymbol{P}_{--}^{-1} = [p_{ij}^{-1}]_{i\in\mathcal{S}_-,j\in\mathcal{S}_-}$.

Adjacent to the boundary at $y_0 = 0$ is the band $\mathcal{D}_0 = [0, \Delta]$ which corresponds to level 0 for the QBD-RAP. We denote the boundary of the QBD-RAP as level -1, which corresponds to $\{0\}$ for the fluid queue. Modelling the behaviour of the fluid queue at boundaries can be broken down into three components. 1) Modelling the time and phase when the fluid level hits the boundary. 2) Modelling the phase whilst the fluid level remains at the boundary. 3) Modelling the fluid level and phase at the exit from the boundary.

We claim that the event $\{L(t) = \ell, \phi(t) = i\}$ models the event $\{X(t) \in \mathcal{D}_{\ell,i}, \varphi(t) = i\}$ and further, that instants u with $L(u^-) \neq L(u), \phi(u) = i$ model the events $\{X(u^-) \in \mathcal{D}_{\ell,i}, X(u) \notin \mathcal{D}_{\ell,i}, \varphi(u) = i\}$. With this in mind, we suppose that when the QBD-RAP is in level 0 in phase $i \in \mathcal{S}_-$ and there is a change of level, this corresponds to approximating the event that the fluid level $\{X(t)\}$ hits the boundary at 0. We refer to this as the QBD-RAP hitting the boundary also. We show later that, using matrix exponential distributions with sufficiently small variance in the construction of the QBD-RAP, then the distribution of time until the QBD-RAP first hits the boundary approximates the time until the fluid level hits the boundary. This addresses component 1).

For a sticky lower boundary at 0, given the fluid level reaches zero at time u in phase $i \in \mathcal{S}_-$, then the phase transitions to some phase $j \in \mathcal{S}_- \cup \mathcal{S}_0$ with probability p_{ij}^{-1}, and the distribution of the phase $\varphi(t)$ (on the event that the process has not left the boundary at or before time $t > u$) is given by the elements of the vector

$$p_{ij}^{-1}\boldsymbol{e}_j \exp\left(\begin{bmatrix} \boldsymbol{T}_{--} & \boldsymbol{T}_{-0} \\ \boldsymbol{T}_{0-} & \boldsymbol{T}_{00} \end{bmatrix}(t-u)\right).$$

The distribution for the time until $\{X(t)\}$ leaves $\{0\}$ for the first time after u is

$$1 - p_{ij}^{-1}\boldsymbol{e}_j \exp\left(\begin{bmatrix} \boldsymbol{T}_{--} & \boldsymbol{T}_{-0} \\ \boldsymbol{T}_{0-} & \boldsymbol{T}_{00} \end{bmatrix}(t-u)\right)\boldsymbol{e}.$$

Therefore, in the case of a sticky boundary, we can capture the behaviour of the process at the boundary exactly. For a reflecting boundary there is no time spent at the boundary. This addresses component 2) of the modelling problem.

For both sticky and reflecting boundaries, given the QBD-RAP is in the correct phase at the instant upon which it hits the boundary, then upon exiting the boundary the phase can be captured exactly too. For a sticky boundary, given the fluid/QBD-RAP hits the boundary in phase $i \in \mathcal{S}_-$ at time t, it then jumps to phase $j \in \mathcal{S}_+ \cup \mathcal{S}_0$ and exits the boundary in phase $k \in \mathcal{S}_+$ at time $(t-u)$ with density

$$p_{ij}^{-1} \boldsymbol{e}_j \exp\left(\begin{bmatrix} \boldsymbol{T}_{--} & \boldsymbol{T}_{-0} \\ \boldsymbol{T}_{0-} & \boldsymbol{T}_{00} \end{bmatrix} (t-u) \right) \begin{bmatrix} \boldsymbol{T}_{-+} \\ \boldsymbol{T}_{0+} \end{bmatrix} \boldsymbol{e}_k.$$

For a reflecting boundary, given the boundary is hit in phase $i \in \mathcal{S}_-$, the phase upon leaving the boundary is $j \in \mathcal{S}_+$ with probability p_{ij}^{-1}. Upon leaving the boundary, the appropriate orbit position for the QBD-RAP is $\boldsymbol{\alpha}$, as this corresponds to the fluid level 0.

To summarise, at time t the rate at which probability mass accumulates at the sticky boundary in phase $j \in \mathcal{S}_- \cup \mathcal{S}_0$ and level -1, upon hitting the boundary in phase $i \in \mathcal{S}_-$ and level 0, is

$$c_i p_{ij}^{-1} \boldsymbol{A}(t) \boldsymbol{s}.$$

The rate at which density from phase $i \in \mathcal{S}_-$ and level 0 transitions to density in phase $j \in \mathcal{S}_+$ and level 0 upon hitting a boundary (a reflection event) is

$$c_i p_{ij}^{-1} \boldsymbol{A}(t) \boldsymbol{s},$$

and upon this transition, the orbit jumps to $\boldsymbol{\alpha}$. The rate at which mass leaves the sticky boundary from phase $i \in \mathcal{S}_- \cup \mathcal{S}_0$ and level -1 into phases $j \in \mathcal{S}_+$ and level 0 is

$$T_{ij},$$

and upon leaving the boundary the orbit is $\boldsymbol{\alpha}$.

This information can be inscribed in the generator of the QBD-RAP. For example, using the augmented state space scheme to account for phases $\mathcal{S}_0$, the generator of the QBD-RAP described above has boundary conditions included as

$$\left[\begin{array}{cc|cc|c} \boldsymbol{T}_{--} & \boldsymbol{T}_{0-} & \boldsymbol{T}_{-+} \otimes \boldsymbol{\alpha} & \boldsymbol{0} & \\ \boldsymbol{T}_{-0} & \boldsymbol{T}_{00} & \boldsymbol{T}_{0+} \otimes \boldsymbol{\alpha} & \boldsymbol{0} & \\ \hline \boldsymbol{0} & \boldsymbol{0} & & & \\ \boldsymbol{0} & \boldsymbol{0} & & & \\ (\boldsymbol{C}_- \boldsymbol{P}_{--}^{-1}) \otimes \boldsymbol{s} & (\boldsymbol{C}_- \boldsymbol{P}_{-0}^{-1}) \otimes \boldsymbol{s} & \multicolumn{2}{c|}{\check{\boldsymbol{B}}_0} & \boldsymbol{B}_{+1} \\ \boldsymbol{0} & \boldsymbol{0} & & & \\ \hline & & \multicolumn{2}{c|}{\boldsymbol{B}_{-1}} & \boldsymbol{B}_0 \\ & & & & \end{array}\right],$$

where

$$\check{\boldsymbol{B}}_0 = \begin{bmatrix} \boldsymbol{C}_+ \otimes \boldsymbol{S} + \boldsymbol{T}_{++} \otimes \boldsymbol{I} & \boldsymbol{T}_{+0} \otimes \boldsymbol{I} & \boldsymbol{T}_{+-} \otimes \boldsymbol{D} & \\ \boldsymbol{T}_{0+} \otimes \boldsymbol{I} & \boldsymbol{T}_{00} \otimes \boldsymbol{I} & \boldsymbol{T}_{0-} \otimes \boldsymbol{D} & \\ (\boldsymbol{C}_- \boldsymbol{P}_{-+}^{-1}) \otimes \boldsymbol{s}\boldsymbol{\alpha} + \boldsymbol{T}_{-+} \otimes \boldsymbol{D} & & \boldsymbol{C}_- \otimes \boldsymbol{S} + \boldsymbol{T}_{--} \otimes \boldsymbol{I} & \boldsymbol{T}_{-0} \otimes \boldsymbol{I} \\ \boldsymbol{T}_{0+} \otimes \boldsymbol{D} & & \boldsymbol{T}_{0-} \otimes \boldsymbol{I} & \boldsymbol{T}_{00} \otimes \boldsymbol{I} \end{bmatrix}.$$

The top-left block represents the point masses due to the sticky boundary. The bottom-left block is the transition of density into point masses due to the sticky boundary. The top-right block is the transition of point masses into density as the process leaves the boundary. The term $(\boldsymbol{C}_- \boldsymbol{P}_{-+}^{-1}) \otimes \boldsymbol{s}\boldsymbol{\alpha}$ in $\check{\boldsymbol{B}}_0$ incorporates the reflecting boundary behaviour.

Upper boundaries can be included in an analogous manner ad we denote the transition probabilities upon hitting a boundary by p_{ij}^{K+1} and define the matrices $\boldsymbol{P}_{+-}^{K+1} = [p_{ij}^{K+1}]_{i \in \mathcal{S}_+, j \in \mathcal{S}_{K+1}}$, $\boldsymbol{P}_{+0}^{-1} = [p_{ij}^{K+1}]_{i \in \mathcal{S}_+, j \in \mathcal{S}_0}$, and $\boldsymbol{P}_{++}^{K+1} = [p_{ij}^{K+1}]_{i \in \mathcal{S}_+, j \in \mathcal{S}_+}$.

The orbit process $\{\boldsymbol{A}(t)\}$ is not required to model the behaviour at the sticky boundary. However, we suppose that $\boldsymbol{A}(t) = 1$ at the boundaries. This choice purely for notational convenience and is arbitrary, but allows us to generalise some notation later when we are describing the evolution of the QBD-RAP. Further, for boundaries at $y_0 = 0$ and $y_{K+1} = (K+1)\Delta$, we let the level process take the value $L(t) = -1$ at the lower boundary, and $L(t) = K + 1$ at the upper boundary. Denote the set of levels, including boundary levels, by $\mathcal{K} = \{-1, 0, \ldots, K, K+1\}$ and without boundary levels by $\mathcal{K}^\circ = \{0, 1, \ldots, K\}$.

Remark 4.3. *It may be possible to extend the QBD-RAP approximation to model jumps into and out of the boundary, provided that the jumps into/out of the boundary happen at an intensity which can be described by a matrix exponential distribution.*

Furthermore, it may be possible to model more general boundary behaviours which can be approximated by a limit of matrix exponential distributions. For example, a boundary condition where the fluid queue spends a deterministic time at the boundary.

4.6 Initial conditions

We argued earlier that we can think of the orbit $\boldsymbol{k}(x)$ as corresponding to the fluid being a distance of x from the left boundary of an interval when $i \in \mathcal{S}_+$, or from the right boundary of an interval when $i \in \mathcal{S}_-$. With this interpretation, an approximation to the initial condition $\boldsymbol{X}(0) = (X(0), \varphi(0)) = (x_0, i)$, $x_0 \in \mathcal{D}_{\ell,i}$, $i \in \mathcal{S}_+ \cup \mathcal{S}_{+0}$ is to set the initial orbit to $\boldsymbol{k}(x_0 - y_\ell)$. Similarly, an approximation to the initial condition $\boldsymbol{X}(0) = (x_0, i)$, $x_0 \in \mathcal{D}_{\ell,i}, x_0 \neq y_{\ell+1}$, $i \in \mathcal{S}_- \cup \mathcal{S}_{-0}$ is to set the initial orbit position to $\boldsymbol{k}(y_{\ell+1} - x_0)$.

We use the notation $\boldsymbol{a}_{\ell,i}(x_0) = \boldsymbol{a}_{\ell,+}(x_0) = \boldsymbol{k}(x_0 - y_\ell)$ for the initial orbit position corresponding to x_0 when the initial phase is $i \in \mathcal{S}_+$ and $x_0 \in \mathcal{D}_{\ell,i}$. Similarly, for $i \in \mathcal{S}_-$ and $x_0 \in \mathcal{D}_{\ell,i}$ define the notation $\boldsymbol{a}_{\ell,i}(x_0) = \boldsymbol{a}_{\ell,-}(x_0) = \boldsymbol{k}(y_{\ell+1} - x_0)$.

More generally, given an initial measure $\mu_i(\cdot) := \mathbb{P}(X(0) \in \cdot, \varphi(0) = i)$, an approximation to this initial condition is to set the orbit to $\int_{x \in \mathcal{D}_{\ell,i}} \boldsymbol{a}_{\ell,i}(x)\, \mathrm{d}\mu_i$, for $i \in \mathcal{S}$, $\ell \in \mathcal{K}^\circ$.

Initial conditions for phases $i \in \mathcal{S}_0$ While the augmented state space scheme described above is a convergent one, the analysis of the QBD-RAP scheme is *greatly* simplified if we allow initial conditions which allocate 0 probability to phases in $i \in \mathcal{S}_{+0} \cup \mathcal{S}_{-0}$. To see why initial mass in $\mathcal{S}_{+0} \cup \mathcal{S}_{-0}$ might introduce further complexity, consider a sample path with $\varphi(0) = k \in \mathcal{S}_{+0}$ and $\boldsymbol{A}(0) = \boldsymbol{a} \in \mathcal{A}$. On $\varphi(u) \in \mathcal{S}_{+0}$, for $u \in [0, t)$, then $\boldsymbol{A}(u) = \boldsymbol{a}$ is constant. Upon the phase leaving $\mathcal{S}_{+0}$, at time t, say, the phase will either jump to $\mathcal{S}_+$ or $\mathcal{S}_-$. If $\varphi(t)$ jumps to $i \in \mathcal{S}_+$, the orbit process at time t will be $\boldsymbol{A}(t) = \boldsymbol{a}$ and the process evolve from time t as if it has started in phase $i \in \mathcal{S}_+$ with initial orbit $\boldsymbol{a}$ – there are no major issues in this case. If, however, $\varphi(t)$ jumps to $i \in \mathcal{S}_-$, then the orbit process will jump to $\boldsymbol{A}(t) = \boldsymbol{aD}$ at time t. From time t onward the QBD-RAP process will evolve as if it has started in phase $i \in \mathcal{S}_-$ with initial orbit $\boldsymbol{aD}$. With the way we have structured the proof of convergence, we have to complete two analyses of the QBD-RAP process – one for the initial condition $\boldsymbol{a}$ and another for the initial condition $\boldsymbol{aD}$.

A solution to this problem is to augment a set of ephemeral states to the QBD-RAP which capture the initial sojourn in $\mathcal{S}_0$ only. Suppose we wish to approximate a fluid queue with the initial condition μ_k, $k \in \mathcal{S}_0$, where μ_k has mass 1. Let us denote the set of ephemeral states by $\mathcal{S}_0^{*,k}$ and each state in $\mathcal{S}_0^{*,k}$ by i^*. Each state $i^* \in \mathcal{S}_0^{*,k}$ is a copy of $i \in \mathcal{S}_0$, and, in $\mathcal{S}_0^{*,k}$ the phase process has the same phase dynamics as it does in $\mathcal{S}_0$. At time 0, we start the QBD-RAP in the ephemeral set of states $\mathcal{S}_0^{*,k}$ in phase $k^* \in \mathcal{S}_0^{*,k}$. While in $\mathcal{S}_0^{*,k}$, the phase process evolves according to the generator $\boldsymbol{T}_{00}$. Upon exiting the ephemeral set $\mathcal{S}_0^{*,k}$ at time t, given the phase at t^- is $j^* \in \mathcal{S}_0^{*,k}$, the QBD-RAP process jumps to $L(t) = \ell$, $\boldsymbol{A}(t) = \int_{\mathcal{D}_{\ell,i}} \boldsymbol{a}_{\ell,i}(x)\, \mathrm{d}\mu_k$, $\varphi(t) = i$, where $\ell \in \mathcal{K}$, $i \in \mathcal{S}_+ \cup \mathcal{S}_-$ with probability

$$\int_{\mathcal{D}_{\ell,i}} \boldsymbol{a}_{\ell,i}(x)\, \mathrm{d}\mu_k \boldsymbol{e} \frac{T_{j^* i}}{-T_{j^* j^*}}.$$

Notice that we have not mentioned anything about the level or orbit process during the sojourn in $\mathcal{S}_0^{*,k}$. This is because all information about the level and phase during this time is contained within the initial condition μ_k, and the only thing that changes is the phase.

To include the ephemeral states within the generator of the QBD-RAP we can do the

following,

$$\left[\begin{array}{c|cccc} \boldsymbol{T}_{00} & \boldsymbol{B}_{-1}^{*,k} & \boldsymbol{B}_{0}^{*,k} & \boldsymbol{B}_{1}^{*,k} & \dots \\ \hline & \check{\boldsymbol{B}}_{0} & \check{\boldsymbol{B}}_{+1} & & \\ & \check{\boldsymbol{B}}_{-1} & \check{\boldsymbol{B}}_{0} & \boldsymbol{B}_{+1} & \\ & & \boldsymbol{B}_{-1} & \boldsymbol{B}_{0} & \ddots \\ & & & \ddots & \ddots \end{array}\right],$$

where

$$\boldsymbol{B}_{-1}^{*,k} = \begin{bmatrix} \boldsymbol{T}_{0-}\mu_k(\{0\}) & \boldsymbol{0} \end{bmatrix} \qquad \boldsymbol{B}_{\ell}^{*,k} = \begin{bmatrix} \boldsymbol{T}_{0+} \otimes \boldsymbol{a}_{\ell}^{+} & \boldsymbol{0} & \boldsymbol{T}_{0-} \otimes \boldsymbol{a}_{\ell}^{-} & \boldsymbol{0} \end{bmatrix},$$

$$\check{\boldsymbol{B}}_{0} = \begin{bmatrix} \boldsymbol{T}_{--} & \boldsymbol{T}_{-0} \\ \boldsymbol{T}_{0-} & \boldsymbol{T}_{00} \end{bmatrix}, \qquad \check{\boldsymbol{B}}_{+1} = \begin{bmatrix} \boldsymbol{T}_{-+} \otimes \boldsymbol{\alpha} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{T}_{0+} \otimes \boldsymbol{\alpha} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \end{bmatrix},$$

$$\check{\boldsymbol{B}}_{-1} = \begin{bmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \\ (\boldsymbol{C}_{-}\boldsymbol{P}_{--}) \otimes \boldsymbol{s} & (\boldsymbol{C}_{-}\boldsymbol{P}_{-0}) \otimes \boldsymbol{s} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix},$$

and

$$\boldsymbol{a}_{\ell}^{r} = \int_{\mathcal{D}_{\ell,r}} \boldsymbol{a}_{\ell,r}(x)\,\mathrm{d}\mu_k,\ \ell \in \mathcal{K}^{\circ},\ r \in \{+,-\}.$$

In general, we need to augment a set of states $\mathcal{S}_0^{*,k}$ to the QBD-RAP for each initial phase, $k \in \mathcal{S}_0$ such that μ_k has positive mass.

4.7 At time t – closing operators

The level process $\{L(t)\}$ of the QBD-RAP approximates the process

$$\left\{\sum_{k\in\mathcal{K}}\sum_{i\in\mathcal{S}} k1(X(t) \in \mathcal{D}_{k,i})\right\}_{t\geq 0}. \tag{4.19}$$

This may be of interest in its own right. However, for some applications, this approximation may be too coarse – the level process tells us nothing about where in the intervals $\mathcal{D}_{k,i}$ the fluid level might be.

We now describe how the orbit position $\boldsymbol{A}(t)$ can be used to approximate the density of the level of the fluid queue $X(t)$. This reconstruction is entirely post-hoc in the sense that it does not affect the dynamics of the QBD-RAP itself, we only take the information contained in $\mathbb{E}[\boldsymbol{A}(t)1(L(t) = \ell, \varphi(t) = i)]$ and rewrite it in a descriptive way.

Suppose that the QBD-RAP is in level $L(t) = \ell$, phase $\phi(t) = j \in \mathcal{S}_+ \cup \mathcal{S}_{+0}$, and the orbit is $\boldsymbol{A}(t) = \boldsymbol{a}$. If the QBD-RAP remains in phase j, the QBD-RAP approximation will transition out of level ℓ in the infinitesimal time interval $t + \mathrm{d}u$ with probability $\boldsymbol{a}e^{|c_j|\boldsymbol{S}u}|c_j|\boldsymbol{s}\,\mathrm{d}u$. At the time of the change of level we estimate the position of $X(t+u)$ by $X(t+u) \approx y_{\ell+1}$. Tracing this back to time t, we estimate the position of $X(t)$ as $X(t) \approx y_{\ell+1} - |c_j|u$. Reverse engineering this logic, the approximation to the probability of $X(t) \in \mathrm{d}x$ is $\boldsymbol{a}e^{|c_j|\boldsymbol{S}(y_{\ell+1}-x)/|c_j|}|c_j|\boldsymbol{s}\,\mathrm{d}x/|c_j|$, since $\mathrm{d}x = |c_j|\,\mathrm{d}u$ where $\mathrm{d}x$ is an infinitesimal interval in space and $\mathrm{d}u$ is an infinitesimal interval with respect to time.

Similarly, for $j \in \mathcal{S}_- \cup \mathcal{S}_{-0}$, if the QBD-RAP remains in phase j then the QBD-RAP approximation will transition out of level ℓ in the infinitesimal time interval $t + \mathrm{d}u$ with probability $\boldsymbol{a}e^{|c_j|\boldsymbol{S}u}|c_j|\boldsymbol{s}\,\mathrm{d}u$. At the time of the transition of level, we estimate the position of $X(t+u)$ by $X(t+u) \approx y_\ell$. Tracing this back to time t, we estimate the position of $X(t)$ as $X(t) \approx y_\ell + |c_j|u$. Reverse engineering this logic, the approximation to the probability $X(t) \in \mathrm{d}x$ is $\boldsymbol{a}e^{|c_j|\boldsymbol{S}(x-y_\ell)/|c_j|}|c_j|\boldsymbol{s}\,\mathrm{d}x/|c_j|$, since $\mathrm{d}x = |c_j|\,\mathrm{d}u$. *

This reasoning leads to an approximation of the distribution of the fluid at time t, $\mathbb{P}(\boldsymbol{X}(t) \in (\mathrm{d}x, j) \mid \boldsymbol{X}(0) = (x_0, i))$, as

$$\int_{\boldsymbol{a}\in\mathcal{A}} \mathbb{P}(\boldsymbol{Y}(t) \in (\ell,\, \mathrm{d}\boldsymbol{a}, j) \mid \boldsymbol{Y}(0) = \boldsymbol{y}_0)\boldsymbol{a}e^{\boldsymbol{S}z}\boldsymbol{s}\,\mathrm{d}x, \tag{4.20}$$

where $x \in \mathcal{D}_{\ell,j}$, $x_0 \in \mathcal{D}_{\ell_0,i}$, $\boldsymbol{y}_0 = (\ell_0, \boldsymbol{a}_{\ell_0,i}(x_0), i)$, and

$$z = \begin{cases} x - y_\ell & i \in \mathcal{S}_-, \\ y_{\ell+1} - x & i \in \mathcal{S}_+. \end{cases}$$

While the estimate (4.20) is appealing, it is, however, defective – it does not integrate to 1 for any finite-order matrix exponential distribution. To see this, compute

$$\begin{aligned}
&\sum_{\ell\in\mathcal{K}}\sum_{j\in\mathcal{S}}\int_{z\in\mathcal{D}_{\ell,j}}\int_{\boldsymbol{a}\in\mathcal{A}} \mathbb{P}(\boldsymbol{Y}(t) \in (\ell,\, \mathrm{d}\boldsymbol{a}, j) \mid \boldsymbol{Y}(0) = \boldsymbol{y}_0)\boldsymbol{a}e^{\boldsymbol{S}z}\boldsymbol{s}\,\mathrm{d}z \\
&= \sum_{\ell\in\mathcal{K}}\sum_{j\in\mathcal{S}}\int_{\boldsymbol{a}\in\mathcal{A}} \mathbb{P}(\boldsymbol{Y}(t) \in (\ell,\, \mathrm{d}\boldsymbol{a}, j) \mid \boldsymbol{Y}(0) = \boldsymbol{y}_0)(1 - \boldsymbol{a}e^{\boldsymbol{S}\Delta}\boldsymbol{e}) \\
&= (1 - \boldsymbol{a}e^{\boldsymbol{S}\Delta}\boldsymbol{e}) \\
&< 1,
\end{aligned}$$

since $\boldsymbol{a}e^{\boldsymbol{S}u}\boldsymbol{e} > 0$ for any $u > 0$. For an orbit position $\boldsymbol{A}(t) = \boldsymbol{a}$, the amount of mass missing from each level and phase is

$$\mathbb{P}(\boldsymbol{Y}(t) \in (\ell,\, \mathrm{d}\boldsymbol{a}, j) \mid \boldsymbol{Y}(0) = \boldsymbol{y}_0)\boldsymbol{a}e^{\boldsymbol{S}\Delta}\boldsymbol{e}\,\mathrm{d}x.$$

*Here we have assumed we are dealing with the augmented state space model. If the non-augmented state space model is used, for phases $j \in \mathcal{S}_0$, if phase j is associated with $\mathcal{S}_+$, we use the same approximation as if $j \in \mathcal{S}_+$, and if phase $j \in \mathcal{S}_0$ is associated with $\mathcal{S}_-$, we use the same approximation as if $j \in \mathcal{S}_-$.

The problem is that the triple $(\boldsymbol{a}, \boldsymbol{S}, \boldsymbol{s})$ defines a distribution on $x \in [0, \infty)$, however, it only makes sense to take $x \in [0, \Delta)$ in the approximation scheme.

To partially rectify this we can instead approximate $\mathbb{P}(\boldsymbol{X}(t) \in (\,\mathrm{d}x, j) \mid \boldsymbol{X}(0) = (x_0, i))$ by

$$\int_{\boldsymbol{a} \in \mathcal{A}} \mathbb{P}(\boldsymbol{Y}(t) \in (\ell,\, \mathrm{d}\boldsymbol{a}, j) \mid \boldsymbol{Y}(0) = \boldsymbol{y}_0) \boldsymbol{a} \boldsymbol{u}_{\ell,j}(x)\, \mathrm{d}x, \tag{4.21}$$

where

$$\boldsymbol{u}_{\ell,j}(x) = \begin{cases} \left(e^{\boldsymbol{S}(y_{\ell+1}-x)}\boldsymbol{s} + e^{\boldsymbol{S}(2\Delta-(y_{\ell+1}-x))}\boldsymbol{s}\right), & j \in \mathcal{S}_+, \\ \left(e^{\boldsymbol{S}(x-y_\ell)}\boldsymbol{s} + e^{\boldsymbol{S}(2\Delta-(x-y_\ell))}\boldsymbol{s}\right), & j \in \mathcal{S}_-. \end{cases} \tag{4.22}$$

Intuitively, we take the density function $\boldsymbol{a}e^{\boldsymbol{S}x}\boldsymbol{s}$ on $x \in [0, 2\Delta)$, and 'fold' it back on itself around Δ, to create a density function on $[0, \Delta)$,

$$\boldsymbol{a}e^{\boldsymbol{S}x}\boldsymbol{s} + \boldsymbol{a}e^{\boldsymbol{S}(2\Delta-x)}\boldsymbol{s},\ x \in [0, \Delta).$$

The missing mass is now proportional to

$$\boldsymbol{a}e^{\boldsymbol{S}2\Delta}\boldsymbol{e},$$

which is less than the quantity $\boldsymbol{a}e^{\boldsymbol{S}\Delta}\boldsymbol{e}$ from the approximation scheme (4.20). There is nothing special about the choice of truncating the distribution at 2Δ in this construction, and we could choose any truncation value greater than Δ.

A third option is to approximate $\mathbb{P}(\boldsymbol{X}(t) \in (\,\mathrm{d}x, j) \mid \boldsymbol{X}(0) = (x_0, i))$ by

$$\int_{\boldsymbol{a} \in \mathcal{A}} \mathbb{P}(\boldsymbol{Y}(t) \in (\ell,\, \mathrm{d}\boldsymbol{a}, j) \mid \boldsymbol{Y}(0) = (\ell_0, \boldsymbol{a}_{\ell_0,i}(x_0), i)) \boldsymbol{a}\boldsymbol{v}_{\ell,j}(x)\, \mathrm{d}x, \tag{4.23}$$

where

$$\boldsymbol{a}\boldsymbol{v}_{\ell,j}(x) = \begin{cases} \boldsymbol{a}\left(e^{\boldsymbol{S}(y_{\ell+1}-x)} + e^{\boldsymbol{S}(2\Delta-(y_{\ell+1}-x))}\right)\left[I - e^{\boldsymbol{S}2\Delta}\right]^{-1}\boldsymbol{s}, & j \in \mathcal{S}_+, \\ \boldsymbol{a}\left(e^{\boldsymbol{S}(x-y_\ell)} + e^{\boldsymbol{S}(2\Delta-(x-y_\ell))}\right)\left[I - e^{\boldsymbol{S}2\Delta}\right]^{-1}\boldsymbol{s}, & j \in \mathcal{S}_-. \end{cases} \tag{4.24}$$

Here, we take the density function $\boldsymbol{a}e^{\boldsymbol{S}x}\boldsymbol{s}$, defined on $[0, \infty)$, and map it to a density function on $[0, \Delta)$ by $\Delta - |(x \mod 2\Delta) - \Delta|$. The resulting density function is

$$\begin{aligned} v(x) &= \sum_{m=0}^{\infty} \boldsymbol{a}\left(e^{\boldsymbol{S}(x+2\Delta m)} + e^{\boldsymbol{S}(2\Delta-x+2\Delta m)}\right)\boldsymbol{s} \\ &= \boldsymbol{a}\left(e^{\boldsymbol{S}x} + e^{\boldsymbol{S}(2\Delta-x)}\right)\sum_{m=0}^{\infty} e^{\boldsymbol{S}2\Delta m}\boldsymbol{s} \end{aligned}$$

$$= \boldsymbol{a}\left(e^{\boldsymbol{S}x} + e^{\boldsymbol{S}(2\Delta - x)}\right)\left[I - e^{S2\Delta}\right]^{-1}\boldsymbol{s}.$$

The difference between the approximation schemes is the way in which the vector $\boldsymbol{a}$ is used. We generalise this idea with the concept of *closing operators* which is a linear operator $\boldsymbol{v}(x) : \mathcal{A} \to \mathbb{R}$, for each $x \in [0, \Delta)$. For example, the closing operator in Equations (4.20) is the operator $\boldsymbol{v}(x)$, $x \in [0, \Delta)$ such that for any $\boldsymbol{a} \in \mathcal{A}$,

$$\boldsymbol{a}\boldsymbol{v}(x) = \boldsymbol{a}e^{\boldsymbol{S}x}\boldsymbol{s}. \tag{4.25}$$

Similarly, in (4.21)

$$\boldsymbol{a}\boldsymbol{v}(x) = \boldsymbol{a}\left(e^{\boldsymbol{S}x}\boldsymbol{s} + e^{\boldsymbol{S}(2\Delta - x)}\boldsymbol{s}\right), \tag{4.26}$$

and in (4.23)

$$\boldsymbol{a}\boldsymbol{v}(x) = \boldsymbol{a}\left(e^{\boldsymbol{S}x}\boldsymbol{s} + e^{\boldsymbol{S}(2\Delta - x)}\right)\left[I - e^{S2\Delta}\right]^{-1}\boldsymbol{s}. \tag{4.27}$$

We will use the notation

$$\boldsymbol{v}_{\ell,j}(x) = \begin{cases} \boldsymbol{v}(y_{\ell+1} - x) & j \in \mathcal{S}_+ \cup \mathcal{S}_{0+}, \\ \boldsymbol{v}(x - y_\ell) & j \in \mathcal{S}_- \cup \mathcal{S}_{0-}. \end{cases} \tag{4.28}$$

The operator $\boldsymbol{v}(x)$ is a choice which forms part of the definition of the approximation scheme. As we shall see, given certain properties of $\boldsymbol{v}(x)$, we can prove that the approximation scheme converges, and ensures positivity. In the cases above, all the closing operators (4.25)-(4.26) lead to an approximation which converges and, due to their interpretation as probability densities, ensure positivity.

Comments on linearity and normalisation We considered taking the closing operators to be

$$\boldsymbol{a}\boldsymbol{v}(x) = \frac{\boldsymbol{a}\left(e^{\boldsymbol{S}x}\boldsymbol{s} + e^{\boldsymbol{S}(2\Delta - x)}\boldsymbol{s}\right)}{1 - \boldsymbol{a}e^{\boldsymbol{S}2\Delta}\boldsymbol{e}}, \tag{4.29}$$

which also ensures that the solution always preserves mass (the result of integrating (4.29) over $[0, \Delta)$ is 1). However, (4.29) is not a linear operator. The importance of this fact is that for a linear operator,

$$\mathbb{E}[\boldsymbol{A}(t)\boldsymbol{v}(x)] = \mathbb{E}[\boldsymbol{A}(t)]\boldsymbol{v}(x),$$

thus, the computation of $\mathbb{E}[\boldsymbol{A}(t)\boldsymbol{v}(x)]$ can be achieved by first computing $\mathbb{E}[\boldsymbol{A}(t)]$ and then applying $\boldsymbol{v}(x)$ to the result. In this sense, $\mathbb{E}[\boldsymbol{A}(t)]$ contains all the information about the history of the process up to time t needed to compute $\mathbb{E}[\boldsymbol{A}(t)\boldsymbol{v}(x)]$. This is much the same as in RAPs and QBD-RAPs where $\mathbb{E}[\boldsymbol{A}(t)]$ is all that is required to compute probabilities about the future of the processes from time t onwards.

When $\boldsymbol{v}(x)$ is not linear then, in general,

$$\mathbb{E}[\boldsymbol{A}(t)\boldsymbol{v}(x)] \neq \mathbb{E}[\boldsymbol{A}(t)]\boldsymbol{v}(x).$$

For example, with (4.29),

$$\mathbb{E}[\boldsymbol{A}(t)\boldsymbol{v}(x)] = \int_{\boldsymbol{a}\in\mathcal{A}} \mathbb{P}(\boldsymbol{A}(t) \in \mathrm{d}\boldsymbol{a}) \frac{\boldsymbol{a}\left(e^{\boldsymbol{S}x}\boldsymbol{s} + e^{\boldsymbol{S}(2\Delta - x)}\boldsymbol{s}\right)}{1 - \boldsymbol{a}e^{\boldsymbol{S}2\Delta}\boldsymbol{e}}. \tag{4.30}$$

While the calculation in (4.30) is theoretically possible, practically it is not. The seemingly innocuous integral over $\boldsymbol{a} \in \mathcal{A}$ actually amounts to an integral over all possible sample paths of the QBD-RAP (i.e. all possible events and event times of the QBD-RAP before time t) and in general, there is no 'nice' way to do this computation.

In the linear case, computation of

$$\mathbb{E}[\boldsymbol{A}(t)] = \int_{\boldsymbol{a}\in\mathcal{A}} \mathbb{P}(\boldsymbol{A}(t) \in \mathrm{d}\boldsymbol{a})\boldsymbol{a},$$

amounts to a matrix exponential calculation.

In practice, we may use

$$\frac{\mathbb{E}\left[\boldsymbol{A}(t)\right]\left(e^{\boldsymbol{S}x}\boldsymbol{s} + e^{\boldsymbol{S}(2\Delta - x)}\boldsymbol{s}\right)}{1 - \mathbb{E}\left[\boldsymbol{A}(t)\right]e^{\boldsymbol{S}2\Delta}\boldsymbol{e}} \tag{4.31}$$

as a normalised version of (4.26), and this seems to work well. However, this is not the same as using the closing operator (4.29).

4.8 Approximating operators and distributions of a fluid-fluid queue

We conclude this chapter with a few remarks about applying the QBD-RAP approximation to approximate operators, particularly those arising in the analysis of fluid-fluid queues from Bean & O'Reilly (2014).

First, a comment on approximating the operator $\mathbb{R}$, defined in Lemma 2.1. Recall from Section 3.5.1, in the context of the DG scheme we could approximate $\mathbb{R}$ as a projection. In the context of the QBD-RAP scheme it is less clear how one can approximate $\mathbb{R}$, in general. If the rate functions $r_i(x)$ are piecewise constant with $r_i(x) = r_i^k$ on cell $\mathcal{D}_{k,j}$, then a suitable approximation for $\mathbb{R}_i^k$ is $1/|r_i^k|\boldsymbol{I}_p$, a diagonal matrix with $1/|r_i^k|$ down the diagonal – this is the case in the numerical experiments in Chapter 7. For more general cases we could consider approximating $r_i(x)$ by a piecewise constant function, then apply the method above.

Second, a comment on the approximation $\boldsymbol{\Psi}$ of the first-return operator Ψ characterised in Theorem 2.2. The approximation $\boldsymbol{\Psi}$ if computed by solving the matrix Riccati equation

$$\boldsymbol{D}^{+-} + \boldsymbol{\Psi}\boldsymbol{D}^{-+}\boldsymbol{\Psi} + \boldsymbol{D}^{++}\boldsymbol{\Psi} + \boldsymbol{\Psi}\boldsymbol{D}^{--} = \mathbf{0}, \tag{4.32}$$

where

$$\boldsymbol{D}^{mn} = \boldsymbol{R}^{m}(\boldsymbol{B}^{mn} + \boldsymbol{B}^{m0}(\boldsymbol{B}^{00})^{-1}\boldsymbol{B}^{0n}),\ m,n \in \{+,-\}.$$

Given the stochastic interpretation of the QBD-RAP scheme, and assuming the rate functions $r_i(x)$ are piecewise constant with $r_i(x) = r_i^k$ on each cell $\mathcal{D}_{k,i}$, the QBD-RAP approximation of the fluid-fluid queue forms a RAP-modulated fluid process (Peralta Gutierrez 2019, Bean et al. 2021). Hence, the approximation $\boldsymbol{\Psi}$ has a stochastic interpretation in terms of the first-return operator of the RAP-modulated fluid process. This is important. Equation (4.32) determines, but does not define $\boldsymbol{\Psi}$. When we use the DG scheme to approximate Ψ we still solve (4.32) (with the DG approximation to the generator $\mathbb{B}$), however, this does not define what the solution means in terms of the fluid-fluid queue (other than it is the solution to (4.32)). In comparison, there is a rigorous definition of $\boldsymbol{\Psi}$ when the QBD-RAP scheme is used in the approximation (Peralta Gutierrez 2019, Bean et al. 2021).

By Proposition 4.3 of Peralta Gutierrez (2019), for the QBD-RAP scheme the approximation $\boldsymbol{\Psi}$ has the property $\boldsymbol{\Psi}\mathbf{1} = \mathbf{1}$ as long as the event that level process of the RAP-modulated fluid returns to 0 in finite time occurs with probability 1.

Lastly, the generator $\boldsymbol{B}$ is a linear operator. Considering the remarks in Section 3.7 this can be advantageous. This is unlike the positivity preserving DG schemes which use slope limiting which are not linear operators in the presence of oscillations.

4.9 Briefly, on expected accuracy

For p odd, Horváth, Horváth & Telek (2020) estimate that the variance of the class of CMEs decreases faster than order $1/p^2$ and estimate that the asymptotic rate of decreases in the variance is of order $1/p^{2.14}$. Hence, we might expect the rate of convergence of the QBD-RAP scheme to be approximately order $1/p^2$ for certain problems. This is in line with the positivity preserving DG scheme mentioned in 2.8.4. We investigate the performance of the QBD-RAP method further in Chapter 7.

Chapter 5

Convergence of the QBD-RAP before the first orbit restart epoch, τ_1

This chapter details a convergence of the approximation scheme constructed in Chapter 4 on the event the first *orbit restart epoch* is yet to occur. Unless the QBD-RAP hits a boundary and is immediately reflected, an orbit restart epoch corresponds to a change of level. If the process hits a boundary and is immediately reflected, then there is no change of level, but the orbit process does 'restart' at this time. We will define orbit restart epochs more precisely, later. From the stochastic interpretation in Chapter 4, the orbit restart epochs approximate the hitting times of the fluid queue on the points $\{y_\ell\}$ when the fluid level is not at a boundary, or the exit times of the boundaries when sticky boundaries are present and the fluid queue is at the boundary. Thus, this chapter proves a convergence of the QBD-RAP scheme to the fluid queue in each of the sets $\{0\}, \mathcal{D}_0, \mathcal{D}_1, \ldots, \mathcal{D}_K, \{y_{K+1}\}$.

In Chapter 6, we use the main results of this chapter to prove further convergence results for the QBD-RAP scheme, and ultimately provide a global result. Conceptually Chapter 6 stitches together the convergence on each of the sets $\{0\}, \mathcal{D}_0, \ldots, \mathcal{D}_K, \{y_{K+1}\}$ proved in this chapter to claim the global convergence. Chapters 5 and 6 differ somewhat in their proof techniques: Chapter 5 relies more heavily on concentrated-matrix-exponential-specific arguments whereas Chapter 6 uses more traditional arguments such as the Markov property, time-homogeneity and the law of total probability. We now detail the QBD-RAP scheme with which we will work throughout this chapter, and detail the structure of the chapter.

Recall that the QBD-RAP is constructed using *matrix exponential distributions* to model, approximately, the sojourn time of the fluid queue in a given interval. This chapter shows a type of convergence under the assumption that the variance of the matrix exponential distribution(s) used in the construction tends to 0. The result applies to any sequence of matrix exponential distributions such that the variance tends to zero. The generality of the result is necessitated by the fact that, in practice, we use the class of *concentrated matrix exponential distributions* found numerically in (Horváth, Horváth &

Telek 2020), for which there are relatively few known properties.

In this chapter we work exclusively with the *augmented state space model* to model phases with rates $c_i = 0$ as described in Section 4.3.

Further, to model an initial condition, $\varphi(0) = k \in \mathcal{S}_0$, we use the ephemeral set of phases $\mathcal{S}_0^{*,k}$ as described in Section 4.6. Approximating the initial condition $\varphi(0) = k \in \mathcal{S}_0$ in this way greatly simplifies some results in this chapter. In Appendix C we provide results which prove convergence of the approximation without the need to model the initial condition $\varphi(0) = k \in \mathcal{S}_0$ by an ephemeral set of phases. Appendix C relies on the fact that $\boldsymbol{a}_{\ell_0,i}(\Delta - x_0)$ and $\boldsymbol{a}_{\ell_0,i}(x_0)\boldsymbol{D}$ are 'close', in some sense, then leverages the results of this chapter to prove various bounds which ultimately show convergence. Beyond that, Appendix C provides limited further insight into the QBD-RAP process, it is also long and somewhat tedious, hence why it is in an appendix.

In this chapter we analyse the distribution of the QBD-RAP scheme up to the first orbit restart epoch. The structure of the analysis is to first partition the distribution of the QBD-RAP at time t (where t is before the first orbit restart epoch) on the number of changes of phase from $\mathcal{S}_+ \cup \mathcal{S}_{+0}$ to $\mathcal{S}_-$ or $\mathcal{S}_- \cup \mathcal{S}_{-0}$ to $\mathcal{S}_+$. We will refer to changes of phase from $\mathcal{S}_+ \cup \mathcal{S}_{+0}$ to $\mathcal{S}_-$ as *up-down* transitions and changes of phase from $\mathcal{S}_- \cup \mathcal{S}_{-0}$ to $\mathcal{S}_+$ as *down-up* transitions. Next, for each term in the partition we take the Laplace transform with respect to time. This is convenient as it enables algebraic manipulations such that we can separate the Laplace transforms into one factor solely about the orbit process of the QBD-RAP and one expression about the phase process and associated rates. Once we have established a convenient algebraic form, we then turn our attention to bounds and convergence, establishing bounds for the difference between the Laplace transforms of the QBD-RAP just described and corresponding Laplace transforms of the fluid queue. Thus, we establish a convergence result for the Laplace transforms with respect to time for each of the distributions in the partition. We then wish to 'undo' the partitioning on the event of a given number of up-down and down-up transitions before the first orbit restart epoch to establish a convergence result for the Laplace transform on the event that the first orbit restart epoch is yet to occur. To 'undo' the partitioning, we use the Dominated Convergence Theorem.

The main steps of the convergence result of this chapter are listed below.

1. Define the partition on the number of up-down and down-up transitions and the collections sample paths of the QBD-RAP and fluid queue with which we will work. Describe the distributions of the processes on theses sample paths and compute their Laplace transforms with respect to time. (Sections 5.1, 5.2 and 5.3).

2. Show error bounds for the difference between the Laplace transforms of the QBD-RAP and the fluid queue for each term in the partition. (Section 5.4).

3. Show a geometric domination condition so that we may apply the Dominated Convergence Theorem to 'undo' the partitioning and hence prove convergence of the

Laplace transforms. (Section 5.5).

First we give some preliminaries and some technical assumptions we will use in the proofs.

Preliminaries

Suppose we have a sequence, $\{Z^{(p)}\}_{p\geq 1}$, of matrix exponential random variables with, $Z^{(p)} \sim ME(\boldsymbol{\alpha}^{(p)}, \boldsymbol{S}^{(p)}, \boldsymbol{s}^{(p)})$, such that $\mathbb{E}[Z^{(p)}] = \Delta$ and $\text{Var}\left(Z^{(p)}\right) \to 0$ as $p \to \infty$. For notational convenience we suppose that p is the order of the representation $(\boldsymbol{\alpha}^{(p)}, \boldsymbol{S}^{(p)}, \boldsymbol{s}^{(p)})$, but strictly speaking, this is not necessary. We use the superscript (p) to denote dependence on the underlying choice of matrix exponential distribution that is used in the construction of the QBD-RAP scheme. To simplify notation, we may omit the superscript (p) when it is not necessary.

In the following we show error bounds for an arbitrary parameter $\varepsilon > 0$. However, keep in mind the ultimate intention is to show convergence, for which we choose this parameter to be $\varepsilon^{(p)} = \text{Var}\left(Z^{(p)}\right)^{1/3}$. Other notations that have been defined, which are functions of $Z^{(p)}$ and therefore also implicitly depend on p, are $\boldsymbol{\alpha}^{(p)}$, $\boldsymbol{S}^{(p)}$, $\boldsymbol{s}^{(p)}$, $\boldsymbol{S}_i^{(p)}$, $\boldsymbol{s}_i^{(p)}$, $\boldsymbol{D}^{(p)}$, $\mathcal{A}^{(p)}$, $\boldsymbol{Y}^{(p)}(t) = (L^{(p)}(t), \boldsymbol{A}^{(p)}(t), \phi^{(p)}(t))$, $\boldsymbol{Y}_{\boldsymbol{\alpha}}^{(p)}(t)$, $\boldsymbol{y}_0^{(p)}$.

In the following we show various results which involve integrating a function g, or a sequence of functions $g_1, g_2, \ldots$. We make the following assumptions about such functions,

Assumptions 5.1. *Let g be a function $g : [0,\infty) \to [0,\infty)$ which is*
(i) non-negative,

$$g(x) \geq 0 \text{ for all } x \geq 0,$$

(ii) bounded,

$$g(x) \leq G < \infty \text{ for all } x \geq 0,$$

(iii) integrable,

$$\int_{x=0}^{\infty} g(x)\,\mathrm{d}x \leq \widehat{G} < \infty,$$

(iv) and Lipschitz continuous

$$|g(x) - g(u)| \leq L|x-u| \text{ for all } x,\, u \geq 0,\, 0 < L < \infty. \tag{5.1}$$

We also need a sequence of closing operators which we denote by $\boldsymbol{v}^{(p)}$. For the convergence results, we require the following properties of the closing operators $\boldsymbol{v}^{(p)}(x)$, $x \in [0, \Delta)$.

Properties 5.2. *Let $\{\boldsymbol{v}^{(p)}(x)\}_{p\geq 1}$ be a sequence of closing operators such that they may be decomposed into $\boldsymbol{v}^{(p)}(x) = \boldsymbol{w}^{(p)}(x) + \widetilde{\boldsymbol{w}}^{(p)}(x)$, where;*
(i) for $x \in [0, \Delta), u, v \geq 0$,

$$\boldsymbol{\alpha}^{(p)} e^{\boldsymbol{S}^{(p)}(u+v)}(-\boldsymbol{S}^{(p)})^{-1}\widetilde{\boldsymbol{w}}^{(p)}(x) \leq \boldsymbol{\alpha}^{(p)} e^{\boldsymbol{S}^{(p)}u}(-\boldsymbol{S}^{(p)})^{-1}\widetilde{\boldsymbol{w}}^{(p)}(x).$$

(ii) for $x \in [0, \Delta), u \geq 0$,

$$\boldsymbol{\alpha}^{(p)} e^{\boldsymbol{S}^{(p)}u}(-\boldsymbol{S}^{(p)})^{-1}\widetilde{\boldsymbol{w}}^{(p)}(x) = \widetilde{G}_{\boldsymbol{v}}^{(p)} \to 0, \quad as\ p \to \infty.$$

(iii) for $x \in [0, \Delta), u \geq 0$,

$$\boldsymbol{\alpha}^{(p)} e^{\boldsymbol{S}^{(p)}u}(-\boldsymbol{S}^{(p)})^{-1}\boldsymbol{w}^{(p)}(x) \leq \boldsymbol{\alpha}^{(p)} e^{\boldsymbol{S}^{(p)}u}\boldsymbol{e}G_{\boldsymbol{v}},$$

for some $0 \leq G_{\boldsymbol{v}} < \infty$ independent of p for $p > p_0$ where $p_0 < \infty$.
(iv) for $\boldsymbol{a} \in \mathcal{A}$, $u \geq 0$,

$$\int_{x\in[0,\Delta)} \boldsymbol{a}^{(p)} e^{\boldsymbol{S}^{(p)}u}\boldsymbol{v}^{(p)}(x)\,\mathrm{d}x \leq \boldsymbol{a}^{(p)} e^{\boldsymbol{S}^{(p)}u}\boldsymbol{e}.$$

(v) Let g be a function satisfying the Assumptions 5.1. For $u \leq \Delta - \varepsilon^{(p)}$, $v \in [0, \Delta)$, then

$$\left| \int_{x=0}^{\infty} \frac{\boldsymbol{\alpha}^{(p)} e^{\boldsymbol{S}^{(p)}(u+x)}}{\boldsymbol{\alpha}^{(p)} e^{\boldsymbol{S}^{(p)}u}\boldsymbol{e}} \boldsymbol{v}^{(p)}(v) g(x)\,\mathrm{d}x - g(\Delta - u - v)1(u + v \leq \Delta - \varepsilon^{(p)}) \right| = |r_{\boldsymbol{v}}^{(p)}(u, v)|,$$

where

$$\int_{u=0}^{\Delta} \left| r_{\boldsymbol{v}}^{(p)}(u, v) \right| \mathrm{d}u \leq R_{\boldsymbol{v},1}^{(p)} \to 0$$

and

$$\int_{v=0}^{\Delta} \left| r_{\boldsymbol{v}}^{(p)}(u, v) \right| \mathrm{d}v \leq R_{\boldsymbol{v},2}^{(p)} \to 0$$

as $\mathrm{Var}(Z^{(p)}) \to 0$.

In Appendix C we provide results which show that the closing operators (4.25) - (4.27) satisfy Properties 5.2.

Though it is a slight abuse of notation, for convenience, let us write

$$\mathbb{P}(\boldsymbol{Y}(t) \in (\ell,\, \mathrm{d}x, j) \mid \boldsymbol{Y}(0) = \boldsymbol{y}_0)$$

in place of

$$\int_{\boldsymbol{a}\in\mathcal{A}} \mathbb{P}(\boldsymbol{Y}(t) \in (\ell,\, \mathrm{d}\boldsymbol{a}, j) \mid \boldsymbol{Y}(0) = \boldsymbol{y}_0)\boldsymbol{a}\boldsymbol{v}_{\ell,j}(x)\,\mathrm{d}x \tag{5.2}$$

where $\boldsymbol{v}_{\ell,j}(x)$ is a closing operator. Expression (5.2) is an approximation to

$$\mathbb{P}(\boldsymbol{X}(t) \in (\,\mathrm{d}x, j) \mid \boldsymbol{X}(0) = (x_0, i)), \tag{5.3}$$

$x \in \mathcal{D}_{\ell,j}$, $x_0 \in \mathcal{D}_{\ell_0,i}$. Further, let us write

$$\mathbb{P}(\boldsymbol{Y}(t) \in (\ell, E, j) \mid \boldsymbol{Y}(0) = \boldsymbol{y}_0)$$

in place of

$$\int_{x\in E} \int_{\boldsymbol{a}\in\mathcal{A}} \mathbb{P}(\boldsymbol{Y}(t) \in (\ell,\, \mathrm{d}\boldsymbol{a}, j) \mid \boldsymbol{Y}(0) = \boldsymbol{y}_0)\boldsymbol{a}\boldsymbol{v}_{\ell,j}(x)\,\mathrm{d}x \tag{5.4}$$

for some measurable set $E \subseteq \mathcal{D}_{\ell,j}$.

Ultimately, in Chapter 6, we will apply the Extended Continuity Theorem for Laplace transforms (Feller 1957, Chapter XIII, Theorem 2a) to claim convergence. The Extended Continuity Theorem for Laplace transforms requires us to show convergence of the Laplace transform pointwise with respect to the transform parameter, λ, on the set $\lambda \in \mathbb{R}$, $\lambda > 0$. Therefore, we can fix $\lambda \in \mathbb{R}$, $\lambda > 0$ in the following.

5.1 Describing the distribution of the QBD-RAP before τ_1

In this chapter we are interested in the QBD-RAP up to the first orbit restart epoch, which we denote by $\tau_1^{(p)}$, which is the random (stopping) time at which the QBD-RAP changes level, or hits the boundary, or exits a boundary, for the first time. More precisely,

$$\tau_1^{(p)} = \inf\left\{t > 0 \mid L^{(p)}(t) \neq L^{(p)}(0),\ \text{or } (L^{(p)}(t), \boldsymbol{A}^{(p)}(t), \varphi(t)) \text{ hits a boundary}\right\}.$$

At this time, the orbit process is restarted at the initial value $\boldsymbol{A}^{(p)}(\tau_1^{(p)}) = \boldsymbol{\alpha}^{(p)}$, unless the QBD-RAP hits a sticky boundary and is temporarily absorbed, in which case the orbit process is restarted at the value $\boldsymbol{A}^{(p)}(\tau_1^{(p)}) = 1$.*

Consider the initial condition $\boldsymbol{y}_0^{(p)} = (L^{(p)}(0), \boldsymbol{A}^{(p)}(0), \varphi(0)) = (\ell_0, \boldsymbol{a}_{\ell_0,i}^{(p)}(x_0), i)$, where $\ell_0 \in \mathcal{K}^\circ$, and $i \in \mathcal{S}$. The value $\boldsymbol{y}_0$ is the approximation to the initial condition $\boldsymbol{X}(0) = (x_0, i)$. We are interested in the quantity, $f^{\ell_0,(p)}(t)(x, j; x_0, i)\,\mathrm{d}x$ given by

$$\int_{\boldsymbol{a}\in\mathcal{A}^{(p)}} \mathbb{P}\Big(\boldsymbol{A}^{(p)}(t) \in\, \mathrm{d}\boldsymbol{a}, \varphi(t) = j, t < \tau_1^{(p)} \mid \boldsymbol{Y}^{(p)}(0) = \boldsymbol{y}_0^{(p)}\Big)\boldsymbol{a}\boldsymbol{v}_{\ell_0,j}^{(p)}(x)\,\mathrm{d}x$$

*Recall from the discussion in Section 4.5 in the paragraph above Remark 4.3, that the orbit process is not actually required to model the behaviour at the boundary. We set it to be $\boldsymbol{A}(t) = 1$ for all times t when the QBD-RAP is at the boundary for notational convenience.

$$= (\boldsymbol{e}_i \otimes \boldsymbol{a}^{(p)}_{\ell_0,i}(x_0)) e^{\boldsymbol{B}^{(p)}t} (\boldsymbol{e}'_j \otimes \boldsymbol{v}^{(p)}_{\ell_0,j}(x))\,\mathrm{d}x, \tag{5.5}$$

which is the QBD-RAP approximation to the distribution

$$\mathbb{P}(\boldsymbol{X}(t) \in (\,\mathrm{d}x, j), t < \tau_1^X \mid \boldsymbol{X}(0) = (x_0, i)).$$

For now consider $\varphi(0) = i \in \mathcal{S}_+ \cup \mathcal{S}_-$. As we shall see, certain Laplace transform expressions for phases $i^* \in \mathcal{S}_0^{*,k}$ with rate $c_{i^*} = 0$ can be written as a linear combination of certain Laplace transforms of phases in $\mathcal{S}_+ \cup \mathcal{S}_-$.

Now, introduce a partition on the number of up-down and down-up transitions of the sample paths. Denote by $\{\Sigma_m\}_{m\geq 1}$ the sequence of (stopping) times at which $\{\varphi(t)\}$ has an up-down transition (i.e. jumps from $\mathcal{S}_+ \cup \mathcal{S}_{+0}$ to $\mathcal{S}_-$) for the mth time. Denote by $\{\Gamma_m\}_{m\geq 1}$ the sequence of (stopping) times at which $\{\varphi(t)\}$ has a down-up transition (i.e. jumps from $\mathcal{S}_- \cup \mathcal{S}_{-0}$ to $\mathcal{S}_+$) for the mth time. More precisely, for sample paths with $\varphi(0) \in \mathcal{S}_+$, let $\Gamma_0 = 0$, then for $m \geq 1$,

$$\Sigma_m := \inf\{t > \Gamma_{m-1} \mid \varphi(t) \in \mathcal{S}_-\}, \tag{5.6}$$
$$\Gamma_m := \inf\{t > \Sigma_m \mid \varphi(t) \in \mathcal{S}_+\}. \tag{5.7}$$

Similarly, for sample paths with $\varphi(0) \in \mathcal{S}_-$, let $\Sigma_0 = 0$, then for $m \geq 1$,

$$\Gamma_m := \inf\{t > \Sigma_{m-1} \mid \varphi(t) \in \mathcal{S}_+\}, \tag{5.8}$$
$$\Sigma_m := \inf\{t > \Gamma_m \mid \varphi(t) \in \mathcal{S}_-\}. \tag{5.9}$$

For times t such that $\Gamma_m \leq t < \Sigma_{m+1}$, then $\varphi(t) \in \mathcal{S}_+ \cup \mathcal{S}_{+0}$. For times t such that $\Sigma_{m+1} \leq t < \Gamma_{m+1}$, then $\varphi(t) \in \mathcal{S}_- \cup \mathcal{S}_{-0}$.

With these stopping times, partition the sample paths of the QBD-RAP by the number of up-down and down-up transition as follows. For $x_0 \in \mathcal{D}_{\ell_0,i}$, $x \in \mathcal{D}_{\ell_0,j}$, $t \geq 0$, $\ell_0 \in \mathcal{K}^\circ$, $m \geq 0$, and for $i \in \mathcal{S}_+, j \in \mathcal{S}_+ \cup \mathcal{S}_{+0}$, let $f^{\ell_0,(p)}_{m,+,+}(t)(x, j; x_0, i)$, be

$$\begin{aligned}
&\int_{\boldsymbol{a}\in\mathcal{A}^{(p)}} \mathbb{P}\Big(\boldsymbol{A}^{(p)}(t) \in \,\mathrm{d}\boldsymbol{a}, \varphi(t) = j, t < \tau_1^{(p)}, \Gamma_m \leq t < \Sigma_{m+1} \mid \boldsymbol{Y}^{(p)}(0) = \boldsymbol{y}_0^{(p)}\Big)\boldsymbol{a}\boldsymbol{v}^{(p)}_{\ell_0,j}(x) \\
&= \int_{\sigma_1=0}^{t} (\boldsymbol{e}_i \otimes \boldsymbol{a}^{(p)}_{\ell_0,i}(x_0)) e^{\boldsymbol{B}^{(p)}_{++}\sigma_1} \boldsymbol{B}^{(p)}_{+-} \int_{\gamma_1=\sigma_1}^{t} e^{\boldsymbol{B}^{(p)}_{--}(\gamma_1-\sigma_1)} \boldsymbol{B}^{(p)}_{-+} \cdots \int_{\gamma_m=\sigma_m}^{t} e^{\boldsymbol{B}^{(p)}_{--}(\gamma_m-\sigma_m)} \boldsymbol{B}^{(p)}_{-+} \\
&\quad \times e^{\boldsymbol{B}^{(p)}_{++}(t-\gamma_m)} \left(\boldsymbol{e}'_j \otimes \boldsymbol{v}^{(p)}_{\ell_0,j}(x)\right) \mathrm{d}\gamma_m \,\mathrm{d}\sigma_m \ldots \mathrm{d}\gamma_1 \,\mathrm{d}\sigma_1,
\end{aligned} \tag{5.10}$$

where $\boldsymbol{e}_k$ is a row-vector of zeros except the kth position which is a 1. Analogously, for $i \in \mathcal{S}_+$, $j \in \mathcal{S}_- \cup \mathcal{S}_{-0}$, let

$$\begin{aligned}
&f^{\ell_0,(p)}_{m+1,+,-}(t)(x, j; x_0, i) \\
&= \int_{\boldsymbol{a}\in\mathcal{A}^{(p)}} \mathbb{P}\Big(\boldsymbol{A}^{(p)}(t) \in \,\mathrm{d}\boldsymbol{a}, \varphi(t) = j, t < \tau_1^{(p)}, \Sigma_{m+1} \leq t < \Gamma_{m+1} \mid \boldsymbol{Y}^{(p)}(0) = \boldsymbol{y}_0^{(p)}\Big)\boldsymbol{a}\boldsymbol{v}^{(p)}_{\ell_0,j}(x)
\end{aligned}$$

$$= \int_{\sigma_1=0}^{t} (\boldsymbol{e}_i \otimes \boldsymbol{a}^{(p)}_{\ell_0,i}(x_0)) e^{\boldsymbol{B}^{(p)}_{++}\sigma_1} \boldsymbol{B}^{(p)}_{+-} \int_{\gamma_1=\sigma_1}^{t} e^{\boldsymbol{B}^{(p)}_{--}(\gamma_1-\sigma_1)} \boldsymbol{B}^{(p)}_{-+} \cdots \int_{\sigma_{m+1}=\gamma_m}^{t} e^{\boldsymbol{B}^{(p)}_{++}(\sigma_{m+1}-\gamma_m)}$$
$$\times \boldsymbol{B}^{(p)}_{+-} e^{\boldsymbol{B}^{(p)}_{--}(t-\sigma_{m+1})} \left(\boldsymbol{e}'_j \otimes \boldsymbol{v}^{(p)}_{\ell_0,j}(x)\right) \mathrm{d}\sigma_1 \, \mathrm{d}\gamma_1 \ldots \mathrm{d}\sigma_m \, \mathrm{d}\gamma_m \, \mathrm{d}\sigma_{m+1} \tag{5.11}$$

for $i \in \mathcal{S}_-$, $j \in \mathcal{S}_+ \cup \mathcal{S}_{+0}$ let

$$f^{\ell_0,(p)}_{m+1,-,+}(t)(x,j;x_0,i)$$
$$= \int_{\boldsymbol{a}\in\mathcal{A}^{(p)}} \mathbb{P}\Big(\boldsymbol{A}^{(p)}(t) \in \mathrm{d}\boldsymbol{a}, \varphi(t) = j, t < \tau^{(p)}_1, \Gamma_{m+1} \le t < \Sigma_{m+1} \mid \boldsymbol{Y}^{(p)}(0) = \boldsymbol{y}^{(p)}_0\Big) \boldsymbol{a}\boldsymbol{v}^{(p)}_{\ell_0,j}(x)$$
$$= \int_{\gamma_1=0}^{t} (\boldsymbol{e}_i \otimes \boldsymbol{a}^{(p)}_{\ell_0,i}(x_0)) e^{\boldsymbol{B}^{(p)}_{--}\gamma_1} \boldsymbol{B}^{(p)}_{-+} \int_{\sigma_1=\gamma_1}^{t} e^{\boldsymbol{B}^{(p)}_{++}(\sigma_1-\gamma_1)} \boldsymbol{B}^{(p)}_{+-} \cdots \int_{\gamma_{m+1}=\sigma_m}^{t} e^{\boldsymbol{B}^{(p)}_{--}(\gamma_{m+1}-\sigma_m)}$$
$$\times \boldsymbol{B}^{(p)}_{-+} e^{\boldsymbol{B}^{(p)}_{++}(t-\gamma_{m+1})} \left(\boldsymbol{e}'_j \otimes \boldsymbol{v}^{(p)}_{\ell_0,j}(x)\right) \mathrm{d}\gamma_1 \, \mathrm{d}\sigma_1 \ldots \mathrm{d}\gamma_m \, \mathrm{d}\sigma_m \, \mathrm{d}\gamma_{m+1} \tag{5.12}$$

and for $i \in \mathcal{S}_-, j \in \mathcal{S}_- \cup \mathcal{S}_{-0}$ let

$$f^{\ell_0,(p)}_{m,-,-}(t)(x,j;x_0,i)$$
$$= \int_{\boldsymbol{a}\in\mathcal{A}^{(p)}} \mathbb{P}\Big(\boldsymbol{A}^{(p)}(t) \in \mathrm{d}\boldsymbol{a}, \varphi(t) = j, t < \tau^{(p)}_1, \Sigma_m \le t < \Gamma_{m+1} \mid \boldsymbol{Y}^{(p)}(0) = \boldsymbol{y}^{(p)}_0\Big) \boldsymbol{a}\boldsymbol{v}^{(p)}_{\ell_0,j}(x)$$
$$= \int_{\gamma_1=0}^{t} (\boldsymbol{e}_i \otimes \boldsymbol{a}^{(p)}_{\ell_0,i}(x_0)) e^{\boldsymbol{B}^{(p)}_{--}\gamma_1} \boldsymbol{B}^{(p)}_{-+} \int_{\sigma_1=\gamma_1}^{t} e^{\boldsymbol{B}^{(p)}_{++}(\sigma_1-\gamma_1)} \boldsymbol{B}^{(p)}_{+-} \cdots \int_{\sigma_m=\gamma_m}^{t} e^{\boldsymbol{B}^{(p)}_{++}(\sigma_m-\gamma_m)} \boldsymbol{B}^{(p)}_{+-}$$
$$\times e^{\boldsymbol{B}^{(p)}_{--}(t-\sigma_m)} \left(\boldsymbol{e}'_j \otimes \boldsymbol{v}^{(p)}_{\ell_0,j}(x)\right) \mathrm{d}\sigma_m \, \mathrm{d}\gamma_m \ldots \mathrm{d}\sigma_1 \, \mathrm{d}\gamma_1. \tag{5.13}$$

Now, for $q, r \in \{+,-\}$, $q \ne r$, define

$$f^{\ell_0,(p)}_{q,q}(t)(x,j;x_0,i) := \sum_{m=0}^{\infty} f^{\ell_0,(p)}_{m,q,q}(t)(x,j;x_0,i) \qquad i \in \mathcal{S}_q,\ j \in \mathcal{S}_q \cup \mathcal{S}_{q0},$$
$$f^{\ell_0,(p)}_{q,r}(t)(x,j;x_0,i) := \sum_{m=1}^{\infty} f^{\ell_0,(p)}_{m,q,r}(t)(x,j;x_0,i) \qquad i \in \mathcal{S}_q,\ j \in \mathcal{S}_0 \cup \mathcal{S}_{r0},$$

so that

$$f^{\ell_0,(p)}(t)(x,j;x_0,i) = \begin{cases} f^{\ell_0,(p)}_{q,q}(t)(x,j;x_0,i) & i \in \mathcal{S}_q,\ j \in \mathcal{S}_q \cup \mathcal{S}_{q0}, \\ f^{\ell_0,(p)}_{q,r}(t)(x,j;x_0,i) & i \in \mathcal{S}_q,\ j \in \mathcal{S}_r \cup \mathcal{S}_{r0}. \end{cases} \tag{5.14}$$

Recall, in this chapter we suppose the QBD-RAP approximation uses ephemeral states $\mathcal{S}^{*,k}_0$ to model the fluid queue whenever the phase starts in $k \in \mathcal{S}_0$. In general, for $r \in \{+,-\}$, $m \ge 0$, we define

$$f^{\ell_0,(p)}_{m,0,r}(t)(x,j;x_0,k) := \sum_{q\in\{+,-\}} \sum_{i\in\mathcal{S}_q} \int_{t_0=0}^{t} \boldsymbol{e}_k e^{\boldsymbol{T}_{00}t_0} \boldsymbol{T}_{0i} f^{\ell_0,(p)}_{m+1(q\ne r),q,r}(t-t_0)(x,j;x_0,i) \, \mathrm{d}t_0. \tag{5.15}$$

Upon taking the Laplace transform of (5.15) the convolutions become products, so the Laplace transform of $f^{\ell_0}_{m,0,r}(t)(x,j;x_0,k)$ is a linear combination of the Laplace transforms of (5.10)-(5.13). Thus, once we show convergence for the Laplace transforms of (5.10)-(5.13) we get convergence of the Laplace transform for starting in $\mathcal{S}_0^{*,k}$ too.

5.2 Describing the distribution of the fluid queue before τ_1

Let τ_1^X be the minimum of the time at which $\{X(t)\}$ hits a boundary, or exits a boundary, or exits $\mathcal{D}_{\ell_0}$, where $X(0) = x_0 \in \mathcal{D}_{\ell_0}$. More precisely,

$$\tau_1^X = \min \left\{ \begin{array}{c} \inf\{t > 0 \mid X(t) = y_\ell, \ell \in \mathcal{K}\}, \\ \inf\{t > 0 \mid X(t) \neq 0, X(0) = 0\}, \\ \inf\{t > 0 \mid X(t) \neq y_{K+1}, X(0) = y_{K+1}\} \end{array} \right\}.$$

Consider the measures on the Borel sets of $\mathcal{D}_{\ell_0,j}$ given by

$$\mu^{\ell_0}(t)(\cdot, j; x_0, i) := \mathbb{P}(\boldsymbol{X}(t) \in (\cdot, j), t < \tau_1^X \mid \boldsymbol{X}(0) = (x_0, i)), \tag{5.16}$$

$\ell_0 \in \mathcal{K}^\circ$, $x_0 \in \mathcal{D}_{\ell_0,i}, i, j \in \mathcal{S}, t \geq 0$. In words, this is the distribution of the fluid queue at time t on the event that the fluid level remains within $\mathcal{D}_{\ell_0}$ up to and including time t and is in phase j at time t, given that is started at $X(0) = x_0 \in \mathcal{D}_{\ell_0,i}$ in phase i.

I do not know of any simple expression for (5.16). There are expressions for the Laplace transform of (5.16) with respect to time. One is in terms of the of first return matrices $\Psi(\lambda)$ and $\Xi(\lambda)$ (Bean et al. 2009*b*). Here we opt for another expression for the Laplace transform which is obtained by partitioning as follows.

As before, we use the sequence of up-down transition times, $\{\Sigma_m\}_{m\geq 1}$, and the sequence of down-up transition times, $\{\Gamma_m\}_{m\geq 1}$, to partition sample paths. The events $\{\Gamma_m \leq t < \Sigma_{m+1}\}$, and $\{\Sigma_{m+1} \leq t < \Gamma_{m+1}\}$, $m \geq 0$, partition the sample paths of (5.16) into periods where the fluid is either non-decreasing or non-increasing, respectively; see Figure 5.1.

For $m \geq 0$, $i \in \mathcal{S}_+$, $j \in \mathcal{S}_+ \cup \mathcal{S}_{+0}$ define

$$\mu^{\ell_0}_{m,+,+}(t)(\cdot, j; x_0, i) = \mathbb{P}(\boldsymbol{X}(t) \in (\cdot, j), t < \tau_1^X, \Gamma_m \leq t < \Sigma_{m+1} \mid \boldsymbol{X}(0) = (x_0, i)), \tag{5.17}$$

for $i \in \mathcal{S}_+$, $j \in \mathcal{S}_- \cup \mathcal{S}_{-0}$ define

$$\mu^{\ell_0}_{m+1,+,-}(t)(\cdot, j; x_0, i) = \mathbb{P}(\boldsymbol{X}(t) \in (\cdot, j), t < \tau_1^X, \Sigma_{m+1} \leq t < \Gamma_{m+1} \mid \boldsymbol{X}(0) = (x_0, i)), \tag{5.18}$$

for $i \in \mathcal{S}_-$, $j \in \mathcal{S}_+ \cup \mathcal{S}_{+0}$ define

$$\mu^{\ell_0}_{m+1,-,+}(t)(\cdot, j; x_0, i) = \mathbb{P}(\boldsymbol{X}(t) \in (\cdot, j), t < \tau_1^X, \Gamma_{m+1} \leq t < \Sigma_{m+1} \mid X(0) = (x_0, i)), \tag{5.19}$$

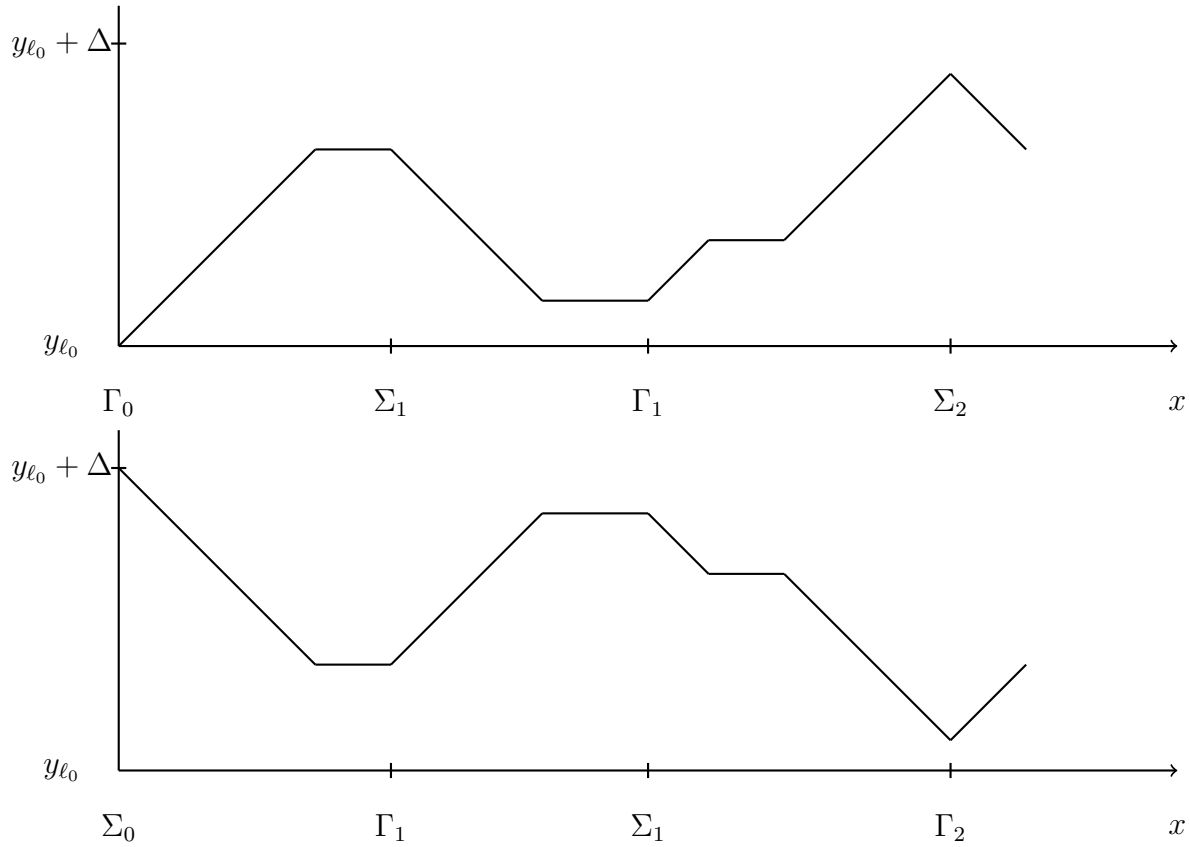

Figure 5.1: Sample paths and times of up-down and down-up transitions for $\varphi(0) \in \mathcal{S}_+$ (top) and $\varphi(0) \in \mathcal{S}_-$ (bottom).

and for $i \in \mathcal{S}_-$, $j \in \mathcal{S}_- \cup \mathcal{S}_{-0}$ define

$$\mu^{\ell_0}_{m,-,-}(t)(\cdot, j; x_0, i) = \mathbb{P}(\boldsymbol{X}(t) \in (\cdot, j), t < \tau_1^X, \Sigma_m \leq t < \Gamma_{m+1} \mid \boldsymbol{X}(0) = (x_0, i)). \quad (5.20)$$

Furthermore, for $q, r \in \{+, -\}$, $q \neq r$, let

$$\mu^{\ell_0}_{q,q}(t)(\cdot, j; x_0, i) := \sum_{m=0}^{\infty} \mu^{\ell_0}_{m,q,q}(t)(\cdot, j; x_0, i) \qquad i \in \mathcal{S}_q,\ j \in \mathcal{S}_q \cup \mathcal{S}_{q0}, \quad (5.21)$$

$$\mu^{\ell_0}_{q,r}(t)(\cdot, j; x_0, i) := \sum_{m=1}^{\infty} \mu^{\ell_0}_{m,q,r}(t)(\cdot, j; x_0, i) \qquad i \in \mathcal{S}_q,\ j \in \mathcal{S}_r \cup \mathcal{S}_{r0}. \quad (5.22)$$

Then we can write (5.16) as

$$\mu^{\ell_0}(t)(\cdot, j; x_0, i) = \begin{cases} \mu^{\ell_0}_{q,q}(t)(\cdot, j; x_0, i) & i \in \mathcal{S}_q,\ j \in \mathcal{S}_q \cup \mathcal{S}_{q0}, \\ \mu^{\ell_0}_{q,r}(t)(\cdot, j; x_0, i) & i \in \mathcal{S}_q,\ j \in \mathcal{S}_r \cup \mathcal{S}_{r0}, \end{cases} \quad (5.23)$$

For states $k \in \mathcal{S}_0$, and $q \in \{+, -\}$, $r \in \{+, -\}$, $m \geq 0$, then

$$\mu^{\ell_0}_{m,0,r}(t)(x, j; x_0, k) := \sum_{q \in \{+,-\}} \sum_{i \in \mathcal{S}_q} \int_{t_0=0}^{t} \boldsymbol{e}_k e^{\boldsymbol{T}_{00} t_0} \boldsymbol{T}_{0i} \mu^{\ell_0}_{m+1(q \neq r),q,r}(t - t_0)(x, j; x_0, i)\, \mathrm{d}t_0. \quad (5.24)$$

5.3 Laplace transforms with respect to time of the distributions before τ_1

In this section we take the Laplace transform with respect to time of the densities $f^{\ell_0,(p)}_{m,q,r}(t)(x, j; x_0, k)$, and measures $\mu^{\ell_0}_{m,q,r}(t)(x, j; x_0, k)$, $q \in \{+, -, 0\}$, $r \in \{+, -\}$. The Laplace transform is convenient as it allows us to manipulate the expressions for the QBD-RAP into one component related to the orbit process and one component related the phase process and the rates c_i, $i \in \mathcal{S}$.

The following matrices play a key role in the analysis of fluid queues (see, for example, Bean et al. (2009*b*), da Silva Soares (2005)). Here, they appear in the Laplace transforms of the QBD-RAP and the fluid queue. Define matrices

$$\begin{aligned}
\boldsymbol{Q}_{+0}(\lambda) &= \boldsymbol{C}_+^{-1} \boldsymbol{T}_{+0} \left[\lambda \boldsymbol{I} - \boldsymbol{T}_{00}\right]^{-1}, \\
\boldsymbol{Q}_{-0}(\lambda) &= \boldsymbol{C}_-^{-1} \boldsymbol{T}_{-0} \left[\lambda \boldsymbol{I} - \boldsymbol{T}_{00}\right]^{-1}, \\
\boldsymbol{Q}_{++}(\lambda) &= \boldsymbol{C}_+^{-1} \left(\boldsymbol{T}_{++} - \lambda \boldsymbol{I} + \boldsymbol{T}_{+0} \left[\lambda \boldsymbol{I} - \boldsymbol{T}_{00}\right]^{-1} \boldsymbol{T}_{0+}\right), \\
\boldsymbol{Q}_{+-}(\lambda) &= \boldsymbol{C}_+^{-1} \left(\boldsymbol{T}_{+-} + \boldsymbol{T}_{+0} \left[\lambda \boldsymbol{I} - \boldsymbol{T}_{00}\right]^{-1} \boldsymbol{T}_{0-}\right), \\
\boldsymbol{Q}_{--}(\lambda) &= \boldsymbol{C}_-^{-1} \left(\boldsymbol{T}_{--} - \lambda \boldsymbol{I} + \boldsymbol{T}_{-0} \left[\lambda \boldsymbol{I} - \boldsymbol{T}_{00}\right]^{-1} \boldsymbol{T}_{0-}\right),
\end{aligned}$$

$$\boldsymbol{Q}_{-+}(\lambda) = \boldsymbol{C}_{-}^{-1}\left(\boldsymbol{T}_{-+} + \boldsymbol{T}_{-0}\left[\lambda\boldsymbol{I} - \boldsymbol{T}_{00}\right]^{-1}\boldsymbol{T}_{0+}\right),$$

and matrix functions,

$$\boldsymbol{H}^{++}(\lambda, x) = \left[h_{ij}^{++}(\lambda, x)\right]_{i\in\mathcal{S}_+, j\in\mathcal{S}_+\cup\mathcal{S}_{+0}} := e^{\boldsymbol{Q}_{++}(\lambda)x}\left[\boldsymbol{C}_{+}^{-1} \quad \boldsymbol{Q}_{+0}(\lambda)\right], \tag{5.25}$$

$$\boldsymbol{H}^{--}(\lambda, x) = \left[h_{ij}^{--}(\lambda, x)\right]_{i\in\mathcal{S}_-, j\in\mathcal{S}_-\cup\mathcal{S}_{-0}} := e^{\boldsymbol{Q}_{--}(\lambda)x}\left[\boldsymbol{C}_{-}^{-1} \quad \boldsymbol{Q}_{-0}(\lambda)\right], \tag{5.26}$$

$$\boldsymbol{H}^{+-}(\lambda, x) = \left[h_{ij}^{+-}(\lambda, x)\right]_{i\in\mathcal{S}_+, j\in\mathcal{S}_-} := e^{\boldsymbol{Q}_{++}(\lambda)x}\boldsymbol{Q}_{+-}(\lambda), \tag{5.27}$$

$$\boldsymbol{H}^{-+}(\lambda, x) = \left[h_{ij}^{-+}(\lambda, x)\right]_{i\in\mathcal{S}_-, j\in\mathcal{S}_+} := e^{\boldsymbol{Q}_{--}(\lambda)x}\boldsymbol{Q}_{-+}(\lambda), \tag{5.28}$$

for $x, \lambda \geq 0$. The function $h_{ij}^{++}(\lambda, x)$ ($h_{ij}^{--}(\lambda, x)$) is the Laplace transform with respect to time of the time taken for the fluid level to shift by an amount x whilst remaining in phases in $\mathcal{S}_+ \cup \mathcal{S}_{+0}$ ($\mathcal{S}_- \cup \mathcal{S}_{-0}$), given the phase was initially $i \in \mathcal{S}_+$ ($i \in \mathcal{S}_-$) (Bean et al. 2005*b*). The function $h_{ij}^{+-}(\lambda, x)$ ($h_{ij}^{-+}(\lambda, x)$) is the Laplace transform with respect to time of the time taken for the fluid level, $\{X(t)\}$ to shift by an amount x whilst remaining in phases in $\mathcal{S}_+ \cup \mathcal{S}_{+0}$ ($\mathcal{S}_- \cup \mathcal{S}_{-0}$), after which time the phase instantaneously changes to $j \in \mathcal{S}_-$ ($\mathcal{S}_+$), given the phase was initially $i \in \mathcal{S}_+$ ($\mathcal{S}_-$) (Bean et al. 2005*b*).

Consider taking the Laplace transform with respect to time of (5.17);

$$\int_{t=0}^{\infty} e^{-\lambda t}\mu_{m,+,+}^{\ell_0}(t)(\cdot, j; x_0, i)\,\mathrm{d}t = \widehat{\mu}_{m,+,+}^{\ell_0}(\lambda)(\cdot, j; x_0, i) \tag{5.29}$$

where we use $\widehat{\mu}_{m,+,+}^{\ell_0}(\lambda)(\cdot, j; x_0, i)$ to denote the Laplace transform with respect to time of (5.17). Throughout, we use the hat $\widehat{\ }$ notation to denote Laplace transforms with respect to time.

From the stochastic interpretations of the Laplace transforms (5.25)-(5.28) the Laplace transforms with respect to time, $\widehat{\mu}_{m,+,+}^{\ell_0}(\lambda)(\,\mathrm{d}x, j; x_0, i)$, of (5.17) are given by

$$\widehat{\mu}_{0,+,+}^{\ell_0}(\lambda)(\,\mathrm{d}x, j; x_0, i)\,\mathrm{d}x = h_{ij}^{++}(\lambda, x - x_0)1(x \geq x_0)\,\mathrm{d}x,$$

for $m = 0$, and

$$\begin{aligned}
&\int_{x_1=0}^{\Delta-(x_0-y_{\ell_0})} \boldsymbol{e}_i\boldsymbol{H}^{+-}(\lambda, \Delta - (x_0 - y_{\ell_0}) - x_1) \\
&\left[\prod_{r=1}^{m-1}\int_{x_{2r}=0}^{\Delta-x_{2r-1}} \boldsymbol{H}^{-+}(\lambda, \Delta - x_{2r} - x_{2r-1})\,\mathrm{d}x_{2r-1}\int_{x_{2r+1}=0}^{\Delta-x_{2r}} \boldsymbol{H}^{+-}(\lambda, \Delta - x_{2r+1} - x_{2r})\,\mathrm{d}x_{2r}\right] \\
&\int_{x_{2m}=0}^{\Delta-x_{2m-1}} \boldsymbol{H}^{-+}(\lambda, \Delta - x_{2m-1} - x_{2m})\,\mathrm{d}x_{2m-1}\boldsymbol{H}^{++}(\lambda, \Delta - x_{2m} - (y_{\ell_0+1} - x))\boldsymbol{e}_j' \\
&1(\Delta - x_{2m} - (y_{\ell_0+1} - x) \geq 0)\,\mathrm{d}x_{2m}\,\mathrm{d}x
\end{aligned} \tag{5.30}$$

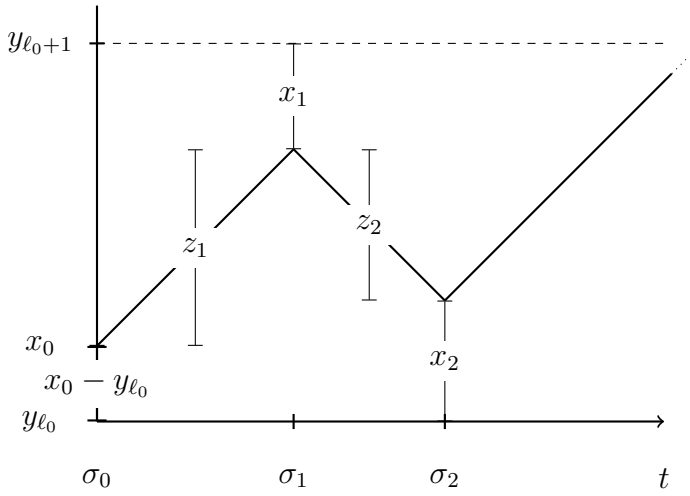

Figure 5.2: Sample paths corresponding to the Laplace transforms (5.30). $z_1 = \Delta - x_1 - (x - y_{\ell_0})$, $z_2 = \Delta - x_2 - x_1$.

for $m \geq 1$. Figure 5.2 shows an example of the sample paths to which these Laplace transforms correspond. Analogously, we can write down similar expressions for the Laplace transforms with respect to time of (5.17)-(5.20) (omitted).

Now consider taking the Laplace transform with respect to time of (5.10);

$$\int_{t=0}^{\infty} e^{-\lambda t} f_{m,+,+}^{\ell_0,(p)}(t)(x,j;x_0,i)\,\mathrm{d}t = \widehat{f}_{m,+,+}^{\ell_0,(p)}(\lambda)(x,j;x_0,i) \tag{5.31}$$

where we use $\widehat{f}_{m,+,+}^{\ell_0,(p)}(\lambda)(x,j;x_0,i)$ to denote the Laplace transform with respect to time of (5.10). Notice that (5.10) is a convolution. Hence, the Laplace transform with respect to time of (5.10) is

$$\begin{aligned}
&\widehat{f}_{m,+,+}^{\ell_0,(p)}(\lambda)(x,j;x_0,i) \\
&= (\boldsymbol{e}_i \otimes \boldsymbol{a}_{\ell_0,i}^{(p)}(x_0)) \int_{t_1=0}^{\infty} e^{-\lambda t_1} e^{\boldsymbol{B}_{++}^{(p)} t_1}\,\mathrm{d}t_1 \boldsymbol{B}_{+-}^{(p)} \int_{t_2=0}^{\infty} e^{-\lambda t_2} e^{\boldsymbol{B}_{--}^{(p)} t_2}\,\mathrm{d}t_2 \boldsymbol{B}_{-+}^{(p)} \\
&\ldots \int_{t_{2m}=0}^{\infty} e^{-\lambda t_{2m}} e^{\boldsymbol{B}_{--}^{(p)} t_{2m}}\,\mathrm{d}t_{2m} \boldsymbol{B}_{-+}^{(p)} \int_{t=0}^{\infty} e^{-\lambda t} e^{\boldsymbol{B}_{++}^{(p)} t}\,\mathrm{d}t \left(\boldsymbol{e}_j' \otimes \boldsymbol{v}_{\ell_0,j}^{(p)}(x)\right).
\end{aligned} \tag{5.32}$$

Analogous expressions can be computed for the Laplace transforms with respect to time of (5.11)-(5.13).

In Corollary D.4 in Appendix D, we show the following relation

$$\left[\boldsymbol{I}_{p|\mathcal{S}_m|} \quad \boldsymbol{0}_{p|\mathcal{S}_m| \times p|\mathcal{S}_0|}\right] \int_{t=0}^{\infty} e^{-\lambda t} e^{\boldsymbol{B}_{mm}^{(p)} t}\,\mathrm{d}t \boldsymbol{B}_{mn}^{(p)}$$

$$= \int_{x=0}^{\infty} \boldsymbol{H}^{mn}(\lambda, x) \otimes e^{\boldsymbol{S}^{(p)}x} \boldsymbol{D}^{(p)} \, \mathrm{d}x \begin{bmatrix} \boldsymbol{I}_{p|\mathcal{S}_n|} & \boldsymbol{0}_{p|\mathcal{S}_n| \times p|\mathcal{S}_0|} \end{bmatrix}, \tag{5.33}$$

for $m, n \in \{+, -\}$, $m \neq n$. Before we can apply this result, observe that, since $i \in \mathcal{S}_+$, we can write the initial vector in (5.32) as

$$\begin{aligned} (\boldsymbol{e}_i)_{1\times|\mathcal{S}_+\cup\mathcal{S}_{+0}|} \otimes \boldsymbol{a}^{(p)}_{\ell_0,i}(x_0) &= \begin{bmatrix} (\boldsymbol{e}_i)_{1\times|\mathcal{S}_+|} & \boldsymbol{0}_{1\times|\mathcal{S}_{+0}|} \end{bmatrix} \otimes \boldsymbol{a}^{(p)}_{\ell_0,i}(x_0) \\ &= \begin{bmatrix} (\boldsymbol{e}_i)_{1\times|\mathcal{S}_+|} \otimes \boldsymbol{a}^{(p)}_{\ell_0,i}(x_0) & \boldsymbol{0}_{1\times p|\mathcal{S}_{+0}|} \end{bmatrix} \\ &= \left((\boldsymbol{e}_i)_{1\times|\mathcal{S}_+|} \otimes \boldsymbol{a}^{(p)}_{\ell_0,i}(x_0)\right) \begin{bmatrix} \boldsymbol{I}_{p|\mathcal{S}_+|} & \boldsymbol{0}_{p|\mathcal{S}_+|\times p|\mathcal{S}_{+0}|} \end{bmatrix}, \end{aligned} \tag{5.34}$$

where explicitly indicate the size of the vectors $(\boldsymbol{e}_i)_{1\times|\mathcal{S}_+\cup\mathcal{S}_{+0}|}$ via the subscript outside the brackets. With this observation, applying (5.33) to the first integral in (5.32) transforms the expression to

$$\begin{aligned} &(\boldsymbol{e}_i \otimes \boldsymbol{a}^{(p)}_{\ell_0,i}(x_0)) \int_{x_1=0}^{\infty} \left(\boldsymbol{H}^{+-}(\lambda, x_1) \otimes e^{\boldsymbol{S}^{(p)}x} \boldsymbol{D}^{(p)}\right) \mathrm{d}x_1 \begin{bmatrix} \boldsymbol{I}_{p|\mathcal{S}_-|} & \boldsymbol{0}_{p|\mathcal{S}_-|\times p|\mathcal{S}_0|} \end{bmatrix} \\ &\int_{t_2=0}^{\infty} e^{-\lambda t_2} e^{\boldsymbol{B}^{(p)}_{--}t_2} \, \mathrm{d}t_2 \boldsymbol{B}^{(p)}_{-+} \cdots \int_{t_{2m}=0}^{\infty} e^{-\lambda t_{2m}} e^{\boldsymbol{B}^{(p)}_{--}t_{2m}} \, \mathrm{d}t_{2m} \boldsymbol{B}^{(p)}_{-+} \int_{t=0}^{\infty} e^{-\lambda t} e^{\boldsymbol{B}^{(p)}_{++}t} \, \mathrm{d}t \\ &\left(\boldsymbol{e}'_j \otimes \boldsymbol{v}^{(p)}_{\ell_0,j}(x)\right). \end{aligned} \tag{5.35}$$

We may now apply (5.33) to the second integral, after which we can apply (5.33) to the third integral and so on. Ultimately, after applying (5.33) to all the integrals in (5.32), we get

$$\begin{aligned} &(\boldsymbol{e}_i \otimes \boldsymbol{a}^{(p)}_{\ell_0,i}(x_0)) \left(\int_{x_1=0}^{\infty} \boldsymbol{H}^{+-}(\lambda, x_1) \otimes e^{\boldsymbol{S}^{(p)}x_1} \boldsymbol{D}^{(p)} \, \mathrm{d}x_1\right) \\ &\left[\prod_{r=1}^{m-1} \left(\int_{x_{2r}=0}^{\infty} \boldsymbol{H}^{-+}(\lambda, x_{2r}) \otimes e^{\boldsymbol{S}^{(p)}x_{2r}} \boldsymbol{D}^{(p)} \, \mathrm{d}x_{2r}\right)\right. \\ &\left.\left(\int_{x_{2r+1}=0}^{\infty} \boldsymbol{H}^{+-}(\lambda, x_{2r+1}) \otimes e^{\boldsymbol{S}^{(p)}x_{2r+1}} \boldsymbol{D}^{(p)} \, \mathrm{d}x_{2r+1}\right)\right] \\ &\left(\int_{x_{2m}=0}^{\infty} \boldsymbol{H}^{-+}(\lambda, x_{2m}) \otimes e^{\boldsymbol{S}^{(p)}x_{2m}} \boldsymbol{D}^{(p)} \, \mathrm{d}x_{2m}\right) \\ &\left(\int_{x_{2m+1}=0}^{\infty} \boldsymbol{H}^{++}(\lambda, x_{2m+1}) \otimes e^{\boldsymbol{S}^{(p)}x_{2m+1}} \, \mathrm{d}x_{2m+1}\right) \left(\boldsymbol{e}'_j \otimes \boldsymbol{v}^{(p)}_{\ell_0,j}(x)\right) \\ &= \int_{x_1=0}^{\infty} \cdots \int_{x_{2m+1}=0}^{\infty} \boldsymbol{e}_i \boldsymbol{M}^{m}_{++}(\lambda, x_1, \ldots, x_{2m+1}) \boldsymbol{e}'_j \\ &\times \boldsymbol{a}^{(p)}_{\ell_0,i}(x_0) \boldsymbol{N}^{2m+1,(p)}(\lambda, x_1, \ldots, x_{2m+1}) \boldsymbol{v}^{(p)}_{\ell_0,j}(x) \, \mathrm{d}x_{2m+1} \ldots \mathrm{d}x_1, \end{aligned} \tag{5.36}$$

by the Mixed Product Rule (see D), where we define matrices

$$\boldsymbol{M}^m_{++}(\lambda, x_1, \ldots, x_{2m+1}) = \prod_{r=1}^{m} \boldsymbol{H}^{+-}(\lambda, x_{2r-1})\boldsymbol{H}^{-+}(\lambda, x_{2r})\boldsymbol{H}^{++}(\lambda, x_{2m+1}),$$

for $m \geq 0$, and

$$\boldsymbol{N}^{n,(p)}(\lambda, x_1, \ldots, x_n) = \prod_{r=1}^{n-1} e^{\boldsymbol{S}^{(p)}x_r}\boldsymbol{D}^{(p)}e^{\boldsymbol{S}^{(p)}x_n},$$

for $n \geq 1$. By convention, we take a product over an empty set to be 1. The relation (5.33) is key to our analysis. It allows us to factorise the integrand of the Laplace transform (5.36) into one factor solely related to the orbit process $\{\boldsymbol{A}^{(p)}(t)\}$,

$$\boldsymbol{a}^{(p)}_{\ell_0,i}(x_0)\boldsymbol{N}^{2m+1,(p)}(\lambda, x_1, \ldots, x_{2m+1})\boldsymbol{v}^{(p)}_{\ell_0,j}(x)$$

and another factor solely related to the fluid queue,

$$\boldsymbol{e}_i\boldsymbol{M}^m_{++}(\lambda, x_1, \ldots, x_{2m+1})\boldsymbol{e}'_j.$$

Further, if we define matrices

$$\boldsymbol{M}^m_{-+}(\lambda, x_1, \ldots, x_{2m}) = \prod_{r=1}^{m-1} \boldsymbol{H}^{-+}(\lambda, x_{2r-1})\boldsymbol{H}^{+-}(\lambda, x_{2r})\boldsymbol{H}^{-+}(\lambda, x_{2m-1})\boldsymbol{H}^{++}(\lambda, x_{2m}),$$

$$\boldsymbol{M}^m_{+-}(\lambda, x_1, \ldots, x_{2m}) = \prod_{r=1}^{m-1} \boldsymbol{H}^{+-}(\lambda, x_{2r-1})\boldsymbol{H}^{-+}(\lambda, x_{2r})\boldsymbol{H}^{+-}(\lambda, x_{2m-1})\boldsymbol{H}^{--}(\lambda, x_{2m}),$$

for $m \geq 1$, and

$$\boldsymbol{M}^m_{--}(\lambda, x_1, \ldots, x_{2m+1}) = \prod_{r=1}^{m} \boldsymbol{H}^{-+}(\lambda, x_{2r-1})\boldsymbol{H}^{+-}(\lambda, x_{2r})\boldsymbol{H}^{--}(\lambda, x_{2m+1}),$$

for $m \geq 1$, then analogous expressions can be shown for the Laplace transforms of (5.11)-(5.13) in terms of these matrices. For $m \geq 0$,

$$\begin{aligned}
\widehat{f}^{\ell_0,(p)}_{m+1,-,+}(\lambda)(x, j; x_0, i) &= \int_{x_1=0}^{\infty} \cdots \int_{x_{2m+2}=0}^{\infty} \boldsymbol{e}_i\boldsymbol{M}^{m+1}_{-+}(\lambda, x_1, \ldots, x_{2m+2})\boldsymbol{e}'_j \\
&\quad \times \boldsymbol{a}^{(p)}_{\ell_0,i}(x_0)\boldsymbol{N}^{2m+2,(p)}(\lambda, x_1, \ldots, x_{2m+2})\boldsymbol{v}^{(p)}_{\ell_0,j}(x)\,\mathrm{d}x_{2m+2}\ldots\,\mathrm{d}x_1, \\
\widehat{f}^{\ell_0,(p)}_{m+1,+,-}(\lambda)(x, j; x_0, i) &= \int_{x_1=0}^{\infty} \cdots \int_{x_{2m+2}=0}^{\infty} \boldsymbol{e}_i\boldsymbol{M}^{m+1}_{+-}(\lambda, x_1, \ldots, x_{2m+2})\boldsymbol{e}'_j
\end{aligned}$$

$$\times \boldsymbol{a}_{\ell_0,i}^{(p)}(x_0)\boldsymbol{N}^{2m+1,(p)}(\lambda,x_1,\ldots,x_{2m+1})\boldsymbol{v}_{\ell_0,j}^{(p)}(x)\,\mathrm{d}x_{2m+2}\ldots\,\mathrm{d}x_1,$$

$$\widehat{f}_{m,-,-}^{\ell_0,(p)}(\lambda)(x,j;x_0,i)=\int_{x_1=0}^{\infty}\cdots\int_{x_{2m+1}=0}^{\infty}\boldsymbol{e}_i\boldsymbol{M}_{--}^{m}(\lambda,x_1,\ldots,x_{2m+1})\boldsymbol{e}_j'$$
$$\times \boldsymbol{a}_{\ell_0,i}^{(p)}(x_0)\boldsymbol{N}^{2m+1,(p)}(\lambda,x_1,\ldots,x_{2m+1})\boldsymbol{v}_{\ell_0,j}^{(p)}(x)\,\mathrm{d}x_{2m+1}\ldots\,\mathrm{d}x_1.$$

In general, for $k\in\mathcal{S}_0$, $q\in\{+,-\}$, $r\in\{+,-\}$, $m\geq 0$,

$$\widehat{f}_{m,0,r}^{\ell_0}(\lambda)(x,j;x_0,k):=\sum_{q\in\{+,-\}}\sum_{i\in\mathcal{S}_q}\boldsymbol{e}_k\left[\lambda\boldsymbol{I}-\boldsymbol{T}_{00}\right]^{-1}\boldsymbol{T}_{0i}\widehat{f}_{m+1(q\neq r),q,r}^{\ell_0}(\lambda)(x,j;x_0,i),\quad(5.37)$$

and

$$\widehat{\mu}_{m,0,r}^{\ell_0}(\lambda)(\,\mathrm{d}x,j;x_0,k):=\sum_{q\in\{+,-\}}\sum_{i\in\mathcal{S}_q}\boldsymbol{e}_k\left[\lambda\boldsymbol{I}-\boldsymbol{T}_{00}\right]^{-1}\boldsymbol{T}_{0i}\widehat{\mu}_{m+1(q\neq r),q,r}^{\ell_0}(\lambda)(\,\mathrm{d}x,j;x_0,i).\quad(5.38)$$

In Section 5.4 we establish that $\widehat{f}_{m,q,r}^{\ell_0,(p)}(\lambda)(x,j;x_0,k)\,\mathrm{d}x\to\widehat{\mu}_{m,q,r}^{\ell_0}(\lambda)(\,\mathrm{d}x,j;x_0,k)$, $q\in\{+,-,0\}$, $r\in\{+,-\}$. To do so we use the fact that the functions $h_{ij}^{qq}(\lambda,x)$, $h_{ij}^{qr}(\lambda,x)$, $q,r\in\{+,-\}$, $i\in\mathcal{S}_q, j\in\mathcal{S}_r\cup\mathcal{S}_{r0}$ and $\lambda>0$ satisfy the Assumptions 5.1 as functions of x. To this end, we observe the following bounds, which follow from the stochastic interpretation of the functions. Let $c_{min}=\min_{i\in\mathcal{S}_-\cup\mathcal{S}_+}|c_i|$ and recall that we fix $\lambda\in\mathbb{R},\lambda>0$. For all $\lambda>0$, there is some $0\leq G<\infty$ such that, for $q,r\in\{+,-\}$, $q\neq r$,

$$0\leq h_{ij}^{qq}(\lambda,x)\leq\max\{1/c_{min},1\}\leq G,\ i\in\mathcal{S}_q,\ j\in\mathcal{S}_q\cup\mathcal{S}_{q0},$$
$$0\leq h_{ij}^{qr}(\lambda,x)\leq\max_{k,\ell}\left[\boldsymbol{Q}_{qr}(0)\right]_{k,\ell}\leq G,\ i\in\mathcal{S}_q,\ j\in\mathcal{S}_r.$$

Furthermore, there exists some $0\leq\widehat{G}<\infty$ such that,

$$\int_{x=0}^{\infty}h_{ij}^{qq}(\lambda,x)\,\mathrm{d}x\leq\int_{x=0}^{\infty}h_{ij}^{qq}(0,x)\,\mathrm{d}x=\left[-\boldsymbol{Q}_{qq}(0)^{-1}\boldsymbol{C}_q\quad-\boldsymbol{Q}_{qq}(0)^{-1}\boldsymbol{Q}_{q0}(0)\right]_{ij}\leq\widehat{G},$$
$$\int_{x=0}^{\infty}h_{ij}^{qr}(\lambda,x)\,\mathrm{d}x\leq\int_{x=0}^{\infty}h_{ij}^{qr}(0,x)\,\mathrm{d}x=\left[-\boldsymbol{Q}_{qq}(0)^{-1}\boldsymbol{Q}_{qr}(0)\right]_{ij}\leq\widehat{G}.$$

Moreover, since $h_{ij}^{qq}(\lambda,x)$ and $h_{ij}^{qr}(\lambda,x)$, are matrix exponential functions with exponent which is a sub-generator matrix, then for every $\lambda>0$, $h_{ij}^{qq}(\lambda,x)$ and $h_{ij}^{qr}(\lambda,x)$ is Lipschitz continuous with respect to x on $x\in[0,\infty)$. Therefore, there exists some $0<L<\infty$ such that $\left|h_{ij}^{qq}(\lambda,x)-h_{ij}^{qq}(\lambda,y)\right|\leq L|x-y|$ and $\left|h_{ij}^{qr}(\lambda,x)-h_{ij}^{qr}(\lambda,y)\right|\leq L|x-y|$.

5.4 Convergence on fixed number of up-down/down-up transitions before τ_1

The main result of this chapter is the following theorem.

Theorem 5.3. *Let $\psi : \mathbb{R} \to \mathbb{R}$, be bounded, $|\psi| \leq F$. As $p \to \infty$, for $m \geq 1$, $q \in \{+,-,0\}$, $r \in \{+,-\}$, and for $m = 0$, $q = 0$, $r \in \{+,-\}$, and for $m = 0$, $q = r$, $q, r \in \{+,-\}$, then*

$$\int_{x\in\mathcal{D}_{\ell_0}} \widehat{f}^{\ell_0,(p)}_{m,q,r}(\lambda)(x,j;x_0,k)\psi(x)\,\mathrm{d}x \to \int_{x\in\mathcal{D}_{\ell_0}} \widehat{\mu}^{\ell_0}_{m,q,r}(\lambda)(\,\mathrm{d}x,j;x_0,k)\psi(x). \tag{5.39}$$

The proof of Theorem 5.3 is at the end of this section as it is the result of numerous other sub-results, which we now proceed to show. Notice that the convergence in Theorem 5.3 is a weak result as we integrate the spatial variable, x, against test functions ψ. This is necessary due to the discontinuity at $x = x_0$ in terms with $m = 0$.

Let $\Delta = \mathbb{E}[Z]$ be the mean of the matrix exponential random variable, Z. The convergence results rely on the fact that integrating a function, g say, against the density function of a matrix exponential random variable conditional on the ME-life-time surviving until some time $u < \Delta - \varepsilon$, approximates integrating said function against a Kronecker delta situated at $\Delta - u$, provided the variance of the ME is sufficiently low.

The next result is used in the proof of Theorem 5.3 to prove convergence on the event that there are no up-down or down-up transitions before, τ_1, the first orbit restart epoch.

Lemma 5.4. *Let $\psi : [0,\Delta) \to \mathbb{R}$ be bounded, $\psi(x) \leq F$. Then, for $x_0 \in \mathcal{D}_{\ell_0,i}$, $x \in \mathcal{D}_{\ell_0,j}$, $\ell_0 \in \mathcal{K}^\circ$, $\lambda > 0$, $q \in \{+,-\}$,*

$$\begin{aligned}&\left|\int_{x\in\mathcal{D}_{\ell_0,j}} \widehat{f}^{\ell_0,(p)}_{0,q,q}(\lambda)(x,j;x_0,i)\psi(x-y_{\ell_0})\,\mathrm{d}x - \int_{x\in\mathcal{D}_{\ell_0,j}} \mu^{\ell_0}_{0,q,q}(\lambda)(\,\mathrm{d}x,j;x_0,i)\psi(x-y_{\ell_0})\right| \\ &\leq \left(R^{(p)}_{\boldsymbol{v},2} + \varepsilon^{(p)}G\right)F.\end{aligned} \tag{5.40}$$

Proof. Let us write $x_0 = y_{\ell_0} + u$, for $u \in [0,\Delta)$. Recalling

$$\widehat{f}^{\ell_0,(p)}_{0,+,+}(\lambda)(v,j;y_{\ell_0}+u,i) = \int_{x_1=0}^{\infty} \frac{\boldsymbol{\alpha}^{(p)}e^{\boldsymbol{S}^{(p)}(u+x_1)}}{\boldsymbol{\alpha}^{(p)}e^{\boldsymbol{S}^{(p)}u}\boldsymbol{e}}\boldsymbol{v}^{(p)}_{\ell_0,j}(x)h^{++}_{ij}(\lambda,x_1)\,\mathrm{d}x_1,$$

and

$$\widehat{\mu}^{\ell_0}_{0,+,+}(\lambda)(\,\mathrm{d}v,j;y_{\ell_0}+u,i) = h^{++}_{ij}(\lambda,\Delta-u-x)1(u+x<\Delta)\,\mathrm{d}v,$$

then (5.40) is equal to

$$\begin{aligned}&\left|\int_{x\in\mathcal{D}_{\ell_0,j}}\int_{x_1=0}^{\infty} \frac{\boldsymbol{\alpha}^{(p)}e^{\boldsymbol{S}^{(p)}(u+x_1)}}{\boldsymbol{\alpha}^{(p)}e^{\boldsymbol{S}^{(p)}u}\boldsymbol{e}}\boldsymbol{v}^{(p)}_{\ell_0,j}(x)h^{++}_{ij}(\lambda,x_1)\,\mathrm{d}x_1\psi(x-y_{\ell_0})\,\mathrm{d}x\right. \\ &\qquad\left. - \int_{x\in\mathcal{D}_{\ell_0,j}} h^{++}_{ij}(\lambda,\Delta-u-x)1(\Delta-u-x\geq 0)\psi(x-y_{\ell_0})\,\mathrm{d}x\right|\end{aligned}$$

$$
\begin{aligned}
&\leq \int_{x\in\mathcal{D}_{\ell_0,j}} \left| \int_{x_1=0}^{\infty} \frac{\boldsymbol{\alpha}^{(p)} e^{\boldsymbol{S}^{(p)}(u+x_1)}}{\boldsymbol{\alpha} e^{\boldsymbol{S}^{(p)}u}\boldsymbol{e}} \boldsymbol{v}_{\ell_0,j}^{(p)}(x) h_{ij}^{++}(\lambda, x_1)\,\mathrm{d}x_1 \right. \\
&\quad \left. - h_{ij}^{++}(\lambda, \Delta-u-x)1(u+x<\Delta) \right| F\,\mathrm{d}x \\
&\leq \int_{x\in\mathcal{D}_{\ell_0,j}} \left| \int_{x_1=0}^{\infty} \frac{\boldsymbol{\alpha}^{(p)} e^{\boldsymbol{S}^{(p)}(u+x_1)}}{\boldsymbol{\alpha}^{(p)} e^{\boldsymbol{S}^{(p)}u}\boldsymbol{e}} \boldsymbol{v}_{\ell_0,j}^{(p)}(x) h_{ij}^{++}(\lambda, x_1)\,\mathrm{d}x_1 \right. \\
&\quad \left. - h_{ij}^{++}(\lambda, \Delta-u-x)1(u+x<\Delta-\varepsilon) \right| F\,\mathrm{d}x \\
&\quad + \int_{x\in\mathcal{D}_{\ell_0,j}} \left| h_{ij}^{++}(\lambda, \Delta-u-x)1(\Delta-\varepsilon\leq u+x<\Delta) \right| F\,\mathrm{d}x, \qquad (5.41)
\end{aligned}
$$

by the triangle inequality and since ψ is bounded. By Property 5.2(v), then

$$
\begin{aligned}
&\left| \int_{x_1=0}^{\infty} \frac{\boldsymbol{\alpha}^{(p)} e^{\boldsymbol{S}^{(p)}(u+x_1)}}{\boldsymbol{\alpha}^{(p)} e^{\boldsymbol{S}^{(p)}u}\boldsymbol{e}} \boldsymbol{v}_{\ell_0,j}^{(p)}(x) h_{ij}^{++}(\lambda, x_1)\,\mathrm{d}x_1 - h_{ij}^{++}(\lambda, \Delta-u-x)1(u+x<\Delta-\varepsilon) \right| \mathrm{d}x \\
&\leq |r_{\boldsymbol{v}}(u,x)|\,. \qquad (5.42)
\end{aligned}
$$

Hence, (5.41) is less than or equal to

$$
\begin{aligned}
&\int_{x\in\mathcal{D}_{\ell_0,j}} |r_{\boldsymbol{v}}^{(p)}(u,x)| F\,\mathrm{d}x + \int_{x\in\mathcal{D}_{\ell_0,j}} \left| h_{ij}^{++}(\lambda, \Delta-u-x)1(\Delta-\varepsilon\leq u+x<\Delta) \right| F\,\mathrm{d}x \\
&\leq R_{\boldsymbol{v},2}^{(p)} F + \varepsilon G F,
\end{aligned}
$$

since $|h_{ij}^{++}| \leq G$. Thus, we have shown (5.40) for $q = +$.

Using analogous arguments we can show (5.40) for $q = -$. □

Upon choosing $\varepsilon^{(p)} = \mathrm{Var}(Z^{(p)})^{1/3}$, then the bounds in (5.40) tend to 0.

Next, we proceed to show results needed to prove convergence on the event that there are one or more up-down or down-up transitions before the first orbit restart epoch. The expressions arising from the QBD-RAP which we wish to show converge have the form

$$
\left| \int_{x_1=0}^{\infty} g_1(x_1)\boldsymbol{k}(x_0) e^{\boldsymbol{S}x_1}\,\mathrm{d}x_1 \boldsymbol{D} \left[\prod_{k=2}^{n-1} \int_{x_k=0}^{\infty} g_k(x_k) e^{\boldsymbol{S}x_k}\,\mathrm{d}x_k \boldsymbol{D} \right] \int_{x_n=0}^{\infty} g_n(x_n) e^{\boldsymbol{S}x_n}\,\mathrm{d}x_n \boldsymbol{v}(x), \right. \qquad (5.43)
$$

where $n \geq 2$, $\boldsymbol{v}(x)$ is a closing operator with the Properties 5.2, $\{g_k\}$ are functions satisfying Assumptions 5.1 and $\boldsymbol{k}(x_0) = \boldsymbol{\alpha} e^{\boldsymbol{S}x_0} / \boldsymbol{\alpha} e^{\boldsymbol{S}x_0} \boldsymbol{e}$.

We now introduce some notation we will use in the sequel. Define $w_n(x_0, x)$ to be the expression (5.43). Define the column vectors

$$\mathcal{I}_{m,k}(u_k) = \left[\prod_{\ell=m}^{k-1} \int_{x_\ell=0}^{\infty} g_\ell(x_\ell) e^{\boldsymbol{S}x_\ell}\, \mathrm{d}x_\ell \boldsymbol{D}\right] \int_{x_k=0}^{\infty} g_k(x_k) e^{\boldsymbol{S}x_k}\, \mathrm{d}x_k e^{\boldsymbol{S}u_k} \boldsymbol{s} \tag{5.44}$$

for $m, k \in \{1, 2, \dots\}$, $m \leq k$, where a product over an empty set is equal to 1. Notice that $\mathcal{I}_{m,k}(u_k)$ can be written as

$$\mathcal{I}_{m,k}(u_k) = \int_{x_m=0}^{\infty} g_m(x_m) e^{\boldsymbol{S}x_m}\, \mathrm{d}x_m \boldsymbol{D} \mathcal{I}_{m+1,k}(u_k). \tag{5.45}$$

Define the row vectors

$$\mathcal{J}_{k+1,k+1}(u_k, x_{k+1}) := g_{k+1}(x_{k+1}) \frac{\boldsymbol{\alpha} e^{\boldsymbol{S}u_k}}{\boldsymbol{\alpha} e^{\boldsymbol{S}u_k} \boldsymbol{e}} e^{\boldsymbol{S}x_{k+1}}, \tag{5.46}$$

and

$$\begin{aligned}\mathcal{J}_{k+1,n}(u_k, x_{k+1}) := g_{k+1}(x_{k+1}) \frac{\boldsymbol{\alpha} e^{\boldsymbol{S}u_k}}{\boldsymbol{\alpha} e^{\boldsymbol{S}u_k} \boldsymbol{e}} e^{\boldsymbol{S}x_{k+1}} \boldsymbol{D} \left[\prod_{m=k+2}^{n-1} \int_{x_m=0}^{\infty} g_m(x_m) e^{\boldsymbol{S}x_m}\, \mathrm{d}x_m \boldsymbol{D}\right] \\ \times \int_{x_n=0}^{\infty} g_n(x_n) e^{\boldsymbol{S}x_n}\, \mathrm{d}x_n \end{aligned} \tag{5.47}$$

for $k, n \in \{0, 1, 2, \dots\}$, $k + 1 < n$. The vectors $\mathcal{J}_{k+1,n}(u_k, x_{k+1})$ can also be written recursively,

$$\mathcal{J}_{k+1,n}(u_k, x_{k+1}) = \mathcal{J}_{k+1,n-1}(u_k, x_{k+1}) \boldsymbol{D} \int_{x_n=0}^{\infty} g_n(x_n) e^{\boldsymbol{S}x_n}\, \mathrm{d}x_n. \tag{5.48}$$

Also define $\boldsymbol{D}(b) = \displaystyle\int_{u=0}^{b} e^{\boldsymbol{S}u} \boldsymbol{s} \frac{\boldsymbol{\alpha} e^{\boldsymbol{S}u}}{\boldsymbol{\alpha} e^{\boldsymbol{S}u} \boldsymbol{e}} \mathrm{d}u$.

We prove that (5.43) converges by writing it as

$$\begin{aligned}&\int_{x_1=0}^{\infty} g_1(x_1) \boldsymbol{k}(x_0) e^{\boldsymbol{S}x_1}\, \mathrm{d}x_1 \boldsymbol{D}(\Delta - \varepsilon) \left[\prod_{k=2}^{n-1} \int_{x_k=0}^{\infty} g_k(x_k) e^{\boldsymbol{S}x_k}\, \mathrm{d}x_k \boldsymbol{D}(\Delta - \varepsilon)\right] \\ &\quad \times \int_{x_n=0}^{\infty} g_n(x_n) e^{\boldsymbol{S}x_n}\, \mathrm{d}x_n \boldsymbol{v}(x) + \sum_{k=1}^{n-1} \int_{x_{k+1}=0}^{\infty} \int_{u_k=\Delta-\varepsilon}^{\infty} \boldsymbol{k}(x_0) \mathcal{I}_{1,k}(u_k) \mathcal{J}_{k+1,n}(u_k, x_{k+1}) \boldsymbol{v}(x).\end{aligned} \tag{5.49}$$

We show that each of the terms in the last summation in (5.49) are bounded by something which can be made arbitrarily small upon choosing the variance of the distribution $(\boldsymbol{\alpha}, \boldsymbol{S})$

to be sufficiently small. Then we show that the difference between the first term in (5.49) and the corresponding expression for the fluid queue is also bounded by something which can be made arbitrarily small. The decomposition in (5.49) is advantageous since in the first term, the matrices $\boldsymbol{D}(\Delta-\varepsilon)$ are the integrals $\int_{u=0}^{\Delta-\varepsilon} e^{\boldsymbol{S}u}\boldsymbol{s}\frac{\boldsymbol{\alpha}e^{\boldsymbol{S}u}}{\boldsymbol{\alpha}e^{\boldsymbol{S}u}\boldsymbol{e}}\mathrm{d}u$, so the variable of integration never exceeds $\Delta-\varepsilon$. As a result, we can use Chebyshev's inequality to bound the denominator in the integrand of $\boldsymbol{D}(\Delta-\varepsilon)$ near 1.

Our next result shows a bound for the terms in the last summation in (5.49).

Recall the row vector function $\boldsymbol{k}(x):[0,\infty)\to\mathcal{A}\subset\mathbb{R}^p$,

$$\boldsymbol{k}(x)=\frac{\boldsymbol{\alpha}e^{\boldsymbol{S}x}}{\boldsymbol{\alpha}e^{\boldsymbol{S}x}\boldsymbol{e}}.$$

Corollary 5.5. *Let $g_1,g_2,\dots,$ be functions satisfying the Assumptions 5.1 and let $\boldsymbol{v}(x)$ be a closing operator with the Properties 5.2, then, for $k,n\in\{1,2,\dots\}$, $k+1\le n$,*

$$\begin{aligned}&\int_{x_{k+1}=0}^{\infty}\int_{u_k=\Delta-\varepsilon}^{\infty}\boldsymbol{k}(x_0)\mathcal{I}_{1,k}(u_k)\mathcal{J}_{k+1,n}(u_k,x_{k+1})\boldsymbol{v}(x)\\&\le\frac{1}{\boldsymbol{\alpha}e^{\boldsymbol{S}x_0}\boldsymbol{e}}\left(\left(2\varepsilon+\frac{\mathrm{Var}(Z)}{\varepsilon}\right)G^2\widehat{G}^{n-2}G_{\boldsymbol{v}}+G\widehat{G}^n\widetilde{G}_{\boldsymbol{v}}\right)=:|r_1(n)|.\end{aligned}\tag{5.50}$$

The structure of the proof is as follows. First, recall that we can decompose $\boldsymbol{v}(x)=\boldsymbol{w}(x)+\widetilde{\boldsymbol{w}}(x)$, by Properties 5.2, hence we can decompose the left-hand side of (5.50) into

$$\begin{aligned}&\int_{x_{k+1}=0}^{\infty}\int_{u_k=\Delta-\varepsilon}^{\infty}\boldsymbol{k}(x_0)\mathcal{I}_{1,k}(u_k)\mathcal{J}_{k+1,n}(u_k,x_{k+1})\boldsymbol{w}(x)\\&+\int_{x_{k+1}=0}^{\infty}\int_{u_k=\Delta-\varepsilon}^{\infty}\boldsymbol{k}(x_0)\mathcal{I}_{1,k}(u_k)\mathcal{J}_{k+1,n}(u_k,x_{k+1})\widetilde{\boldsymbol{w}}(x).\end{aligned}\tag{5.51}$$

Next, we bound $\boldsymbol{k}(x_0)\mathcal{I}_{1,k}(u_k)$ and $\mathcal{J}_{k+1,n}(u_k,x_{k+1})\boldsymbol{w}(x)$. With these two bounds we can derive a bound for the first term in (5.51). A bound on the second term of (5.51) follows from the bound on $\boldsymbol{k}(x_0)\mathcal{I}_{1,n-1}(u_{n-1})$ along with Properties 5.2(i) and 5.2(ii) of $\widetilde{\boldsymbol{w}}$.

Proof. Step 1: Decompose the left-hand side of (5.50) as (5.51). Referring to the Properties 5.2, we can decompose the closing operator $\boldsymbol{v}(x)=\boldsymbol{w}(x)+\widetilde{\boldsymbol{w}}(x)$, and therefore, due to the linearity of the decomposition, we can decompose the left-hand side of (5.50) as (5.51).

Step 2: Show the following bound.

$$\boldsymbol{k}(x_0)\mathcal{I}_{1,k}(u_k)\le\frac{1}{\boldsymbol{\alpha}e^{\boldsymbol{S}x_0}\boldsymbol{e}}G\widehat{G}^{k-1}\boldsymbol{\alpha}e^{\boldsymbol{S}u_k}\boldsymbol{e}.\tag{5.52}$$

Recall the definition of $\boldsymbol{D} := \int_{u=0}^{\infty} e^{\boldsymbol{S}u}\boldsymbol{s}\frac{\boldsymbol{\alpha}e^{\boldsymbol{S}u}}{\boldsymbol{\alpha}e^{\boldsymbol{S}u}\boldsymbol{e}}\mathrm{d}u$ and substitute it into the left-hand side of (5.52),

$$\begin{aligned}\boldsymbol{k}(x_0)\mathcal{I}_{1,k}(u_k) &= \boldsymbol{k}(x_0)\int_{x_1=0}^{\infty} g_1(x_1)e^{\boldsymbol{S}x_1}\,\mathrm{d}x_1\boldsymbol{D}\mathcal{I}_{2,k}(u_k)\\ &= \boldsymbol{k}(x_0)\int_{x_1=0}^{\infty} g_1(x_1)e^{\boldsymbol{S}x_1}\,\mathrm{d}x_1\int_{u_1=0}^{\infty} e^{\boldsymbol{S}u_1}\boldsymbol{s}\frac{\boldsymbol{\alpha}e^{\boldsymbol{S}u_1}}{\boldsymbol{\alpha}e^{\boldsymbol{S}u_1}\boldsymbol{e}}\mathrm{d}u_1\mathcal{I}_{2,k}(u_k). \qquad (5.53)\end{aligned}$$

Since $|g_1| \le G$, then (5.53) is less than or equal to

$$\boldsymbol{k}(x_0)\int_{x_1=0}^{\infty} Ge^{\boldsymbol{S}x_1}\,\mathrm{d}x_1\int_{u_1=0}^{\infty} e^{\boldsymbol{S}u_1}\boldsymbol{s}\frac{\boldsymbol{\alpha}e^{\boldsymbol{S}u_1}}{\boldsymbol{\alpha}e^{\boldsymbol{S}u_1}\boldsymbol{e}}\mathrm{d}u_1\mathcal{I}_{2,k}(u_k). \qquad (5.54)$$

Computing the integral with respect to x_1 in (5.54) gives

$$\begin{aligned}&G\boldsymbol{k}(x_0)(-\boldsymbol{S})^{-1}\int_{u_1=0}^{\infty} e^{\boldsymbol{S}u_1}\boldsymbol{s}\frac{\boldsymbol{\alpha}e^{\boldsymbol{S}u_1}}{\boldsymbol{\alpha}e^{\boldsymbol{S}u_1}\boldsymbol{e}}\mathrm{d}u_1\mathcal{I}_{2,k}(u_k)\\ &= \frac{G}{\boldsymbol{\alpha}e^{\boldsymbol{S}x_0}\boldsymbol{e}}\int_{u_1=0}^{\infty}\boldsymbol{\alpha}e^{\boldsymbol{S}(x_0+u_1)}\boldsymbol{e}\frac{\boldsymbol{\alpha}e^{\boldsymbol{S}u_1}}{\boldsymbol{\alpha}e^{\boldsymbol{S}u_1}\boldsymbol{e}}\mathrm{d}u_1\mathcal{I}_{2,k}(u_k), \qquad (5.55)\end{aligned}$$

since $(-\boldsymbol{S})^{-1}$ and $e^{\boldsymbol{S}t}$ commute, $\boldsymbol{s} = -\boldsymbol{S}\boldsymbol{e}$ and $e^{\boldsymbol{S}(t+u)} = e^{\boldsymbol{S}t}e^{\boldsymbol{S}u}$. Since $\boldsymbol{\alpha}e^{\boldsymbol{S}(x_0+u_1)}\boldsymbol{e} \le \boldsymbol{\alpha}e^{\boldsymbol{S}u_1}\boldsymbol{e}$, then (5.55) is less than or equal to

$$G\frac{1}{\boldsymbol{\alpha}e^{\boldsymbol{S}x_0}\boldsymbol{e}}\int_{u_1=0}^{\infty}\boldsymbol{\alpha}e^{\boldsymbol{S}u_1}\boldsymbol{e}\frac{\boldsymbol{\alpha}e^{\boldsymbol{S}u_1}}{\boldsymbol{\alpha}e^{\boldsymbol{S}u_1}\boldsymbol{e}}\mathrm{d}u_1\mathcal{I}_{2,k}(u_k) = G\frac{1}{\boldsymbol{\alpha}e^{\boldsymbol{S}x_0}\boldsymbol{e}}\int_{u_1=0}^{\infty}\boldsymbol{\alpha}e^{\boldsymbol{S}u_1}\,\mathrm{d}u_1\mathcal{I}_{2,k}(u_k),$$

where we have cancelled the terms $\boldsymbol{\alpha}e^{\boldsymbol{S}u_1}\boldsymbol{e}$ on the numerator and denominator.[†]

Now integrate with respect to u_1 and use the facts that $(-\boldsymbol{S})^{-1}$ and $e^{\boldsymbol{S}x}$ commute, and $\boldsymbol{s} = -\boldsymbol{S}\boldsymbol{e}$, to get

$$\begin{aligned}&G\frac{1}{\boldsymbol{\alpha}e^{\boldsymbol{S}x_0}\boldsymbol{e}}\boldsymbol{\alpha}(-\boldsymbol{S})^{-1}\mathcal{I}_{2,k}(u_k) \qquad (5.56)\\ &= G\frac{1}{\boldsymbol{\alpha}e^{\boldsymbol{S}x_0}\boldsymbol{e}}\boldsymbol{\alpha}(-\boldsymbol{S})^{-1}\int_{x_2=0}^{\infty} g_2(x_2)e^{\boldsymbol{S}x_2}\,\mathrm{d}x_2\int_{u_2=0}^{\infty} e^{\boldsymbol{S}u_2}\boldsymbol{s}\frac{\boldsymbol{\alpha}e^{\boldsymbol{S}u_2}}{\boldsymbol{\alpha}e^{\boldsymbol{S}u_2}\boldsymbol{e}}\mathrm{d}u_2\mathcal{I}_{3,k}(u_k)\\ &= G\frac{1}{\boldsymbol{\alpha}e^{\boldsymbol{S}x_0}\boldsymbol{e}}\int_{x_2=0}^{\infty} g_2(x_2)\boldsymbol{\alpha}e^{\boldsymbol{S}x_2}\,\mathrm{d}x_2\int_{u_2=0}^{\infty} e^{\boldsymbol{S}u_2}\boldsymbol{e}\frac{\boldsymbol{\alpha}e^{\boldsymbol{S}u_2}}{\boldsymbol{\alpha}e^{\boldsymbol{S}u_2}\boldsymbol{e}}\mathrm{d}u_2\mathcal{I}_{3,k}(u_k) \qquad (5.57)\end{aligned}$$

Since $\boldsymbol{\alpha}e^{\boldsymbol{S}x_2}e^{\boldsymbol{S}u_2}\boldsymbol{e} \le \boldsymbol{\alpha}e^{\boldsymbol{S}u_2}\boldsymbol{e}$, then (5.57) is less than or equal to

$$G\frac{1}{\boldsymbol{\alpha}e^{\boldsymbol{S}x_0}\boldsymbol{e}}\int_{x_2=0}^{\infty} g_2(x_2)\,\mathrm{d}x_2\int_{u_2=0}^{\infty}\boldsymbol{\alpha}e^{\boldsymbol{S}u_2}\boldsymbol{e}\frac{\boldsymbol{\alpha}e^{\boldsymbol{S}u_2}}{\boldsymbol{\alpha}e^{\boldsymbol{S}u_2}\boldsymbol{e}}\mathrm{d}u_2\mathcal{I}_{3,k}(u_k)$$

[†]The cancellation of terms is important as, for $u_1 > \Delta$, then $\boldsymbol{\alpha}^{(p)}e^{\boldsymbol{S}^{(p)}u_1}\boldsymbol{e}$ becomes small as $p \to \infty$.

$$= G\frac{1}{\boldsymbol{\alpha}e^{\boldsymbol{S}x_0}\boldsymbol{e}}\int_{x_2=0}^{\infty} g_2(x_2)\,\mathrm{d}x_2\int_{u_2=0}^{\infty}\boldsymbol{\alpha}e^{\boldsymbol{S}u_2}\boldsymbol{e}\,\mathrm{d}u_2\mathcal{I}_{3,k}(u_k), \tag{5.58}$$

where we have cancelled the terms $\boldsymbol{\alpha}e^{\boldsymbol{S}u_2}\boldsymbol{e}$ on the numerator and denominator.[‡]

Now, since $\displaystyle\int_{x_2=0}^{\infty} g_2(x_2)\,\mathrm{d}x_2 \leq \widehat{G}$, then (5.58) is less than or equal to

$$G\frac{1}{\boldsymbol{\alpha}e^{\boldsymbol{S}x_0}\boldsymbol{e}}\widehat{G}\int_{u_2=0}^{\infty}\boldsymbol{\alpha}e^{\boldsymbol{S}u_2}\,\mathrm{d}u_2\mathcal{I}_{3,k}(u_k) = G\frac{1}{\boldsymbol{\alpha}e^{\boldsymbol{S}x_0}\boldsymbol{e}}\widehat{G}\boldsymbol{\alpha}(-\boldsymbol{S})^{-1}\mathcal{I}_{3,k}(u_k). \tag{5.59}$$

Repeating the arguments which got us from (5.56) to (5.59) another $k-2$ times gives the result.

Step 3: Show the bound

$$\mathcal{J}_{k+1,n}(u_k,x_{k+1})\boldsymbol{w}(x) \leq g_{k+1}(x_{k+1})\widehat{G}^{n-k-2}GG_{\boldsymbol{v}}. \tag{5.60}$$

The argument is much the same as that we used to bound (5.52).

Starting with the left-hand side, upon substituting $\boldsymbol{D}$,

$$\begin{aligned}
&\mathcal{J}_{k+1,n}(u_k,x_{k+1})\boldsymbol{w}(x)\\
&= \mathcal{J}_{k+1,n-1}(u_k,x_{k+1})\boldsymbol{D}\int_{x_n=0}^{\infty} g_n(x_n)e^{\boldsymbol{S}x_n}\,\mathrm{d}x_n\boldsymbol{w}(x)\\
&= \mathcal{J}_{k+1,n-1}(u_k,x_{k+1})\int_{u_{n-1}=0}^{\infty} e^{\boldsymbol{S}u_{n-1}}\boldsymbol{s}\frac{\boldsymbol{\alpha}e^{\boldsymbol{S}u_{n-1}}}{\boldsymbol{\alpha}e^{\boldsymbol{S}u_{n-1}}\boldsymbol{e}}\mathrm{d}u_{n-1}\int_{x_n=0}^{\infty} g_n(x_n)e^{\boldsymbol{S}x_n}\,\mathrm{d}x_n\boldsymbol{w}(x)\\
&\leq \mathcal{J}_{k+1,n-1}(u_k,x_{k+1})\int_{u_{n-1}=0}^{\infty} e^{\boldsymbol{S}u_{n-1}}\boldsymbol{s}\frac{\boldsymbol{\alpha}e^{\boldsymbol{S}u_{n-1}}}{\boldsymbol{\alpha}e^{\boldsymbol{S}u_{n-1}}\boldsymbol{e}}\mathrm{d}u_{n-1}\int_{x_n=0}^{\infty} Ge^{\boldsymbol{S}x_n}\,\mathrm{d}x_n\boldsymbol{w}(x),
\end{aligned} \tag{5.61}$$

since $|g_n| \leq G$. By Property 5.2(*iii*) of $\boldsymbol{w}(x)$, $\boldsymbol{\alpha}e^{\boldsymbol{S}u_{n-1}}\displaystyle\int_{x_n=0}^{\infty} e^{\boldsymbol{S}x_n}\boldsymbol{w}(x)\,\mathrm{d}x_n \leq \boldsymbol{\alpha}e^{\boldsymbol{S}u_{n-1}}\boldsymbol{e}G_{\boldsymbol{v}}$, hence (5.61) is less than or equal to

$$\begin{aligned}
&\mathcal{J}_{k+1,n-1}(u_k,x_{k+1})\int_{u_{n-1}=0}^{\infty} e^{\boldsymbol{S}u_{n-1}}\boldsymbol{s}\frac{\boldsymbol{\alpha}e^{\boldsymbol{S}u_{n-1}}\boldsymbol{e}}{\boldsymbol{\alpha}e^{\boldsymbol{S}u_{n-1}}\boldsymbol{e}}\mathrm{d}u_{n-1}GG_{\boldsymbol{v}}\\
&= \mathcal{J}_{k+1,n-1}(u_k,x_{k+1})\int_{u_{n-1}=0}^{\infty} e^{\boldsymbol{S}u_{n-1}}\boldsymbol{s}\,\mathrm{d}u_{n-1}GG_{\boldsymbol{v}}
\end{aligned} \tag{5.62}$$

where the terms $\boldsymbol{\alpha}e^{\boldsymbol{S}u_{n-1}}\boldsymbol{e}$ cancel from the numerator and denominator.[§]

[‡]As I mentioned in the previous footnote, this is important as, for $u_k > \Delta$, then $\boldsymbol{\alpha}^{(p)}e^{\boldsymbol{S}^{(p)}u_k}\boldsymbol{e}$ becomes small as $p \to \infty$. Deriving a bound in such a way that this cancellation occurs was one of the main challenges I encountered with this proof – in retrospect it is somewhat obvious once we accept that g_1 is bounded and g_k, $k > 1$, are integrable.

[§]Once again, this cancellation is important. In this case Property 5.2(*iii*) of $\boldsymbol{w}(x)$ and that g_n is bounded are key the deriving an expression where this term cancels.

Computing the integral with respect to u_{n-1} in (5.62), gives

$$\mathcal{J}_{k+1,n-1}(u_k, x_{k+1})\boldsymbol{e}GG_{\boldsymbol{v}} = \mathcal{J}_{k+1,n-2}(u_k, x_{k+1}) \int_{u_{n-2}=0}^{\infty} e^{\boldsymbol{S}u_{n-2}}\boldsymbol{s}\frac{\boldsymbol{\alpha}e^{\boldsymbol{S}u_{n-2}}}{\boldsymbol{\alpha}e^{\boldsymbol{S}u_{n-2}}\boldsymbol{e}}\,\mathrm{d}u_{n-2} \times \int_{x_{n-1}=0}^{\infty} g_{n-1}(x_{n-1})e^{\boldsymbol{S}x_{n-1}}\,\mathrm{d}x_{n-1}\boldsymbol{e}GG_{\boldsymbol{v}}. \tag{5.63}$$

Since $\boldsymbol{\alpha}e^{\boldsymbol{S}(x_{n-1}+u_{n-2})}\boldsymbol{e} \leq \boldsymbol{\alpha}e^{\boldsymbol{S}(u_{n-2})}\boldsymbol{e}$, then (5.63) is less than or equal to

$$\begin{aligned}&\mathcal{J}_{k+1,n-2}(u_k, x_{k+1}) \int_{u_{n-2}=0}^{\infty} e^{\boldsymbol{S}u_{n-2}}\boldsymbol{s}\frac{\boldsymbol{\alpha}e^{\boldsymbol{S}u_{n-2}}\boldsymbol{e}}{\boldsymbol{\alpha}e^{\boldsymbol{S}u_{n-2}}\boldsymbol{e}}\,\mathrm{d}u_{n-2} \int_{x_{n-1}=0}^{\infty} g_{n-1}(x_{n-1})\,\mathrm{d}x_{n-1}GG_{\boldsymbol{v}}\\ &= \mathcal{J}_{k+1,n-2}(u_k, x_{k+1}) \int_{u_{n-2}=0}^{\infty} e^{\boldsymbol{S}u_{n-2}}\boldsymbol{s}\,\mathrm{d}u_{n-2} \int_{x_{n-1}=0}^{\infty} g_{n-1}(x_{n-1})\,\mathrm{d}x_{n-1}GG_{\boldsymbol{v}}\end{aligned} \tag{5.64}$$

where $\boldsymbol{\alpha}e^{\boldsymbol{S}u_{n-2}}\boldsymbol{e}$ cancels in the numerator and denominator.¶ Since $\int_{x_{n-1}=0}^{\infty} g_{x_{n-1}}\,\mathrm{d}x_{n-1} \leq \widehat{G}$, then (5.64) is less than or equal to

$$\mathcal{J}_{k+1,n-2}(u_k, x_{k+1}) \int_{u_{n-2}=0}^{\infty} e^{\boldsymbol{S}u_{n-2}}\boldsymbol{s}\,\mathrm{d}u_{n-2}\widehat{G}GG_{\boldsymbol{v}} = \mathcal{J}_{k+1,n-2}(u_k, x_{k+1})\boldsymbol{e}\widehat{G}GG_{\boldsymbol{v}}. \tag{5.65}$$

This is of the same form as the left-hand side of (5.63), hence repeating the same arguments which took us from (5.63) to (5.65) another $n-k-3$ more times gives

$$\begin{aligned}\mathcal{J}_{k+1,n}(u_k, x_{k+1})\boldsymbol{w}(x) &\leq \mathcal{J}_{k+1,k+1}(u_k, x_{k+1})\boldsymbol{e}\widehat{G}^{n-k-2}GG_{\boldsymbol{v}}\\ &= g_{k+1}(x_{k+1})\frac{\boldsymbol{\alpha}e^{\boldsymbol{S}(u_k+x_{k+1})}}{\boldsymbol{\alpha}e^{\boldsymbol{S}u_k}\boldsymbol{e}}\boldsymbol{e}\widehat{G}^{n-k-2}GG_{\boldsymbol{v}}\\ &\leq g_{k+1}(x_{k+1})\widehat{G}^{n-k-2}GG_{\boldsymbol{v}}.\end{aligned}$$

Step 4: Combine the bounds on $\boldsymbol{k}(x_0)\mathcal{I}_{1,k}(u_k)$ and $\mathcal{J}_{k+1,n}(u_k, x_{k+1})\boldsymbol{w}(x)$ to bound the first term in (5.51).

First consider $k+1 < n$. With the bounds (5.52) and (5.60), the first term of (5.51) is less than or equal to

$$\begin{aligned}&\frac{1}{\boldsymbol{\alpha}e^{\boldsymbol{S}x_0}\boldsymbol{e}}G\widehat{G}^{k-1} \int_{x_{k+1}=0}^{\infty}\int_{u_k=\Delta-\varepsilon}^{\infty} \boldsymbol{\alpha}e^{\boldsymbol{S}u_k}\boldsymbol{e}g_{k+1}(x_{k+1})\,\mathrm{d}u_k\,\mathrm{d}x_{k+1}\widehat{G}^{n-k-2}GG_{\boldsymbol{v}}\\ &\leq \frac{1}{\boldsymbol{\alpha}e^{\boldsymbol{S}x_0}\boldsymbol{e}}G\widehat{G}^{k-1} \int_{u_k=\Delta-\varepsilon}^{\infty} \boldsymbol{\alpha}e^{\boldsymbol{S}u_k}\boldsymbol{e}\,\mathrm{d}u_k\widehat{G}\widehat{G}^{n-k-2}GG_{\boldsymbol{v}}.\end{aligned} \tag{5.66}$$

¶In this case the fact that the g_k are integrable helps us cancel these terms.

Now, observe that

$$\begin{aligned}\int_{u_k=\Delta-\varepsilon}^{\infty} \boldsymbol{\alpha} e^{\boldsymbol{S}u_k}\boldsymbol{e}\,\mathrm{d}u_k &= \int_{u_k=\Delta-\varepsilon}^{\Delta+\varepsilon} \mathbb{P}(Z>u_k)\,\mathrm{d}u_k + \int_{u_k=\Delta+\varepsilon}^{\infty} \mathbb{P}(Z>u_k)\,\mathrm{d}u_k \\ &\leq \int_{u_k=\Delta-\varepsilon}^{\Delta+\varepsilon} \mathrm{d}u_k + \int_{u_k=\Delta+\varepsilon}^{\infty} \frac{\mathrm{Var}(Z)}{(u_k-\Delta)^2}\,\mathrm{d}u_k \\ &= 2\varepsilon + \frac{\mathrm{Var}(Z)}{\varepsilon}, \end{aligned} \tag{5.67}$$

where in the second integrand we have used Chebyshev's inequality to bound the tail probability,

$$\mathbb{P}(Z>u_k) \leq \mathbb{P}(|Z-\Delta|>|u_k-\Delta|) \leq \frac{\mathrm{Var}(Z)}{(u_k-\Delta)^2},$$

for $u_k \geq \Delta+\varepsilon$. Hence, (5.66) is less than or equal to

$$\frac{1}{\boldsymbol{\alpha}e^{\boldsymbol{S}x_0}\boldsymbol{e}} G\widehat{G}^{k-1}\left(2\varepsilon + \frac{\mathrm{Var}(Z)}{\varepsilon}\right)\widehat{G}^{n-k-1}GG_{\boldsymbol{v}}.$$

Now consider $k+1=n$. By the bound (5.52), the first term of (5.51),

$$\int_{x_{k+1}=0}^{\infty}\int_{u_k=\Delta-\varepsilon}^{\infty} \boldsymbol{k}(x_0)\mathcal{I}_{1,n-1}(u_{n-1})g_n(x_n)\frac{\boldsymbol{\alpha}e^{\boldsymbol{S}u_{n-1}}}{\boldsymbol{\alpha}e^{\boldsymbol{S}u_{n-1}}\boldsymbol{e}}e^{\boldsymbol{S}x_n}\boldsymbol{w}(x) \tag{5.68}$$

$$\leq \frac{1}{\boldsymbol{\alpha}e^{\boldsymbol{S}x_0}\boldsymbol{e}} G\widehat{G}^{k-1}\int_{x_{k+1}=0}^{\infty}\int_{u_k=\Delta-\varepsilon}^{\infty} \boldsymbol{\alpha}e^{\boldsymbol{S}u_k}\boldsymbol{e}g_{k+1}(x_{k+1})\frac{\boldsymbol{\alpha}e^{\boldsymbol{S}(u_k+x_{k+1})}}{\boldsymbol{\alpha}e^{\boldsymbol{S}u_k}\boldsymbol{e}}\boldsymbol{w}(x)\,\mathrm{d}u_k\,\mathrm{d}x_{k+1}. \tag{5.69}$$

Since $g_{k+1}\leq G$, and upon integrating over x_{k+1}, then (5.69) is less than or equal to

$$\frac{1}{\boldsymbol{\alpha}e^{\boldsymbol{S}x_0}\boldsymbol{e}} G^2\widehat{G}^{k-1}\int_{u_k=\Delta-\varepsilon}^{\infty} \boldsymbol{\alpha}e^{\boldsymbol{S}u_k}(-\boldsymbol{S})^{-1}\boldsymbol{w}(x)\,\mathrm{d}u_k \leq \frac{1}{\boldsymbol{\alpha}e^{\boldsymbol{S}x_0}\boldsymbol{e}} G^2\widehat{G}^{k-1}\int_{u_k=\Delta-\varepsilon}^{\infty} \boldsymbol{\alpha}e^{\boldsymbol{S}u_k}\boldsymbol{e}G_{\boldsymbol{v}}\,\mathrm{d}u_k, \tag{5.70}$$

where we have used Property 5.2(*iii*) to get the upper bound on the right-hand side of (5.70). Using (5.67) again, then (5.70) is less than or equal to

$$\frac{1}{\boldsymbol{\alpha}e^{\boldsymbol{S}x_0}\boldsymbol{e}} G\widehat{G}^{n-2}GG_{\boldsymbol{v}}\left(2\varepsilon + \frac{\mathrm{Var}\,(Z)}{\varepsilon}\right). \tag{5.71}$$

Thus, we have shown the desired bound.

Step 5: Bound the second term in (5.51).

To bound the second term in (5.51) we instead bound

$$\begin{aligned}&\int_{x_{k+1}=0}^{\infty}\int_{u_k=0}^{\infty}\boldsymbol{k}(x_0)\mathcal{I}_{1,k}(u_k)\mathcal{J}_{k+1,n}(u_k,x_{k+1})\widetilde{\boldsymbol{w}}(x)\\ &=\int_{x_{k+1}=0}^{\infty}\int_{u=0}^{\infty}\mathcal{I}_{1,n}(u)\frac{\boldsymbol{\alpha}e^{\boldsymbol{S}u}}{\boldsymbol{\alpha}e^{\boldsymbol{S}u}\boldsymbol{e}}\int_{x_n=0}^{\infty}g_n(x_n)e^{\boldsymbol{S}x_n}\,\mathrm{d}x_n\widetilde{\boldsymbol{w}}(x)\end{aligned}\tag{5.72}$$

which is the same as the second term in (5.51) except that in (5.72) the integral with respect to u_k is over a larger interval. Using the bound in (5.52), then (5.72) is less than or equal to

$$\frac{1}{\boldsymbol{\alpha}e^{\boldsymbol{S}x_0}\boldsymbol{e}}G\widehat{G}^{n-1}\int_{u=0}^{\infty}\boldsymbol{\alpha}e^{\boldsymbol{S}u}\boldsymbol{e}\frac{\boldsymbol{\alpha}e^{\boldsymbol{S}u}}{\boldsymbol{\alpha}e^{\boldsymbol{S}u}\boldsymbol{e}}\mathrm{d}u\int_{x_n=0}^{\infty}g_n(x_n)e^{\boldsymbol{S}x_n}\,\mathrm{d}x_n\widetilde{\boldsymbol{w}}(x).$$

Integrating over u gives

$$\begin{aligned}&\frac{1}{\boldsymbol{\alpha}e^{\boldsymbol{S}x_0}\boldsymbol{e}}G\widehat{G}^{n-1}\boldsymbol{\alpha}(-\boldsymbol{S})^{-1}\int_{x_n=0}^{\infty}g_n(x_n)e^{\boldsymbol{S}x_n}\,\mathrm{d}x_n\widetilde{\boldsymbol{w}}(x)\\ &\leq\frac{1}{\boldsymbol{\alpha}e^{\boldsymbol{S}x_0}\boldsymbol{e}}G\widehat{G}^{n-1}\boldsymbol{\alpha}(-\boldsymbol{S})^{-1}\int_{x_n=0}^{\infty}g_n(x_n)\,\mathrm{d}x_n\widetilde{\boldsymbol{w}}(x),\end{aligned}$$

where the inequality holds by Property 5.2(i). Integrating over x_n, gives

$$\frac{1}{\boldsymbol{\alpha}e^{\boldsymbol{S}x_0}\boldsymbol{e}}G\widehat{G}^{n}\boldsymbol{\alpha}(-\boldsymbol{S})^{-1}\widetilde{\boldsymbol{w}}(x)=\frac{1}{\boldsymbol{\alpha}e^{\boldsymbol{S}x_0}\boldsymbol{e}}G\widehat{G}^{n}\widetilde{G}_{\boldsymbol{v}},\tag{5.73}$$

by Property 5.2(ii).

Combining all the bounds proves the result. □

Next we wish to prove a bound on the difference between the first term in (5.49) and $g^*_{1,n}(x_0,x)$, where we define

$$\begin{aligned}g^*_{2,n}(u_1,x):=&\int_{u_2=0}^{\Delta-u_1}g_2(\Delta-u_2-u_1)\,\mathrm{d}u_1\ldots\int_{u_{n-1}=0}^{\Delta-u_{n-2}}g_{n-1}(\Delta-u_{n-1}-u_{n-2})\,\mathrm{d}u_{n-2}\\ &g_n(\Delta-x-u_{n-1})1(\Delta-x-u_{n-1}\geq 0)\,\mathrm{d}u_{n-1},\end{aligned}\tag{5.74}$$

and

$$g^*_{1,n}(x_0,x):=\int_{u_1=0}^{\Delta-x_0}g_1(\Delta-u_1-x_0)g^*_{2,n}(u_1,x)\,\mathrm{d}u_1.\tag{5.75}$$

The expression $g^*_{1,n}$ is of a similar form as $\widehat{\mu}^{\ell_0}_{n,q,r}(\lambda)$ except that the functions in $\widehat{\mu}^{\ell_0}_{n,q,r}(\lambda)$ are matrices. Since the functions in $\widehat{\mu}^{\ell_0}_{n,q,r}(\lambda)$ are matrices, then $\widehat{\mu}^{\ell_0}_{n,q,r}(\lambda)$ can be written as a linear combination of terms with the form $g^*_{1,n}$.

The idea of the proof is to first show a bound for the difference between the first term in (5.49) and the expression $g_{1,n}^{*,\varepsilon}(x_0, x)$ given by

$$\int_{u_1=0}^{\Delta-\varepsilon-x_0} g_1(\Delta - u_1 - x_0) \int_{u_2=0}^{\Delta-\varepsilon-u_1} g_2(\Delta - u_2 - u_1)\,\mathrm{d}u_1 \\ \ldots \int_{u_{n-1}=0}^{\Delta-\varepsilon-u_{n-2}} g_{n-1}(\Delta - u_{n-1} - u_{n-2})\,\mathrm{d}u_{n-2} g_n(\Delta - x - u_{n-1}) 1(\Delta - x - u_{n-1} \geq \varepsilon). \tag{5.76}$$

We then establish a bound on the difference between $g_{1,n}^{*,\varepsilon}(x_0, x)$ and $g_{1,n}^{*}(x_0, x)$ which can be made arbitrarily small by choosing ε sufficiently small.

Recall that the first term in (5.49) looks like

$$\int_{x_1=0}^{\infty} g_1(x_1)\boldsymbol{k}(x_0)e^{\boldsymbol{S}x_1}\,\mathrm{d}x_1 \boldsymbol{D}(\Delta-\varepsilon) \left[\prod_{k=2}^{n-1} \int_{x_k=0}^{\infty} g_k(x_k)e^{\boldsymbol{S}x_k}\,\mathrm{d}x_k \boldsymbol{D}(\Delta-\varepsilon)\right] \\ \times \int_{x_n=0}^{\infty} g_n(x_n)e^{\boldsymbol{S}x_n}\,\mathrm{d}x_n \boldsymbol{v}(x) \tag{5.77}$$

which, upon substituting $\boldsymbol{D}(\Delta-\varepsilon) = \displaystyle\int_{u=0}^{\Delta-\varepsilon} e^{\boldsymbol{S}u}\boldsymbol{s}\frac{\boldsymbol{\alpha}e^{\boldsymbol{S}u}}{\boldsymbol{\alpha}e^{\boldsymbol{S}u}\boldsymbol{e}}\mathrm{d}u$, can be written as

$$\int_{u_1=0}^{\Delta-\varepsilon}\int_{x_1=0}^{\infty} \frac{\boldsymbol{\alpha}e^{\boldsymbol{S}(x_0+x_1+u_1)}\boldsymbol{s}}{\boldsymbol{\alpha}e^{\boldsymbol{S}x_0}\boldsymbol{e}} g_1(x_1)\,\mathrm{d}x_1 \left[\prod_{k=2}^{n-1}\int_{u_k=0}^{\Delta-\varepsilon}\int_{x_k=0}^{\infty} \frac{\boldsymbol{\alpha}e^{\boldsymbol{S}(u_{k-1}+x_k+u_k)}\boldsymbol{s}}{\boldsymbol{\alpha}e^{\boldsymbol{S}u_{k-1}}\boldsymbol{e}} g_k(x_k)\,\mathrm{d}x_k\,\mathrm{d}u_{k-1}\right] \\ \times \int_{x_n=0}^{\infty} \frac{\boldsymbol{\alpha}e^{\boldsymbol{S}(u_{n-1}+x_n)}}{\boldsymbol{\alpha}e^{\boldsymbol{S}u_{n-1}}\boldsymbol{e}} \boldsymbol{v}(x) g_n(x_n)\,\mathrm{d}x_n\,\mathrm{d}u_{n-1}. \tag{5.78}$$

The last integral in (5.78) (with respect to x_n) is close to $g_n(\Delta - x - u_{n-1})$ by Property 5.2(v).

Also, appearing in (5.78) are integrals of the form

$$\int_{x_\ell=0}^{\infty} \frac{\boldsymbol{\alpha}e^{\boldsymbol{S}(u_{\ell-1}+x_\ell+u_\ell)}\boldsymbol{s}}{\boldsymbol{\alpha}e^{\boldsymbol{S}u_{\ell-1}}\boldsymbol{e}} g_\ell(x_\ell)\,\mathrm{d}x_\ell. \tag{5.79}$$

Intuitively, if the variance of Z is sufficiently small and Δ is the expected value of Z, then the distribution of Z will be concentrated around Δ and the integral in (5.79) should be approximately equal to $g_\ell(\Delta - u_\ell - u_{\ell-1})$, provided $u_{\ell-1} \leq \Delta - \varepsilon$. Our first step towards showing a bound for the difference between the first term in (5.49) and the expression $g_{1,n}^{*,\varepsilon}(x_0, x)$ is to prove this intuition. We start with a result about with a simpler integral than that in (5.79), from which the result we require follows as a corollary.

Lemma 5.6. *Let g be a function satisfying Assumptions 5.1, then, for $u \leq \Delta - \varepsilon$,*

$$\int_{x=0}^{\infty} g(x)\,\boldsymbol{\alpha}e^{\boldsymbol{S}(x+u)}\boldsymbol{s}\,\mathrm{d}x = g(\Delta - u) + r_2, \tag{5.80}$$

where

$$|r_2| \leq 2G\frac{\mathrm{Var}(Z)}{\varepsilon^2} + 2L\varepsilon.$$

The proof follows closely that of (Horváth, Horváth & Telek 2020, Appendix A, Theorem 4). The idea of the proof is to recognise that (assuming the variance of Z is small) the largest contribution to the integral on the left-hand side of (5.80) will come from integrating over the interval $x \in (\Delta - u - \varepsilon, \Delta - u + \varepsilon)$. Since g is non-negative and bounded, then the rest of the integral is bounded by

$$\int_{\substack{x\in[0,\infty)\\ x\notin(\Delta-u-\varepsilon,\Delta-u+\varepsilon)}} G\boldsymbol{\alpha}e^{\boldsymbol{S}(x+u)}\boldsymbol{s}\,\mathrm{d}x,$$

which can be shown to be small by Chebyshev's inequality provided the variance of Z is small.

Proof. With a change of variables,

$$\begin{aligned}
&\left|\int_{x=0}^{\infty} g(x)\,\boldsymbol{\alpha}e^{\boldsymbol{S}(x+u)}\boldsymbol{s}\,\mathrm{d}x - g(\Delta-u)\right| \\
&= \left|\int_{x=u}^{\infty} g(x-u)\,\boldsymbol{\alpha}e^{\boldsymbol{S}x}\boldsymbol{s}\,\mathrm{d}x - g(\Delta-u)\right| \\
&= \left|\int_{x=u}^{\infty} g(x-u)\,\boldsymbol{\alpha}e^{\boldsymbol{S}x}\boldsymbol{s}\,\mathrm{d}x - \int_{x=u}^{\infty} g(\Delta-u)\,\boldsymbol{\alpha}e^{\boldsymbol{S}x}\boldsymbol{s}\,\mathrm{d}x - g(\Delta-u)\left(1-\boldsymbol{\alpha}e^{\boldsymbol{S}u}\boldsymbol{e}\right)\right|.
\end{aligned}$$

By the triangle inequality this is less than or equal to

$$\begin{aligned}
&\left|\int_{x=u}^{\infty} (g(x-u)-g(\Delta-u))\,\boldsymbol{\alpha}e^{\boldsymbol{S}x}\boldsymbol{s}\,\mathrm{d}x\right| + \left|g(\Delta-u)\left(1-\boldsymbol{\alpha}e^{\boldsymbol{S}u}\boldsymbol{e}\right)\right| \\
&= \left|\int_{x=u}^{\infty} (g(x-u)-g(\Delta-u))\,\boldsymbol{\alpha}e^{\boldsymbol{S}x}\boldsymbol{s}\,\mathrm{d}x\right| + \left|\int_{x=0}^{u} g(\Delta-u)\,\boldsymbol{\alpha}e^{\boldsymbol{S}x}\boldsymbol{s}\,\mathrm{d}x\right| \\
&\leq d_1 + d_2
\end{aligned}$$

where

$$d_1 = \left|\int_{x=0}^{u} g(\Delta-u)\,\boldsymbol{\alpha}e^{\boldsymbol{S}x}\boldsymbol{s}\,\mathrm{d}x\right| + \left|\int_{x=u}^{\Delta-\varepsilon} (g(x-u)-g(\Delta-u))\,\boldsymbol{\alpha}e^{\boldsymbol{S}x}\boldsymbol{s}\,\mathrm{d}x\right|$$

$$+\left|\int_{x=\Delta+\varepsilon}^{\infty}\left(g\left(x-u\right)-g\left(\Delta-u\right)\right)\boldsymbol{\alpha}e^{\boldsymbol{S}x}\boldsymbol{s}\,\mathrm{d}x\right|,$$

$$d_2=\left|\int_{x=\Delta-\varepsilon}^{\Delta+\varepsilon}\left(g\left(x-u\right)-g\left(\Delta-u\right)\right)\boldsymbol{\alpha}e^{\boldsymbol{S}x}\boldsymbol{s}\,\mathrm{d}x\right|.$$

By the triangle inequality for integrals, d_2 is less than or equal to

$$\begin{aligned}\int_{x=\Delta-\varepsilon}^{\Delta+\varepsilon}\left|g\left(x-u\right)-g\left(\Delta-u\right)\right|\boldsymbol{\alpha}e^{\boldsymbol{S}x}\boldsymbol{s}\,\mathrm{d}x &\leq \int_{x=\Delta-\varepsilon}^{\Delta+\varepsilon}2L\varepsilon\boldsymbol{\alpha}e^{\boldsymbol{S}x}\boldsymbol{s}\,\mathrm{d}x\\ &=2L\varepsilon\mathbb{P}(Z\in(\Delta-\varepsilon,\Delta+\varepsilon))\\ &\leq 2L\varepsilon,\end{aligned}$$

where we have used the Lipschitz property of g from Assumption 5.1(iv) in the first line.

Applying the triangle inequality to d_1,

$$\begin{aligned}d_1 &\leq \int_{x=u}^{\Delta-\varepsilon}\left|g\left(x-u\right)-g\left(\Delta-u\right)\right|\boldsymbol{\alpha}e^{\boldsymbol{S}x}\boldsymbol{s}\,\mathrm{d}x+\int_{x=\Delta+\varepsilon}^{\infty}\left|g\left(x-u\right)-g\left(\Delta-u\right)\right|\boldsymbol{\alpha}e^{\boldsymbol{S}x}\boldsymbol{s}\,\mathrm{d}x\\ &\quad+\left|\int_{x=0}^{u}g\left(\Delta-u\right)\boldsymbol{\alpha}e^{\boldsymbol{S}x}\boldsymbol{s}\,\mathrm{d}x\right|\\ &\leq 2G\left(\int_{x=u}^{\Delta-\varepsilon}\boldsymbol{\alpha}e^{\boldsymbol{S}x}\boldsymbol{s}\,\mathrm{d}x+\int_{x=\Delta+\varepsilon}^{\infty}\boldsymbol{\alpha}e^{\boldsymbol{S}x}\boldsymbol{s}\,\mathrm{d}x+\int_{x=0}^{u}\boldsymbol{\alpha}e^{\boldsymbol{S}x}\boldsymbol{s}\,\mathrm{d}x\right)=2G\mathbb{P}\left(\left|Z-\Delta\right|>\varepsilon\right),\end{aligned}$$

where the second inequality holds since $|g\left(x\right)|\leq G$. By Chebyshev's inequality,

$$2G\mathbb{P}\left(\left|Z-\Delta\right|>\varepsilon\right)\leq 2G\frac{\mathrm{Var}\left(Z\right)}{\varepsilon^2}. \tag{5.81}$$

Hence, there is some r_2 such that

$$\left|\int_{x=0}^{\infty}g\left(x\right)\boldsymbol{\alpha}e^{\boldsymbol{S}(x+u)}\boldsymbol{s}\,\mathrm{d}x-g\left(\Delta-u\right)\right|=|r_2|\leq 2G\frac{\mathrm{Var}(Z)}{\varepsilon^2}+2L\varepsilon,$$

and this completes the proof. □

Corollary 5.7. *Let g be a function satisfying the Assumptions 5.1. For $u\leq\Delta-\varepsilon$, $v\geq 0$,*

$$\int_{x=0}^{\infty}\frac{\boldsymbol{\alpha}e^{\boldsymbol{S}(x+u+v)}\boldsymbol{s}}{\boldsymbol{\alpha}e^{\boldsymbol{S}u}\boldsymbol{e}}g(x)\,\mathrm{d}x=g(\Delta-u-v)1(u+v\leq\Delta-\varepsilon)+r_3(u+v), \tag{5.82}$$

where

$$|r_3(u+v)|\leq\begin{cases}3G\delta+2L\varepsilon & u+v\leq\Delta-\varepsilon,\\ G & u+v\in(\Delta-\varepsilon,\Delta+\varepsilon),\\ G\dfrac{\delta}{1-\delta} & u+v\geq\Delta+\varepsilon.\end{cases},$$

and $\delta=\dfrac{\mathrm{Var}(Z)}{\varepsilon^2}$.

Proof. First consider $u+v \leq \Delta-\varepsilon$. Observe that, since $u \leq \Delta-\varepsilon$, Chebyshev's inequality gives

$$\begin{aligned}\boldsymbol{\alpha}e^{\boldsymbol{S}u}\boldsymbol{e} &= \mathbb{P}\left(Z > u\right)\\ &\geq \mathbb{P}\left(|Z-\Delta| \leq \varepsilon\right)\\ &\geq 1 - \frac{\mathrm{Var}\left(Z\right)}{\varepsilon^2}\\ &=: 1-\delta,\end{aligned}$$

thus we have a bound for the denominator in the integrand on the left-hand side of (5.82).

Now, since $1-\delta \leq \boldsymbol{\alpha}e^{\boldsymbol{S}u}\boldsymbol{e} \leq 1$, then

$$\int_{x=0}^{\infty} \boldsymbol{\alpha}e^{\boldsymbol{S}(x+u+v)}\boldsymbol{s}g(x)\,\mathrm{d}x \leq \int_{x=0}^{\infty} \frac{\boldsymbol{\alpha}e^{\boldsymbol{S}(x+u+v)}\boldsymbol{s}}{\boldsymbol{\alpha}e^{\boldsymbol{S}u}\boldsymbol{e}}g(x)\,\mathrm{d}x \leq \frac{1}{1-\delta}\int_{x=0}^{\infty} \boldsymbol{\alpha}e^{\boldsymbol{S}(x+u+v)}\boldsymbol{s}g(x)\,\mathrm{d}x.$$

By Lemma 5.6

$$g(\Delta-u-v)+r_2 \leq \int_{x=0}^{\infty} \frac{\boldsymbol{\alpha}e^{\boldsymbol{S}(x+u+v)}\boldsymbol{s}}{\boldsymbol{\alpha}e^{\boldsymbol{S}u}\boldsymbol{e}}g(x)\,\mathrm{d}x \leq \frac{g(\Delta-u-v)+r_2}{1-\delta}.$$

Multiplying by $1-\delta$, then subtracting $g(\Delta-u-v)$ and adding $\displaystyle\int_{x=0}^{\infty} \frac{\boldsymbol{\alpha}e^{\boldsymbol{S}(x+u+v)}\boldsymbol{s}}{\boldsymbol{\alpha}e^{\boldsymbol{S}u}\boldsymbol{e}}g(x)\,\mathrm{d}x\delta$ gives

$$\begin{aligned}&r_2(1-\delta) - g(\Delta-u-v)\delta + \int_{x=0}^{\infty} \frac{\boldsymbol{\alpha}e^{\boldsymbol{S}(x+u+v)}\boldsymbol{s}}{\boldsymbol{\alpha}e^{\boldsymbol{S}u}\boldsymbol{e}}g(x)\,\mathrm{d}x\delta\\ &\leq \int_{x=0}^{\infty} \frac{\boldsymbol{\alpha}e^{\boldsymbol{S}(x+u+v)}\boldsymbol{s}}{\boldsymbol{\alpha}e^{\boldsymbol{S}u}\boldsymbol{e}}g(x)\,\mathrm{d}x - g(\Delta-u-v)\\ &\leq r_2 + \int_{x=0}^{\infty} \frac{\boldsymbol{\alpha}e^{\boldsymbol{S}(x+u+v)}\boldsymbol{s}}{\boldsymbol{\alpha}e^{\boldsymbol{S}u}\boldsymbol{e}}g(x)\,\mathrm{d}x\delta.\end{aligned}$$

The last line is bounded above by

$$r_2 + \int_{x=0}^{\infty} \frac{\boldsymbol{\alpha}e^{\boldsymbol{S}(x+u+v)}\boldsymbol{s}}{\boldsymbol{\alpha}e^{\boldsymbol{S}u}\boldsymbol{e}}g(x)\,\mathrm{d}x\delta \leq r_2 + G\delta.$$

The first line is bounded below by

$$r_2(1-\delta) - g(\Delta-u-v)\delta + \int_{x=0}^{\infty} \frac{\boldsymbol{\alpha}e^{\boldsymbol{S}(x+u+v)}\boldsymbol{s}}{\boldsymbol{\alpha}e^{\boldsymbol{S}u}\boldsymbol{e}}g(x)\,\mathrm{d}x\delta \geq r_2(1-\delta) - g(\Delta-u-v)\delta.$$

Therefore,

$$\int_{x=0}^{\infty} \frac{\boldsymbol{\alpha}e^{\boldsymbol{S}(x+u+v)}\boldsymbol{s}}{\boldsymbol{\alpha}e^{\boldsymbol{S}u}\boldsymbol{e}}g(x)\,\mathrm{d}x = g(\Delta-u-v)+r_3, \tag{5.83}$$

where

$$\begin{aligned}|r_3| &\leq \max\left(|r_2|(1-\delta)+g(\Delta-u-v)\delta, |r_2|+G\delta\right)\\ &\leq |r_2|+G\delta\\ &\leq 3G\delta+2L\varepsilon,\end{aligned} \tag{5.84}$$

as required.

For $u+v\in(\Delta-\varepsilon,\Delta+\varepsilon)$,

$$\int_{x=0}^{\infty}\frac{\boldsymbol{\alpha}e^{\boldsymbol{S}(x+u+v)}\boldsymbol{s}}{\boldsymbol{\alpha}e^{\boldsymbol{S}u}\boldsymbol{e}}g(x)\,\mathrm{d}x\leq G\mathbb{P}(Z>u+v\mid Z>u)\leq G. \tag{5.85}$$

For $u+v\geq\Delta+\varepsilon$,

$$\int_{x=0}^{\infty}\frac{\boldsymbol{\alpha}e^{\boldsymbol{S}(x+u+v)}\boldsymbol{s}}{\boldsymbol{\alpha}e^{\boldsymbol{S}u}\boldsymbol{e}}g(x)\,\mathrm{d}x\leq G\frac{\mathbb{P}(Z>u+v)}{\mathbb{P}(Z>u)}\leq G\frac{\mathrm{Var}(Z)/\varepsilon^2}{1-\mathrm{Var}(Z)/\varepsilon^2}. \tag{5.86}$$

□

The error term $r_3^{(p)}$ depends on p, as it is defined by $Z^{(p)}$ and $\varepsilon^{(p)}$, but we have omitted the superscript p here. Choosing $\varepsilon=\mathrm{Var}(Z^{(p)})^{1/3}$ then, outside the vanishingly small interval $u\in(\Delta-\varepsilon^{(p)},\Delta+\varepsilon^{(p)})$, the error term $|r_3^{(p)}(u)|$ is bounded by $O\left(\mathrm{Var}\left(Z^{(p)}\right)^{1/3}\right)$, which tends to 0 as $p\to\infty$. On $u\in(\Delta-\varepsilon^{(p)},\Delta+\varepsilon^{(p)})$ the error term $|r_3^{(p)}(u)|$ is bounded by a constant which does not tend to 0 as $p\to\infty$. However, when we integrate a bounded function against $r_3^{(p)}(u)$, then the resulting integral tends to 0, i.e. for $|\psi(x)|\leq F$, $M<\infty$, $\displaystyle\int_0^M\psi(u)|r_3^{(p)}(u)|\,\mathrm{d}u\leq F\Delta(3G\delta^{(p)}+2L\varepsilon)+2GF\varepsilon^{(p)}+(M-\Delta)GF\delta^{(p)}/(1-\delta^{(p)})=O\left(\mathrm{Var}\left(Z^{(p)}\right)^{1/3}\right)\to 0$ as $p\to\infty$. This is the context in which we apply Corollary 5.7 and thus the error bound is sufficient.

We are now in a position to prove the desired bound on the difference between the first term in (5.49) and $g_{1,n}^*(x_0,x)$.

Lemma 5.8. *Let $g_1,g_2,\ldots,$ be functions satisfying the Assumptions 5.1 and let $\boldsymbol{v}(x)$ be a closing operator with the Properties 5.2. Then, for $n\geq 2$,*

$$\begin{aligned}&\int_{x_1=0}^{\infty}g_1(x_1)\boldsymbol{k}(x_0)e^{\boldsymbol{S}x_1}\,\mathrm{d}x_1\boldsymbol{D}(\Delta-\varepsilon)\left[\prod_{k=2}^{n-1}\int_{x_k=0}^{\infty}g_k(x_k)e^{\boldsymbol{S}x_k}\,\mathrm{d}x_k\right]\boldsymbol{D}(\Delta-\varepsilon)\\ &\qquad\times\int_{x_n=0}^{\infty}g_n(x_n)e^{\boldsymbol{S}x_n}\,\mathrm{d}x_n\boldsymbol{v}(x)\\ &=g_{1,n}^*(x_0,x)+r_4(n)+r_5(n),\end{aligned} \tag{5.87}$$

where

$$|r_4(n)| = O\left(\max\left\{\delta, \varepsilon, \frac{\delta}{1-\delta}, R_{\boldsymbol{v},1}\right\}\right),$$
$$|r_5(n)| \le \varepsilon^{n-1} G^{n-1}$$

Proof. Rewriting the left-hand side of (5.87) as in (5.78), then we see that we can apply Corollary 5.7 to all the integrals over x_k, $k = 1, \ldots, n-1$ and use Property 5.2(v) of $\boldsymbol{v}(x)$ for the integral over x_n, to get

$$\begin{aligned}
&\int_{u_1=0}^{\Delta-\varepsilon} [g_1(\Delta - u_1 - x_0)1(u_1 + x_0 \le \Delta - \varepsilon) + r_3(u_1 + x_0)] \\
&\quad \times \int_{u_2=0}^{\Delta-\varepsilon} [g_2(\Delta - u_2 - u_1)1(u_2 + u_1 \le \Delta - \varepsilon) + r_3(u_2 + u_1)]\, \mathrm{d}u_1 \\
&\quad \ldots \int_{u_{n-1}=0}^{\Delta-\varepsilon} [g_{n-1}(\Delta - u_{n-1} - u_{n-2})1(u_{n-1} + u_{n-2} \le \Delta - \varepsilon) + r_3(u_{n-1} + u_{n-2})]\, \mathrm{d}u_{n-2} \\
&\quad \times [g_n(\Delta - u_{n-1} - x)1(u_{n-1} + x \le \Delta - \varepsilon) + r_{\boldsymbol{v}}(u_{n-1}, x)]\, \mathrm{d}u_{n-1} \\
&= g_{1,n}^{*,\varepsilon}(x_0, x) + r_4(n)
\end{aligned}$$

where $r_4(n)$ is an error term. The leading terms of $r_4(n)$ are of the form

$$\begin{aligned}
&\int_{u_1=0}^{\Delta-\varepsilon-x_0} g_1(\Delta - u_1 - x_0) \int_{u_2=0}^{\Delta-\varepsilon-u_1} g_2(\Delta - u_2 - u_1)\, \mathrm{d}u_1 \\
&\quad \ldots \int_{u_{k-1}=0}^{\Delta-\varepsilon-u_{k-2}} g_{k-1}(\Delta - u_{k-1} - u_{k-2})\, \mathrm{d}u_{k-2} \int_{u_k=0}^{\Delta-\varepsilon} r_3(u_k + u_{k-1})\, \mathrm{d}u_{k-1} \\
&\quad \times \int_{u_{k+1}=0}^{\Delta-\varepsilon-u_k} g_{k+1}(\Delta - u_{k+1} - u_k)\, \mathrm{d}u_k \ldots \int_{u_{n-1}=0}^{\Delta-\varepsilon-u_{n-2}} g_{n-1}(\Delta - u_{n-1} - u_{n-2})\, \mathrm{d}u_{n-2} \\
&\quad \times g_n(\Delta - u_{n-1} - x)1(u_{n-1} + x \le \Delta - \varepsilon)\, \mathrm{d}u_{n-1}
\end{aligned} \tag{5.88}$$

or

$$\begin{aligned}
&\int_{u_1=0}^{\Delta-\varepsilon-x_0} g_1(\Delta - u_1 - x_0) \int_{u_2=0}^{\Delta-\varepsilon-u_1} g_2(\Delta - u_2 - u_1)\, \mathrm{d}u_1 \\
&\ldots \int_{u_{n-1}=0}^{\Delta-\varepsilon-u_{n-2}} g_{n-1}(\Delta - u_{n-1} - u_{n-2})\, \mathrm{d}u_{n-2} r_{\boldsymbol{v}}(u_{n-1}, x)\, \mathrm{d}u_{n-1},
\end{aligned} \tag{5.89}$$

whichever is larger. Since $|g_\ell| \le G$, $\ell \ge 1$, then (5.89) and (5.88) are less than or equal to

$$G^{k-1}\Delta^{k-2} \int_{u_{k-1}=0}^{\Delta-\varepsilon} \int_{u_k=0}^{\Delta-\varepsilon} r_3(u_k + u_{k-1})\, \mathrm{d}u_k\, \mathrm{d}u_{k-1} G^{n-k}\Delta^{n-k-1},$$

and

$$G^{n-1}\Delta^{n-2}\int_{u_{n-1}=0}^{\Delta-\varepsilon} r_{\boldsymbol{v}}(u_{n-1},x)\,\mathrm{d}u_{n-1},$$

respectively.

Recall that we have a bound on $|r_3|$ which is piecewise constant. Breaking up the integral of $|r_3|$ above into three intervals over which the bound is constant and using the triangle inequality, then

$$\begin{aligned}
&\left|\int_{u_{k-1}=0}^{\Delta-\varepsilon}\int_{u_k=0}^{\Delta-\varepsilon} r_3(u_k+u_{k-1})\,\mathrm{d}u_k\,\mathrm{d}u_{k-1}\right| \\
&\leq \int_{u_{k-1}=0}^{\Delta-\varepsilon}\left[\int_{u_k=u_{k-1}}^{\Delta-\varepsilon}|r_3(u_k)|\,\mathrm{d}u_k+\int_{u_k=\Delta-\varepsilon}^{\min(\Delta+\varepsilon,\Delta-\varepsilon+u_{k-1})}|r_3(u_k)|\,\mathrm{d}u_k\right. \\
&\qquad\left.+\int_{u_k=\Delta+\varepsilon}^{\Delta-\varepsilon+u_{k-1}}|r_3(u_k)|\,\mathrm{d}u_k 1(u_{k-1}>2\varepsilon)\right]\mathrm{d}u_{k-1}. \qquad (5.90)
\end{aligned}$$

Using the piecewise upper bounds on $|r_3|$ then, in the square brackets in (5.90), the first integral is less than or equal to

$$(\Delta-\varepsilon-u_{k-1})(3G\Delta+2L\varepsilon),$$

the second integral is less than or equal to

$$\int_{u_k=\Delta-\varepsilon}^{\Delta+\varepsilon}|r_3(u_k)|\,\mathrm{d}u_k\leq 2\varepsilon G$$

and the third integral is less than or equal to

$$G\frac{\delta}{1-\delta}(u_{k-1}-2\varepsilon)1(u_{k-1}>2\varepsilon).$$

With these bounds (5.90) is less than or equal to

$$\begin{aligned}
&\left[\int_{u_{k-1}=0}^{\Delta-\varepsilon}(\Delta-\varepsilon-u_{k-1})(3G\delta+2L\varepsilon)+2\varepsilon G+G\frac{\delta}{1-\delta}(u_{k-1}-2\varepsilon)1(u_{k-1}>2\varepsilon)\right]\mathrm{d}u_{k-1} \\
&\leq \frac{1}{2}\Delta^2(3G\delta+2L\varepsilon)+2\Delta\varepsilon G+\frac{1}{2}\Delta^2 G\frac{\delta}{1-\delta} \\
&= O\left(\delta,\varepsilon,\frac{\delta}{1-\delta}\right).
\end{aligned}$$

Thus, (5.88) is, at worst, $O\left(\delta, \varepsilon, \dfrac{\delta}{1-\delta}\right)$.

As for (5.89), by Property 5.2(v),

$$\int_{u_{n-1}=0}^{\Delta-\varepsilon} |r_{\boldsymbol{v}}(u_{n-1}, x)| \, \mathrm{d}u_{n-1} \leq R_{\boldsymbol{v},1},$$

hence (5.89) is $O(R_{\boldsymbol{v},1})$.

Therefore, the error term $|r_4(n)| = O\left(\max\left\{\delta, \varepsilon, \dfrac{\delta}{1-\delta}, R_{\boldsymbol{v},1}\right\}\right)$.

Now,

$$\begin{aligned}
&\left| g_{1,n}^{*,\varepsilon}(x_0, x) - g_{1,n}^{*}(x_0, x) \right| \\
&= \int_{u_1=\Delta-\varepsilon-x_0}^{\Delta-x_0} g_1(\Delta - u_1 - x_0) \int_{u_2=\Delta-\varepsilon-u_1}^{\Delta-u_1} g_2(\Delta - u_2 - u_1) \, \mathrm{d}u_1 \\
&\quad \ldots \int_{u_{n-1}=\Delta-\varepsilon-u_{n-2}}^{\Delta-u_{n-2}} g_{n-1}(\Delta - u_{n-1} - u_{n-2}) \, \mathrm{d}u_{n-2} g_n(\Delta - x - u_{n-1}) \\
&\quad \times 1(\Delta - x - u_{n-1} \geq 0) \, \mathrm{d}u_{n-1} \\
&\leq \int_{u_1=\Delta-\varepsilon-x_0}^{\Delta-x_0} G \int_{u_2=\Delta-\varepsilon-u_1}^{\Delta-u_1} G \, \mathrm{d}u_1 \ldots \int_{u_{n-1}=\Delta-\varepsilon-u_{n-2}}^{\Delta-u_{n-2}} G \, \mathrm{d}u_{n-2} G \, \mathrm{d}u_{n-1} \\
&= \varepsilon^{n-1} G^n,
\end{aligned}$$

where the inequality holds since $|g_\ell|$ are bounded. Therefore, the left-hand side of (5.87) is equal to

$$g_{1,n}^{*,\varepsilon}(x_0, x) + r_4(n) + r_5(n),$$

where $r_5(n) = g_{1,n}^{*}(x_0, x) - g_{1,n}^{*,\varepsilon}(x_0, x) \leq \varepsilon^{n-1} G^n$. □

Combining the results obtained so far in this chapter we have a bound on the difference between $w_n(x_0, x)$ and $g_{1,n}^{*}(x_0, x)$ which we state formally as the following corollary.

Corollary 5.9. *Let $g_1, g_2, \ldots,$ be functions satisfying Assumptions 5.1 and let $\boldsymbol{v}(x)$, $x \in [0, \Delta)$, be a closing operator with Properties 5.2. Then, for $n \geq 2$, $x_0 \in [0, \Delta)$,*

$$\left| w_n(x_0, x) - g_{1,n}^{*}(x_0, x) \right| \leq (n-1)|r_1(n)| + |r_4(n)| + |r_5(n)|, \tag{5.91}$$

Proof. Substitute the expression for $w_n(x_0, x)$ in (5.49) into the left-hand side of (5.91), apply the triangle inequality and Corollary 5.5 and Lemma 5.8 to get the result. □

A direct corollary is the following.

Corollary 5.10. *Let $g_1, g_2, \ldots,$ be functions satisfying Assumptions 5.1, $\psi : [0, \Delta) \to \mathbb{R}$ be bounded, $|\psi| \leq F$, and let $\boldsymbol{v}(x)$, $x \in [0, \Delta)$, be a closing operator with Properties 5.2. Then, for $n \geq 2$, $x_0 \in [0, \Delta)$,*

$$\left| \int_{x=0}^{\Delta} w_n(x_0, x)\psi(x) - g^*_{1,n}(x_0, x)\psi(x)\,\mathrm{d}x \right| \leq ((n-1)|r_1(n)| + |r_4(n)| + |r_5(n)|)\Delta F. \tag{5.92}$$

Proof. The left-hand side of (5.92) is less than or equal to

$$\int_{x=0}^{\Delta} \left| w_n(x_0, x) - g^*_{1,n}(x_0, x) \right| |\psi(x)|\,\mathrm{d}x. \tag{5.93}$$

Applying Corollary 5.9, and since $|\psi| \leq F$, then (5.93) is less than or equal to

$$\int_{x=0}^{\Delta} ((n-1)|r_1(n)| + |r_4(n)| + |r_5(n)|)F\,\mathrm{d}x = ((n-1)|r_1(n)| + |r_4(n)| + |r_5(n)|)\Delta F$$

which is the desired result. □

We have assumed throughout this section that the functions g and $\{g_k\}$ are scalar functions, however, we are ultimately interested in expressions of the form (5.36), which contain matrix functions. Conveniently, the matrix-function expression (5.36) can be written as a linear combination of the scalar case. Hence, we can obtain a bound for the matrix-function case from the scalar case, which we state in the following result.

Lemma 5.11. *Let $\boldsymbol{G}_k(x)$, $k \in \{1, 2, ...\}$, be matrix functions with dimensions $N_k \times N_{k+1}$, and let $\psi : [0, \Delta) \to \mathbb{R}$ be bounded, $|\psi| \leq F$. Further, suppose the scalar function $[\boldsymbol{G}_k(x)]_{ij}$, $i \in \{1, ..., N_k\}$, $j \in \{1, ..., N_{k+1}\}$, $k \in \{1, 2, ...\}$ satisfy Assumptions 5.1. Then,*

$$\begin{aligned}
&\left| \int_{x=0}^{\Delta} \int_{x_1=0}^{\infty} \boldsymbol{G}_1(x_1) \otimes \boldsymbol{k}(x_0) e^{\boldsymbol{S}x_1} \boldsymbol{D}(x_1)\,\mathrm{d}x_1 \left[\prod_{k=2}^{n-1} \int_{x_k=0}^{\infty} \boldsymbol{G}_k(x_k) \otimes e^{\boldsymbol{S}x_k}\,\mathrm{d}x_k \boldsymbol{D} \right] \right. \\
&\qquad \int_{x_n=0}^{\infty} \boldsymbol{G}_n(x_n) \otimes e^{\boldsymbol{S}x_n}\,\mathrm{d}x_n \boldsymbol{v}(x)\psi(x)\,\mathrm{d}x \\
&- \int_{x=0}^{\Delta} \int_{u_1=0}^{\Delta - x_0} \boldsymbol{G}_1(\Delta - u_1 - x_0) \left[\prod_{k=2}^{n-1} \int_{u_k=0}^{\Delta - u_{k-1}} \boldsymbol{G}_k(\Delta - u_k - u_{k-1})\,\mathrm{d}u_{k-1} \right] \\
&\qquad \left. \boldsymbol{G}_n(\Delta - x - u_{n-1}) 1(\Delta - x - u_{n-1} \geq 0)\,\mathrm{d}u_{n-1}\psi(x)\,\mathrm{d}x \right|
\end{aligned}$$

$$\leq ((n-1)|r_1(n)| + |r_4(n)| + |r_5(n)|)\Delta F \prod_{k=2}^{n} N_k, \tag{5.94}$$

where the inequality is an element-wise inequality. Moreover, choosing $\varepsilon = \mathrm{Var}(Z)$*, then, for each* n*, the bound (5.94) is* $\mathcal{O}(\mathrm{Var}(Z)^{1/3})$.

Proof. By the Mixed Product Rule the (i,j)th element of the first term on the left-hand side of (5.94) is

$$\begin{aligned}
&\int_{x=0}^{\Delta}\int_{x_1=0}^{\infty}\cdots\int_{x_n=0}^{\infty}[\boldsymbol{G}_1(x_1)\ldots\boldsymbol{G}_n(x_n)]_{i,j}\,\boldsymbol{k}(x_0)e^{\boldsymbol{S}x_1}\boldsymbol{D}\ldots e^{\boldsymbol{S}x_{n-1}}\boldsymbol{D}\\
&\qquad e^{\boldsymbol{S}x_n}\,\mathrm{d}x_n\ldots\,\mathrm{d}x_1\boldsymbol{v}(x)\psi(x)\,\mathrm{d}x\\
&=\int_{x=0}^{\Delta}\int_{x_1=0}^{\infty}\cdots\int_{x_n=0}^{\infty}\sum_{j_1=1}^{N_2}[\boldsymbol{G}_1(x_1)]_{i,j_1}\sum_{j_2=1}^{N_3}[\boldsymbol{G}_2(x_2)]_{j_1,j_2}\cdots\sum_{j_{n-1}=1}^{N_n}[\boldsymbol{G}_n(x_n)]_{j_{n-1},j}\\
&\qquad \boldsymbol{k}(x_0)e^{\boldsymbol{S}x_1}\boldsymbol{D}\ldots e^{\boldsymbol{S}x_{n-1}}\boldsymbol{D}e^{\boldsymbol{S}x_n}\,\mathrm{d}x_n\ldots\,\mathrm{d}x_1\boldsymbol{v}(x)\psi(x)\,\mathrm{d}x,
\end{aligned}\tag{5.95}$$

from which we see that (5.95) is a linear combination of the scalar function case in Corollary 5.10. Applying the bound for the scalar case, Corollary 5.10, to each term in the linear combination then summing the bounds obtained gives the bound (5.94).

The fact that the error bound is $\mathcal{O}(\mathrm{Var}(Z)^{1/3})$ follows by substituting $\varepsilon = \mathrm{Var}(Z)^{1/3}$ into each term and observing that each term is at most $\mathcal{O}(\mathrm{Var}(Z)^{1/3})$. □

Finally, we are in a position to prove the main result of this section.

Proof of Theorem 5.3. Cases $q = r \in \{+,-\}$ *and* $m = 0$. Lemma 5.4 bounds the absolute difference

$$\left|\int_{x\in\mathcal{D}_{\ell_0}}\widehat{f}^{\ell_0,(p)}_{0,r,r}(\lambda)(x,j;x_0,k)\psi(x)\,\mathrm{d}x - \int_{x\in\mathcal{D}_{\ell_0}}\widehat{\mu}^{\ell_0}_{0,r,r}(\lambda)(\,\mathrm{d}x,j;x_0,k)\psi(x)\right|.$$

Since the bounds from Lemma 5.4 are $\mathcal{O}(\mathrm{Var}(Z^{(p)})^{1/3})$ then, as we take $p\to\infty$, the bounds becomes arbitrarily small which gives the required convergence.

Cases $q, r \in \{+,-\}$, *and* $m \geq 1$. Given the properties of the functions $h^{u,v}_{ij}$, $u,v \in \{+,-\}$, then $\int_{x\in\mathcal{D}_{\ell_0}}\widehat{f}^{\ell_0,(p)}_{0,q,r}(\lambda)(x,j;x_0,k)\psi(x)\,\mathrm{d}x$ satisfies the assumptions of Lemma 5.11. To see this, let q' be the opposite sign to q, i.e. $q' \in \{+,-\}$, $q \neq q'$. Then, in Equation (5.94), take $n = 2m+1 (q = r)$, $\boldsymbol{G}_1(x_1) = \boldsymbol{e}_i\boldsymbol{H}^{qq'}(\lambda,x_1)$, $\boldsymbol{G}_{2k}(x_{2k}) = \boldsymbol{H}^{q'q}(\lambda,x_{2k})$, $\boldsymbol{G}_{2k+1}(x_{2k}) = \boldsymbol{H}^{qq'}(\lambda,x_{2k+1})$, $k = 1,\ldots,m-1$; if $q \neq r$ then take $\boldsymbol{G}_{2m}(x_{2m}) = \boldsymbol{H}^{rr}(x_{2m})\boldsymbol{e}'_j$, otherwise, take $\boldsymbol{G}_{2m}(x_{2m}) = \boldsymbol{H}^{q'r}(x_{2m})$ and $\boldsymbol{G}_{2m+1} = \boldsymbol{H}^{rr}(\lambda,x_{2m+1})\boldsymbol{e}'_j$. Thus, Lemma 5.11, establishes a bound on (5.39) which is $\mathcal{O}(\mathrm{Var}(Z^{(p)})^{1/3})$, thus taking $p\to\infty$ gives the stated convergence.

Cases $q = 0$, $r \in \{+, -\}$ *and* $m \geq 0$. Since

$$\widehat{f}^{\ell_0}_{m,0,r}(\lambda)(x, j; x_0, k)\,\mathrm{d}x = \sum_{q\in\{+,-\}} \sum_{i\in\mathcal{S}_q} \boldsymbol{e}_k \left[\lambda \boldsymbol{I} - \boldsymbol{T}_{00}\right]^{-1} \boldsymbol{T}_{0i} \widehat{f}^{\ell_0}_{m+1(r\neq q),q,r}(\lambda)(x, j; x_0, i)\,\mathrm{d}x, \tag{5.96}$$

is a linear combination of terms which are treated in the two cases above, then (5.96) converges to

$$\widehat{\mu}^{\ell_0}_{m,0,r}(\lambda)(\,\mathrm{d}x, j; x_0, k) = \sum_{q\in\{+,-\}} \sum_{i\in\mathcal{S}_q} \boldsymbol{e}_k \left[\lambda \boldsymbol{I} - \boldsymbol{T}_{00}\right]^{-1} \boldsymbol{T}_{0i} \widehat{\mu}^{\ell_0}_{m+1(r\neq q),q,r}(\lambda)(\,\mathrm{d}x, j; x_0, i), \tag{5.97}$$

as required. □

5.5 Convergence before the first orbit restart epoch, τ_1

Recall that the goal in this chapter is to show a convergence of

$$\widehat{f}^{\ell_0,(p)}_{q,r}(\lambda)(x, j; x_0, i)\,\mathrm{d}x \to \widehat{\mu}^{\ell_0}_{q,r}(\lambda)(\,\mathrm{d}x, j; x_0, i),$$

where

$$\widehat{f}^{\ell_0,(p)}(\lambda)(x, j; x_0, i) = \int_{t=0}^{\infty} \sum_{m=0}^{\infty} e^{-\lambda t} f^{\ell_0,(p)}_{m+1(p\neq q),q,r}(t)(x, j; x_0, k)\,\mathrm{d}t.$$

Since $f^{\ell_0,(p)}_{m+1(p\neq q),q,r}$ are positive, as is $e^{-\lambda t}$, then we can use the Fubini-Tonelli Theorem to justify a swap of the integral and infinite sum to get

$$\widehat{f}^{\ell_0,(p)}(\lambda)(x, j; x_0, i) = \sum_{m=0}^{\infty} \widehat{f}^{\ell_0,(p)}_{m+1(p\neq q),q,r}(\lambda)(x, j; x_0, k). \tag{5.98}$$

Similarly, we can write

$$\widehat{\mu}^{\ell_0}(\lambda)(\,\mathrm{d}x, j; x_0, i) = \sum_{m=0}^{\infty} \widehat{\mu}^{\ell_0}_{m+1(p\neq q),q,r}(\lambda)(\,\mathrm{d}x, j; x_0, k).$$

The previous section proved that the Laplace transforms

$$\widehat{f}^{\ell_0,(p)}_{m,q,r}(\lambda)(x, j; x_0, k)\,\mathrm{d}x \to \widehat{\mu}^{\ell_0}_{m,q,r}(\lambda)(\,\mathrm{d}x, j; x_0, k),$$

for $q \in \{+, -, +0, -0\}$, $r \in \{+, -\}$. Thus, all we need to show is that, upon taking the limit of (5.98), we can swap the limit and the summation. Here we apply the Dominated

Convergence Theorem to justify the swap. To this end, we show a domination condition in Lemma 5.12 below.

Recall $c_{min} = \min_{i \in \mathcal{S}_+ \cup \mathcal{S}_-} |c_i|$, and let E^λ be an independent exponential random variable with rate λ and $\gamma = \max_{i \in \mathcal{S}_+ \cup \mathcal{S}_-} -T_{ii}/|c_i|$. In the following we use the stochastic interpretation of the Laplace transform of a probability distribution with non-negative support and real, non-negative parameter λ. For a random variable W with distribution function $F_W(w) = \mathbb{P}(W < w)$, then $\displaystyle\int_{w=0}^{\infty} e^{-\lambda w} \, \mathrm{d}F_W(w) = \mathbb{P}(W < E^\lambda)$. That is, the Laplace transform with parameter $\lambda > 0$ is the probability that W occurs before E^λ, an independent random exponential time with rate λ, occurs.

Lemma 5.12. *For all $M \geq 0$, $x \in \mathcal{D}_{\ell_0,j}$, $x_0 \in \mathcal{D}_{\ell_0,i}$, $\ell_0 \in \mathcal{K}$, $\lambda > 0$, $r \in \{+,-\}$, $j \in \mathcal{S}_r \cup \mathcal{S}_{r0}$, and either $q \in \{+,-\}$, $i \in \mathcal{S}_q$, or $q = 0$, $i \in \mathcal{S}_0^*$, for any bounded function ψ, $|\psi| < F$,*

$$\sum_{m=M+1}^{\infty} \left| \int_{x \in \mathcal{D}_{\ell_0}} \widehat{f}^{\ell_0,(p)}_{m,q,r}(\lambda)(x, j; x_0, i)\psi(x) \, \mathrm{d}x - \int_{x \in \mathcal{D}_{\ell_0}} \widehat{\mu}^{\ell_0}_{m,q,r}(\lambda)(\, \mathrm{d}x, j; x_0, i)\psi(x) \right| \leq r_6^M \tag{5.99}$$

where

$$r_6^M = F(\Delta G + \widehat{G}) \left(\frac{\gamma}{\gamma + \lambda} \right)^{2M+1+1(q=r)} \left(1 - \left(\frac{\gamma}{\gamma + \lambda} \right)^2 \right)^{-1}.$$

Note that the bound r_6^M is independent of p.

We prove the result for $q = r = +$ first, with the proof for the other cases $q, r \in \{+,-\}$ following analogously. The proof for $q = 0$, $i \in \mathcal{S}_0^*$, follows after noting that the Laplace transforms $\widehat{f}^{\ell_0,(p)}_{m,0,r}(\lambda)(x, j; x_0, i)$ and $\widehat{\mu}^{\ell_0}_{m,0,r}(\lambda)(\, \mathrm{d}x, j; x_0, i)$ are linear combinations of $\widehat{f}^{\ell_0,(p)}_{m,q,r}(\lambda)(x, j; x_0, i)$ and $\widehat{\mu}^{\ell_0}_{m,q,r}(\lambda)(\, \mathrm{d}x, j; x_0, i)$ for $q \in \{+,-\}$, respectively.

Essentially, the result for $q = r = +$ follows from noting the probabilistic interpretation of the Laplace transforms $\widehat{f}^{\ell_0}_{m,+,+}(\lambda)(x, j; x_0, i)$, as the probability that,

- there are m up-down and down-up transitions,
- the orbit process $\{\boldsymbol{A}(t)\}$ evolves accordingly,
- and an independent exponential random variable with rate λ, E^λ, has not yet occurred,
- before the first orbit restart epoch.

We obtain an upper bound by ignoring the behaviour of the orbit process $\{\boldsymbol{A}(t)\}$, then, by a uniformisation argument, we bound the probability that there are m up-down and down-up transitions before E^λ occurs, by the event that there are m independent exponential events before an E^λ occurs.

Similarly, the stochastic interpretation of the Laplace transforms $\widehat{\mu}^{\ell_0}_{m,+,+}(\lambda)(\,\mathrm{d}x, j; x_0, i)$, is the probability that,

- there are m up-down and down-up transitions,
- the fluid level $X(t)$ remains in $\mathcal{D}_{\ell_0}$,
- and an independent exponential random variable with rate λ, E^λ, has not yet occurred,
- before the first orbit restart epoch.

We obtain an upper bound by removing the requirement that the fluid level $X(t)$ remain in $\mathcal{D}_{\ell_0}$, then applying the same uniformisation argument as we do for $\widehat{f}^{\ell_0}_{m,+,+}(\lambda)(x, j; x_0, i)$.

Proof. The same arguments and results apply for all p, so let us drop the dependence on p.

Consider $i \in \mathcal{S}_+, j \in \mathcal{S}_+ \cup \mathcal{S}_{+0}$. By the triangle inequality,

$$\begin{aligned}
&\sum_{m=M+1}^{\infty} \left| \int_{x\in\mathcal{D}_{\ell_0}} \widehat{f}^{\ell_0}_{m,+,+}(\lambda)(x, j; x_0, i)\psi(x)\,\mathrm{d}x - \int_{x\in\mathcal{D}_{\ell_0}} \widehat{\mu}^{\ell_0}_{m,+,+}(\lambda)(\,\mathrm{d}x, j; x_0, i)\psi(x) \right| \\
&\leq \sum_{m=M+1}^{\infty} \int_{x\in\mathcal{D}_{\ell_0}} \widehat{f}^{\ell_0}_{m,+,+}(\lambda)(x, j; x_0, i)|\psi(x)|\,\mathrm{d}x \\
&\qquad + \sum_{m=M+1}^{\infty} \int_{x\in\mathcal{D}_{\ell_0}} \widehat{\mu}^{\ell_0}_{m,+,+}(\lambda)(\,\mathrm{d}x, j; x_0, i)|\psi(x)|,
\end{aligned}$$

since all terms are non-negative.

Consider $\int_{x\in\mathcal{D}_{\ell_0}} \widehat{f}^{\ell_0}_{m,+,+}(\lambda)(x, j; x_0, i)|\psi(x)|\,\mathrm{d}x$, which is given by

$$\begin{aligned}
&\int_{x\in\mathcal{D}_{\ell_0}} \boldsymbol{e}_i \left[\prod_{r=1}^{m} \int_{x_{2r-1}=0}^{\infty} \boldsymbol{H}^{+-}(\lambda, x_{2r-1}) \int_{x_{2r}=0}^{\infty} \boldsymbol{H}^{-+}(\lambda, x_{2r})\right] \int_{x_{2m+1}=0}^{\infty} \boldsymbol{H}^{++}(\lambda, x_{2m+1})\boldsymbol{e}'_j \\
&\boldsymbol{a}_{\ell_0,i}(x_0)\boldsymbol{N}^{2m+1}(\lambda, x_1, \ldots, x_{2m+1})\boldsymbol{v}_{\ell_0,j}(x)\,\mathrm{d}x_{2m+1}\ldots\,\mathrm{d}x_1\psi(x)\,\mathrm{d}x \\
&\leq \int_{x\in\mathcal{D}_{\ell_0}} \boldsymbol{e}_i \left[\prod_{r=1}^{m} \int_{x_{2r-1}=0}^{\infty} \boldsymbol{H}^{+-}(\lambda, x_{2r-1}) \int_{x_{2r}=0}^{\infty} \boldsymbol{H}^{-+}(\lambda, x_{2r})\right] \int_{x_{2m+1}=0}^{\infty} \boldsymbol{H}^{++}(\lambda, x_{2m+1})\boldsymbol{e}'_j \\
&\boldsymbol{a}_{\ell_0,i}(x_0)\boldsymbol{N}^{2m+1}(\lambda, x_1, \ldots, x_{2m+1})\boldsymbol{v}_{\ell_0,j}(x)\,\mathrm{d}x_{2m+1}\ldots\,\mathrm{d}x_1\,\mathrm{d}xF \qquad (5.100)
\end{aligned}$$

since $|\psi| \leq F$. To bound the last-line of (5.100) we first observe that for $\boldsymbol{a} \in \mathcal{A}$,

$$\begin{aligned}\boldsymbol{a}\int_{x\in\mathcal{D}_{\ell_0}} \boldsymbol{D}e^{\boldsymbol{S}x_{2m+1}}\boldsymbol{v}_{\ell_0,j}(x) &= \boldsymbol{a}\int_{x\in\mathcal{D}_{\ell_0}}\int_{u=0}^{\infty} e^{\boldsymbol{S}u}\boldsymbol{s}\frac{\boldsymbol{\alpha}e^{\boldsymbol{S}u}}{\boldsymbol{\alpha}e^{\boldsymbol{S}u}\boldsymbol{e}}e^{\boldsymbol{S}x_{2m+1}}\boldsymbol{v}_{\ell_0,j}(x)\,\mathrm{d}u\,\mathrm{d}x\\ &\leq \boldsymbol{a}\int_{u=0}^{\infty} e^{\boldsymbol{S}u}\boldsymbol{s}\frac{\boldsymbol{\alpha}e^{\boldsymbol{S}u}}{\boldsymbol{\alpha}e^{\boldsymbol{S}u}\boldsymbol{e}}e^{\boldsymbol{S}x_{2m+1}}\boldsymbol{e}\,\mathrm{d}u\\ &= \boldsymbol{a}\boldsymbol{D}e^{\boldsymbol{S}x_{2m+1}}\boldsymbol{e}\,\mathrm{d}u, \end{aligned} \tag{5.101}$$

where the inequality holds from Property 5.2(iv). By definition, the last-line of (5.100) is

$$\begin{aligned}&\int_{x\in\mathcal{D}_{\ell_0}} \boldsymbol{a}_{\ell_0,i}(x_0)\boldsymbol{N}^{2m+1}(\lambda,x_1,\ldots,x_{2m+1})\boldsymbol{v}_{\ell_0,j}(x)\,\mathrm{d}x_{2m+1}\ldots\,\mathrm{d}x_1\\ &= \int_{x\in\mathcal{D}_{\ell_0}} \boldsymbol{a}_{\ell_0,i}(x_0)e^{\boldsymbol{S}x_1}\boldsymbol{D}e^{\boldsymbol{S}x_2}\boldsymbol{D}\ldots e^{\boldsymbol{S}x_{2m}}\boldsymbol{D}e^{\boldsymbol{S}x_{2m+1}}\boldsymbol{v}_{\ell_0,j}(x). \end{aligned} \tag{5.102}$$

Now, using (5.101), then (5.102) is less than or equal to

$$\begin{aligned}&\boldsymbol{a}_{\ell_0,i}(x_0)e^{\boldsymbol{S}x_1}\boldsymbol{D}e^{\boldsymbol{S}x_2}\boldsymbol{D}\ldots e^{\boldsymbol{S}x_{2m}}\boldsymbol{D}e^{\boldsymbol{S}x_{2m+1}}\boldsymbol{e}\\ &= \boldsymbol{a}_{\ell_0,i}(x_0)e^{\boldsymbol{S}x_1}\boldsymbol{D}e^{\boldsymbol{S}x_2}\boldsymbol{D}\ldots e^{\boldsymbol{S}x_{2m}}\int_{u=0}^{\infty} e^{\boldsymbol{S}u}\frac{\boldsymbol{\alpha}e^{\boldsymbol{S}u}}{\boldsymbol{\alpha}e^{\boldsymbol{S}u}\boldsymbol{e}}\mathrm{d}u e^{\boldsymbol{S}x_{2m+1}}\boldsymbol{e}\\ &\leq \boldsymbol{a}_{\ell_0,i}(x_0)e^{\boldsymbol{S}x_1}\boldsymbol{D}e^{\boldsymbol{S}x_2}\boldsymbol{D}\ldots e^{\boldsymbol{S}x_{2m}}\int_{u=0}^{\infty} e^{\boldsymbol{S}u}\frac{\boldsymbol{\alpha}e^{\boldsymbol{S}u}}{\boldsymbol{\alpha}e^{\boldsymbol{S}u}\boldsymbol{e}}\mathrm{d}u\boldsymbol{e}\\ &= \boldsymbol{a}_{\ell_0,i}(x_0)e^{\boldsymbol{S}x_1}\boldsymbol{D}e^{\boldsymbol{S}x_2}\boldsymbol{D}\ldots e^{\boldsymbol{S}x_{2m}}\boldsymbol{e}. \end{aligned}$$

Repeating m more times gives the bound $\boldsymbol{a}_{\ell_0,i}(x_0)\boldsymbol{e} = 1$.[‖] Hence, we have the bound

$$\boldsymbol{a}_{\ell_0,i}(x_0)e^{\boldsymbol{S}x_1}\boldsymbol{D}e^{\boldsymbol{S}x_2}\boldsymbol{D}\ldots e^{\boldsymbol{S}x_{2m}}\boldsymbol{D}e^{\boldsymbol{S}x_{2m+1}}\boldsymbol{e} \leq 1.$$

Therefore, (5.100) is less than or equal to

$$\begin{aligned}&\boldsymbol{e}_i\left[\prod_{r=1}^{m}\int_{x_{2r-1}=0}^{\infty}\boldsymbol{H}^{+-}(\lambda,x_{2r-1})\int_{x_{2r}=0}^{\infty}\boldsymbol{H}^{-+}(\lambda,x_{2r})\right]\int_{x_{2m+1}=0}^{\infty}\boldsymbol{H}^{++}(\lambda,x_{2m+1})\boldsymbol{e}'_j\\ &\mathrm{d}x_{2m+1}\ldots\,\mathrm{d}x_1 F. \end{aligned} \tag{5.103}$$

Now, for any row-vector of non-negative numbers $\boldsymbol{b}$, since the elements of $\boldsymbol{H}^{++}$ are non-negative and integrable, then

$$\boldsymbol{b}\int_{x_{2m+1}=0}^{\infty}\boldsymbol{H}^{++}(\lambda,x_{2m+1})\,\mathrm{d}x_{2m+1}\boldsymbol{e}'_j \leq \boldsymbol{b}\boldsymbol{e}'_j\widehat{G} \leq \boldsymbol{b}\boldsymbol{e}\widehat{G}.$$

[‖] In fact, this bound holds for any initial vector $\boldsymbol{a} \in \mathcal{A}$.

Observing that

$$\boldsymbol{e}_i \left[\prod_{r=1}^{m} \int_{x_{2r-1}=0}^{\infty} \boldsymbol{H}^{+-}(\lambda, x_{2r-1}) \int_{x_{2r}=0}^{\infty} \boldsymbol{H}^{-+}(\lambda, x_{2r}) \right] \mathrm{d}x_{2m} \ldots \mathrm{d}x_1$$

is a vector of row-vector non-negative numbers, then (5.103) is less than or equal to

$$\boldsymbol{e}_i \left[\prod_{r=1}^{m} \int_{x_{2r-1}=0}^{\infty} \boldsymbol{H}^{+-}(\lambda, x_{2r-1}) \int_{x_{2r}=0}^{\infty} \boldsymbol{H}^{-+}(\lambda, x_{2r}) \right] \boldsymbol{e}\, \mathrm{d}x_{2m} \ldots \mathrm{d}x_1 \widehat{G}F \tag{5.104}$$

The stochastic interpretation of the ith element of the vector $\boldsymbol{H}^{+-}(\lambda, x)\boldsymbol{e}$ is that it is the probability density of an up-down transition at the time when the in-out fluid has increased by $\mathrm{d}x$ and before an exponential random variable with rate λ occurs, given the phase is initially i. There may be multiple changes of phase within $\mathcal{S}_+ \cup \mathcal{S}_{+0}$ before the first up-down transition. The first change of phase occurs at rate (with respect to the in-out level) $-T_{ii}/|c_i|$ and this is the lowest in-out fluid level at which it may be possible to see an up-down transition. Consider a uniformised version of the in-out fluid process with uniformisation parameter $\gamma = \max\limits_{i \in \mathcal{S}_+ \cup \mathcal{S}_-} -T_{ii}/|c_i|$. Then the first event of the phase process of the uniformised version of the in-out fluid process occurs at rate γ and occurs at, or before, the first change of phase of the uniformised process. Therefore, the first uniformisation event occurs at, or before, the first up-down transition of the uniformised version of the in-out process. Hence, the first uniformisation event occurs at, or before, the first up-down transition of the original process (since they are versions of each other). This gives the bound $\boldsymbol{H}^{+-}(\lambda, x)\boldsymbol{e} \leq \gamma e^{-(\lambda+\gamma)x}\boldsymbol{e}$ where the inequality is understood elementwise. Similarly, for $\boldsymbol{H}^{-+}(\lambda, x)\boldsymbol{e} \leq \gamma e^{-(\lambda+\gamma)x}\boldsymbol{e}$.

From the stochastic interpretation above, (5.104) is less than or equal to

$$\begin{aligned}
&\boldsymbol{e}_i \boldsymbol{H}^{+-}(\lambda, x_1)\, \mathrm{d}x_1 \int_{x_2=0}^{\infty} \boldsymbol{H}^{-+}(\lambda, x_2)\boldsymbol{e}\, \mathrm{d}x_2 \ldots \int_{x_{2m}=0}^{\infty} \gamma e^{(-\gamma-\lambda)x_{2m}}\, \mathrm{d}x_{2m} \widehat{G}F \\
&\leq \int_{x_1=0}^{\infty} \gamma e^{(-\gamma-\lambda)x_1}\, \mathrm{d}x_1 \int_{x_2=0}^{\infty} \gamma e^{(-\gamma-\lambda)x_2}\, \mathrm{d}x_2 \ldots \int_{x_{2m}=0}^{\infty} \gamma e^{(-\gamma-\lambda)x_{2m}}\, \mathrm{d}x_{2m} \widehat{G}F \\
&= \left(\frac{\gamma}{\gamma+\lambda} \right)^{2m} \widehat{G}F.
\end{aligned} \tag{5.105}$$

Note, for the cases with $q \neq r$ then the equivalent bound to (5.105) is $\left(\frac{\gamma}{\gamma+\lambda}\right)^{2m-1} \widehat{G}F$ as there is one less factor in the expression.

Hence,

$$\sum_{m=M+1}^{\infty} \int_{x \in \mathcal{D}_{\ell_0}} \widehat{f}^{\ell_0}_{m,+,+}(\lambda)(x, j; x_0, i)|\psi(x)|\, \mathrm{d}x \leq \widehat{G}F \sum_{m=M+1}^{\infty} \left(\frac{\gamma}{\gamma+\lambda} \right)^{2m}$$

$$\leq \widehat{G}F\left(\frac{\gamma}{\gamma+\lambda}\right)^{2M+2}\left(1-\left(\frac{\gamma}{\gamma+\lambda}\right)^2\right)^{-1}. \tag{5.106}$$

Now consider $\widehat{\mu}^{\ell_0}_{m,+,+}(\lambda)(\,\mathrm{d}x, j; x_0, i)$ which is given by

$$\begin{aligned}
&\int_{x_1=0}^{\Delta-(x_0-y_{\ell_0})} \boldsymbol{e}_i\boldsymbol{H}^{+-}(\lambda,\Delta-(x_0-y_{\ell_0})-x_1)\int_{x_2=0}^{\Delta-x_1}\boldsymbol{H}^{-+}(\lambda,\Delta-x_2-x_1)\,\mathrm{d}x_1\\
&\cdots\int_{x_{2m}=0}^{\Delta-x_{m-1}}\boldsymbol{H}^{-+}(\lambda,\Delta-x_{2m-1}-x_{2m})\,\mathrm{d}x_{2m-1}\boldsymbol{H}^{++}(\lambda,\Delta-x_{2m}-(y_{\ell_0+1}-x))\boldsymbol{e}'_j\,\mathrm{d}x_{2m}\,\mathrm{d}x\\
&=\int_{x_1=(x_0-y_{\ell_0})}^{\Delta}\boldsymbol{e}_i\boldsymbol{H}^{+-}(\lambda,\Delta-x_1)\int_{x_2=x_1}^{\Delta}\boldsymbol{H}^{-+}(\lambda,\Delta-x_2)\,\mathrm{d}x_1\ldots\\
&\int_{x_{2m}=x_{2m-1}}^{\Delta}\boldsymbol{H}^{-+}(\lambda,\Delta-x_{2m})\boldsymbol{H}^{++}(\lambda,\Delta-x_{2m}-x_{2m-1}-(y_{\ell_0+1}-x))\boldsymbol{e}'_j\,\mathrm{d}x_{2m-1}\,\mathrm{d}x_{2m}\,\mathrm{d}x.
\end{aligned} \tag{5.107}$$

Using the bound $\boldsymbol{H}^{++}(\lambda, x_{m+1}) \leq G$ elementwise, then (5.107) is less than or equal to

$$\begin{aligned}
&\int_{x_1=(x_0-y_{\ell_0})}^{\Delta}\boldsymbol{e}_i\boldsymbol{H}^{+-}(\lambda,\Delta-x_1)\int_{x_2=x_1}^{\Delta}\boldsymbol{H}^{-+}(\lambda,\Delta-x_2)\,\mathrm{d}x_1\\
&\quad\cdots\int_{x_{2m}=x_{2m-1}}^{\Delta}\boldsymbol{H}^{-+}(\lambda,\Delta-x_{2m})\,\mathrm{d}x_{2m-1}\,\mathrm{d}x_{2m}\boldsymbol{e}G\,\mathrm{d}x.
\end{aligned} \tag{5.108}$$

The expression (5.108) differs from (5.104) only by a constant factor and that the integrals in the (5.108) are finite, hence we may bound it in the same way. Therefore,

$$\begin{aligned}
\sum_{m=M+1}^{\infty}\int_{x\in\mathcal{D}_{\ell_0}}\widehat{\mu}^{\ell_0}_{m,+,+}(\lambda)(\,\mathrm{d}x,j;x_0,i)|\psi(x)| &\leq G\sum_{m=M+1}^{\infty}\left(\frac{\gamma}{\gamma+\lambda}\right)^{2m}\int_{x\in\mathcal{D}_{\ell_0}}\psi(x)\,\mathrm{d}x\\
&\leq G\Delta F\left(\frac{\gamma}{\gamma+\lambda}\right)^{2M+2}\left(1-\left(\frac{\gamma}{\gamma+\lambda}\right)^2\right)^{-1}.
\end{aligned} \tag{5.109}$$

Once again, for $q \neq r$ the bound on $\int_{x\in\mathcal{D}_{\ell_0}}\widehat{\mu}^{\ell_0}_{m,+,+}(\lambda)(\,\mathrm{d}x,j;x_0,i)|\psi(x)|$ is $(\frac{\gamma}{\gamma+\lambda})^{2m-1}$ as there is one less factor.

Analogous arguments show the same bounds for any $i \in \mathcal{S}_q$, $j \in \mathcal{S}_r \cup \mathcal{S}_{r0}$, where $q \in \{+,-\}$, $r \in \{+,-\}$.

The result for $q = 0$, $i \in \mathcal{S}_0^*$, $j \in \mathcal{S}_r \cup \mathcal{S}_{r0}$, $r \in \{+, -\}$ holds after noting that

$$\widehat{f}^{\ell_0}_{m,0,r}(\lambda)(x, j; x_0, i)\,\mathrm{d}x = \sum_{q\in\{+,-\}} \sum_{k\in\mathcal{S}_q} \boldsymbol{e}_i \left[\lambda \boldsymbol{I} - \boldsymbol{T}_{00}\right]^{-1} \boldsymbol{T}_{0k} \widehat{f}^{\ell_0}_{m+1(r\neq q),q,r}(\lambda)(x, j; x_0, k)\,\mathrm{d}x, \tag{5.110}$$

$$\widehat{\mu}^{\ell_0}_{m,0,r}(\lambda)(\,\mathrm{d}x, j; x_0, i) = \sum_{q\in\{+,-\}} \sum_{k\in\mathcal{S}_q} \boldsymbol{e}_i \left[\lambda \boldsymbol{I} - \boldsymbol{T}_{00}\right]^{-1} \boldsymbol{T}_{0k} \widehat{\mu}^{\ell_0}_{m+1(r\neq q),q,r}(\lambda)(\,\mathrm{d}x, j; x_0, k), \tag{5.111}$$

are a linear combination of terms which are treated the cases above, hence we may use the bounds we have established above. First, use the triangle inequality to write,

$$\begin{aligned}
&\sum_{m=M+1}^{\infty} \left| \int_{x\in\mathcal{D}_{\ell_0}} \widehat{f}^{\ell_0}_{m,0,+}(\lambda)(x, j; x_0, i)\psi(x)\,\mathrm{d}x - \int_{x\in\mathcal{D}_{\ell_0}} \widehat{\mu}^{\ell_0}_{m,0,+}(\lambda)(\,\mathrm{d}x, j; x_0, i)\psi(x) \right| \\
&\leq \sum_{m=M+1}^{\infty} \int_{x\in\mathcal{D}_{\ell_0}} \widehat{f}^{\ell_0}_{m,0,+}(\lambda)(x, j; x_0, i)|\psi(x)|\,\mathrm{d}x \\
&\quad + \sum_{m=M+1}^{\infty} \int_{x\in\mathcal{D}_{\ell_0}} \widehat{\mu}^{\ell_0}_{m,0,+}(\lambda)(\,\mathrm{d}x, j; x_0, i)|\psi(x)|.
\end{aligned} \tag{5.112}$$

Upon substituting (5.110) and (5.111) into (5.112) we get

$$\begin{aligned}
&\sum_{m=M+1}^{\infty} \int_{x\in\mathcal{D}_{\ell_0}} \sum_{q\in\{+,-\}} \sum_{k\in\mathcal{S}_q} \boldsymbol{e}_i \left[\lambda \boldsymbol{I} - \boldsymbol{T}_{00}\right]^{-1} \boldsymbol{T}_{0k} \widehat{f}^{\ell_0}_{m+1(r\neq q),q,r}(\lambda)(x, j; x_0, k)\,\mathrm{d}x|\psi(x)|\,\mathrm{d}x \\
&\quad + \sum_{m=M+1}^{\infty} \int_{x\in\mathcal{D}_{\ell_0}} \sum_{q\in\{+,-\}} \sum_{k\in\mathcal{S}_q} \boldsymbol{e}_i \left[\lambda \boldsymbol{I} - \boldsymbol{T}_{00}\right]^{-1} \boldsymbol{T}_{0k} \widehat{\mu}^{\ell_0}_{m+1(r\neq q),q,r}(\lambda)(\,\mathrm{d}x, j; x_0, k)|\psi(x)| \\
&= \sum_{q\in\{+,-\}} \sum_{k\in\mathcal{S}_q} \sum_{m=M+1}^{\infty} \int_{x\in\mathcal{D}_{\ell_0}} \boldsymbol{e}_i \left[\lambda \boldsymbol{I} - \boldsymbol{T}_{00}\right]^{-1} \boldsymbol{T}_{0k} \widehat{f}^{\ell_0}_{m+1(r\neq q),q,r}(\lambda)(x, j; x_0, k)\,\mathrm{d}x|\psi(x)|\,\mathrm{d}x \\
&\quad + \sum_{q\in\{+,-\}} \sum_{k\in\mathcal{S}_q} \sum_{m=M+1}^{\infty} \int_{x\in\mathcal{D}_{\ell_0}} \boldsymbol{e}_i \left[\lambda \boldsymbol{I} - \boldsymbol{T}_{00}\right]^{-1} \boldsymbol{T}_{0k} \widehat{\mu}^{\ell_0}_{m+1(r\neq q),q,r}(\lambda)(\,\mathrm{d}x, j; x_0, k)|\psi(x)|
\end{aligned} \tag{5.113}$$

where the swap of the sums and integrals is justified by Tonelli's Theorem. Now, using the bounds we found earlier, then (5.113) is less than or equal to

$$\sum_{q\in\{+,-\}} \sum_{k\in\mathcal{S}_q} \boldsymbol{e}_i \left[\lambda \boldsymbol{I} - \boldsymbol{T}_{00}\right]^{-1} \boldsymbol{T}_{0k} F\widehat{G} \left(\frac{\gamma}{\gamma+\lambda}\right)^{2M+2} \left(1 - \left(\frac{\gamma}{\gamma+\lambda}\right)^2\right)^{-1}$$

$$
\begin{aligned}
&+ \sum_{q\in\{+,-\}} \sum_{k\in\mathcal{S}_q} \boldsymbol{e}_i \left[\lambda \boldsymbol{I} - \boldsymbol{T}_{00}\right]^{-1} \boldsymbol{T}_{0k} F \Delta G \left(\frac{\gamma}{\gamma+\lambda}\right)^{2M+2} \left(1 - \left(\frac{\gamma}{\gamma+\lambda}\right)^2\right)^{-1} \\
&\leq \sum_{q\in\{+,-\}} \sum_{k\in\mathcal{S}_q} \boldsymbol{e}_i \left[\lambda \boldsymbol{I} - \boldsymbol{T}_{00}\right]^{-1} \boldsymbol{T}_{0k} F \widehat{G} \left(\frac{\gamma}{\gamma+\lambda}\right)^{2M+1} \left(1 - \left(\frac{\gamma}{\gamma+\lambda}\right)^2\right)^{-1} \\
&\quad + \sum_{q\in\{+,-\}} \sum_{k\in\mathcal{S}_q} \boldsymbol{e}_i \left[\lambda \boldsymbol{I} - \boldsymbol{T}_{00}\right]^{-1} \boldsymbol{T}_{0k} F \Delta G \left(\frac{\gamma}{\gamma+\lambda}\right)^{2M+1} \left(1 - \left(\frac{\gamma}{\gamma+\lambda}\right)^2\right)^{-1} \\
&\leq F(\widehat{G} + \Delta G) \left(\frac{\gamma}{\gamma+\lambda}\right)^{2M+1} \left(1 - \left(\frac{\gamma}{\gamma+\lambda}\right)^2\right)^{-1}
\end{aligned}
$$

since

$$
\begin{aligned}
\sum_{q\in\{+,-\}} \sum_{k\in\mathcal{S}_q} \boldsymbol{e}_i \left[\lambda \boldsymbol{I} - \boldsymbol{T}_{00}\right]^{-1} \boldsymbol{T}_{0k} &= \sum_{q\in\{+,-\}} \sum_{k\in\mathcal{S}_q} \boldsymbol{e}_i \int_{t=0}^{\infty} e^{(\boldsymbol{T}_{00} - \lambda \boldsymbol{I})t} \,\mathrm{d}t \boldsymbol{T}_{0k} \\
&\leq \sum_{q\in\{+,-\}} \sum_{k\in\mathcal{S}_q} \boldsymbol{e}_i \int_{t=0}^{\infty} e^{\boldsymbol{T}_{00} t} \,\mathrm{d}t \boldsymbol{T}_{0k} \\
&= \sum_{q\in\{+,-\}} \sum_{k\in\mathcal{S}_q} \boldsymbol{e}_i (-\boldsymbol{T}_{00})^{-1} \boldsymbol{T}_{0k} \\
&= \boldsymbol{e}_i (-\boldsymbol{T}_{00})^{-1} (-\boldsymbol{T}_{00}) \boldsymbol{e}' \\
&= 1.
\end{aligned}
$$

□

Combining the domination in Lemma 5.12 and the convergence in Theorem 5.3 via the Dominated Convergence Theorem gives the following result.

Lemma 5.13. *For all $x \in \mathcal{D}_{\ell_0,j}$, $x_0 \in \mathcal{D}_{\ell_0,i}$, $i, j \in \mathcal{S}$, $\ell_0 \in \mathcal{K}$, $\lambda > 0$,*

$$
\left| \int_{x\in\mathcal{D}_{\ell_0}} \widehat{f}^{\ell_0,(p)}(\lambda)(x, j; x_0, i)\psi(x) \,\mathrm{d}x - \int_{x\in\mathcal{D}_{\ell_0}} \widehat{\mu}^{\ell_0}(\lambda)(\,\mathrm{d}x, j; x_0, i)\psi(x) \right| \to 0 \tag{5.114}
$$

as $p \to \infty$.

Remark 5.14. *For a fixed $\lambda > 0$, convergence of*

$$
\left| \widehat{f}^{\ell_0,(p)}(\lambda)(x, j; x_0, i) \,\mathrm{d}x - \widehat{\mu}^{\ell_0}(\lambda)(\,\mathrm{d}x, j; x_0, i) \right| \tag{5.115}
$$

actually holds pointwise for each $\ell_0 \in \mathcal{K}^\circ$, and each $i, j \in \mathcal{S}$, $x_0 \in \mathcal{D}_{\ell_0,i}$, $x \in \mathcal{D}_{\ell_0,j}$ except at the set of points where $x = x_0$. Specifically, the lack of pointwise convergence at this point

occurs due to terms with the index $m = 0$, that is, terms where there are no up-down or down-up transitions. On these sample paths the relevant Laplace transforms of the fluid queue are discontinuous at this point. For example,

$$\widehat{\mu}^{\ell_0}_{0,+,+}(\lambda)(\,\mathrm{d}x, j; x_0, i) = h^{++}_{ij}(\lambda, x - x_0)1(x \geq x_0)\,\mathrm{d}x,$$

is discontinuous at $x = x_0$.

5.6 Convergence at the time of the first orbit restart epoch, τ_1

We conclude this chapter with a statement about a convergence of the QBD-RAP to the fluid queue *at the time of the first orbit restart epoch*, $\tau_1^{(p)}$.

Corollary 5.15. *Recall $\boldsymbol{y}_0^{(p)} = (\ell_0, \boldsymbol{a}^{(p)}_{\ell_0,j}(x_0), i)$. For $\ell_0 \in \mathcal{K}$ $x_0 \in \mathcal{D}_{\ell_0,i}$, $i \in \mathcal{S}_+ \cup \mathcal{S}_- \cup \mathcal{S}_0^*$, $j \in \mathcal{S}_+ \cup \mathcal{S}_-$,*

$$\begin{aligned}&\mathbb{P}(L^{(p)}(\tau_1^{(p)}) = \ell(\ell_0, j), \varphi(\tau_1^{(p)}) = j, \tau_1^{(p)} \leq E^\lambda \mid \boldsymbol{Y}^{(p)}(0) = \boldsymbol{y}_0^{(p)})\\ &\to \mathbb{P}(\boldsymbol{X}(\tau_1^X) = (y_{\ell(\ell_0,j)+1(j\in\mathcal{S}_-)}, j), \tau_1^X \leq E^\lambda \mid \boldsymbol{X}(0) = (x_0, i))\end{aligned} \tag{5.116}$$

where $\ell(\ell_0, j)$ can take values

$$\ell(\ell_0, j) = \begin{cases} \ell_0 - 1, & \text{if } \ell_0 \in \{0, 1, \ldots, K+1\},\ j \in \mathcal{S}_- \\ \ell_0, & \text{if } \ell_0 = 0, j \in \mathcal{S}_+, \text{ or } \ell_0 = K, j \in \mathcal{S}_-, \\ \ell_0 + 1, & \text{if } \ell_0 \in \{-1, 0, 1, \ldots, K\},\ j \in \mathcal{S}_+. \end{cases}$$

Proof. The proof follows the same structure as the proof of Theorem 5.3 however, changes are required in all the results used in the proof, as here we do not need to integrate a function ψ. We give an outline of the proof only.

At a boundary we can model the fluid queue exactly, hence (5.116) holds for $\ell_0 = -1$ and $\ell_0 = K + 1$.

Now consider $i \in \mathcal{S}_+, j \in \mathcal{S}_+$. Partition the probability (5.116) on the times $\{\Sigma_n\}_{n\geq 1}$ and $\{\Gamma_n\}_{n\geq 1}$ and, specifically, partition on the event that there are exactly m events $\{\Sigma_n\}_{n=1}^m$ and exactly m events $\{\Gamma_n\}_{n=1}^m$. The resulting partitioned probabilities are

$$\begin{aligned}&\int_{x_1=0}^\infty \left(\boldsymbol{e}_i \boldsymbol{H}^{+-}(\lambda, x_1) \otimes \boldsymbol{a}^{(p)}_{\ell_0,i}(x_0) e^{\boldsymbol{S}^{(p)} x_1} \boldsymbol{D}^{(p)}\right) \mathrm{d}x_1 \\ &\left[\prod_{r=1}^{m-1} \int_{x_{2r}=0}^\infty \left(\boldsymbol{H}^{-+}(\lambda, x_{2r}) \otimes e^{\boldsymbol{S}^{(p)} x_{2r}} \boldsymbol{D}^{(p)}\right) \mathrm{d}x_{2r}\right.\end{aligned}$$

$$\int_{x_{2r+1}=0}^{\infty} \left(\boldsymbol{H}^{+-}(\lambda, x_{2r+1}) \otimes e^{\boldsymbol{S}^{(p)} x_{2r+1}} \boldsymbol{D}^{(p)} \right) \mathrm{d}x_{2r+1} \Bigg]$$
$$\int_{x_{2m}=0}^{\infty} \left(\boldsymbol{H}^{-+}(\lambda, x_{2m}) \otimes e^{\boldsymbol{S}^{(p)} x_{2m}} \boldsymbol{D}^{(p)} \right) \mathrm{d}x_{2m}$$
$$\int_{x_{2m+1}=0}^{\infty} \left(\boldsymbol{H}^{++}(\lambda, x_{2m+1}) \boldsymbol{e}'_j \otimes e^{\boldsymbol{S}^{(p)} x_{2m+1}} \boldsymbol{s}^{(p)} \right) \mathrm{d}x_{2m+1}. \tag{5.117}$$

To show that the terms (5.117) converge to

$$\mathbb{P}(\boldsymbol{X}(\tau_1^X) = (y_{\ell_0+1}, j), \tau_1^X \leq E^\lambda, \Sigma_m \leq \tau_1^X < \Gamma_{m+1}, \mid \boldsymbol{X}(0) = (x_0, i)) \tag{5.118}$$

we can use the bounds from Corollary 5.7 and Corollary 5.9. For $m = 0$ we recognise (5.117) as the same form as that appearing in Corollary 5.7 upon choosing $v = 0$. For $m \geq 1$, choose the closing operator to be $\boldsymbol{v}(x) = e^{\boldsymbol{S}x}\boldsymbol{s}$ and set $x = 0$ in Corollary 5.9. Now take the bound from Corollary 5.9 and extend it to the case of matrix functions in the same way we extended Corollary 5.10 to the matrix case in Lemma 5.11. In this way, we have a bound for (5.117) which tends to 0 as $p \to \infty$. Analogous arguments give the convergence for all terms $i \in \mathcal{S}_+ \cup \mathcal{S}_- \cup \mathcal{S}_0^*$, $j \in \mathcal{S}_+ \cup \mathcal{S}_{+0} \cup \mathcal{S}_- \cup \mathcal{S}_{-0}$ on a given number of up-down/down-up transitions.

What remains is a domination condition so that we may apply the Dominated Convergence Theorem to claim that the sum over the number of up-down and down-up transition converges (i.e. the sum over m in (5.117) converges). After algebraic manipulation, (5.117) is

$$\boldsymbol{e}_i \left[\prod_{r=1}^{m} \int_{x_{2r-1}=0}^{\infty} \boldsymbol{H}^{+-}(\lambda, x_{2r-1}) \int_{x_{2r}=0}^{\infty} \boldsymbol{H}^{-+}(\lambda, x_{2r}) \right] \int_{x_{2m+1}=0}^{\infty} \boldsymbol{H}^{++}(\lambda, x_{2m+1}) \boldsymbol{e}'_j$$
$$\boldsymbol{a}_{\ell_0,i}(x_0) \boldsymbol{N}^{2m}(\lambda, x_1, \ldots, x_{2m}) \boldsymbol{D} e^{\boldsymbol{S}x_{2m+1}} \boldsymbol{s} \, \mathrm{d}x_{2m+1} \ldots \mathrm{d}x_1 \tag{5.119}$$

Now, since $[\boldsymbol{H}^{++}(\lambda, x_{2m+1})]_{ij} \leq G$ and

$$\boldsymbol{e}_i \left[\prod_{r=1}^{m} \int_{x_{2r-1}=0}^{\infty} \boldsymbol{H}^{+-}(\lambda, x_{2r-1}) \int_{x_{2r}=0}^{\infty} \boldsymbol{H}^{-+}(\lambda, x_{2r}) \right]$$

is a row-vector of positive numbers, then (5.119) is less than or equal to

$$\boldsymbol{e}_i \left[\prod_{r=1}^{m} \int_{x_{2r-1}=0}^{\infty} \boldsymbol{H}^{+-}(\lambda, x_{2r-1}) \int_{x_{2r}=0}^{\infty} \boldsymbol{H}^{-+}(\lambda, x_{2r}) \right] \int_{x_{2m+1}=0}^{\infty} \boldsymbol{e}G$$
$$\boldsymbol{a}_{\ell_0,i}(x_0) \boldsymbol{N}^{2m}(\lambda, x_1, \ldots, x_{2m}) \boldsymbol{D} e^{\boldsymbol{S}x_{2m+1}} \boldsymbol{s} \, \mathrm{d}x_{2m+1} \ldots \mathrm{d}x_1.$$

Integrating with respect to x_{2m+1} gives

$$\begin{aligned}
&\boldsymbol{e}_i \left[\prod_{r=1}^{m} \int_{x_{2r-1}=0}^{\infty} \boldsymbol{H}^{+-}(\lambda, x_{2r-1}) \int_{x_{2r}=0}^{\infty} \boldsymbol{H}^{-+}(\lambda, x_{2r}) \right] \boldsymbol{e}G \\
&\quad \boldsymbol{a}_{\ell_0,i}(x_0) \boldsymbol{N}^{2m}(\lambda, x_1, \ldots, x_{2m}) \boldsymbol{D}\boldsymbol{e} \, \mathrm{d}x_{2m} \ldots \mathrm{d}x_1 \\
&\leq \boldsymbol{e}_i \left[\prod_{r=1}^{m} \int_{x_{2r-1}=0}^{\infty} \boldsymbol{H}^{+-}(\lambda, x_{2r-1}) \int_{x_{2r}=0}^{\infty} \boldsymbol{H}^{-+}(\lambda, x_{2r}) \right] \boldsymbol{e}G \, \mathrm{d}x_{2m} \ldots \mathrm{d}x_1 \qquad (5.120)
\end{aligned}$$

the last inequality holds since, $\boldsymbol{D}\boldsymbol{e} = \boldsymbol{e}$, and $\boldsymbol{a}_{\ell_0,i}(x_0)\boldsymbol{N}^{2m}(\lambda, x_1, \ldots, x_{2m})\boldsymbol{e} \leq 1$, as we claimed previously in the discussion after (5.103) in the proof of Lemma 5.12. Equation (5.120) is of a similar form to (5.104) (they differ only by a constant), hence the same arguments used to bound (5.104) can be applied to get the desired domination result.

Analogous arguments give a suitable geometric bound for $i, j \in \mathcal{S}_+ \cup \mathcal{S}_-$. For $i \in \mathcal{S}_0^*$, the left-hand side of (5.116) can be written as a linear combination of the left-hand side of (5.116) for initial phases in $\mathcal{S}_+ \cup \mathcal{S}_-$. Hence, a bound follows along similar arguments to those used to prove the case $i \in \mathcal{S}_0^*$ in Lemma 5.12.

Ultimately, we can apply the Dominated Convergence Theorem to prove that the sum of the partitioned probabilities (5.117) converges as $p \to \infty$. The sum of the limits is

$$\mathbb{P}(\boldsymbol{X}(\tau_1^X) = (y_{\ell_0+1}, j), \tau_1 \leq E^\lambda \mid \boldsymbol{X}(0) = (x_0, i)).$$

□

Chapter 6

Global convergence results

In this chapter we prove convergence results which build on the main results of Chapter 5. We consider the discrete-time embedded processes formed by observing the QBD-RAP at the orbit restart epochs and by observing the fluid queue at the hitting times of the level process at the on points $\{y_\ell\}$, which are the boundaries of the intervals $\{\mathcal{D}_\ell\}$. In Corollary 6.2 we prove that the transition probability of the embedded process of the QBD-RAP converge those of the embedded process of the fluid queue. In Corollary 6.3 we prove that the distribution of the sojourn time of the QBD-RAP in a given level converges to the distribution of the sojourn time of the fluid queue in the corresponding interval. In Theorem 6.12, we state global results on the weak convergence (in space and time) of the QBD-RAP approximation scheme to the fluid queue.

In this chapter we work with the augmented state space scheme to model phases with rates $c_i = 0$ as described in Section 4.3. The results of this chapter rely on results of Chapter 5 and therefore apply to the QBD-RAP scheme which uses ephemeral states to model a fluid queue which starts in a phase with rate 0, as described in Section 4.6. However, supplementing the results of Chapter 5 with the results from Appendix C and using the same arguments from this chapter with only slight modifications, then the results of this chapter can be extended to the augmented state space QBD-RAP scheme without the initial ephemeral phases.

6.1 Convergence of an embedded process

Consider the embedded process formed by observing the QBD-RAP at the orbit restart epochs. Let $\{\tau_n^{(p)}\}_{n\geq 0, n\in\mathbb{Z}}$, $\tau_0^{(p)} = 0$, and

$$\tau_n^{(p)} = \inf\left\{t \geq \tau_{n-1}^{(p)} \mid L^{(p)}(t) \neq L^{(p)}(\tau_{n-1}^{(p)}), \text{or } \{\boldsymbol{Y}(t)\} \text{ hits a boundary}\right\},$$

be orbit restart epochs. These are the (stopping) times at which the level process of the QBD-RAP, $\{L^{(p)}(t)\}$, changes, or the boundary is hit, or, if the process is at

the boundary, the process leaves the boundary. To simplify notation, we may drop the superscript p where it is not explicitly needed. Further, let $\{\boldsymbol{Y}_{\boldsymbol{\alpha}}^{(p)}(n)\}_{n\geq 0,n\in\mathbb{Z}} = \{(L^{(p)}(\tau_n^{(p)}),\varphi(\tau_n^{(p)}))\}_{n\geq 0,n\in\mathbb{Z}}$ be the level and phase of the discrete-time process embedded at the orbit restart epochs, $\{\tau_n^{(p)}\}_{n\geq 0}$. The subscript $\boldsymbol{\alpha}$ refers to the fact that $\boldsymbol{A}^{(p)}(\tau_n^{(p)}) = \boldsymbol{\alpha}^{(p)}$ for $n \geq 1$, unless the process is temporarily absorbed in a sticky boundary, in which case $\boldsymbol{A}^{(p)}(\tau_n^{(p)}) = 1$. The process $\{\boldsymbol{Y}_{\boldsymbol{\alpha}}^{(p)}(n)\}_{n\geq 0}$ is a discrete-time Markov chain, which is time-homogeneous for $n \geq 1$.

Let $\{\tau_n^X\}_{n\geq 0}$, be the sequence of (stopping) times with $\tau_0^X = 0$, and

$$\tau_{n+1}^X = \min\left\{\begin{array}{c}\inf\left\{t > \tau_n^X \mid X(t) = y_\ell, \ell \in \mathcal{K}\right\},\\ \inf\left\{t > \tau_n^X \mid X(t) \neq 0, X(\tau_{n-1}^X) = 0\right\},\\ \inf\left\{t > \tau_n^X \mid X(t) \neq y_{K+1}, X(\tau_{n-1}^X) = y_{K+1}\right\}\end{array}\right\},$$

for $n \geq 0$. For $n \geq 1$, τ_n^X is the time at which $X(t)$ either changes band, or hits a boundary, or leaves a boundary, for the nth time. The embedded process $\{\boldsymbol{X}(\tau_n)\}$ is a discrete-time Markov chain which is time-homogeneous for $n \geq 1$.

We have the following result on the convergence of the embedded processes $\{\boldsymbol{Y}_{\boldsymbol{\alpha}}^{(p)}(n)\}$ and $\{\boldsymbol{X}(\tau_n^X)\}$, which we also will utilise later to prove a global result.

Corollary 6.1. *For $\ell_0 \in \mathcal{K}$, $x_0 \in \mathcal{D}_{\ell_0,i}$, $i \in \mathcal{S}$, for $n = 0$, then*

$$\begin{aligned}&\mathbb{P}(\boldsymbol{Y}_{\boldsymbol{\alpha}}^{(p)}(1) = (\ell(\ell_0,j),j), \tau_1^{(p)} \leq E^\lambda \mid \boldsymbol{Y}^{(p)}(0) = (\ell_0, \boldsymbol{a}_{\ell_0,i}^{(p)}(x_0), i))\\ &\to \mathbb{P}(\boldsymbol{X}(\tau_1^X) = (y_{\ell(\ell_0,j)+1(j\in\mathcal{S}_-)}, j), \tau_1^X \leq E^\lambda \mid \boldsymbol{X}(0) = (x_0,i)),\end{aligned} \tag{6.1}$$

and for $n \geq 1$,

$$\begin{aligned}&\mathbb{P}(\boldsymbol{Y}_{\boldsymbol{\alpha}}^{(p)}(n+1) = (\ell(\ell_0,j),j), \tau_{n+1}^{(p)} \leq E^\lambda \mid \boldsymbol{Y}_{\boldsymbol{\alpha}}^{(p)}(n) = (\ell_0,i), \tau_n^{(p)} \leq E^\lambda)\\ &\to \mathbb{P}(\boldsymbol{X}(\tau_{n+1}^X) = (y_{\ell(\ell_0,j)+1(j\in\mathcal{S}_-)}, j), \tau_{n+1}^X \leq E^\lambda \mid \boldsymbol{X}(\tau_n^X) = (y_{\ell_0+1(i\in\mathcal{S}_-)}, i), \tau_n^X \leq E^\lambda).\end{aligned} \tag{6.2}$$

where $\ell(\ell_0,j)$ can take values

$$\ell(\ell_0,j) = \begin{cases}\ell_0 - 1, & \ell_0 \in \{0,1,\ldots,K+1\},\ j \in \mathcal{S}_-\\ \ell_0, & \ell_0 = 0, j \in \mathcal{S}_+,\ \text{or } \ell_0 = K, j \in \mathcal{S}_-,\\ \ell_0 + 1, & \ell_0 \in \{-1,0,1,\ldots,K\},\ j \in \mathcal{S}_+.\end{cases} \tag{6.3}$$

Proof. The convergence for $\ell_0 \in \{-1, K+1\}$ holds trivially as the QBD-RAP and fluid queue have the same behaviour at the sticky boundary.

The case for $n = 0$ is a direct result of Corollary 5.15.

All that is left is to prove (6.2). For $n \geq 1$, consider the transition probabilities of the embedded process from the QBD-RAP,

$$\mathbb{P}(\boldsymbol{Y}_{\boldsymbol{\alpha}}^{(p)}(n+1) = (\ell(\ell_0,j),j), \tau_{n+1}^{(p)} \leq E^\lambda \mid \boldsymbol{Y}_{\boldsymbol{\alpha}}^{(p)}(n) = (\ell_0,i), \tau_n^{(p)} \leq E^\lambda)$$

$$= \mathbb{P}(\boldsymbol{Y}_{\boldsymbol{\alpha}}^{(p)}(1) = (\ell(\ell_0, j), j), \tau_1^{(p)} \leq E^\lambda \mid \boldsymbol{Y}^{(p)}(0) = (\ell_0, \boldsymbol{\alpha}, i)), \tag{6.4}$$

since the QBD-RAP is time-homogeneous and the exponential random variable E^λ is memoryless. Applying Corollary 5.15 to (6.4) then

$$\begin{aligned} &\mathbb{P}(\boldsymbol{Y}_{\boldsymbol{\alpha}}^{(p)}(n+1) = (\ell(\ell_0, j), j), \tau_{n+1}^{(p)} \leq E^\lambda \mid \boldsymbol{Y}_{\boldsymbol{\alpha}}^{(p)}(n) = (\ell_0, i), \tau_n^{(p)} \leq E^\lambda) \\ &\to \mathbb{P}(\boldsymbol{X}(\tau_1^X) = (y_{\ell(\ell_0,j)+1(j\in\mathcal{S}_-)}, j), \tau_1^X \leq E^\lambda \mid \boldsymbol{X}(0) = (y_{\ell_0+1(i\in\mathcal{S}_-)}, i)). \end{aligned} \tag{6.5}$$

Since the fluid queue is time-homogeneous, and E^λ memoryless, then (6.5) is equal to

$$\begin{aligned} &\mathbb{P}(\boldsymbol{X}(\tau_1^X) = (y_{\ell(\ell_0,j)+1(j\in\mathcal{S}_-)}, j), \tau_1^X \leq E^\lambda \mid \boldsymbol{X}(\tau_1^X) = (y_{\ell_0+1(i\in\mathcal{S}_-)}, i)) \\ &= \mathbb{P}(\boldsymbol{X}(\tau_{n+1}^X) = (y_{\ell(\ell_0,j)+1(j\in\mathcal{S}_-)}, j), \tau_{n+1}^X \leq E^\lambda \mid \boldsymbol{X}(\tau_n^X) = (y_{\ell_0+1(i\in\mathcal{S}_-)}, i), \tau_n^X \leq E^\lambda), \end{aligned} \tag{6.6}$$

which proves the result. □

A direct corollary of Corollary 6.1 is the convergence of the transition probabilities of the embedded process.

Corollary 6.2. *For $\ell_0 \in \mathcal{K}$, $x_0 \in \mathcal{D}_{\ell_0,i}$, $i \in \mathcal{S}$, for $n = 0$, then*

$$\begin{aligned} &\mathbb{P}(\boldsymbol{Y}_{\boldsymbol{\alpha}}^{(p)}(1) = (\ell(\ell_0, j), j) \mid \boldsymbol{Y}^{(p)}(0) = (\ell_0, \boldsymbol{a}_{\ell_0,i}^{(p)}(x_0), i)) \\ &\to \mathbb{P}(\boldsymbol{X}(\tau_1^X) = (y_{\ell(\ell_0,j)+1(j\in\mathcal{S}_-)}, j) \mid \boldsymbol{X}(0) = (x_0, i)). \end{aligned} \tag{6.7}$$

and for $n \geq 1$,

$$\begin{aligned} &\mathbb{P}(\boldsymbol{Y}_{\boldsymbol{\alpha}}^{(p)}(n+1) = (\ell(\ell_0, j), j) \mid \boldsymbol{Y}_{\boldsymbol{\alpha}}^{(p)}(n) = (\ell_0, i)) \\ &\to \mathbb{P}(\boldsymbol{X}(\tau_{n+1}^X) = (y_{\ell(\ell_0,j)+1(j\in\mathcal{S}_-)}, j) \mid \boldsymbol{X}(\tau_n) = (y_{\ell_0+1(i\in\mathcal{S}_-)}, i)). \end{aligned} \tag{6.8}$$

Proof. Since $\tau_1^{(p)} < \infty$ almost surely, as is τ_1^X, then taking $\lambda \to 0$ in Corollary 6.1 yields the result. □

Corollary 6.2 states that the transition probabilities of the embedded processes converge. Thus, the finite-dimensional distributions of $\{\boldsymbol{Y}_{\boldsymbol{\alpha}}^{(p)}(n)\}$ converge, and if the space $\mathcal{K} \times \mathcal{S}$ is finite, then the sequence of distributions of $\{\boldsymbol{Y}_{\boldsymbol{\alpha}}^{(p)}(n)\}$ is tight. Thus, we can establish the convergence in distribution of $\{\boldsymbol{Y}_{\boldsymbol{\alpha}}^{(p)}(n)\}$ and $\{\boldsymbol{X}(\tau_n^X)\}$.

Another direct corollary of Corollary 5.15 is the convergence in distribution of the random variables $\{\tau_1^{(p)}\}_{p\geq 1}$ to τ_1^X.

Corollary 6.3. *The random variables $\{\tau_1^{(p)}\}_{p\geq 1}$ converge in distribution to τ_1^X.*

Proof. By Corollary 5.15 the probabilities

$$\begin{aligned}&\mathbb{P}(\boldsymbol{Y}_{\alpha}^{(p)}(1) = (\ell(\ell_0, j), j), \tau_1^{(p)} \leq E^{\lambda} \mid \boldsymbol{Y}^{(p)}(0) = \boldsymbol{y}_0)\\ &\to \mathbb{P}(\boldsymbol{X}(\tau_1^X) = (y_{\ell(\ell_0,j)+1(j\in\mathcal{S}_-)}, j), \tau_1^{(p)} \leq E^{\lambda} \mid \boldsymbol{X}(0) = (x_0, i)).\end{aligned}$$

By the law of total probability and the convergence above,

$$\begin{aligned}&\mathbb{P}(\tau_1^{(p)} \leq E^{\lambda} \mid \boldsymbol{Y}^{(p)}(0) = \boldsymbol{y}_0)\\ &= \sum_{\ell\in\{\ell_0-1,\ell_0,\ell_0+1\}\cap\mathcal{K}} \sum_{j\in\mathcal{S}} \mathbb{P}(\boldsymbol{Y}_{\alpha}^{(p)}(1) = (\ell, j), \tau_1^{(p)} \leq E^{\lambda} \mid \boldsymbol{Y}^{(p)}(0) = \boldsymbol{y}_0)\\ &\to \sum_{\ell\in\{\ell_0,\ell_0+1\}\cap\{0,1,\ldots,K+1\}} \sum_{j\in\mathcal{S}} \mathbb{P}(\boldsymbol{X}(\tau_1^X) = (y_\ell, j), \tau_1^{(p)} \leq E^{\lambda} \mid \boldsymbol{X}(0) = (x_0, i)).\\ &= \mathbb{P}(\tau^X \leq E^{\lambda} \mid \boldsymbol{X}(0) = (x_0, i)). \qquad (6.9)\end{aligned}$$

Thus, the Laplace transform of $\tau_1^{(p)}$ converges to the Laplace transform of τ_1^X. By the Continuity Theorem for Laplace transforms (Feller 1957, Chapter XIII, Theorem 2a), then $\{\tau_1^{(p)}\}$ converges in distribution to τ_1^X. □

6.2 Convergence of the QBD-RAP scheme

This section is dedicated to proving a global convergence results of the QBD-RAP approximation scheme to the fluid queue. Ultimately, we prove the weak convergence (in space and time) of the QBD-RAP approximation scheme to the distribution of the fluid queue. This is done in stages. First, by considering the Laplace transform with respect to time of the distribution of the QBD-RAP *at the time of the nth orbit restart epoch*, τ_n. Second, considering the Laplace transform with respect to time of the distribution of the QBD-RAP *between* the nth and $(n+1)$th orbit restart epoch. Third, summing over the number of orbit restart epochs, n, and via the Dominated Convergence Theorem, we claim a convergence of the Laplace transforms with respect to time of the QBD-RAP and fluid queue. Lastly, we apply the Extended Continuity Theorem for Laplace transforms (Feller 1957, Chapter XIII, Theorem 2a) to claim a weak convergence (in space and time) of the QBD-RAP approximation scheme to the fluid queue.

6.2.1 At the nth orbit restart epoch

The idea for the analysis of the QBD-RAP at the nth orbit restart epoch is to partition the distribution of the QBD-RAP on the phase and level at the times $\tau_1^{(p)}$ and $\tau_2^{(p)}, ..., \tau_{n-1}^{(p)}$ and then use the strong Markov property to write the expression in terms of the distributions of the QBD-RAP at time τ_k^p given the distribution of the QBD-RAP at time $\tau_{k-1}^{(p)}$, $k = 1, ..., n$.

The distribution at time $\tau_1^{(p)}$ depends on the initial condition, so we treat it separately and use Corollary 5.15 to claim convergence to the fluid queue. By time-homogeneity, the distributions of the QBD-RAP at times τ_k^p given the position of the QBD-RAP at time $\tau_{k-1}^{(p)}$ is equal to the distribution of the QBD-RAP at time $\tau_1^{(p)}$ given an initial distribution of the QBD-RAP. Hence, we may also use Corollary 5.15 to claim convergence to the fluid queue for each of these distributions. Putting all the elements together yields the convergence at the nth orbit restart epoch.

Formally, for $n \geq 1$, consider the Laplace transform

$$\begin{aligned}&\int_{t=0}^{\infty} e^{-\lambda t}\mathbb{P}(\boldsymbol{Y}_{\boldsymbol{\alpha}}^{(p)}(n) = (\ell, j_n), \tau_n^{(p)} \in \mathrm{d}t \mid \boldsymbol{Y}^{(p)}(0) = \boldsymbol{y}_0^{(p)})\,\mathrm{d}t \\ &= \mathbb{P}(\boldsymbol{Y}_{\boldsymbol{\alpha}}^{(p)}(n) = (\ell, j_n), \tau_n^{(p)} \leq E^{\lambda} \mid \boldsymbol{Y}^{(p)}(0) = \boldsymbol{y}_0^{(p)}),\end{aligned} \tag{6.10}$$

which is the Laplace transform of the time until the nth orbit restart epoch of the QBD-RAP on the event that the level and phase at the nth orbit restart epoch are ℓ and j_n, respectively, given that the initial level and phases are ℓ_0 and i, respectively, and the initial orbit is $\boldsymbol{a}_{\ell_0,i}^{(p)}(x_0)$. Partitioning on the time of the first orbit restart epoch, τ_1, and the level and phase at this time, then (6.10) is equal to

$$\begin{aligned}&\sum_{j_1 \in \mathcal{S}} \sum_{\ell_1 \in \{\ell_0+1, \ell_0, \ell_0-1\} \cap \mathcal{K}} \mathbb{P}(\boldsymbol{Y}_{\boldsymbol{\alpha}}^{(p)}(n) = (\ell, j_n), \tau_n^{(p)} \leq E^{\lambda} \mid \boldsymbol{Y}_{\boldsymbol{\alpha}}^{(p)}(1) = (\ell_1, j_1), \tau_1^{(p)} \leq E^{\lambda}) \\ &\times \mathbb{P}(\boldsymbol{Y}_{\boldsymbol{\alpha}}^{(p)}(1) = (\ell_1, j_1), \tau_1^{(p)} \leq E^{\lambda} \mid \boldsymbol{Y}(0) = \boldsymbol{y}_0^{(p)}).\end{aligned} \tag{6.11}$$

An application of Corollary 5.15 to the expression on the second line of (6.11) states,

$$\begin{aligned}&\lim_{p\to\infty} \mathbb{P}(\boldsymbol{Y}_{\boldsymbol{\alpha}}^{(p)}(1) = (\ell_1, j_1), \tau_1^{(p)} \leq E^{\lambda} \mid \boldsymbol{Y}^{(p)}(0) = \boldsymbol{y}_0^{(p)}) \\ &\to \mathbb{P}(\boldsymbol{X}(\tau_1^X) = (y_{\ell+1(j_1 \in \mathcal{S}_-)}, j_1), \tau_1^X \leq E^{\lambda} \mid \boldsymbol{X}(0) = (x_0, i))\end{aligned} \tag{6.12}$$

for $i \in \mathcal{S}$, $j_1 \in \mathcal{S}_+ \cup \mathcal{S}_-$, $\ell_0 \in \mathcal{K}$, $x_0 \in \mathcal{D}_{\ell_0}$.

We now turn our attention to the first factor in the summands of (6.11). For a given $j_1 \in \mathcal{S}_+ \cup \mathcal{S}_-$ and $\ell_1 \in \mathcal{K}$, consider

$$\begin{aligned}&\mathbb{P}(\boldsymbol{Y}_{\boldsymbol{\alpha}}^{(p)}(n) = (\ell, j_n), \tau_n^{(p)} \leq E^{\lambda} \mid \boldsymbol{Y}_{\boldsymbol{\alpha}}^{(p)}(1) = (\ell_1, j_1), \tau_1^{(p)} \leq E^{\lambda}) \\ &= \mathbb{P}(\boldsymbol{Y}_{\boldsymbol{\alpha}}^{(p)}(n-1) = (\ell, j_n), \tau_{n-1}^{(p)} \leq E^{\lambda} \mid \boldsymbol{Y}(0) = (\ell_1, \boldsymbol{\alpha}, j_1))\end{aligned} \tag{6.13}$$

by the time-homogeneous property of the QBD-RAP and the memoryless property of the exponential distribution.

Let

$$\mathcal{P}_{\ell_0,\ell_n}^n = \left\{ (\ell_1, \ldots, \ell_{n-1}) \in \mathcal{K}^{n-1} \,\middle|\, \begin{array}{c} \ell_{r-1} = \ell_r = 0 \text{ or } \ell_{r-1} = \ell_r = K \\ \text{or } |\ell_{r-1} - \ell_r| = 1, r = 1, \ldots, n \end{array} \right\}. \tag{6.14}$$

The set $\mathcal{P}^n_{\ell_0,\ell}$ contains all the possible values which $\{L(\tau_m)\}_{m=2}^{n-1}$ may take on a sample path which starts in level ℓ_0, ends in level ℓ and has n orbit restart epochs.

Partitioning on the phases and the levels at the times τ_m, $m = 2, \ldots, n-1$, and using the strong Markov property of the QBD-RAP, then (6.13) is

$$\sum_{\substack{j_2,\ldots,j_{n-1}\in\mathcal{S}\\(\ell_2,\ldots,\ell_{n-1})\in\mathcal{P}^{n-1}_{\ell_1,\ell}}} \prod_{m=2}^{n} \mathbb{P}(\boldsymbol{Y}^{(p)}_{\boldsymbol{\alpha}}(m) = (\ell_m, j_m), \tau^{(p)}_m \le E^\lambda \mid \boldsymbol{Y}^{(p)}_{\boldsymbol{\alpha}}(m-1) = (\ell_{m-1}, j_{m-1}),$$
$$\tau^{(p)}_{m-1} \le E^\lambda)$$
$$= \sum_{\substack{j_2,\ldots,j_{n-1}\in\mathcal{S}\\(\ell_2,\ldots,\ell_{n-1})\in\mathcal{P}^{n-1}_{\ell_1,\ell}}} \prod_{m=2}^{n} \mathbb{P}(\boldsymbol{Y}^{(p)}_{\boldsymbol{\alpha}}(1) = (\ell_m, j_m), \tau^{(p)}_1 \le E^\lambda \mid \boldsymbol{Y}^{(p)}(0) = (\ell_{m-1}, \boldsymbol{\alpha}, j_{m-1})), \quad (6.15)$$

by the time-homogeneity property of the QBD-RAP and the memoryless property of E^λ and where we define $\ell_n = \ell$ for notational convenience. We can apply Corollary 5.15 to the factors of the summands in (6.15) and conclude

$$\mathbb{P}(\boldsymbol{Y}^{(p)}_{\boldsymbol{\alpha}}(1) = (\ell_m, j_m), \tau^{(p)}_1 \le E^\lambda \mid \boldsymbol{Y}^{(p)}(0) = (\ell_{m-1}, \boldsymbol{\alpha}, j_{m-1}))$$
$$\to \mathbb{P}(\boldsymbol{X}(\tau^X_1) = (y_{\ell_m+1(j_m\in\mathcal{S}_-)}, j_m), \tau^X_1 \le E^\lambda \mid \boldsymbol{X}(0) = (y_{\ell_{m-1}+1(j_{m-1}\in\mathcal{S}_-)}, j_{m-1})). \quad (6.16)$$

By the time-homogeneous property of the fluid queue and the memoryless property of the exponential distribution, (6.16) is equal to

$$\mathbb{P}(\boldsymbol{X}(\tau^X_m) = (y_{\ell_m+1(j_m\in\mathcal{S}_-)}, j_m), \tau^X_m \le E^\lambda \mid \boldsymbol{X}(\tau_{m-1}) = (y_{\ell_{m-1}+1(j_{m-1}\in\mathcal{S}_-)}, j_{m-1}),$$
$$\tau^X_{m-1} \le E^\lambda). \quad (6.17)$$

Thus, returning to (6.15), taking the limit as $p \to \infty$, and using the convergence established in (6.12) and (6.17), then

$$\lim_{p\to\infty} \sum_{\substack{j_2,\ldots,j_{n-1}\in\mathcal{S}\\(\ell_2,\ldots,\ell_{n-1})\in\mathcal{P}^{n-1}_{\ell_1,\ell}}} \prod_{m=2}^{n} \mathbb{P}(\boldsymbol{Y}^{(p)}_{\boldsymbol{\alpha}}(1) = (\ell_m, j_m), \tau^{(p)}_1 \le E^\lambda \mid \boldsymbol{Y}^{(p)}(0) = (\ell_{m-1}, \boldsymbol{\alpha}, j_{m-1}))$$
$$= \sum_{\substack{j_2,\ldots,j_{n-1}\in\mathcal{S}\\(\ell_2,\ldots,\ell_{n-1})\in\mathcal{P}^{n-1}_{\ell_1,\ell}}} \prod_{m=2}^{n} \lim_{p\to\infty} \mathbb{P}(\boldsymbol{Y}^{(p)}_{\boldsymbol{\alpha}}(1) = (\ell_m, j_m), \tau^{(p)}_1 \le E^\lambda \mid \boldsymbol{Y}^{(p)}(0) = (\ell_{m-1}, \boldsymbol{\alpha}, j_{m-1})), \quad (6.18)$$

where we may swap the limit and the sums as they are finite, and we can swap the limit and the product since all the limits exist and the product is finite. Substituting the limits

(6.17) into (6.18) gives

$$\sum_{\substack{j_2,\ldots,j_{n-1}\in\mathcal{S}\\(\ell_2,\ldots,\ell_{n-1})\in\mathcal{P}^{n-1}_{\ell_1,\ell}}}\prod_{m=2}^{n}\mathbb{P}(\boldsymbol{X}(\tau_m^X)=(y_{\ell_m+1(j_m\in\mathcal{S}_-)},j_m),\tau_m^X\le E^\lambda\mid$$
$$\boldsymbol{X}(\tau_{m-1}^X)=(y_{\ell_{m-1}+1(j_{m-1}\in\mathcal{S}_-)},j_{m-1}),\tau_{m-1}^X\le E^\lambda),$$
$$=\mathbb{P}(\boldsymbol{X}(\tau_n^X)=(y_{\ell+1(j_n\in\mathcal{S}_-)},j_n),\tau_n^X\le E^\lambda\mid\boldsymbol{X}(0)=(y_{\ell_1+1(j_1\in\mathcal{S}_-)},j_1),\tau_1^X\le E^\lambda),\quad(6.19)$$

by the strong Markov property of the fluid queue and the Law of total probability.

Returning now to (6.11) and taking the limit as $p\to\infty$,

$$\lim_{p\to\infty}\sum_{j_1\in\mathcal{S}}\sum_{\ell_1\in\{\ell_0+1,\ell_0,\ell_0-1\}\cap\mathcal{K}}\mathbb{P}(\boldsymbol{Y}_{\boldsymbol{\alpha}}^{(p)}(n)=(\ell,j_n),\tau_n^{(p)}\le E^\lambda\mid\boldsymbol{Y}_{\boldsymbol{\alpha}}^{(p)}(1)=(\ell_1,j_1),\tau_1^{(p)}\le E^\lambda)$$
$$\mathbb{P}(\boldsymbol{Y}_{\boldsymbol{\alpha}}^{(p)}(1)=(\ell_1,j_1),\tau_1^{(p)}\le E^\lambda\mid\boldsymbol{Y}^{(p)}(0)=\boldsymbol{y}_0^{(p)})$$
$$=\sum_{j_1\in\mathcal{S}}\sum_{\ell_1\in\{\ell_0+1,\ell_0-1\}\cap\mathcal{K}}\lim_{p\to\infty}\mathbb{P}(\boldsymbol{Y}_{\boldsymbol{\alpha}}^{(p)}(n)=(\ell,j_n),\tau_n^{(p)}\le E^\lambda\mid\boldsymbol{Y}_{\boldsymbol{\alpha}}^{(p)}(1)=(\ell_1,j_1),\tau_1^{(p)}\le E^\lambda)$$
$$\lim_{p\to\infty}\mathbb{P}(\boldsymbol{Y}_{\boldsymbol{\alpha}}^{(p)}(1)=(\ell_1,j_1),\tau_1^{(p)}\le E^\lambda\mid\boldsymbol{Y}^{(p)}(0)=\boldsymbol{y}_0^{(p)})$$
$$=\sum_{j_1\in\mathcal{S}}\sum_{\ell_1\in\{\ell_0+1,\ell_0-1\}\cap\mathcal{K}}\mathbb{P}(\boldsymbol{X}(\tau_n^X)=(y_{\ell+1(j_n\in\mathcal{S}_-)},j_n),\tau_n^X\le E^\lambda\mid\boldsymbol{X}(\tau_1^X)=(y_{\ell_1+1(j_1\in\mathcal{S}_-)},j_1),$$
$$\tau_1^X\le E^\lambda)\mathbb{P}(\boldsymbol{X}(\tau_1^X)=(y_{\ell_1+1(j_1\in\mathcal{S}_-)},j_1),\tau_1^X\le E^\lambda\mid\boldsymbol{X}(0)=(x_0,i))$$
$$=\mathbb{P}(\boldsymbol{X}(\tau_n^X)=(y_{\ell+1(j_n\in\mathcal{S}_-)},j_n),\tau_n^X\le E^\lambda\mid\boldsymbol{X}(0)=(x_0,i))\quad(6.20)$$

where the swapping of limits and sums in the first equality is justified as the sums are finite, the swapping limits and products in the first equality is justified as the product is finite and all limits exist, and the last inequality is the Law of total probability.

Hence, we have proved the following result

Lemma 6.4. *For all $\ell,\ell_0\in\mathcal{K}$, $i,j_n\in\mathcal{S}$, $x_0\in\mathcal{D}_{\ell_0,i}$, $n\ge1$, then, as $p\to\infty$,*

$$\mathbb{P}(\boldsymbol{Y}_{\boldsymbol{\alpha}}^{(p)}(n)=(\ell,j_n),\tau_n^{(p)}\le E^\lambda\mid\boldsymbol{Y}^{(p)}(0)=\boldsymbol{y}_0^{(p)}),$$
$$\to\mathbb{P}(\boldsymbol{X}(\tau_n^X)=(y_{\ell+1(j_n\in\mathcal{S}_-)},j_n),\tau_n^X\le E^\lambda\mid\boldsymbol{X}(0)=(x_0,i)).\quad(6.21)$$

Lemma 6.4 is what is required to proceed with the proof of convergence of the QBD-RAP scheme. However, at this point we also point out the following corollary, which may be of interest in other contexts.

Corollary 6.5. *The random variables $\{\tau_n^{(p)}\}_{p\ge1}$ converge in distribution to τ_n^X.*

Proof. The proof follows the same arguments as the proof of Corollary 6.3. In Lemma 6.4 sum over ℓ and j_n, then apply the Extended Continuity Theorem (Feller 1957, Chapter XIII, Theorem 2a). □

6.2.2 Between the nth and $(n+1)$th orbit restart epochs

In the last section we proved a convergence of the QBD-RAP to the fluid queue *at* the nth orbit restart epoch. However, we ultimately want to make a convergence statement about the QBD-RAP scheme at any time. The next step is to show a convergence of the QBD-RAP to the fluid queue between the nth and $(n+1)$th orbit restart epoch, after which we sum over n to prove a convergence result independent of the number of orbit restart epochs.

The idea is to partition the distribution of the QBD-RAP between time $\tau_n^{(p)}$ and τ_{n+1}^{p} on the position of the QBD-RAP at time $\tau_n^{(p)}$, then use the strong Markov property to write the distribution of the QBD-RAP between time $\tau_n^{(p)}$ and τ_{n+1}^{p} in terms of the distribution of the QBD-RAP at time $\tau_n^{(p)}$ given the initial position of the QBD-RAP, and the distribution of the QBD-RAP at between time $\tau_n^{(p)}$ and $\tau_{n+1}^{(p)}$ given the position of the QBD-RAP at time $\tau_n^{(p)}$. In the previous section we showed convergence of the distribution of the QBD-RAP at time $\tau_n^{(p)}$ given the initial position of the QBD-RAP. By a time-homogeneity argument, the distribution of the QBD-RAP between time $\tau_n^{(p)}$ and $\tau_{n+1}^{(p)}$ given the position of the QBD-RAP at time $\tau_n^{(p)}$ is equal to the distribution of the QBD-RAP before time $\tau_1^{(p)}$ given an initial condition on the QBD-RAP. Thus, Theorem 5.3 establishes the convergence of this term. Together, we have convergence of the QBD-RAP between the nth and $(n+1)$th orbit restart epoch.

More formally, let $\mathcal{T}_n^{(p)} = (\tau_n^{(p)}, \tau_{n+1}^{(p)}]$ and $\mathcal{T}_n^{X} = (\tau_n^{X}, \tau_{n+1}^{X}]$. Consider the Laplace transform

$$\int_{t=0}^{\infty} e^{-\lambda t} \int_{x \in \mathcal{D}_{\ell,j}} \mathbb{P}(\boldsymbol{Y}^{(p)}(t) = (\ell, \mathrm{d}x, j), t \in \mathcal{T}_n^{(p)} \mid \boldsymbol{Y}^{(p)}(0) = \boldsymbol{y}_0^{(p)}) \psi(x) \, \mathrm{d}t, \tag{6.22}$$

where $\psi : \mathbb{R} \to \mathbb{R}$ is a bounded function.

Partitioning (6.22) on the time of the nth orbit restart epoch and the phase and level at this time gives

$$\begin{aligned}
&\int_{t=0}^{\infty} e^{-\lambda t} \int_{x \in \mathcal{D}_{\ell,j}} \int_{u_n=0}^{t} \sum_{j_n \in \mathcal{S}} \mathbb{P}(\boldsymbol{Y}^{(p)}(t) = (\ell, \mathrm{d}x, j), t \in (u_n, \tau_{n+1}^{(p)}] \mid \boldsymbol{Y}_{\boldsymbol{\alpha}}^{(p)}(n) = (\ell, j_n), \\
&\quad \tau_n^{(p)} = u_n) \psi(x) \mathbb{P}(\boldsymbol{Y}_{\boldsymbol{\alpha}}^{(p)}(n) = (\ell, j_n), \tau_n^{(p)} \in \mathrm{d}u_n \mid \boldsymbol{Y}^{(p)}(0) = \boldsymbol{y}_0^{(p)}) \, \mathrm{d}t \\
&= \sum_{j_n \in \mathcal{S}} \int_{x \in \mathcal{D}_{\ell,j}} \int_{t=0}^{\infty} e^{-\lambda t} \mathbb{P}(\boldsymbol{Y}^{(p)}(t) = (\ell, \mathrm{d}x, j), t \in \mathcal{T}_0^{(p)} \mid \boldsymbol{Y}(0) = (\ell, \boldsymbol{\alpha}, j_n)) \, \mathrm{d}t \psi(x) \\
&\quad \mathbb{P}(\boldsymbol{Y}_{\boldsymbol{\alpha}}^{(p)}(n) = (\ell, j_n), \tau_n^{(p)} \leq E^{\lambda} \mid \boldsymbol{Y}^{(p)}(0) = \boldsymbol{y}_0^{(p)})
\end{aligned} \tag{6.23}$$

by the time homogenous property of the QBD-RAP, the memoryless property of the exponential distribution, and the convolution theorem of Laplace transforms. The swap of integrals and sums is justified by the Fubini-Tonelli Theorem. We recognise the probability

$$\mathbb{P}(\boldsymbol{Y}_{\boldsymbol{\alpha}}^{(p)}(n) = (\ell, j_n), \tau_n^{(p)} \leq E^{\lambda} \mid \boldsymbol{Y}^{(p)}(0) = \boldsymbol{y}_0^{(p)}) \tag{6.24}$$

as that appearing in Lemma 6.4, hence (6.24) converges to

$$\mathbb{P}(\boldsymbol{X}(\tau_n^X) = (y_{\ell+1(j_n\in\mathcal{S}_-)}, j_n), \tau_n^X \leq E^\lambda \mid \boldsymbol{X}(0) = (x_0, i)) \tag{6.25}$$

as $p \to \infty$.

Now consider the expression

$$\int_{x\in\mathcal{D}_{\ell,j}} \int_{t=0}^{\infty} e^{-\lambda t}\mathbb{P}(\boldsymbol{Y}^{(p)}(t) = (\ell, \mathrm{d}x, j), t \in \mathcal{T}_0^{(p)} \mid \boldsymbol{Y}(0) = (\ell, \boldsymbol{\alpha}, j_n))\,\mathrm{d}t\psi(x) \tag{6.26}$$

which appears as part of (6.23). We can rewrite (6.26) as

$$\begin{aligned}&\int_{x\in\mathcal{D}_{\ell,j}} \int_{t=0}^{\infty} e^{-\lambda t}\mathbb{P}(\boldsymbol{Y}^{(p)}(t) = (\ell, \mathrm{d}x, j), t \in \mathcal{T}_0^{(p)} \mid \boldsymbol{Y}_{\boldsymbol{\alpha}}^{(p)}(0) = (\ell, j_n))\,\mathrm{d}t\psi(x)\\ &= \int_{x\in\mathcal{D}_{\ell,j}} \widehat{f}^{\ell,(p)}(\lambda)(x, j; y_{\ell+1(j_n\in\mathcal{S}_-)}, j_n)\psi(x)\,\mathrm{d}x\end{aligned} \tag{6.27}$$

Applying Theorem 5.3, then (6.27) converges to

$$\begin{aligned}&\int_{x\in\mathcal{D}_{\ell,j}} \widehat{\mu}^{\ell}(\lambda)(\,\mathrm{d}x, j; y_{\ell+1(j_n\in\mathcal{S}_-)}, j_n)\psi(x)\\ &= \int_{x\in\mathcal{D}_{\ell,j}} \int_{t=0}^{\infty} e^{-\lambda t}\mathbb{P}(\boldsymbol{X}(t) \in (\,\mathrm{d}x, j), t \in \mathcal{T}_0^X \mid \boldsymbol{X}(0) = (y_{\ell+1(j_n\in\mathcal{S}_-)}, j_n))\psi(x).\end{aligned} \tag{6.28}$$

Since the fluid queue is time-homogeneous and E^λ memoryless, then (6.28) is equal to

$$\int_{x\in\mathcal{D}_{\ell,j}} \int_{t=0}^{\infty} e^{-\lambda t}\mathbb{P}(\boldsymbol{X}(t) \in (\,\mathrm{d}x, j), t \in \mathcal{T}_n^X \mid \boldsymbol{X}(\tau_n^X) = (y_{\ell+1(j_n\in\mathcal{S}_-)}, j_n), \tau_n^X \leq E^\lambda)\psi(x). \tag{6.29}$$

Therefore, we have shown (6.26) converges to (6.29) as $p \to \infty$.

Returning to the right-hand side of (6.23), taking the limit as $p \to \infty$, and recalling the limits in (6.25) and (6.29), then

$$\begin{aligned}&\lim_{p\to\infty} \sum_{j_n\in\mathcal{S}} \int_{x\in\mathcal{D}_{\ell,j}} \int_{t=0}^{\infty} e^{-\lambda t}\mathbb{P}(\boldsymbol{Y}^{(p)}(t) = (\ell, \mathrm{d}x, j), t \in \mathcal{T}_0^{(p)} \mid \boldsymbol{Y}(0) = (\ell, \boldsymbol{\alpha}, j_n))\,\mathrm{d}t\psi(x)\\ &\qquad \mathbb{P}(\boldsymbol{Y}_{\boldsymbol{\alpha}}^{(p)}(n) = (\ell, j_n), \tau_n^{(p)} \leq E^\lambda \mid \boldsymbol{Y}^{(p)}(0) = \boldsymbol{y}_0^{(p)})\\ &= \sum_{j_n\in\mathcal{S}} \lim_{p\to\infty} \int_{x\in\mathcal{D}_{\ell,j}} \int_{t=0}^{\infty} e^{-\lambda t}\mathbb{P}(\boldsymbol{Y}^{(p)}(t) = (\ell, \mathrm{d}x, j), t \in \mathcal{T}_0^{(p)} \mid \boldsymbol{Y}(0) = (\ell, \boldsymbol{\alpha}, j_n))\,\mathrm{d}t\psi(x)\\ &\qquad \lim_{p\to\infty} \mathbb{P}(\boldsymbol{Y}_{\boldsymbol{\alpha}}^{(p)}(n) = (\ell, j_n), \tau_n^{(p)} \leq E^\lambda \mid \boldsymbol{Y}^{(p)}(0) = \boldsymbol{y}_0^{(p)})\end{aligned}$$

$$= \sum_{j_n \in \mathcal{S}} \int_{x \in \mathcal{D}_{\ell,j}} \int_{t=0}^{\infty} e^{-\lambda t} \mathbb{P}(\boldsymbol{X}(t) \in (\,\mathrm{d}x, j), t \in \mathcal{T}_n^X \mid \boldsymbol{X}(\tau_n^X) = (y_{\ell+1(j_n \in \mathcal{S}_-)}, j_n), \tau_n^X \leq E^\lambda)$$
$$\psi(x) \mathbb{P}(\boldsymbol{X}(\tau_n^X) = (y_{\ell+1(j_n \in \mathcal{S}_-)}, j_n), \tau_n^X \leq E^\lambda \mid \boldsymbol{X}(0) = (x_0, i))$$
$$= \int_{t=0}^{\infty} e^{-\lambda t} \int_{x \in \mathcal{D}_{\ell,j}} \mathbb{P}(\boldsymbol{X}(t) \in (\,\mathrm{d}x, j), t \in \mathcal{T}_n^X \mid X(0) = x_0, \varphi(0) = i) \psi(x)\,\mathrm{d}t, \tag{6.30}$$

where the first equality holds since the sum is finite and the limits exist, and the last equality holds from the law of total probability.

Hence, we have shown the following result

Lemma 6.6. *For $\ell, \ell_0 \in \mathcal{K}$, $i, j \in \mathcal{S}$, $n \geq 0$, then, as $p \to \infty$,*

$$\int_{t=0}^{\infty} e^{-\lambda t} \int_{x \in \mathcal{D}_{\ell,j}} \mathbb{P}(\boldsymbol{Y}^{(p)}(t) = (\ell,\, \mathrm{d}x, j), t \in \mathcal{T}_n^{(p)} \mid \boldsymbol{Y}(0) = \boldsymbol{y}_0) \psi(x)\,\mathrm{d}t$$
$$\to \int_{t=0}^{\infty} e^{-\lambda t} \int_{x \in \mathcal{D}_{\ell,j_n}} \mathbb{P}(\boldsymbol{X}(t) \in (\,\mathrm{d}x, j), t \in \mathcal{T}_n^X \mid \boldsymbol{X}(0) = (x_0, i)) \psi(x)\,\mathrm{d}t. \tag{6.31}$$

6.2.3 Intermezzo: A domination condition

Our aim is to prove convergence of

$$\int_{t=0}^{\infty} e^{-\lambda t} \int_{x \in \mathcal{D}_{\ell,j}} \mathbb{P}(\boldsymbol{Y}^{(p)}(t) = (\ell,\, \mathrm{d}x, j) \mid \boldsymbol{Y}(0) = \boldsymbol{y}_0) \psi(x)\,\mathrm{d}t$$
$$= \int_{t=0}^{\infty} e^{-\lambda t} \int_{x \in \mathcal{D}_{\ell,j}} \sum_{n=0}^{\infty} \mathbb{P}(\boldsymbol{Y}^{(p)}(t) = (\ell,\, \mathrm{d}x, j), t \in \mathcal{T}_n^{(p)} \mid \boldsymbol{Y}(0) = \boldsymbol{y}_0) \psi(x)\,\mathrm{d}t$$
$$= \sum_{n=0}^{\infty} \int_{t=0}^{\infty} e^{-\lambda t} \int_{x \in \mathcal{D}_{\ell,j}} \mathbb{P}(\boldsymbol{Y}^{(p)}(t) = (\ell,\, \mathrm{d}x, j), t \in \mathcal{T}_n^{(p)} \mid \boldsymbol{Y}(0) = \boldsymbol{y}_0) \psi(x)\,\mathrm{d}t \tag{6.32}$$

where the swap of the intergals and sums is justified by the Fubini-Tonelli Theorem. Now, (6.32) is an infinite sum of terms appearing in Lemma 6.6. Hence, upon taking the limit of (6.32), if we can justify the swap of the sum and the limit, we will obtain the desired result. To this end, we show a domination condition in Corollary 6.9 so that we may apply the Dominated Convergence Theorem. First, we need a couple of intermediate results.

Lemma 6.7. *For all $i \in \mathcal{S}_+ \cup \mathcal{S}_-$, and $n \geq 2$,*

$$\mathbb{P}(\tau_n \leq E^\lambda \mid \phi(\tau_{n-1}) = i, \tau_{n-1} \leq E^\lambda) \leq b, \tag{6.33}$$

where

$$b = \max\left\{1 - e^{-\gamma(\Delta+\varepsilon)}\left[1 - e^{\gamma\varepsilon - \lambda\Delta/|c_{min}|}\right] + \frac{\mathrm{Var}(Z)}{\varepsilon^2} + |r_2|, \frac{\gamma}{\gamma+\lambda}\right\},$$

$$|r_2| \leq 2G\frac{\mathrm{Var}\,(Z)}{\varepsilon^2} + 2L\varepsilon,$$

and $\gamma = \max_{i\in\mathcal{S}_+\cup\mathcal{S}_-} -T_{ii}/|c_i|$.

Note that b and r_1 depend on p which has been suppressed to simplify notation. When explicitly needed, we use a superscript p to denote this dependence.

Proof. For the QBD-RAP, orbit restart epochs can only occur when $i \in \mathcal{S}_+ \cup \mathcal{S}_-$.

Suppose that the phase at time τ_{n-1} is $i \in \mathcal{S}_+$ and that at time τ_{n-1} the QBD-RAP is not at a boundary. The arguments for an initial phase $i \in \mathcal{S}_-$ are analogous. For an orbit restart epochs to occur, either;

1. the QBD-RAP remains in phase i until the orbit restart epoch occurs, or
2. the QBD-RAP changes phase before there is an orbit restart epoch, after which the orbit restart epoch occurs eventually.

Hence, for sample paths which contribute to the Laplace transform, one of two things must happen, either;

1. the phase remains i until there is an orbit restart epoch and E^λ does not occur before the orbit restart epoch, or,
2. the phase changes before there is an orbit restart epoch and E^λ does not occur before the orbit restart epoch.

The probability of 1 is

$$\int_{x=0}^{\infty} \boldsymbol{\alpha} e^{\boldsymbol{S}x} \boldsymbol{s} e^{(T_{ii}-\lambda)x/|c_i|}\,\mathrm{d}x = e^{(T_{ii}-\lambda)\Delta/|c_i|} + r_2, \tag{6.34}$$

by Lemma 5.6.

The probability of 2 is

$$\begin{aligned}\int_{x=0}^{\infty} \boldsymbol{\alpha} e^{\boldsymbol{S}x} \boldsymbol{e} e^{(T_{ii}-\lambda)x/|c_i|}(-T_{ii}/|c_i|)\,\mathrm{d}x = &\int_{x=0}^{\Delta+\varepsilon} \boldsymbol{\alpha} e^{\boldsymbol{S}x} \boldsymbol{e} e^{(T_{ii}-\lambda)x/|c_i|}(-T_{ii}/|c_i|)\,\mathrm{d}x \\ &+ \int_{x=\Delta+\varepsilon}^{\infty} \boldsymbol{\alpha} e^{\boldsymbol{S}x} \boldsymbol{e} e^{(T_{ii}-\lambda)x/|c_i|}(-T_{ii}/|c_i|)\,\mathrm{d}x. \end{aligned}\tag{6.35}$$

Now, since $\boldsymbol{\alpha} e^{\boldsymbol{S}x}\boldsymbol{e} \leq 1$ for $x \leq \Delta + \varepsilon$ then the first term on the right-hand side of (6.35) is less than or equal to

$$\int_{x=0}^{\Delta+\varepsilon} e^{(T_{ii}-\lambda)/|c_i|x}(-T_{ii}/|c_i|)\,\mathrm{d}x \leq \int_{x=0}^{\Delta+\varepsilon} e^{T_{ii}/|c_i|x}(-T_{ii}/|c_i|)\,\mathrm{d}x = 1 - e^{T_{ii}/|c_i|(\Delta+\varepsilon)}.$$

By Chebyshev's inequality, $\boldsymbol{\alpha} e^{\boldsymbol{S}x}\boldsymbol{e} \leq \dfrac{\mathrm{Var}(Z)}{\varepsilon^2}$ for $x > \Delta + \varepsilon$, hence the second term on the right-hand side of (6.35) is less than or equal to

$$\int_{x=\Delta+\varepsilon}^{\infty} \frac{\mathrm{Var}(Z)}{\varepsilon^2} e^{(T_{ii}-\lambda)x/|c_i|}(-T_{ii}/|c_i|)\,\mathrm{d}x \leq \frac{\mathrm{Var}(Z)}{\varepsilon^2}.$$

Putting these together, then the right-hand side of (6.35) is less than or equal to

$$1 - e^{T_{ii}(\Delta+\varepsilon)/|c_i|} + \frac{\mathrm{Var}(Z)}{\varepsilon^2}. \tag{6.36}$$

Combining (6.34) and (6.36), then $\mathbb{P}(\tau_n \leq E^\lambda \mid \phi(\tau_{n-1}) = i, \tau_{n-1} \leq E^\lambda)$ is less than or equal to

$$\begin{aligned}
&e^{(T_{ii}-\lambda)\Delta/|c_i|} + |r_2| + 1 - e^{T_{ii}(\Delta+\varepsilon)/|c_i|} + \frac{\mathrm{Var}(Z)}{\varepsilon^2} \\
&= 1 - e^{T_{ii}(\Delta+\varepsilon)/|c_i|}\left[1 - e^{(-T_{ii}\varepsilon-\lambda\Delta)/|c_i|}\right] + \frac{\mathrm{Var}(Z)}{\varepsilon^2} + |r_2| \\
&= 1 - e^{-\gamma(\Delta+\varepsilon)}\left[1 - e^{\gamma\varepsilon-\lambda\Delta/|c_{min}|}\right] + \frac{\mathrm{Var}(Z)}{\varepsilon^2} + |r_2|,
\end{aligned} \tag{6.37}$$

since $-T_{ii}/|c_i| \leq \gamma$ and $\lambda\Delta/|c_i| \leq \lambda\Delta/c_{min}$ for all $i \in \mathcal{S}_+ \cup \mathcal{S}_-$.

Now consider the QBD-RAP at a boundary. To leave the boundary there must be at-least one change of phase before E^λ. By a uniformisation argument, the probability of at-least one change of phase before E^λ is less than or equal to $\gamma/(\gamma+\lambda)$. □

Lemma 6.8. *For $n \geq 2$, $i \in \mathcal{S}_+ \cup \mathcal{S}_-$,*

$$\mathbb{P}(\tau_n \leq E^\lambda \mid \phi(\tau_1) = i, \tau_1 \leq E^\lambda) \leq b^{n-1}. \tag{6.38}$$

Proof. The proof is by induction.

For the base case, set $n = 2$ and apply Lemma 6.7.

Now, assume the induction hypothesis $\mathbb{P}(\tau_{n-1} \leq E^\lambda \mid \phi(\tau_1) = i, \tau_1 \leq E^\lambda) \leq b^{n-2}$ for arbitrary $n \geq 3$.

Since $\{\tau_{n-1} \leq E^\lambda\}$ is a subset of $\{\tau_n \leq E^\lambda\}$, then

$$\mathbb{P}(\tau_n \leq E^\lambda \mid \phi(\tau_1) = i, \tau_1 \leq E^\lambda) = \mathbb{P}(\tau_n \leq E^\lambda, \tau_{n-1} \leq E^\lambda \mid \phi(\tau_1) = i, \tau_1 \leq E^\lambda). \tag{6.39}$$

Now partition (6.39) on the phase and level at time τ_{n-1},

$$\sum_{\ell_{n-1}\in\mathcal{K}} \sum_{j_{n-1}\in\mathcal{S}} \mathbb{P}(\tau_n \leq E^\lambda, \tau_{n-1} \leq E^\lambda, L(\tau_{n-1}) = \ell_{n-1}, \phi(\tau_{n-1}) = j_{n-1} \mid \phi(\tau_1) = i, \tau_1 \leq E^\lambda)$$

$$= \sum_{\ell_{n-1}\in\mathcal{K}} \sum_{j_{n-1}\in\mathcal{S}} \mathbb{P}(\tau_n \le E^\lambda \mid L(\tau_{n-1}) = \ell_{n-1}, \phi(\tau_{n-1}) = j_{n-1}, \tau_{n-1} \le E^\lambda)$$
$$\times \mathbb{P}(L(\tau_{n-1}) = \ell_{n-1}, \phi(\tau_{n-1}) = j_{n-1}, \tau_{n-1} \le E^\lambda \mid \phi(\tau_1) = i, \tau_1 \le E^\lambda), \tag{6.40}$$

by the strong Markov property of the QBD-RAP and the fact that $\boldsymbol{A}(\tau_{n-1}) = \boldsymbol{\alpha}$.

By Lemma 6.7 (6.40) is less than or equal to

$$\sum_{\ell_{n-1}\in\mathcal{K}} \sum_{j_{n-1}\in\mathcal{S}} b\mathbb{P}(L(\tau_{n-1}) = \ell_{n-1}\phi(\tau_{n-1}) = j_{n-1}, \tau_{n-1} \le E^\lambda \mid \phi(\tau_1) = i, \tau_1 \le E^\lambda)$$
$$= b\mathbb{P}(\tau_{n-1} \le E^\lambda \mid \phi(\tau_1) = i, \tau_1 \le E^\lambda)$$
$$\le b \cdot b^{n-2}, \tag{6.41}$$

by the induction hypothesis, and this completes the proof. □

Corollary 6.9. *For $\boldsymbol{y}_0 = (\ell_0, \boldsymbol{a}_{\ell_0,i}(x_0), i)$, $\lambda > 0$, $\ell_0, \ell \in \mathcal{K}$, $i, j \in \mathcal{S}$ and any bounded function $\psi \le F$, then*

$$\left| \int_{t=0}^{\infty} e^{-\lambda t} \int_{x\in\mathcal{D}_{\ell,j}} \mathbb{P}(\boldsymbol{Y}(t) \in (\ell, \mathrm{d}x, j), t \in \mathcal{T}_n \mid \boldsymbol{Y}(0) = \boldsymbol{y}_0)\psi(x)\,\mathrm{d}t \right| \le \frac{Fb^{n-1}}{\lambda}. \tag{6.42}$$

Proof. First, since $|\psi(x)| \le F$, then the left-hand side of (6.42) is less than or equal to

$$\int_{t=0}^{\infty} e^{-\lambda t} \int_{x\in\mathcal{D}_{\ell,j}} \mathbb{P}(\boldsymbol{Y}(t) \in (\ell, \mathrm{d}x, j), t \in \mathcal{T}_n \mid \boldsymbol{Y}(0) = \boldsymbol{y}_0)F\,\mathrm{d}t$$
$$= \int_{t=0}^{\infty} e^{-\lambda t}\mathbb{P}(\boldsymbol{Y}(t) \in (\ell, \mathcal{D}_{\ell,j}, j), t \in \mathcal{T}_n \mid \boldsymbol{Y}(0) = \boldsymbol{y}_0)F\,\mathrm{d}t. \tag{6.43}$$

Partitioning on the time of the first orbit restart epoch, τ_1, and the phase and level at time τ_1, and using the strong Markov property, then (6.43) is equal to

$$\int_{t=0}^{\infty} e^{-\lambda t} \int_{u_1=0}^{t} \sum_{\substack{j_1\in\mathcal{S} \\ \ell_1\in\{\ell_0+1,\ell_0,\ell_0-1\}\cap\mathcal{K}}} \mathbb{P}(\boldsymbol{Y}(t) \in (\ell, \mathcal{D}_{\ell,j}, j), t \in \mathcal{T}_n \mid \boldsymbol{Y}_{\boldsymbol{\alpha}}(1) = (\ell_1, j_1), \tau_1 = u_1)$$
$$\mathbb{P}(\boldsymbol{Y}_{\boldsymbol{\alpha}}(1) = (\ell_1, j_1), \tau_1 \in \mathrm{d}u_1 \mid \boldsymbol{Y}(0) = \boldsymbol{y}_0)F\,\mathrm{d}t$$
$$= \int_{t=0}^{\infty} e^{-\lambda t} \int_{u_1=0}^{t} \sum_{\substack{j_1\in\mathcal{S} \\ \ell_1\in\{\ell_0+1,\ell_0,\ell_0-1\}\cap\mathcal{K}}} \mathbb{P}(\boldsymbol{Y}(t) \in (\ell, \mathcal{D}_{\ell,j}, j), t - u_1 \in \mathcal{T}_{n-1} \mid \boldsymbol{Y}(0) = (\ell_1, \boldsymbol{\alpha}, j_1))$$
$$\mathbb{P}(\boldsymbol{Y}_{\boldsymbol{\alpha}}(1) = (\ell_1, j_1), \tau_1 \in \mathrm{d}u_1 \mid \boldsymbol{Y}(0) = \boldsymbol{y}_0)F\,\mathrm{d}t, \tag{6.44}$$

by the time-homogeneous property of the QBD-RAP. By the convolution theorem for Laplace transforms, (6.44) is equal to

$$\sum_{\substack{j_1\in\mathcal{S} \\ \ell_1\in\{\ell_0+1,\ell_0,\ell_0-1\}\cap\mathcal{K}}} \int_{t=0}^{\infty} e^{-\lambda t}\mathbb{P}(\boldsymbol{Y}(t) \in (\ell, \mathcal{D}_{\ell,j}, j), t \in \mathcal{T}_{n-1} \mid \boldsymbol{Y}(0) = (\ell_1, \boldsymbol{\alpha}, j_1))\,\mathrm{d}t$$

$$\times \int_{u_1=0}^{\infty} e^{-\lambda u_1} \mathbb{P}(\boldsymbol{Y}_{\boldsymbol{\alpha}}(1) = (\ell_1, j_1), \tau_1 \in \mathrm{d}u_1 \mid \boldsymbol{Y}(0) = \boldsymbol{y}_0) F$$

$$= \sum_{\substack{j_1 \in \mathcal{S} \\ \ell_1 \in \{\ell_0+1, \ell_0, \ell_0-1\} \cap \mathcal{K}}} \int_{t=0}^{\infty} e^{-\lambda t} \mathbb{P}(\boldsymbol{Y}(t) \in (\ell, \mathcal{D}_{\ell,j}, j), t \in \mathcal{T}_{n-1} \mid \boldsymbol{Y}(0) = (\ell_1, \boldsymbol{\alpha}, j_1)) \, \mathrm{d}t$$

$$\times \mathbb{P}(\boldsymbol{Y}_{\boldsymbol{\alpha}}(1) = (\ell_1, j_1), \tau_1 \leq E^{\lambda} \mid \boldsymbol{Y}(0) = \boldsymbol{y}_0) F. \tag{6.45}$$

Since $\{\boldsymbol{Y}(t) \in (\ell, \mathcal{D}_{\ell,j}, j), t \in \mathcal{T}_n\} \subseteq \{t \in \mathcal{T}_n\}$, then the expression

$$\begin{aligned}
&\int_{t=0}^{\infty} e^{-\lambda t} \mathbb{P}(\boldsymbol{Y}(t) \in (\ell, \mathcal{D}_{\ell,j}, j), t \in \mathcal{T}_n \mid \boldsymbol{Y}(0) = (\ell_1, \boldsymbol{\alpha}, j_1)) \, \mathrm{d}t \\
&\leq \int_{t=0}^{\infty} e^{-\lambda t} \mathbb{P}(\tau_n \leq t \mid \boldsymbol{Y}(0) = (\ell_1, \boldsymbol{\alpha}, j_1)) \, \mathrm{d}t \\
&\leq b^{n-1} \int_{t=0}^{\infty} e^{-\lambda t} \, \mathrm{d}t \\
&= b^{n-1} \frac{1}{\lambda},
\end{aligned} \tag{6.46}$$

where the second inequality holds by Lemma 6.8.

Using the bound (6.46) in (6.45) gives

$$\sum_{j_1 \in \mathcal{S}} \sum_{\ell_1 \in \{\ell_0+1, \ell_0-1\} \cap \mathcal{K}} b^{n-1} \frac{1}{\lambda} \mathbb{P}(\boldsymbol{Y}_{\boldsymbol{\alpha}}(1) = (\ell_1, j_1), \tau_1 \leq E^{\lambda} \mid \boldsymbol{Y}(0) = \boldsymbol{y}_0) F \leq b^{n-1} \frac{1}{\lambda} F, \tag{6.47}$$

by the law of total probability. This concludes the proof. □

6.2.4 Global convergence

Finally, we combine the convergence result of Lemma 6.6 and the domination condition from Corollary 6.9 via the Dominated Convergence Theorem to claim convergence of the Laplace transform of the QBD-RAP given by

$$\int_{t=0}^{\infty} e^{-\lambda t} \int_{x \in \mathcal{D}_{\ell,j}} \mathbb{P}(\boldsymbol{Y}^{(p)}(t) = (\ell, \, \mathrm{d}x, j) \mid \boldsymbol{Y}^{(p)}(0) = \boldsymbol{y}_0^{(p)}) \psi(x) \, \mathrm{d}t. \tag{6.48}$$

We start by partitioning (6.48) on the number of orbit restart epochs by time t to get

$$\begin{aligned}
&\int_{t=0}^{\infty} e^{-\lambda t} \int_{x \in \mathcal{D}_{\ell,j}} \sum_{n=0}^{\infty} \mathbb{P}(\boldsymbol{Y}^{(p)}(t) \in (\ell, \, \mathrm{d}x, j), t \in \mathcal{T}_n^{(p)} \mid \boldsymbol{Y}^{(p)}(0) = \boldsymbol{y}_0^{(p)}) \, \mathrm{d}t \psi(x) \\
&= \sum_{n=0}^{\infty} \int_{t=0}^{\infty} e^{-\lambda t} \int_{x \in \mathcal{D}_{\ell,j}} \mathbb{P}(\boldsymbol{Y}^{(p)}(t) \in (\ell, \, \mathrm{d}x, j), t \in \mathcal{T}_n^{(p)} \mid \boldsymbol{Y}^{(p)}(0) = \boldsymbol{y}_0^{(p)}) \, \mathrm{d}t \psi(x).
\end{aligned} \tag{6.49}$$

We can justify the swap of the sum and integrals since

$$\begin{aligned}
&\int_{t=0}^{\infty} e^{-\lambda t} \int_{x\in\mathcal{D}_{\ell,j}} \sum_{n=0}^{\infty} \mathbb{P}(\boldsymbol{Y}^{(p)}(t) \in (\ell,\, \mathrm{d}x, j), t \in \mathcal{T}_n^{(p)} \mid \boldsymbol{Y}^{(p)}(0) = \boldsymbol{y}_0^{(p)})\, \mathrm{d}t |\psi(x)| \\
&\leq \int_{t=0}^{\infty} e^{-\lambda t} \int_{x\in\mathcal{D}_{\ell,j}} \sum_{n=0}^{\infty} \mathbb{P}(\boldsymbol{Y}^{(p)}(t) \in (\ell,\, \mathrm{d}x, j), t \in \mathcal{T}_n^{(p)} \mid \boldsymbol{Y}^{(p)}(0) = \boldsymbol{y}_0^{(p)})\, \mathrm{d}t F \\
&\leq \int_{t=0}^{\infty} e^{-\lambda t}\, \mathrm{d}t F \\
&\leq \frac{1}{\lambda} F < \infty
\end{aligned}$$

so the Fubini-Tonelli Theorem applies.

By Lemma 6.6, each term in the sum (6.49) converges. Furthermore, for $n \geq 1$, each term is dominated by $\left(b^{(p)}\right)^{n-1} F/\lambda$, from Corollary 6.9. The dominating terms $\left(b^{(p)}\right)^{n-1} F/\lambda$ depend on p and may not be summable. However, for p sufficiently large, there exists a $p_0 < \infty$ and a B with $B < 1$ such that $b^{(p)} < B$ for all $p > p_0$. Hence, we can apply the Dominated Convergence Theorem to (6.49) and claim that

$$\begin{aligned}
&\lim_{p\to\infty} \sum_{n=0}^{\infty} \int_{x\in\mathcal{D}_{\ell,j}} \int_{t=0}^{\infty} e^{-\lambda t} \mathbb{P}(\boldsymbol{Y}^{(p)}(t) \in (\ell,\, \mathrm{d}x, j), t \in \mathcal{T}_n^{(p)} \mid \boldsymbol{Y}^{(p)}(0) = \boldsymbol{y}_0^{(p)})\, \mathrm{d}t \psi(x) \\
&= \sum_{n=0}^{\infty} \int_{t=0}^{\infty} e^{-\lambda t} \int_{x\in\mathcal{D}_{\ell,j}} \mathbb{P}(\boldsymbol{X}(t) \in (\,\mathrm{d}x, j), t \in \mathcal{T}_n^X \mid \boldsymbol{X}(0) = (x_0, i)) \psi(x)\, \mathrm{d}t. \qquad (6.50)
\end{aligned}$$

where Lemma 6.6 gives the limits of the terms in the sum. Swapping the sum and integrals and by the law of total probability, then (6.50) is equal to

$$\begin{aligned}
&\int_{t=0}^{\infty} e^{-\lambda t} \int_{x\in\mathcal{D}_{\ell,j}} \sum_{n=0}^{\infty} \mathbb{P}(\boldsymbol{X}(t) \in (\,\mathrm{d}x, j), t \in \mathcal{T}_n^X \mid \boldsymbol{X}(0) = (x_0, i)) \psi(x)\, \mathrm{d}t \\
&= \int_{t=0}^{\infty} e^{-\lambda t} \int_{x\in\mathcal{D}_{\ell,j}} \mathbb{P}(\boldsymbol{X}(t) \in (\,\mathrm{d}x, j) \mid \boldsymbol{X}(0) = (x_0, i)) \psi(x)\, \mathrm{d}t.
\end{aligned}$$

The swap of the sum and integrals is justified as

$$\begin{aligned}
&\int_{t=0}^{\infty} e^{-\lambda t} \int_{x\in\mathcal{D}_{\ell,j}} \sum_{n=0}^{\infty} \mathbb{P}(\boldsymbol{X}(t) \in (\,\mathrm{d}x, j), t \in \mathcal{T}_n^X \mid \boldsymbol{X}(0) = (x_0, i)) |\psi(x)|\, \mathrm{d}t \\
&\leq \int_{t=0}^{\infty} e^{-\lambda t} \int_{x\in\mathcal{D}_{\ell,j}} \sum_{n=0}^{\infty} \mathbb{P}(\boldsymbol{X}(t) \in (\,\mathrm{d}x, j), t \in \mathcal{T}_n^X \mid \boldsymbol{X}(0) = (x_0, i))\, \mathrm{d}t F \\
&\leq \int_{t=0}^{\infty} e^{-\lambda t}\, \mathrm{d}t F
\end{aligned}$$

$$= \frac{1}{\lambda} F < \infty \tag{6.51}$$

so the Fubini-Tonelli Theorem applies.

Thus, we have shown the following result.

Lemma 6.10. *For all $\ell_0, \ell \in \mathcal{K}$, $i, j \in \mathcal{S}$, $x_0 \in \mathcal{D}_{\ell_0,i}$, as $p \to \infty$,*

$$\begin{aligned}
&\int_{t=0}^{\infty} e^{-\lambda t} \int_{x \in \mathcal{D}_{\ell,j}} \mathbb{P}(\boldsymbol{Y}^{(p)}(t) = (\ell, \, \mathrm{d}x, j) \mid \boldsymbol{Y}^{(p)}(0) = \boldsymbol{y}_0^{(p)}) \psi(x) \, \mathrm{d}t \\
&\to \int_{t=0}^{\infty} e^{-\lambda t} \int_{x \in \mathcal{D}_{\ell,j}} \mathbb{P}(\boldsymbol{X}(t) \in (\, \mathrm{d}x, j) \mid \boldsymbol{X}(0) = (x_0, i)) \psi(x) \, \mathrm{d}t.
\end{aligned}$$

Lemma 6.10 establishes a convergence for a given interval $\mathcal{D}_\ell$, $\ell \in \mathcal{K}$, and phase $j \in \mathcal{S}$. We now formally extend this to a global result. To do so, we find it convenient to re-write the problem in terms of expectations.

Let $\mathcal{R}(L(t), \boldsymbol{A}(t), \varphi(t))$ be the random variable with density function $\boldsymbol{A}(t)\boldsymbol{v}_{L(t),\varphi(t)}(x)$, $x \in \mathcal{D}_{L(t),\varphi(t)}$, then

$$\begin{aligned}
&\int_{t=0}^{\infty} e^{-\lambda t} \mathbb{E}\left[\psi(\mathcal{R}(L(t), \boldsymbol{A}(t), \varphi(t)), \varphi(t)) \mid \boldsymbol{Y}(0) = \boldsymbol{y}_0 \right] \mathrm{d}t \\
&= \int_{t=0}^{\infty} e^{-\lambda t} \sum_{\substack{\ell \in \mathcal{K} \\ j \in \mathcal{S}}} \int_{\boldsymbol{a} \in \mathcal{A}} \int_{x \in \mathcal{D}_{\ell,j}} \mathbb{P}(\boldsymbol{Y}(t) \in (\ell, \, \mathrm{d}\boldsymbol{a}, j) \mid \boldsymbol{Y}(0) = \boldsymbol{y}_0) \boldsymbol{a} \boldsymbol{v}_{\ell,j}(x) \psi(x) \, \mathrm{d}x \, \mathrm{d}t \\
&= \sum_{\substack{\ell \in \mathcal{K} \\ j \in \mathcal{S}}} \int_{t=0}^{\infty} e^{-\lambda t} \int_{x \in \mathcal{D}_{\ell,j}} \mathbb{P}(\boldsymbol{Y}(t) \in (\ell, \, \mathrm{d}x, j) \mid \boldsymbol{Y}(0) = \boldsymbol{y}_0) \psi(x) \, \mathrm{d}t,
\end{aligned} \tag{6.52}$$

and the terms in the last line are those in Lemma 6.10.

Corollary 6.11. *Let $\psi : \mathbb{R} \times \mathcal{S} \to \mathbb{R}$ be an arbitrary, bounded function with $|\psi(\cdot, \cdot)| \leq F$. For each $i \in \mathcal{S}$, $\ell_0 \in \mathcal{K}$, $x_0 \in \mathcal{D}_{\ell_0,i}$,*

$$\begin{aligned}
&\int_{t=0}^{\infty} e^{-\lambda t} \mathbb{E}\left[\psi(\mathcal{R}(L^{(p)}(t), \boldsymbol{A}^{(p)}(t), \varphi(t)), \varphi(t)) \mid \boldsymbol{Y}^{(p)}(0) = \boldsymbol{y}_0^{(p)} \right] \mathrm{d}t \\
&\to \int_{t=0}^{\infty} e^{-\lambda t} \mathbb{E}\left[\psi(\boldsymbol{X}(t)) \mid \boldsymbol{X}(0) = (x_0, i) \right] \mathrm{d}t.
\end{aligned}$$

Proof. Consider the left-hand side

$$\int_{t=0}^{\infty} e^{-\lambda t} \mathbb{E}\left[\psi(\mathcal{R}(L^{(p)}(t), \boldsymbol{A}^{(p)}(t), \varphi(t)), \varphi(t)) \mid \boldsymbol{Y}^{(p)}(0) = \boldsymbol{y}_0^{(p)} \right] \mathrm{d}t$$

$$= \int_{t=0}^{\infty} e^{-\lambda t} \sum_{\ell \in \mathcal{K}} \sum_{j \in \mathcal{S}} \mathbb{E}\left[\psi(\mathcal{R}(\ell, \boldsymbol{A}^{(p)}(t), j), j) 1(L^{(p)}(t) = \ell, \varphi(t) = j)) \mid \boldsymbol{Y}^{(p)}(0) = \boldsymbol{y}_0^{(p)} \right] \mathrm{d}t$$

$$= \sum_{\ell \in \mathcal{K}} \sum_{j \in \mathcal{S}} \int_{t=0}^{\infty} e^{-\lambda t} \mathbb{E}\left[\psi(\mathcal{R}(\ell, \boldsymbol{A}^{(p)}(t), j), j) 1(L^{(p)}(t) = \ell, \varphi(t) = j)) \mid \boldsymbol{Y}^{(p)}(0) = \boldsymbol{y}_0^{(p)} \right] \mathrm{d}t, \tag{6.53}$$

where the swap of the summations and integrals is on the last line justified since ψ is bounded and by the Fubini-Tonelli Theorem. By Lemma 6.10, for each $\ell \in \mathcal{K}$, $j \in \mathcal{S}$, the terms

$$\int_{t=0}^{\infty} e^{-\lambda t} \mathbb{E}\left[\psi(\mathcal{R}(\ell, \boldsymbol{A}^{(p)}(t), j), j) \, \mathrm{d}x 1(L^{(p)}(t) = \ell, \varphi(t) = j)) \mid \boldsymbol{Y}^{(p)}(0) = \boldsymbol{y}_0^{(p)} \right] \mathrm{d}t$$

converge to

$$\int_{t=0}^{\infty} e^{-\lambda t} \mathbb{E}\left[\psi(\boldsymbol{X}(t)) 1(\boldsymbol{X}(t) \in (\mathcal{D}_{\ell_j}, j)) \mid \boldsymbol{X}(0) = (x_0, i) \right] \mathrm{d}t.$$

If $\mathcal{K}$ is finite, we are done upon taking the limit of (6.53) as $p \to \infty$ and swapping the limit and the sums.

If $\mathcal{K}$ is countably infinite, then for a given $k \in \mathcal{K}$, since ψ is bounded,

$$\left| \sum_{j \in \mathcal{S}} \int_{t=0}^{\infty} e^{-\lambda t} \mathbb{E}\left[\psi(\mathcal{R}(\ell, \boldsymbol{A}^{(p)}(t), j), \varphi(t)) 1(L^{(p)}(t) = \ell, \varphi(t) = j)) \mid \boldsymbol{Y}^{(p)}(0) = \boldsymbol{y}_0^{(p)} \right] \mathrm{d}t \right|$$

$$\leq F \sum_{j \in \mathcal{S}} \int_{t=0}^{\infty} e^{-\lambda t} \mathbb{E}\left[1(L^{(p)}(t) = \ell, \varphi(t) = j)) \mid \boldsymbol{Y}^{(p)}(0) = \boldsymbol{y}_0^{(p)} \right] \mathrm{d}t$$

$$= F \int_{t=0}^{\infty} e^{-\lambda t} \mathbb{P}(\boldsymbol{Y}^{(p)}(t) \in (\ell, \mathcal{D}_{\ell,j}, j) \mid \boldsymbol{Y}^{(p)}(0) = \boldsymbol{y}_0^{(p)}) \, \mathrm{d}t$$

$$\leq F \mathbb{P}(\tau_{|\ell - \ell_0|}^{(p)} \leq E^{\lambda} \mid \boldsymbol{Y}^{(p)}(0) = \boldsymbol{y}_0^{(p)}), \tag{6.54}$$

since, to be in level ℓ after starting in level ℓ_0, there must be at least $|\ell_0 - \ell|$ orbit restart epochs. By Lemma 6.8 then (6.54) is bounded by $\left(b^{(p)}\right)^{|\ell - \ell_0| - 1}$ for $|\ell - \ell_0| \geq 2$ and by 1 otherwise. Now, choose p_0 sufficiently large so that $b^{(p)} < B < 1$ for all $p > p_0$. Therefore, for all $p > p_0$, the terms in (6.53) are dominated by $F \min\{B^{|\ell - \ell_0| - 1}, 1\}$. Moreover,

$$F \sum_{\ell \in \mathcal{K}} \min\{B^{|\ell - \ell_0| - 1}, 1\} \leq 2 \sum_{n=1}^{\infty} B^{n-1} + 1$$

$$= \frac{2}{1 - B} + 1$$

$$< \infty,$$

hence the dominating terms are summable. Hence, we may apply the Dominated Convergence Theorem to swap the necessary limits and sums, from which the result follows. □

The Extended Continuity Theorem for Laplace transforms (Feller 1957, Chapter XIII, Theorem 2a) can now be used to claim that the QBD-RAP approximation scheme converges weakly (in space and time) to the fluid queue.

Theorem 6.12. *For all* $\boldsymbol{y}_0^{(p)} = (\ell, \boldsymbol{a}_{\ell_0,i}^{(p)}(x_0), i)$, $\ell_0 \in \mathcal{K}$, $x_0 \in \mathcal{D}_{\ell_0,i}$, $i \in \mathcal{S}$, *and any bounded function* $\psi : \mathbb{R} \times \mathcal{S} \to \mathbb{R}$, *then*

$$\mathbb{E}\left[\psi(\mathcal{R}(L^{(p)}(t), \boldsymbol{A}^{(p)}(t), \varphi(t)), \varphi(t)) \mid \boldsymbol{Y}^{(p)}(0) = \boldsymbol{y}_0^{(p)}\right] \to \mathbb{E}\left[\psi(\boldsymbol{X}(t)) \mid \boldsymbol{X}(0) = (x_0, i)\right]$$

weakly in t *as* $p \to \infty$.

Proof. Combine the Extended Continuity Theorem for Laplace transforms (Feller 1957, Chapter XIII, Theorem 2a) with the convergence of Laplace transforms in Corollary 6.11. □

Ideally, we would like to obtain a point-wise convergence result in the variable t. However, to date, this has eluded me. Although

$$\mathbb{E}\left[\psi(\boldsymbol{X}(t)) \mid \boldsymbol{X}(0) = (x_0, i)\right]$$

is a continuous function of t (it is a Feller semigroup), and, for $p < \infty$,

$$\mathbb{E}\left[\psi(\mathcal{R}(L^{(p)}(t), \boldsymbol{A}^{(p)}(t), \varphi(t)), \varphi(t)) \mid \boldsymbol{Y}^{(p)}(0) = \boldsymbol{y}_0^{(p)}\right]$$

is continuous in t, the limit

$$\lim_{p\to\infty} \mathbb{E}\left[\psi(\mathcal{R}(L^{(p)}(t), \boldsymbol{A}^{(p)}(t), \varphi(t)), \varphi(t)) \mid \boldsymbol{Y}^{(p)}(0) = \boldsymbol{y}_0^{(p)}\right]$$

need not be continuous in t. Nonetheless, we can claim that

$$\mathbb{E}\left[\psi(\mathcal{R}(L^{(p)}(t), \boldsymbol{A}^{(p)}(t), \varphi(t)), \varphi(t)) \mid \boldsymbol{Y}^{(p)}(0) = \boldsymbol{y}_0^{(p)}\right] \to \mathbb{E}\left[\psi(\boldsymbol{X}(t)) \mid \boldsymbol{X}(0) = (x_0, i)\right]$$

for almost all $t \geq 0$. At such values of t, since ψ is arbitrary and bounded, then The Portmanteau Theorem (Theorem 13.16 of Klenke (2014)) states that the QBD-RAP approximation scheme converges in distribution to the fluid queue.

A sufficient condition to upgrade the convergence from weak to point-wise (in the variable t) is to show that for $t \geq 0$

$$\sup_p \mathbb{E}\left[\psi(\mathcal{R}(L^{(p)}(t), \boldsymbol{A}^{(p)}(t), \varphi(t)), \varphi(t)) \mid \boldsymbol{Y}^{(p)}(0) = \boldsymbol{y}_0^{(p)}\right] \leq M(t) < \infty$$

and the sequence $\mathbb{E}\left[\psi(\mathcal{R}(L^{(p)}(t), \boldsymbol{A}^{(p)}(t), \varphi(t)), \varphi(t)) \mid \boldsymbol{Y}^{(p)}(0) = \boldsymbol{y}_0^{(p)}\right]$ is eventually equicontinuous in t. That is, for every $\varepsilon > 0$ there exists a $\delta(t,\varepsilon) > 0$ and a $p_0(t,\varepsilon)$ such that $|t-u| < \delta(t,\varepsilon)$ implies that

$$\Bigg| \mathbb{E}\left[\psi(\mathcal{R}(L^{(p)}(t), \boldsymbol{A}^{(p)}(t), \varphi(t)), \varphi(t)) \mid \boldsymbol{Y}^{(p)}(0) = \boldsymbol{y}_0^{(p)}\right] - \mathbb{E}\left[\psi(\mathcal{R}(L^{(p)}(u), \boldsymbol{A}^{(p)}(u), \varphi(u)), \varphi(u)) \mid \boldsymbol{Y}^{(p)}(0) = \boldsymbol{y}_0^{(p)}\right] \Bigg| < \varepsilon$$

for all $p \geq p_0(t,\varepsilon)$.

6.3 Extension to arbitrary (but fixed) discretisation structures

To conclude this chapter we include some remarks on how to extend the convergence results to arbitrary discretisation structures.

Throughout, we have assumed that all intervals are of width Δ, i.e. $|y_{\ell+1} - y_\ell| = \Delta$, and that on every interval the dynamics of the fluid queue are modelled based on the same matrix exponential representation $(\boldsymbol{\alpha}, \boldsymbol{S}, \boldsymbol{s})$. These assumptions are, in fact, not necessary, but they do serve to simplify the presentation slightly. The convergence results can be extended to use different sequences of matrix exponential representations on each interval, provided that for each sequence of matrix exponential distributions, the variance tends to 0. Moreover, we can extend the results to intervals of arbitrary width, provided that the width of the intervals is not arbitrarily small. Here we describe how one would prove such results.

The arguments which prove Theorem 5.3 are independent of all other levels/intervals, i.e. the hypotheses of the theorem depend only on the interval/level ℓ_0, and the sequence of matrix exponential distributions used to model the behaviour of the fluid queue on this interval/level, and not on any other interval/level. Thus, Lemma 5.3 holds independently on each interval, as does Corollary 5.15.

Let the width of interval ℓ_0 be $\Delta_{\ell_0} = y_{\ell_0+1} - y_{\ell_0}$ and suppose that sequence of matrix exponential random variables used to model the dynamics of the fluid queue on the interval ℓ_0 is $Z_{\ell_0}^{(p)}$. Regarding the domination condition in Lemma 6.7, we can extend it the following version,

Lemma 6.13. *Assume* $\inf_{\ell_0} \Delta_{\ell_0} > 0$ *and* $\sup_{\ell_0} \mathrm{Var}(Z_{\ell_0}) < \infty$. *Then, for all* $i \in \mathcal{S}_+ \cup \mathcal{S}_-$, $\ell_0 \in \mathcal{K}^\circ$, *and* $n \geq 2$,

$$\mathbb{P}(\tau_n^{(p)} \leq E^\lambda \mid \boldsymbol{Y}_{\boldsymbol{\alpha}}^{(p)}(n-1) = (\ell_0, i), \tau_{n-1}^{(p)} \leq E^\lambda) \leq b_{\ell_0}^{(p)}, \tag{6.55}$$

where

$$b_{\ell_0}^{(p)} = 1 - e^{-q(\Delta_{\ell_0}+\varepsilon_{\ell_0}^{(p)})}\left[1 - e^{q\varepsilon_{\ell_0}^{(p)} - \lambda\Delta_{\ell_0}/|c_{min}|}\right] + \frac{\mathrm{Var}(Z_{\ell_0}^{(p)})}{\left(\varepsilon^{(p)}\right)^2} + |r_{1,\ell_0}^{(p)}|$$

and

$$|r_{1,\ell_0}^{(p)}| \leq 2G\frac{\mathrm{Var}\left(Z_{\ell_0}^{(p)}\right)}{\left(\varepsilon_{\ell_0}^{(p)}\right)^2} + 2L\varepsilon_{\ell_0}^{(p)}.$$

Hence, for all $i \in \mathcal{S}_+ \cup \mathcal{S}_-$*, and* $n \geq 2$*,*

$$\mathbb{P}(\tau_n^{(p)} \leq E^\lambda \mid \phi^{(p)}(\tau_{n-1}^{(p)}) = i, \tau_{n-1}^{(p)} \leq E^\lambda) \leq b^{(p)}, \tag{6.56}$$

where

$$b^{(p)} = \max\left\{\sup_{\ell_0} b_{\ell_0}^{(p)}, \frac{q}{\lambda+q}\right\}.$$

Proof. For the proof of (6.55) follow the same arguments as in the proof of Lemma 6.7. The bound in (6.56) follows by the assumptions in the statement of the lemma, since

$$\begin{aligned}
&\mathbb{P}(\tau_n^{(p)} \leq E^\lambda \mid \phi^{(p)}(\tau_{n-1}^{(p)}) = i, \tau_{n-1}^{(p)} \leq E^\lambda) \\
&\leq \sup_{\ell_0} \mathbb{P}(\tau_n^{(p)} \leq E^\lambda \mid \boldsymbol{Y}_{\boldsymbol{\alpha}}^{(p)}(n-1) = (\ell_0, i), \tau_{n-1}^{(p)} \leq E^\lambda) \\
&\leq \max\left\{\sup_{\ell_0} b_{\ell_0}^{(p)}, \frac{q}{\lambda+q}\right\}.
\end{aligned}$$

□

Given Lemma 6.13, then an equivalent of Lemma 6.8 remains true, the proof of which follows verbatim except with the use of Lemma 6.7 replaced by Lemma 6.13. Corollary 6.9 remains true without modification. Lemma 6.10 and Corollary 6.11 remain true without modification provided that $\lim_{p\to\infty} \mathrm{Var}(Z_\ell^{(p)}) \to 0$ for all ℓ.

Remark 6.14. *I suspect that the convergence results can also be extended to approximating so-called* multi-layer fluid queues*, as described in Bean & O'Reilly (2008).*

Chapter 7

Numerical investigations

We have now established multiple schemes for approximating fluid queues which are suitable for approximating the performance measures of fluid-fluid queues derived in Bean & O'Reilly (2014). This chapter investigates, numerically, various aspects of the approximation schemes. Throughout, we compare the QBD-RAP scheme (from Chapter 4), the discontinuous Galerkin scheme (as described in Chapter 3, with and without slope limiting), and the *spatially-coherent uniformisation* scheme introduced by Bean & O'Reilly (2013*a*).

The performance of the discontinuous Galerkin scheme has been well-studied in some contexts (Cockburn (1999), Hesthaven & Warburton (2007) Section 5.5), and it is well-known that the discontinuous Galerkin scheme performs remarkably well on problems with smooth solutions. Here, we mostly focus on investigating the numerical performance of the schemes on problems with non-smooth solutions. The purpose is to emphasise the probabilistic interpretation and positivity-preserving properties of the QBD-RAP scheme. In the stochastic modelling community it is common to have problems with discontinuities, such as non-smooth initial conditions, hence the type of problems we investigate in this chapter are of relevance. As we will see (in Section 7.5), even if a fluid-queue is initialised with a smooth initial density, the boundary dynamics may induce transient discontinuities or non-smooth behaviour into the problem. A very specific set of conditions on the initial density and point masses must hold for the transient distribution to remain continuous as it evolves over time (Bean & O'Reilly 2014, Bean et al. 2020). Further, it is possible that discontinuities are present in the performance measures of fluid-fluid queues (see, for example, Section 7.7). Limiting distributions of fluid queues are smooth, however.

The experiments demonstrate that the DG scheme can perform extremely well when the true solution is sufficiently smooth. However, when discontinuities are present in the true solution, the DG scheme can lead to approximate solutions which are oscillatory and this may lead to approximate solutions which violate the axioms of probability, such as approximations to PDFs which take negative values (equivalently, approximations to CDFs which are non-increasing), and in the worst cases, approximations to CDFs which

take values outside $[0, 1]$. Hence, there is a need for schemes which ensure the axioms of probability are maintained.

By design, the QBD-RAP scheme, uniformisation scheme, and DG schemes with slope limiting, produce approximations which obey the axioms of probability, i.e. produce approximations to PDFs which are non-negative, and CDFs which are non-decreasing, and $[0, 1]$-valued.

We implement two slope limited DG schemes, the DG-lim and DG-lin-lim scheme, which we describe in more detail in the next section. The numerical experiments demonstrate that the DG-lim scheme does not lead to a viable approximation scheme in the presence of discontinuities, while the DG-lin-lim scheme results in a relatively performant approximation scheme for the discontinuous problems considered here. In general, the QBD-RAP scheme performs similarly to the DG-lin-lim scheme, while the uniformisation scheme is a viable approximation method, but produces larger errors than the QBD-RAP and DG-lin-lim scheme, for the examples considered here.

Furthermore, when slope limiting is used, the resultant approximation is at best linear around discontinuities. Moreover, there is a computational cost in applying a limiter (or filter), which could be significant, particularly if many initial conditions are to be considered. In contrast, for the uniformisation and QBD-RAP, the approximate solution is guaranteed to produce positive probabilities without any post-processing or extra computation.

Although we have been unable to supply a rigorous argument to show that we can use the DG approximation to the generator of a fluid queue to approximate performance measures of fluid-fluid queues, such as the operator Ψ, we can rely on the intuition that substituting the DG approximation for the true operator *should* give reasonable results, and further, *assume* that the resulting approximations are polynomial.

On the other hand, using the stochastic approximations (QBD-RAP or uniformisation), results in rigorously defined operators as proved in Bean et al. (2021).

In summary, the QBD-RAP and uniformisation schemes are the only two schemes which result in rigorously defined approximations, preserve the axioms of probability, and, of the two, the numerical experiments demonstrate that the QBD-RAP method is superior.

7.1 Preliminaries

Before presenting the results of the experiments, we need to describe some aspects of the approximation schemes in more detail, which we now proceed to do.

The main focus of the numerical experiments is on the error as the number of basis functions per cell is increased. In the numerical experiments the number of basis functions used to approximate the solutions is the same on all cells and the number of basis functions on a cell (and hence all cells) is referred to as the *dimension* of the scheme. For the

discontinuous Galerkin scheme, the dimension of the scheme means is the number of polynomial basis functions used to approximate the solution within each cell. For example, if we use 3 basis functions in the discontinuous Galerkin scheme, we approximate the solution by 3 linearly independent quadratics on each cell and hence the approximation is a quadratic on each cell. For the QBD-RAP scheme, the dimension of the scheme is the order of the CME distribution used to construct the scheme. To make a comparable equivalent for the uniformisation scheme, we divide each cell into smaller sub-cells over which we approximate the solution. That is, for a dimension p uniformisation scheme we divide each cell into p sub-cells. Equivalently, we may think of a dimension p uniformisation scheme as using p piecewise constant functions to approximate the solution on each cell. For all schemes (DG, QBD-RAP and uniformisation), if we construct a dimension p approximation, there are K cells, N phases and $c_i \neq 0$ for all $i \in \mathcal{S}$, then the resulting approximation to the generator $\mathbb{B}$ is a square matrix of dimension $pKN+N$.* In this sense each approximation scheme leads to matrices of the same order (although not necessarily the same number of non-zero elements).

We investigate two slope-limited schemes. In the scheme we refer to the *DG-lim* scheme, we use a DG scheme of dimension p and apply the Generalised MUSCL slope limiter as described in Section 2.8.2. In the *DG-lin-lim* scheme, we follow the construction described in Section 2.8.4, which is effectively a DG scheme with cells of width $2\Delta/(p+1)$, for $p = 1, 3, 5, ..., 21$, and a basis of dimension 2 (linear) on each cell, with an upwind flux, and we apply the Generalised MUSCL limiter. Thus, the DG-lin-lim scheme uses $p+1$ basis functions to represent the solution on each of the original cells $\mathcal{D}_k$, and we define the dimension of the scheme to be $p+1$. In the presence of discontinuities, slope-limiting forces the approximation to be linear, i.e. an order 1 polynomial, around the discontinuity. For DG schemes of dimension greater than 2 with a slope limiter and in the presence of discontinuities, the ability of the scheme to represent higher-order polynomials is not well-utilised as the slope limiter forces the solution to be linear anyway. The DG-lin-lim scheme recognises this and instead of using higher-order polynomials, uses piecewise-linear functions and smaller cell widths to approximate the solution.

The concept of slope-limiting originates in the context of numerically integrating (typically with respect to a *time* variable) a so-called *semi-discrete system*, as we do in Section 7.5. However, approximating the operator Ψ of a fluid-fluid queue amounts to solving an implicit matrix-equation and there is no concept of numerical integration with respect to time. Hence, the concept of slope-limiting, it is not directly transferable. One could process the initial condition and post-process the solution with a limiter. However, there are possibly better choices for this post-processing, such as filtering, (Hesthaven & Warburton 2007, Section 5.6.1). One could also view the composition of the DG operator and

*There are pK basis functions for each phase which approximate the solution on the interior of the domain, hence pKN basis functions in total, plus an additional N dimensions to capture boundary conditions.

the limiting operator as one, but the result is a non-linear operator which is of no use for simplifying the Riccati Equation which characterises Ψ. Considering there is no theory on the use of the DG scheme to approximate Ψ and hence the resultant approximation is not rigorously-defined, the author does not know if the use of a limiter in this context is possible.

To keep the content of this chapter contained, we do not investigate all aspects of the schemes. For each numerical experiment we keep the number of cells, K (equivalently, the cell size, Δ), fixed for the DG and QBD-RAP schemes. As part of deriving a discontinuous Galerkin scheme, one needs to choose the *numerical flux* which is used to approximate the transition of density from one cell to the next (Hesthaven & Warburton 2007). We investigate schemes with an upwind flux only. We do not investigate filtering for the DG scheme.

For schemes which require us to integrate over time we do not investigate the stability of the schemes with respect to the t-step-size, or the time-integration scheme itself (where required). Instead, we fix the time-integration step size for each numerical experiment at a suitably small value to obey a certain stability criterion (a CFL-like condition, (Hesthaven & Warburton 2007, Section 4.8)). Moreover, we always implement the strong stability preserving Runge-Kutta method of order 4 with 5 stages (Spiteri & Ruuth (2002), Hesthaven & Warburton (2007) Section 5.7, see also Section 2.8.1), which is claimed to introduce no more oscillations into the solution as we integrate over time.

Where a slope limiter is implemented, we implement the *Generalised MUSCL* limiter (Cockburn (1999), Hesthaven & Warburton (2007) Section 5.6.2, see also Section 2.8.2).

In this chapter, all the error plots are on a $\log_{10}$-$\log_{10}$ scale. If the error function takes the form $error = \beta_0^* p^{\beta_1}$, where p is the dimension of the scheme, then the $\log_{10}$-$\log_{10}$ will show a linear trend since $\log_{10}(error) = \beta_0 + \beta_1 \log_{10}(p)$, where $\beta_0 = \log_{10}(\beta_0^*)$. Where relevant, to estimate the asymptotic rate of convergence, β_1, we estimate the slope of the line $\log_{10}(error) = \beta_0 + \beta_1 \log_{10}(p)$ using ordinary least squares and the last eight data points of each line. The estimated linear trends are plotted and the slope of the line is displayed to the right of the plot.

The structure of this chapter is as follows. In Section 7.2 we compute approximations to various initial conditions for the different schemes and observe their performance which allows us to investigate the performance of the reconstruction methods without considering a specific model or any dynamics. In Section 7.3 we investigate travelling wave problems with various initial conditions. For the travelling wave problem the dynamics are deterministic, which means that the solution is known, and we do not need to resort to simulation as a *ground-truth*. Furthermore, the travelling wave problems allow us to investigate the ability of the schemes to approximate the flow of probability across cells without any stochastic dynamics. Next, we investigate a simple fluid queue with two phases. Section 7.4 investigates approximations to the limiting distribution of the queue, Section 7.5 investigates approximations to the transient distribution of the queue for two

initial conditions, and Section 7.6 investigates approximations the first hitting time of the fluid level on the set $\{0, 1\}$. Lastly, we apply the approximation methods to a fluid-fluid queue in Section 7.7, compute approximations to Ψ and ultimately, approximate the distribution of $\{(X(\zeta_Y(\{0\})), \varphi(\zeta_Y(\{0\})))\}$ at time $\zeta_Y(\{0\})$ at which $\{Y(t)\}$ first returns to 0.

7.2 Function approximation/reconstruction

We start our investigation by looking at how well the schemes perform at approximating an initial condition. By approximating the initial condition only, we aim to investigate the performance of the approximation schemes without any dynamics.

For the discontinuous Galerkin method, we project the initial condition on to the set of polynomial basis functions which define the scheme, for the spatially-coherent uniformisation scheme, we look at the sub-cell averages of the initial condition, and for the QBD-RAP scheme, we compute the initial vector for the approximation, then reconstruct the solutions as described in Sections 4.6 and 4.7. For the purposes of approximation and reconstruction for the QBD-RAP scheme we must orientate the initial condition in a positive or negative direction; here we suppose that the initial conditions belong to a positive phase. First we investigate the three closing vectors from Section 4.7 which we can use for the reconstruction in the QBD-RAP scheme. From the investigation, we decide to use the closing vector (4.27) throughout the rest of the chapter.

In this section we consider approximating initial conditions over a single cell of width $\Delta = 1$. To numerically evaluate integrals arising in the approximation step (inner products and cell averages) we use a trapezoidal rule with 10,001 function evaluations. Similarly, we use 10,001 function evaluations to approximate L^p errors, and also as a finite set of points over which to compute the Kolmogorov-Smirnov (KS) statistics.

7.2.1 QBD-RAP closing operators

We investigate the three closing operators introduced in 4.7. For convenience, we restate the three closing operators from Section 4.7 here. The unnormalised closing vector,

$$\boldsymbol{av}(x) = \boldsymbol{a}e^{\boldsymbol{S}x}\boldsymbol{s}. \tag{4.25}$$

A normalised version of the closing operator in Equation (4.26), with the normalised version given by

$$\boldsymbol{av}(x) = \frac{\boldsymbol{a}\left(e^{\boldsymbol{S}x}\boldsymbol{s} + e^{\boldsymbol{S}(2\Delta - x)}\boldsymbol{s}\right)}{1 - \boldsymbol{a}e^{\boldsymbol{S}2\Delta}\boldsymbol{e}}. \tag{4.29}$$

We refer to the closing operator in (4.29) as the *non-linear normalised* closing operator.[†] The third closing vector we investigate is given by

$$\boldsymbol{av}(x) = \boldsymbol{a}\left(e^{\boldsymbol{S}x}\boldsymbol{s} + e^{\boldsymbol{S}(2\Delta - x)}\right)\left[I - e^{S2\Delta}\right]^{-1}\boldsymbol{s}, \tag{4.27}$$

which we refer to as the *normalised* closing operator.

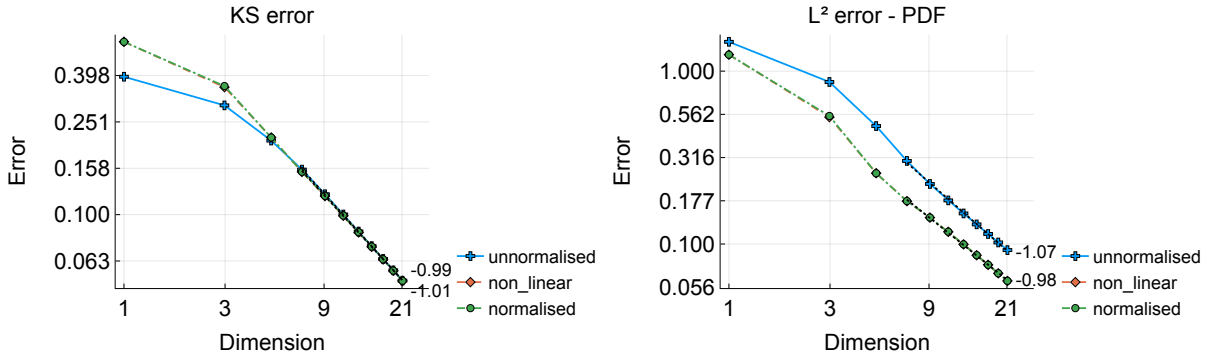

Figure 7.1: KS error (left) and L^2 error of the PDF (right) of Example 7.1 for the three closing vectors considered; unnormalised (blue solid line), non-linear normalised (orange dashed line) and normalised (green dash-dotted line). The error curves of the non-linear normalised (orange) and normalised closing vectors are almost coincident. The black dotted lines are linear least-squares fits to the last 8 data points and the slopes of the least square lines are written next to the last point.

Example 7.1. First consider approximating the initial condition with density function $2 \times 1(x < 0.5)$, where $1(\cdot)$ is the indicator function. In the left-hand side of Figure 7.1 we plot the Kolmogorov-Smirnov error between the true CDF and the CDF constructed via the QBD-RAP approximation with the three different closing vectors. Interestingly, in this case, the unnormalised closing operator outperforms the other two at low dimensions. In this case, it just so happens that the error due to truncation of the unnormalised reconstruction constructively interferes with other errors to cause the error to be lower at low dimensions for the unnormalised scheme.

Figure 7.1 also shows that the non-linear normalised and normalised closing operators perform very similarly – this is the case throughout much of this section. The difference between the non-linear normalised and normalised closing operators is how they treat mass in the tail of a matrix exponential distribution (from 2Δ onward). Intuitively, there should be very little mass in the tail of the distribution as we are using concentrated matrix exponential distributions in the reconstruction.

[†]Technically we have not yet proved that the *non-linear* normalised closing operator leads to a convergent approximation scheme.

If we instead look at the L^2 error between the true PDF and the reconstructions in Figure 7.1 (right), then we see that the unnormalised reconstruction performs the poorest. Figure 7.1 also suggests the non-linear normalised and normalised reconstruction are very similar.

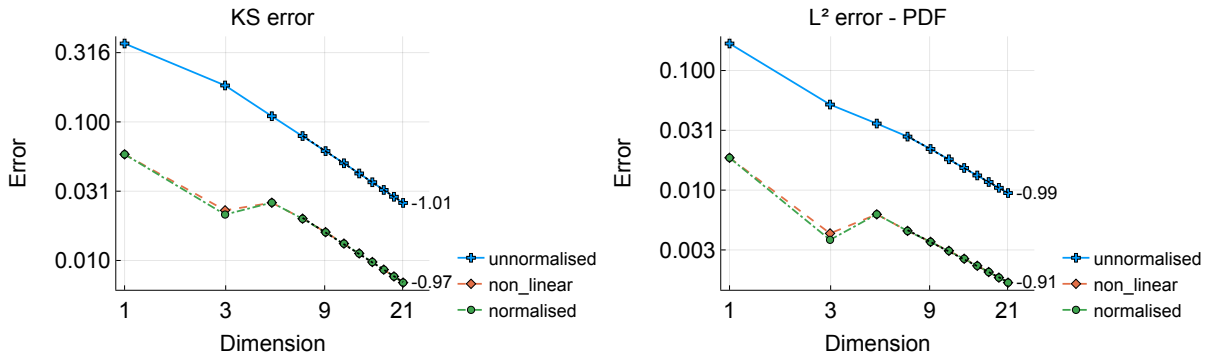

Figure 7.2: KS error (left) and L^2 error of the PDF (right) of Example 7.2 for the three closing vectors considered; unnormalised (blue solid line), non-linear normalised (orange dashed line) and normalised (green dash-dotted line). The error curves of the error curves of the non-linear normalised (orange) and normalised closing vectors are almost coincident. The black dotted lines are linear least-squares fits to the last 8 data points and the slopes of the least square lines are written next to the last point.

Example 7.2. Now consider approximating the initial density $f(x) = 1$ for $x \in [0, 1)$. Observing Figure 7.2 of the KS error and L^2 error of the PDF we now see that the normalised reconstructions outperform the unnormalised reconstruction. This suggests that, in this case, the 'folding' of the closing operator about Δ has greatly increased the ability of the reconstruction to approximate this initial distribution.

Some insight is gained by looking at Figure 7.3 where we plot the reconstructed PDFs for the unnormalised (4.25) and normalised (4.27) closing operators for dimension 1, 3, 5 and 7, as well as the true PDF. Observing Figure 7.3 notice that the unnormalised reconstruction fails to capture the density at the left of each of the plots. This feature is due to a significant amount of mass being lost due to the truncation. In comparison, the reconstruction with the normalised closing operator is much better in this region due to the 'folding' around Δ in the construction of the closing operator. The 'folding' in the normalised closing operator manifests itself as more mass at the left of these plots compared to the unnormalised reconstructions.

Figure 7.3 also shows that both closing operators fail to approximate the initial condition well at the right-hand side of the interval. A potential area for further research is to determine if there is a different reconstruction method or another closing operator which could alleviate this issue.

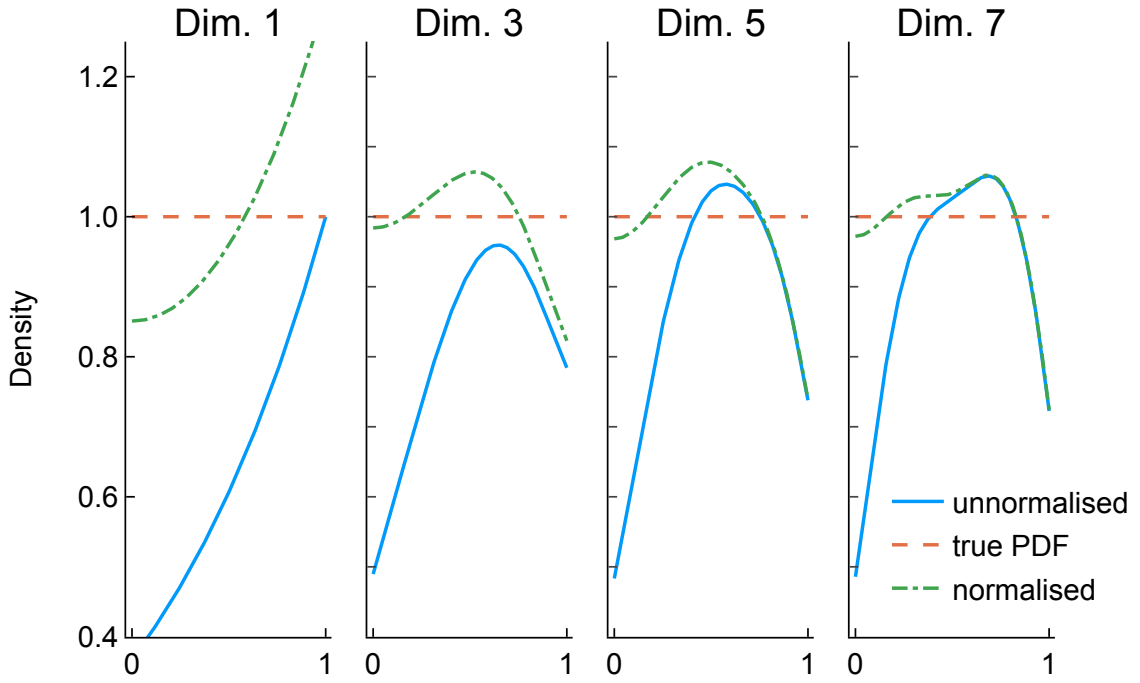

Figure 7.3: Approximations to the PDF $f(x) = 1$, $x \in [0, 1)$ (orange dotted line) for Example 7.2, constructed using matrix exponential distributions of various dimensions, and using the unnormalised closing operator (blue solid line) and normalised closing operator (green dash-dotted line).

Example 7.3. Next we consider approximating the initial distribution with density $-6x^2 + 6x$ on $[0, 1)$. Observing the left-hand panel of Figure 7.4 which plots the KS error against the dimension of the matrix exponential approximation for the three closing operators, we once again see that the unnormalised closing operator performs the worst, while the performance of the two normalised closing operators is indistinguishable. However, if we instead look at the right-hand panel of Figure 7.4 which plots the L^2 error between the reconstructed PDF and the true PDF, then the unnormalised closing operator performs better than the two normalised ones for dimension 5 and above.

The fact that the unnormalised closing operator outperforms the two normalised ones can be explained by the 'folding' of the normalised operators around Δ. In Figure 7.5 we plot the reconstructions obtained using the unnormalised and normalised closing operators, along with the true PDF, $-6x^2 + 6x$. In Figure 7.5 at the left of the plots, we observe that the normalised closing operator over-estimate the density function, whereas the unnormalised closing operator looks to be doing better. The 'folding' in the normalised closing operator manifests itself as more mass at the left of these plots compared to the unnormalised closing operator.

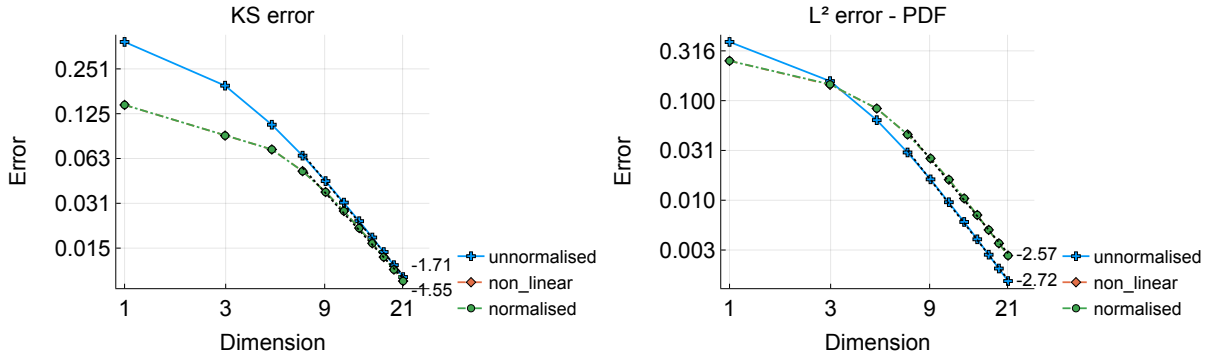

Figure 7.4: KS error (left) and L^2 error for the PDFs (right) of Example 7.3, for the three closing vectors considered; unnormalised (blue solid line), non-linear normalised (orange dashed line) and normalised (green dash-dotted line). The error curves for the non-linear normalised (orange) and normalised closing vectors are almost coincident. The black dotted lines are linear least-squares fits to the last 8 data points and the slopes of the least square lines are written next to the last point.

Example 7.4. Lastly, we consider approximating the initial distribution with PDF $3e^{-3x}/(1-e^{-3})$. This density function is at a maximum at the left of the region. Considering what we have learnt so far about the unnormalised operator underestimating in this region, we expect that the unnormalised closing operator will perform relatively poorly in this case. Indeed, observing Figures 7.6 we see that this is the case; the normalised closing operators perform relatively well compared to the unnormalised closing operator, as measured by both error metrics (KS statistic and L^2 norm on the PDF). Plots of the PDFs (omitted), confirm that this is due to the loss of mass at the left-hand side of the region due to the truncation.

Table 7.1 summarises the rates of convergence of the L^2 error on the PDFs for unnormalised and normalised closing operators in the examples above (i.e. the slopes of the black dotted lines in the error plots). In general, the errors of the approximations obtained using normalised closing operators converge at a rate approximately as fast, or faster, than the approximations obtained using the unnormalised closing operator.

In summary, the non-linear normalised and normalised closing operators perform almost identically. In some cases the unnormalised closing operator can outperform the normalised closing operators, typically when the mass of the function to be approximated at the left-hand edge of the cell is low. Further, the rates of convergence for all three schemes were similar to each other, except for in Example 7.4 where we approximated the initial condition $3e^{-3x}/(1-e^{-3})$; in this case the rate of convergence of the normalised closing operators was approximately twice as large as the unnormalised version. Given the results above and its theoretical tractability, we will opt to use the normalised closing

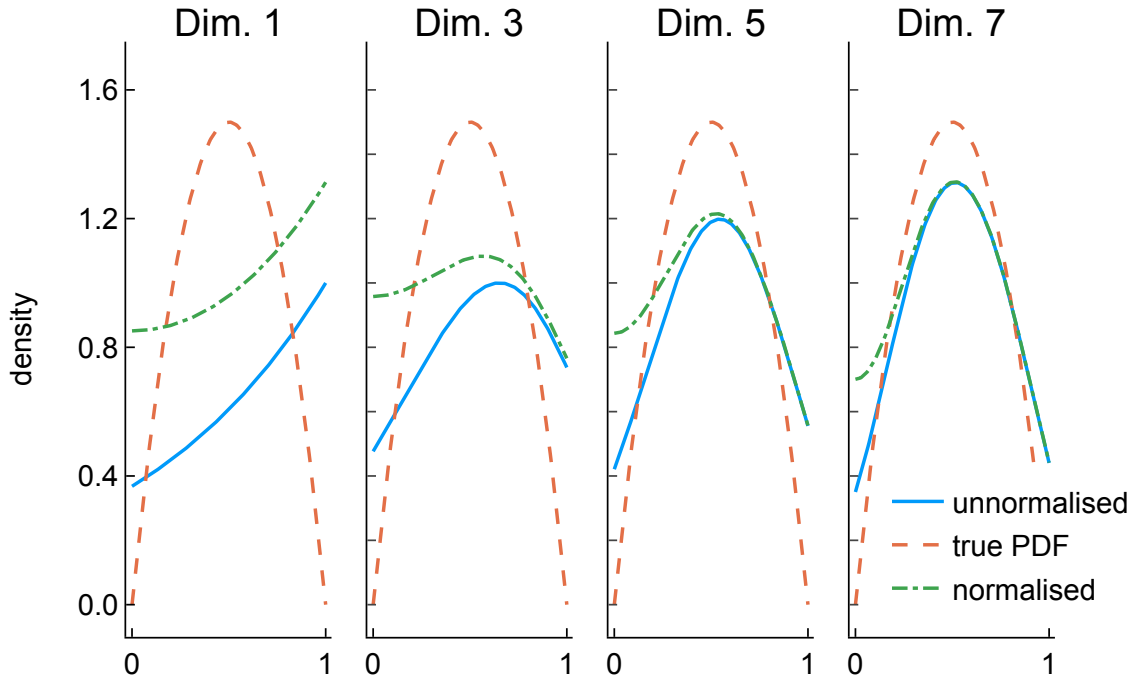

Figure 7.5: Approximations to the PDF $f(x) = -6x^2 + 6x$ (orange dotted line) for Example 7.3, constructed using matrix exponential distributions of various dimensions, and using the unnormalised closing operator (blue solid line) and normalised closing operator (green dash-dotted line).

operator in (4.27) for the rest of the numerical experiments.

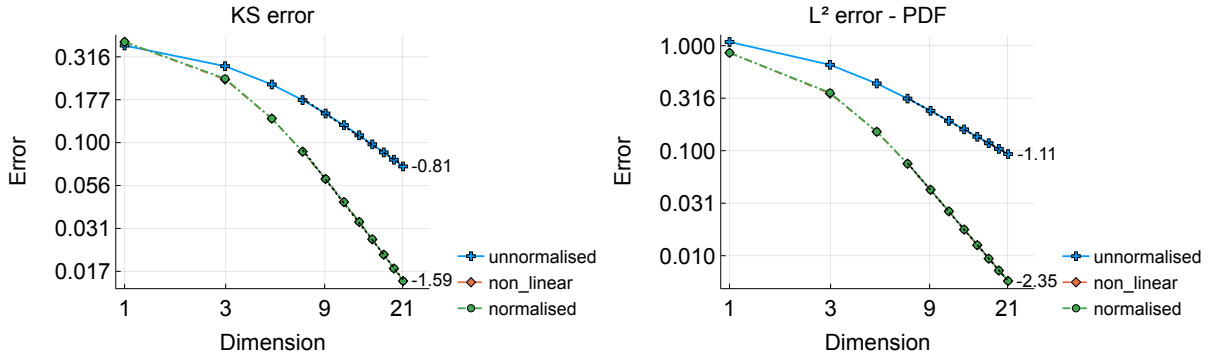

Figure 7.6: KS error (left) and L^2 error of the PDF (right) for Example 7.4 for the three closing vectors considered; unnormalised (blue solid line), non-linear normalised (orange dashed line) and normalised (green dash-dotted line). The error curves for the non-linear normalised (orange) and normalised closing vectors are almost coincident. The black dotted lines are linear least-squares fits to the last 8 data points and the slopes of the least square lines are written next to the last point.

Example	PDF on $[0, 1)$	β_1, unnormalised	β_1, normalised
7.1	$2 \times 1(x < 0.5)$	-1.07	-0.98
7.2	1	-0.99	-0.91
7.3	$-6x^2 + 6x$	-2.72	-2.57
7.4	$3e^{-3x}/(1 - e^{-3})$	-1.11	-2.35

Table 7.1: Estimated rates of convergence, β_1, of the L^2 error on the PDFs for unnormalised and normalised closing operators in the examples above (i.e. the slopes of the black dotted lines in the error plots).

7.2.2 Comparison of schemes

Here we compare the ability of the QBD-RAP, uniformisation, and DG schemes to reconstruct initial conditions. With this section we aim to investigate the accuracy of the reconstruction methods for the numerical schemes without any dynamics.

Example 7.5. First we consider the initial condition with CDF $1(x \geq 0.5)$, that is, a point mass of 1 at 0.5. This distribution does not have a PDF, so we compare the CDFs only. In Figure 7.7 we plot the KS metric (left) and L^1 metric (right) between the true CDF and the reconstructed approximations. Observing the KS metric, it appears that none of the schemes converge and the KS error sits around 0.5 for all schemes. This reflects the fact that convergence in distribution implies point-wise convergence of the CDFs except, perhaps, at points of discontinuity. None of the schemes converge at the discontinuity at $x = 0.5$. Observing the L^1 error between the true CDF and reconstructed

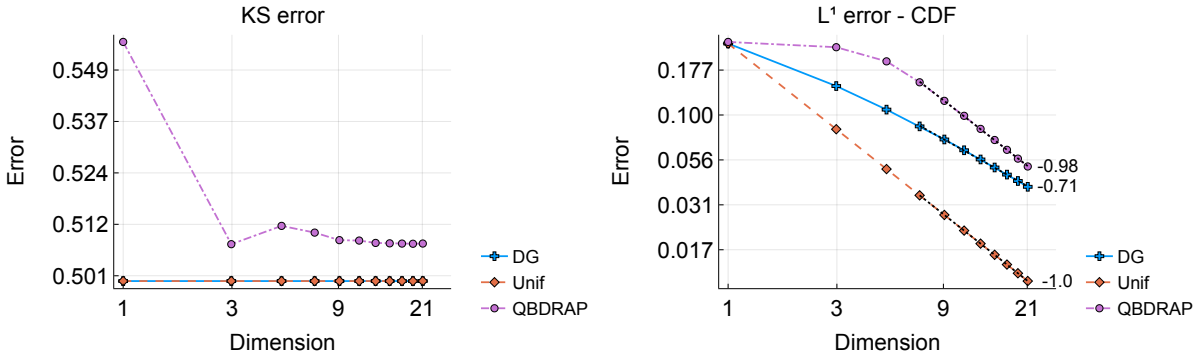

Figure 7.7: KS error (left) and L^1 error of the CDF (right) of Example 7.5 for the DG (blue solid line), uniformisation (orange dashed line) and QBD-RAP (purple dashed line) schemes. The black dotted lines are linear least-squares fits to the last 8 data points and the slopes of the least square lines are written next to the last point.

approximation (which is the area between the two CDFs) we now see that the schemes appear to show the convergent behaviour we expect. Here, the uniformisation scheme appears to perform the best, while the QBD-RAP scheme performs the worst.

Perhaps it is no surprise that the uniformisation scheme performs best. In the uniformisation scheme, as we increase the dimension, we partition the cell $[0, 1)$ into smaller sub-cells, and use constant functions on each sub-cell to construct an approximate initial density. As such, the uniformisation scheme can produce a piecewise continuous, linear approximation to the CDF. In contrast, both the DG and QBD-RAP schemes result in a smooth approximation to the CDF over the entire cell. Given the initial distribution is far from smooth (it is a point mass), then we might expect that the uniformisation scheme will perform relatively well.

In Figure 7.8 we plot approximate CDFs from the DG, uniformisation and QBD-RAP schemes alongside the true CDF. The DG scheme displays undesirable features for an approximation to a CDF – it is not monotonically increasing and is not constrained to the bounds of probability, $[0, 1]$. On the other hand, although the QBD-RAP scheme converges slowest, it displays good properties in that it results in a monotonically increasing CDF, starting at 0 and ending at 1. The uniformisation scheme approximates the CDF exceptionally well.

Example 7.6. Now consider approximating the initial distribution with PDF $2 \times 1(x \leq 0.5)$. Figure 7.9 plots the KS error (left) and L^2 error between the true and approximate PDFs (right). Figure 7.9 suggests that all schemes converge at a similar rate for this problem. Here, the QBD-RAP scheme performs worst, the uniformisation scheme second, and the DG scheme the best. However, once again the DG scheme exhibits undesirable properties (plots not shown) – the approximation to the CDF is at some points above 1

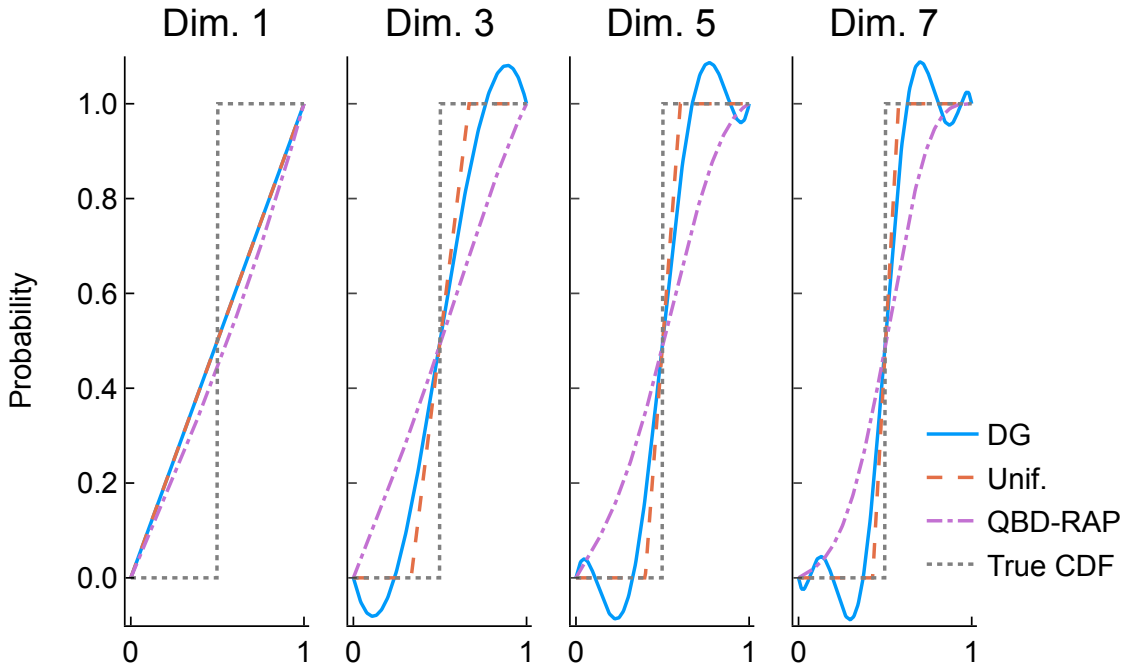

Figure 7.8: Reconstructed CDFs using the DG (blue solid line), uniformisation (orange dashed line) and QBD-RAP (purple dashed line) schemes for Example 7.5. The true distribution function is $1(x \geq 0.5)$ (grey dotted line).

and is not monotonic (although these violations do not appear to be as severe in this case as they were for Example 7.5).

Example 7.7. So far we have considered two problems which exhibit discontinuities. At the other extreme we now consider an initial distribution with density $-6x^2+6x$ on $[0, 1)$. In Figure 7.10 we plot the KS error and the L^2 error between the true and approximate PDFs. Since the DG method projects the initial condition onto a polynomial basis, then, for a dimension 3 approximation and above, the DG scheme can represent the initial condition exactly. This is reflected in Figure 7.10, where the blue curve drops sharply from dimension 1 to 3, then plateaus. Due to numerical integration errors, for example in the evaluation of the integral in the L^2 norm, the errors for the DG scheme are not 0. Regarding the other two schemes, they too appear to be convergent at approximately the same rate, with the uniformisation scheme performing better for the KS error, and both performing similarly in terms of the L^2 error.

Example 7.8. Consider now the initial distribution with PDF $cos(4\pi(x + 0.5)) + 1$. Figure 7.11 shows the KS and L^2 errors for the PDF. Both the KS error (left) and L^2 norm between the true and approximated PDFs (right) tell a similar story; for sufficiently high

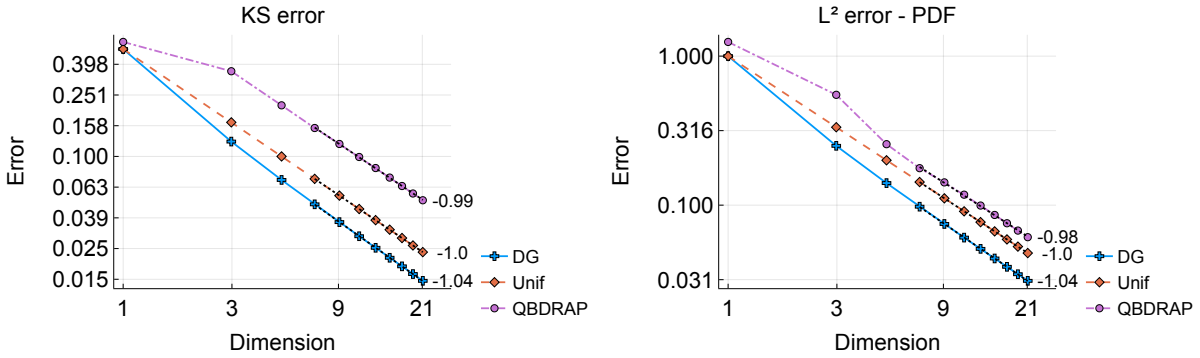

Figure 7.9: KS error (left) and L^2 error of the PDF (right) for Example 7.6 for the DG (blue solid line), uniformisation (orange dashed line) and QBD-RAP (purple dashed line) schemes. The black dotted lines are linear least-squares fits to the last 8 data points and the slopes of the least square lines are written next to the last point.

order, the DG scheme approximates the initial condition very well. The uniformisation scheme performs second best, while the QBD-RAP scheme performs poorest.

Summary

First, we investigated the three proposed closing operators for the QBD-RAP scheme proposed in Section 4.7. We concluded that the normalised closing operator is most appropriate, hence we will use it for the rest of this chapter.

Next, we investigated reconstructing approximations to initial conditions for the DG, QBD-RAP and uniformisation schemes. We observed that the DG scheme performs well with respect to the error metrics considered, but can display oscillations and result in approximations which violate the axioms of probability. On the other hand, the uniformisation and QBD-RAP schemes result in non-decreasing approximations to the CDFs, but can converge slower with respect to the error metrics considered, with the QBD-RAP scheme often converging the slowest.

This section investigated the performance of the schemes without considering any of the dynamics of the problem. By doing so, we have gained insights into the performance of the approximations independent of the dynamics. In the remaining sections we introduce dynamics into the problem.

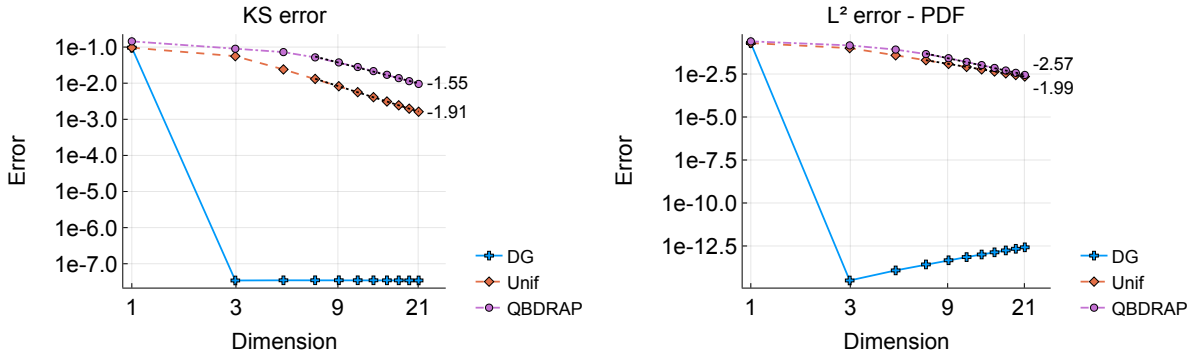

Figure 7.10: KS error (left) and L^2 error of the PDF (right) for Example 7.7 for the DG (blue solid line), uniformisation (orange dashed line) and QBD-RAP (purple dashed line) schemes. The black dotted lines are linear least-squares fits to the last 8 data points and the slopes of the least square lines are written next to the last point.

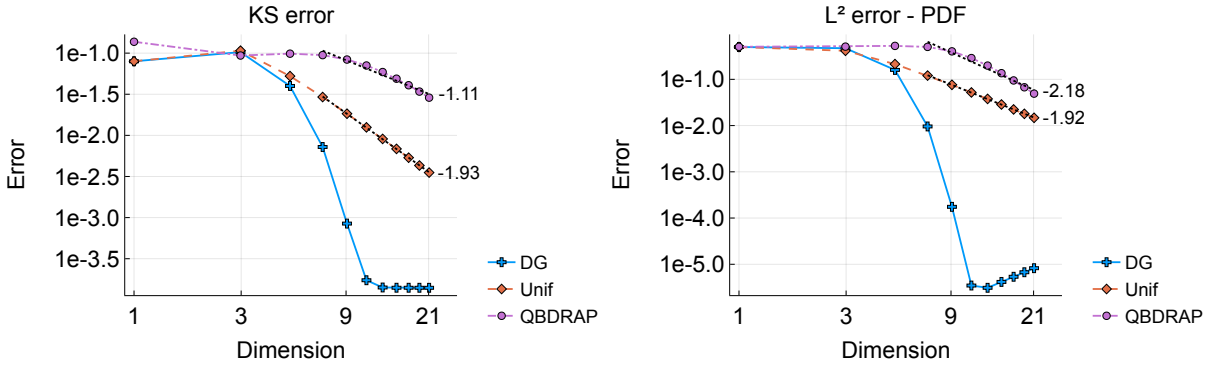

Figure 7.11: KS error (left) and L^2 error of the PDF (right) for Example 7.8 for the DG scheme (blue solid line), uniformisation scheme (orange dashed line) and QBD-RAP scheme (purple dashed line). The black dotted lines are linear least-squares fits to the last 8 data points and the slopes of the least square lines are written next to the last point.

7.3 Travelling wave problems

Here we investigate the performance of the different schemes for approximating transient distributions of one-dimensional travelling wave problems with various initial conditions. Consider a (trivial) fluid queue with one phase, generator $\boldsymbol{T} = [0]$ and rate $c = 1$. The PDE (if/when it exists) which describes this system is

$$\frac{\partial}{\partial t} f(x,t) = -\frac{\partial}{\partial x} f(x,t),$$

where $f(x,t)$ is the density (if/when it exists) at time t. Given an initial condition, $f(x,0)$, solutions to this problem are given by

$$f(x,t) = f(x-t,0)$$

so the solution at time t is just a shift in the initial condition t units to the right. We suppose that the fluid queue is bounded, with a lower boundary $x = 0$ and upper boundary $x = 10$. This example is convenient as it has a known solution and no stochastic dynamics, hence we can investigate the ability of the schemes to approximate the flow of mass, without any stochastic dynamics.

We use the QBD-RAP, uniformisation and DG schemes to discretise the solution in space and discretising the interval $[0, 10]$ into 10 cells, each of width 1. We use $10,001$ points to approximate the integrals which appear in the construction of the initial conditions, to approximate the integrals appearing in the error metrics, and also as a set of discrete points on which to evaluate the CDFs to approximate the KS metric. Furthermore, we use the SSPRK4 method to integrate over time with a t-step size of 0.005, and we evolve the system until time $t = 4$.[‡] For the DG scheme we also implement the Generalised MUSCL slope limiter in the form of the DG-lim and DG-lin-lim schemes (see also Section 2.8.2).

To investigate the performance of the schemes without the need to reconstruct the function within each cell we use the cell-wise error metric obtained by computing

$$\begin{aligned}&\sum_{\ell=1}^{10} |\mathbb{P}(X(4) \in \mathcal{D}_{\ell,1}, \varphi(4) = 1 \mid X(0) = 0.5, \varphi(0) = 1) - p(4,\ell,1)| \\ &+ |\mathbb{P}(X(4) \in \{10\}, \varphi(4) = 1 \mid X(0) = 0.5, \varphi(0) = 1) - p(4,11,1)| \end{aligned} \tag{7.1}$$

where

$$\mathbb{P}(X(4) \in \{10\}, \varphi(4) = 1 \mid X(0) = 0.5, \varphi(0) = 1) \tag{7.2}$$

is the mass at the boundary, $p(4,\ell,1)$ is an approximation to $\mathbb{P}(X(4) \in \mathcal{D}_{\ell,1}, \varphi(4) = 1 \mid X(0) = 0.5, \varphi(0) = 1)$, and $p(4,11,1)$ is an approximation to $\mathbb{P}(X(4) \in \{10\}, \varphi(4) = 1 \mid X(0) = 0.5, \varphi(0) = 1)$.

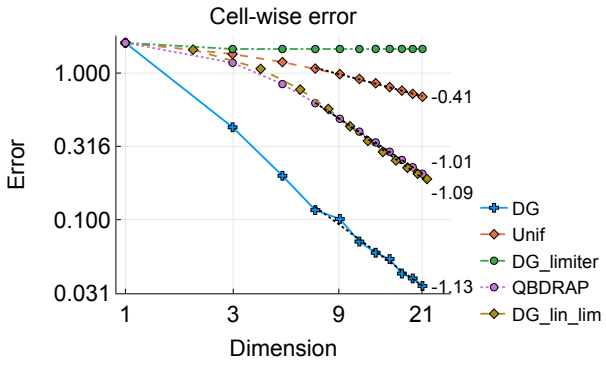

Figure 7.12: Cell-wise error defined in (7.1) for the travelling wave in Example 7.9. Plotted are the cell-wise errors for the DG (blue solid line), DG-lim scheme (green dashed line), uniformisation (orange dashed line), QBD-RAP (purple dotted line) and DG-lin-lim scheme (gold dashed line) schemes, versus the dimension of the approximation. The black dotted lines are linear least-squares fits to the last 8 data points and the slopes of the least square lines are written next to the last point.

Example 7.9. First consider the initial condition with PDF $1(x < 1)$. The level of the fluid queue is uniformly distributed over the first cell. For the DG-based schemes (including the uniformisation scheme) the initial condition can be represented exactly whereas, for the QBD-RAP scheme, it cannot. Thus, in this case, there is no discretisation error in constructing the initial condition for the DG and uniformisation schemes. Moreover, at time $t = 4$, the solution is $f(x, 4) = 1(x \in [4, 5))$. The projections related to the DG-based schemes can represent this solution exactly too, hence we might expect these schemes to work well here.

We plot the cell-wise error metric in Figure 7.12 and observe that the DG scheme clearly produces the lowest errors, followed by the QBD-RAP and DG-lin-lim schemes which perform comparably, then the uniformisation scheme. The worst performer is the DG-lim scheme which does not appear to converge; during the integration of the solution over time the DG-lim scheme has detected oscillations and reduced the scheme to linear where the oscillations were detected, thereby decreasing the accuracy of the scheme to linear where oscillations are present.

Figure 7.13 plots the KS error between the approximate and true CDFs and L^2 error between the approximate and true PDFs. Comparing Figure 7.12 with Figure 7.13, each scheme seems to show a similar rate of decay for these two error metrics as they did for the cell-wise errors.

Continuing with Figure 7.13. The DG-lim scheme does not appear to converge with

[‡]The t-step size must be chosen to ensure that numerical integration over time is stable up to dimension 21, adhering to a CFL-like condition (Hesthaven & Warburton 2007, Section 4.8).

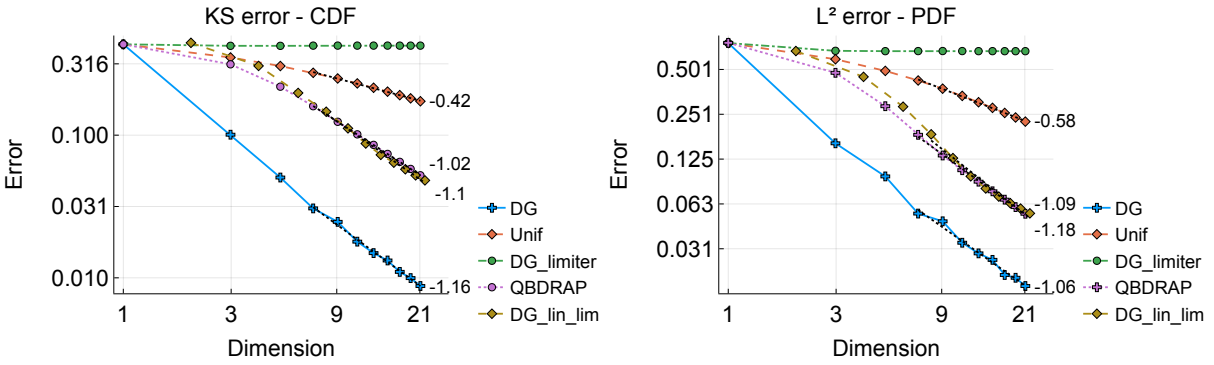

Figure 7.13: KS error (left) and L^2 error of the PDF (right) for the travelling wave problem in Example 7.9 using the approximations from the DG (blue solid line), DG-lim scheme (green dash-dotted line), uniformisation (orange dashed line), QBD-RAP (purple dotted line) and DG-lin-lim (gold dashed line) schemes. The black dotted lines are linear least-squares fits to the last 8 data points and the slopes of the least square lines are written next to the last point.

these error metrics. Of the other three positivity preserving schemes, the uniformisation, DG-lin-lim and QBD-RAP schemes, all appear to be converging with the QBD-RAP and DG-lin-lim schemes converging similarly and faster than the uniformisation scheme. With these error metrics, the DG scheme is converging fastest. However, if we observe the approximations resulting from the DG scheme (Figure 7.14 top-row), we find them to be unsatisfactory due to oscillations and negative densities.

Figure 7.14 plots the density functions reconstructed using the DG, uniformisation, DG-lim, QBD-RAP and DG-lin-lim schemes for various dimension schemes. In the first row observe that, when the dimension is greater than one, the DG scheme (sans limiter) produces an approximation to the PDF which is negative in some places. In the third row of Figure 7.14 observe that, with the limiter, the DG-lim approximation does not change significantly after dimension 3. This is due to the fact that the DG-lim scheme is at best linear in the presence of oscillations. In the second row of Figure 7.14 is the solution approximated using the uniformisation scheme, in the fourth row is the solution approximated using the QBD-RAP scheme, and in the last row is the approximation from the DG-lin-lim scheme.

This is a particularly interesting example. For the DG scheme, even though there is no discretisation error for the initial condition, we use a strong stability preserving time-integration method, and the projection on which the DG method is based can represent the transient distribution at time $t = 4$ exactly, there is still the possibility of badly behaved solutions as shown in the top row of Figure 7.14.

Example 7.10. Another interesting example occurs with the initial condition with CDF

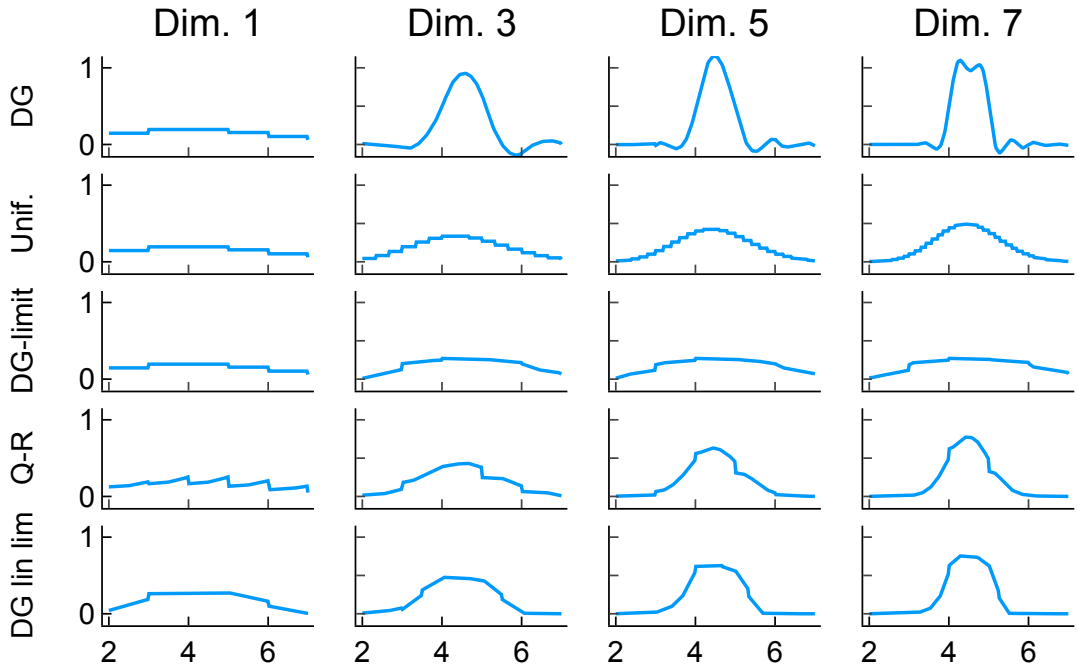

Figure 7.14: Reconstructed PDFs using the DG (top row), uniformisation (second row), DG-lim (third row), QBD-RAP (fourth row) and DG-lin-lim schemes (fifth row), for dimension 1, 3, 5, and 7 (columns) for the travelling wave problem in Example 7.9. The true density function is $1(4 \leq x < 5)$.

$1(x \geq 0.5)$, i.e. a point mass of 1 at 0.5. The exact solution at time $t = 4$ is therefore a point mass at 4.5. No PDF exists for the true distribution, so instead we compare the CDFs. Also recall that, when we analysed reconstruction of this initial condition we saw that using the KS metric may be uninformative due to the lack of point-wise convergence at the discontinuity, so for this example, we measure errors by looking at the L^1 error between the CDFs (the area between the CDFs) instead, and also with the cell-wise error from (7.1)-(7.2).

Figure 7.15 (left) plots the L^1 error between the true and approximated CDFs. The L^1 metric tells a similar story to the previous analysis: the DG-lim scheme does not perform well, the other three positivity preserving schemes (the uniformisation, DG-lin-lim and QBD-RAP schemes) appear to converge, with the QBD-RAP and DG-lin-lim schemes performing similarly and better than the uniformisation scheme. The DG scheme performs the best. However, if we plot the approximations to the CDFs from the DG scheme (not shown) we once again see an oscillating, non-monotonic function.

Another interesting observation is to compare the performance of the uniformisation and QBD-RAP schemes with respect to the L^1 metric on the CDFs in Figure 7.15 (left)

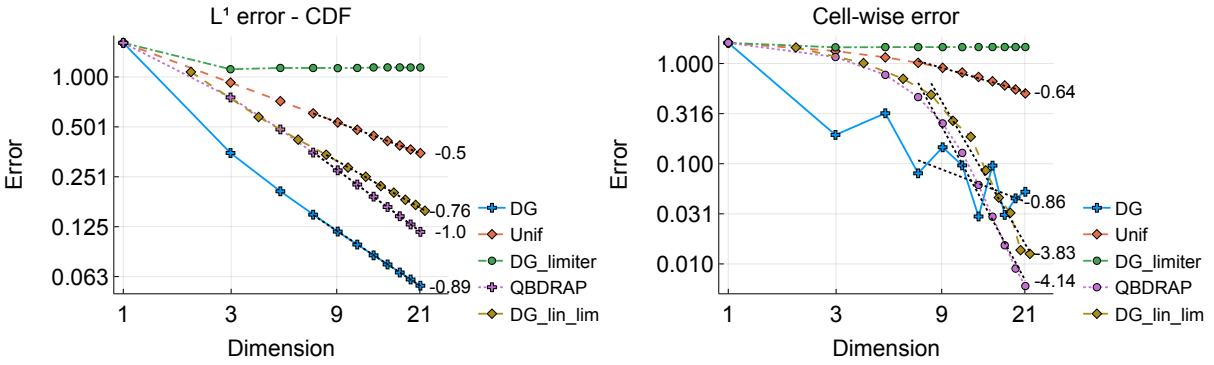

Figure 7.15: L^1 error of CDFs (left) and the cell-wise error metric in (7.1)-(7.2) (right) for the travelling wave problem in Example 7.10 where approximation were constructed via the DG (blue solid line), DG-lin-lim (green dash-dotted line), uniformisation (orange dashed line), QBD-RAP (purple dotted line) and DG-lin-lim (gold dashed line) schemes. The black dotted lines are linear least-squares fits to the last 8 data points and the slopes of the least square lines are written next to the last point.

to that in Figure 7.7 (right). Recall, in Figure 7.7 we investigated the ability of the approximation schemes to represent the initial distribution with CDF $1(x \geq 0.5)$, which is the initial condition of the problem we are considering here. In Figure 7.7 (right) the uniformisation scheme out-performed the QBD-RAP scheme at reconstructing the initial condition. However, in Figure 7.15 (left) we see that the QBD-RAP scheme out-performs the uniformisation scheme. This suggests that the QBD-RAP scheme is better able to resolve movement of mass across the domain when integrating over time compared to the uniformisation scheme.

Figure 7.15 (right) plots the cell-based error metric in (7.1)-(7.2). Once again, the DG-lim scheme does not appear to converge. Interestingly, the error curve for the DG scheme is not monotonic. Furthermore, the DG scheme can result in negative estimates of the cell probabilities in (7.1)-(7.2) (not shown). Of the other three positivity preserving schemes, the uniformisation, DG-lin-lim and QBD-RAP scheme all appear to converge, with the QBD-RAP and DG-lin-lim schemes performing similarly; even performing better than the DG scheme when the dimension is sufficiently large. The estimated rate of convergence of the QBD-RAP and DG-lin-lim schemes is approximately -4 for this example, which is significantly faster than the rate of convergence of any scheme in any of the other examples in this section (typically around -2 to -1).

Example 7.11. Now consider an initial distribution which is truncated Gaussian with mean 2.5 and standard deviation 0.5, and truncated at the boundaries, 0 and 10;

$$\mu([0,x)) := \mathbb{P}(X(0) \leq x, \varphi(0) = 1) = \frac{\Phi((x-2.5)/0.5) - \Phi(-2.5/0.5)}{\Phi(7.5/0.5) - \Phi(-2.5/0.5)}, \tag{7.3}$$

for $x \in [0, 10]$ and where $\Phi(x)$ is the CDF of the standard normal distribution.

At time $t = 4$ the distribution is truncated Gaussian with mean 6.5, standard deviation parameter 0.5, and is truncated below at 4.0 and above at 10, and there is also a small mass at the upper boundary;

$$\begin{aligned}&\mathbb{P}(X(4) \leq x, \varphi(4) = 1 \mid X(0) \sim \mu, \varphi(0) = 1)\\ =&\frac{\Phi((x-4-2.5)/0.5) - \Phi(-2.5/0.5)}{\Phi(7.5/0.5) - \Phi(-2.5/0.5)} 1(x \in [4, 10)) + 1(x \geq 10). \end{aligned} \tag{7.4}$$

At time $t = 4$ the mass at the boundary is approximately 1.28×10^{-12} and there is also a small discontinuity in the density at $x = 4$.

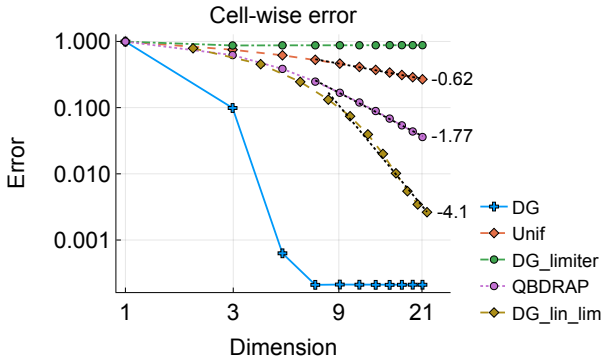

Figure 7.16: Cell-wise error defined in (7.1)-(7.2) for the travelling wave problem in Example 7.11. Plotted are the cell-wise errors for the DG (blue solid line), DG-lim (green dashed line), uniformisation (orange dashed line), QBD-RAP (purple dotted line) and DG-lin-lim (gold dashed line) schemes, versus the dimension of the approximation. The black dotted lines are linear least-squares fits to the last 8 data points and the slopes of the least square lines are written next to the last point.

Since the magnitude of the discontinuity at $t = 4$, $x = 4$ is small (much smaller than numerical integration errors in the evaluation of the error metrics), and the distribution is otherwise smooth, we expect that the DG scheme will perform well for this example. Figures 7.16 and 7.17 confirms that this is indeed the case. For all three error metrics the error obtained by the DG scheme rapidly decreases to a point where it is swamped by other numerical errors. This is characteristic of the DG scheme for the smooth problems we investigate throughout this chapter. For low order DG schemes there are regions where the approximate PDF is negative, however, as the order of the DG scheme increases, these regions disappear.

Interestingly, even though this is a relatively smooth problem, the DG-lim scheme does not perform well. It must be that the initial and transient distributions are 'sufficiently

pointy' that oscillations in the numerical solutions occur and the limiter reduces the order of the scheme to linear.

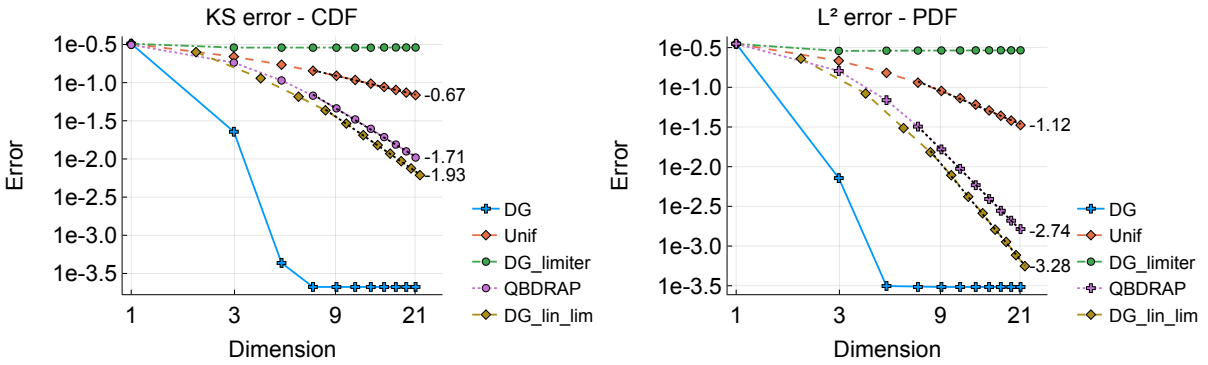

Figure 7.17: KS error (left) and L^2 error for the PDFs (right) for Example 7.11 where approximations were constructed via the DG (blue solid line), DG-lim (green dash-dotted line), uniformisation (orange dashed line), QBD-RAP (purple dotted line) and DG-lin-lim (gold dashed line) schemes. The black dotted lines are linear least-squares fits to the last 8 data points and the slopes of the least square lines are written next to the last point.

The other three positivity preserving schemes appear to converge, with the uniformisation scheme performing the worst and the DG-lin-lim scheme performing the best of the three for all error metrics. For the KS and L^2 error on the PDFs in Figure 7.17 the DG-lin-lim and QBD-RAP schemes perform similarly, however, for the cell-wise error in Figures 7.16 the DG-lin-lim scheme performs significantly better than the QBD-RAP scheme. Given that the DG-lin-lim scheme performed similarly to the QBD-RAP scheme for the previous examples, and also for the other error metrics for this example, this is unexpected.

To investigate this unexpected fact, in Figure 7.18 we plot the dimension 21 QBD-RAP and dimension 22 DG-lin-lim approximations to the PDF. Figure 7.18 shows that the DG-lin-lim scheme approximates the PDF well, except at the peak located at $x = 6.5$ (the limiter has affected the solution in this region). Specifically, the DG-lin-lim scheme underestimates the density in the peak, but overestimates the density either side of the peak. As a result, when we integrate over $x \in [6, 7]$ to compute the cell-wise probabilities, errors cancel over this region and the DG-lin-lim scheme produces an accurate estimate of the probability in this cell. In comparison, the QBD-RAP scheme also underestimates the density at the peak, but also within the entire interval $[6, 7]$, albeit less so at the edges than the middle. The QBD-RAP scheme also overestimates the density in the adjacent cells, $[5, 7]$ and $[7, 8]$. Hence, the overall cell-wise error is larger for the QBD-RAP scheme.

Example 7.12. We now want to look at how the schemes might handle a mass at the boundary. We introduce an ephemeral second phase into the model with phase transition

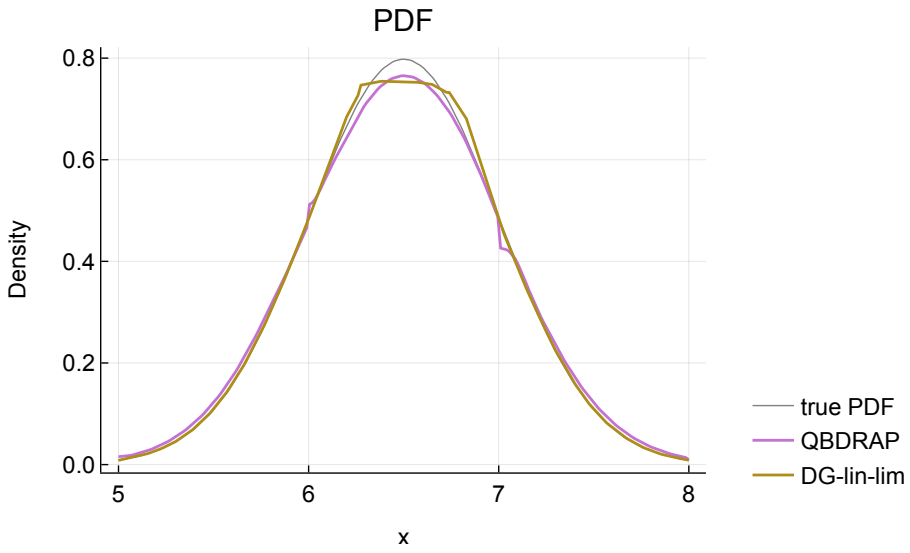

Figure 7.18: Dimension 21 QBD-RAP (purple) and dimension 22 DG-lin-lim (gold) approximations to the transient PDF at time $t = 4$ for Example 7.11. The true PDF is plotted in grey.

rate $T_{22} = -1$, and fluid rate $c_2 = 0$. The generator is therefore

$$T = \begin{bmatrix} 0 & 0 \\ 1 & -1 \end{bmatrix}.$$

We suppose that the initial condition is a point mass of 1 at the boundary $x = 0$ in Phase 2, i.e.

$$\mathbb{P}(X(0) \leq x, \varphi(0) = 2) = 1(0 \leq x).$$

With this initial condition, the transient distribution at time $t = 4$ is

$$\begin{aligned} &\mathbb{P}(X(4) \leq x, \varphi(4) = 1 \mid X(0) = 0, \varphi(0) = 2) \\ &= \left(e^{T_{22}(4-x)} - e^{T_{22}4}\right) 1(4 < x) + (1 - e^{T_{22}4}) 1(4 \geq x) \\ &= \left(e^{-(4-x)} - e^{-4}\right) 1(4 < x) + (1 - e^{-4}) 1(4 \geq x) \end{aligned} \tag{7.5}$$

and

$$\mathbb{P}(X(4) \leq x, \varphi(4) = 2 \mid X(0) = 0, \varphi(0) = 2) = e^{-4} 1(0 \leq x). \tag{7.6}$$

The PDF at $t = 4$ is discontinuous at $x = 4$, which is at the edge of a cell. For this problem, all schemes can represent the initial condition exactly.

Figure 7.19 plots the cell-wise error metric and Figure 7.20 plots the KS error (left) and the L^2 error between the PDFs (right). Observing Figures 7.19 and 7.20, the DG-lim

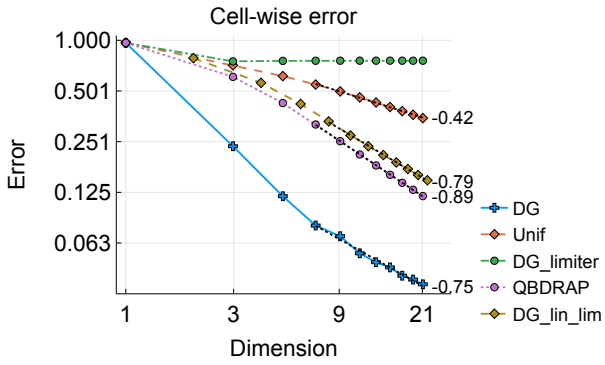

Figure 7.19: Cell-wise error for the travelling wave problem in Example 7.12. Plotted are the cell-wise errors for the DG (blue solid line), DG-lim (green dashed line), uniformisation (orange dashed line), QBD-RAP (purple dotted line) and DG-lin-lim (gold dashed line) schemes, versus the dimension of the approximation. The black dotted lines are linear least-squares fits to the last 8 data points and the slopes of the least square lines are written next to the last point.

scheme does not perform well, which might be expected due to the discontinuity in the transient distribution at $x = t < 10$ in Phase 1. The uniformisation, DG-lin-lim and QBD-RAP schemes appear to converge, with the QBD-RAP scheme converging fastest, but performing similarly to the DG-lin-lim scheme, and the uniformisation scheme performs the poorest of the three. The DG scheme performs the best, however, produces approximations which are negative and oscillatory as we might expect given the discontinuity. A selection of approximations to the transient PDF are shown in Figure 7.21.

Interestingly, this solution is discontinuous at a cell-edge and even though the DG scheme[§] allows for discontinuities in the PDF at cell edges, it is unable to capture this discontinuity without producing oscillations.

Summary

We applied the DG, DG-lim, DG-lin-lim, uniformisation and QBD-RAP schemes to travelling wave problems with various initial conditions. All the problems induced oscillations in the DG scheme at some point and as a result the DG-lim scheme reduced the order of the approximation to linear. In some cases, the oscillations in the DG scheme appeared to be transient and were not material in the final approximation. In other cases, the oscillations were present in the final DG approximation causing the approximation to violate the axioms of probability. Of the three positivity preserving alternatives, the DG-lin-lim, uniformisation and QBD-RAP schemes, in general, the uniformisation scheme performs

[§]All the schemes considered allow discontinuities at a cell-edges.

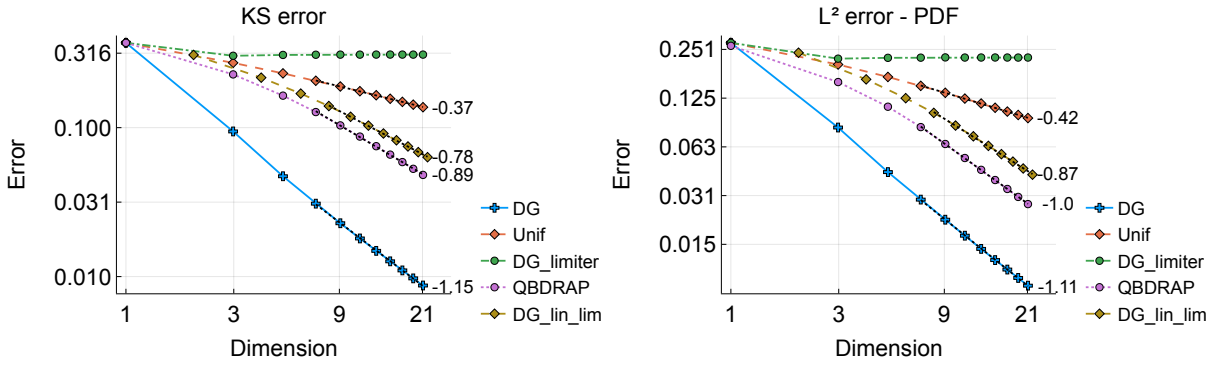

Figure 7.20: KS error (left) and L^2 error of the PDF (right) for the travelling wave problem in Example 7.12 where approximations are constructed via the DG (blue solid line), DG-lim (green dash-dotted line), uniformisation (orange dashed line), QBD-RAP (purple dotted line) and DG-lin-lim (gold dashed line) schemes. The black dotted lines are linear least-squares fits to the last 8 data points and the slopes of the least square lines are written next to the last point.

worse than the other two and converges slowly to the true solution. The DG-lin-lim and QBD-RAP schemes perform similarly with the DG-lin-lim scheme performing slightly better for the smoother problems and the QBD-RAP scheme performing slightly better for the problems with more significant discontinuities and point masses.

This section is the first step in analysing the ability of the approximation schemes to capture the dynamics fluid queues. This section demonstrates that, although the QBD-RAP scheme may have certain deficiencies for solution reconstruction (as demonstrated in Section 7.2), it can resolve the movement of probability between cells better than the uniformisation scheme, and similarly compared DG-lin-lim scheme. In the following sections we consider more complicated fluid queue dynamics by introducing some simple stochastic phase changes.

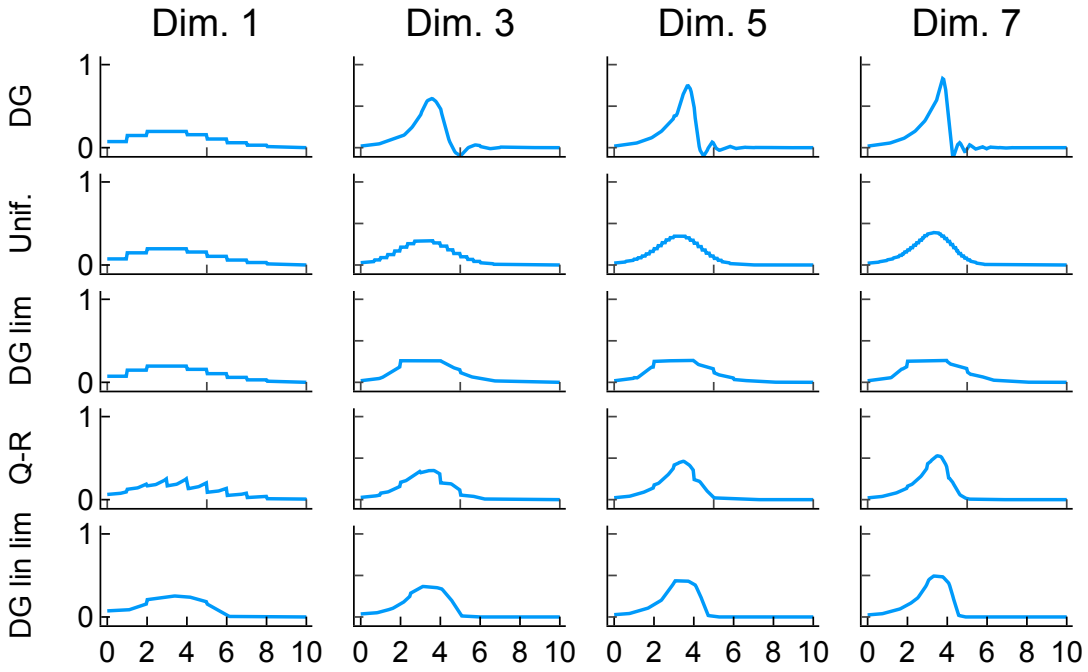

Figure 7.21: Approximate transient PDFs at time $t = 4$ for Example 7.12 using the DG (top row), uniformisation (second row), DG-lim (third row), QBD-RAP (fourth row), and DG-lin-lim (fifth row) schemes of dimensions 1, 3, 5, and 7 (columns). The true density function is $e^{-(x-4)}1(x < 4)$.

7.4 Limiting distributions

We briefly turn our attention to limiting distributions of fluid queues. Since the limiting distribution is smooth we expect that the DG scheme will work well here. Moreover, since the solution is smooth, the slope limiter has no effect which means that the DG-lim and DG schemes are equivalent. Analysing the performance of the approximation schemes with the limiting distribution allows us to explore the ability of the schemes to capture the stochastic dynamics, without numerical integration with respect to time. In addition, the limiting distribution is convenient for numerical experiments as we can evaluate it analytically (Sonenberg 2017), and it does not require us to approximate initial conditions.

Here we analyse a simple model which is based on Example 2 in Bean et al. (2009*b*), except here we add a lower boundary to the model (no lower boundary is specified in Example 2 of Bean et al. (2009*b*) as it is inconsequential to their analysis).

Model 7.13. *Consider a fluid queue where the driving process is a CTMC with state*

space $\mathcal{S} = \{1, 2\}$*, generator*

$$\boldsymbol{T} = \begin{bmatrix} -1.1 & 1.1 \\ 1 & -1 \end{bmatrix},$$

and there are associated rates $c_1 = 1, c_2 = -1$*, and boundaries at* $x = 0$ *and* $x = 10$*. We specify two types of behaviour at the boundary.*

Upon hitting the lower boundary, the process immediately transitions from Phase 2 *to Phase* j *with probability* p_{2j} *where*

$$\begin{bmatrix} p_{21} & p_{22} \end{bmatrix} = \begin{bmatrix} 1 & 0 \end{bmatrix}.$$

Upon hitting the upper boundary, the process immediately transitions from Phase 1 *to Phase* j *with probability* p_{1j} *where*

$$\begin{bmatrix} p_{11} & p_{12} \end{bmatrix} = \begin{bmatrix} 0 & 1 \end{bmatrix}.$$

Thus, upon hitting the boundary, the fluid queue immediately transitions to the other phase and is reflected.

The model was discretised using the DG, uniformisation and QBD-RAP schemes using ten cells of width 1. We compute the coefficients for the limiting distribution in the following way. Suppose that a given discretisation scheme results in an approximation to the generator of the fluid queue as a matrix $\boldsymbol{B}$. Then the limiting coefficients are found by solving

$$\boldsymbol{b}\boldsymbol{B} = 0, \tag{7.7}$$

$$\text{such that } \boldsymbol{b}\boldsymbol{1} = 1, \tag{7.8}$$

for the coefficients $\boldsymbol{b}$.

Figure 7.22 plots KS errors between the true limiting CDF and the approximations and the L^2 error between the true limiting PDF and the approximations. Clearly the DG scheme is superior here as its error rapidly decreases to a point where it becomes insignificant compared to other numerical errors. The QBD-RAP and uniformisation schemes both appear to be converging, with the errors for the QBD-RAP scheme decreasing faster.

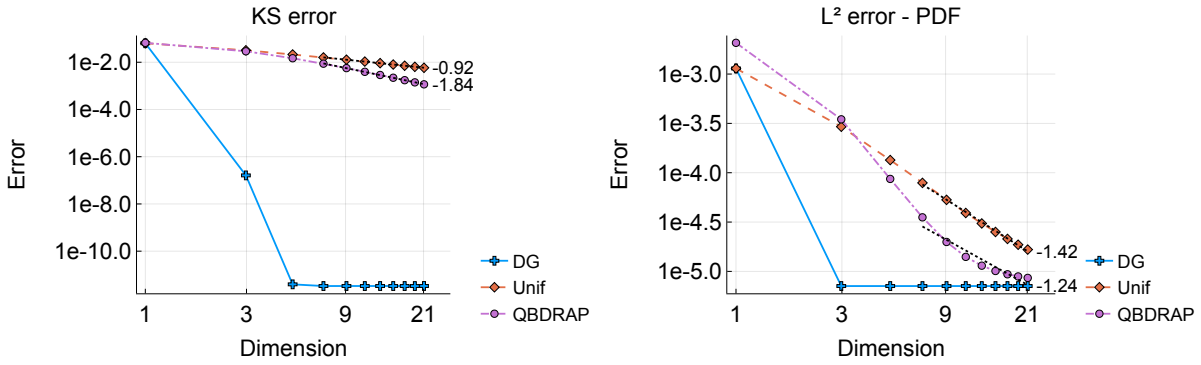

Figure 7.22: KS error (left) and L^2 error of the PDF (right) for Model 7.13 using the DG (blue solid line), uniformisation (orange dashed line) and QBD-RAP (purple dotted line) schemes. The black dotted lines are linear least-squares fits to the last 8 data points and the slopes of the least square lines are written next to the last point.

7.5 Transient distributions

Once again we consider Model 7.13 and use the same spatial discretisation as described in Section 7.4 (ten cells with width $\Delta = 1$). Two initial conditions are considered, a point mass with mass 1 at 0 in Phase 1, and the initial distribution with PDF

$$\frac{1}{2}e^{-x}/(1-e^{-10}) \tag{7.9}$$

in Phases 1 and 2, with no mass at the boundaries. We numerically integrate over time until time $t = 2.0$ using the SSPRK4 method with t-step size 0.005.[¶] Here we use the DG scheme without a limiter and the DG-lin-lim scheme. Error plots for the DG-lim scheme are included for completeness, but we do not comment on them in detail in this section due to their poor performance for discontinuous problems which we noted previously.

To obtain a *ground truth* 5×10^{10} realisations of the fluid queue were simulated until $t = 2$, then the empirical CDF and the masses within each cell were computed from the simulations. We denote the empirical probability distribution by $\mathbb{P}_{sim}$. We then compute the KS and L^1 error metrics between the approximated and simulated CDFs, as well as the cell-wise error metric, given as follows.

$$\sum_{j\in\{1,2\}}\sum_{\ell=1}^{10}|\mathbb{P}_{sim}(X(2)\in\mathcal{D}_{\ell,j},\varphi(2)=j\mid X(0)=0.5,\varphi(0)=1)-p(2,\ell,j)|$$
$$+|\mathbb{P}_{sim}(X(2)\in\{10\},\varphi(2)=4\mid X(0),\varphi(0))-p(2,11,4)|$$

[¶]The t-step size must be chosen to ensure that numerical integration over time is stable up to dimension 21, adhering to a CFL-like condition (Hesthaven & Warburton 2007, Section 4.8).

$$+ |\mathbb{P}_{sim} X(2) \in \{0\}, \varphi(2) = 3 \mid X(0), \varphi(0)) - p(2, 0, 3)| \tag{7.10}$$

where $p(2, \ell, j)$ is an approximation to $\mathbb{P}_{sim}(X(2) \in \mathcal{D}_{\ell,j}, \varphi(2) = j \mid X(0), \varphi(0))$, $p(2, 11, 4)$ is an approximation to $\mathbb{P}_{sim}(X(2) \in \{10\}, \varphi(2) = 4 \mid X(0), \varphi(0))$, and $p(2, 0, 3)$ is an approximation to $\mathbb{P}_{sim}(X(2) \in \{0\}, \varphi(2) = 3 \mid X(0), \varphi(0))$.

To account for possible Monte-Carlo error, we used a bootstrap with 1,000 bootstrap samples. For the bootstrap we sample 5×10^{10} realisations of the fluid queue with replacement from the original 5×10^{10} samples, then compute error metrics with the resampled data. We resample 1,000 times, resulting in 1,000 estimates of the errors. Via the bootstrap, we report the 5th and 95th percentile of the sampling distribution of the errors.

To evaluate error metrics, we use a grid of 10,001 evenly spaced points for each phase.

To approximate the point mass initial condition we compute the initial coefficients for each scheme exactly. For the exponential initial condition in (7.9) we compute the initial coefficients via Gauss-Lobatto quadrature for the DG scheme, by using the mid-point rule for the uniformisation scheme, and by using a trapezoidal rule with 2,001 points on each cell for the QBD-RAP scheme.

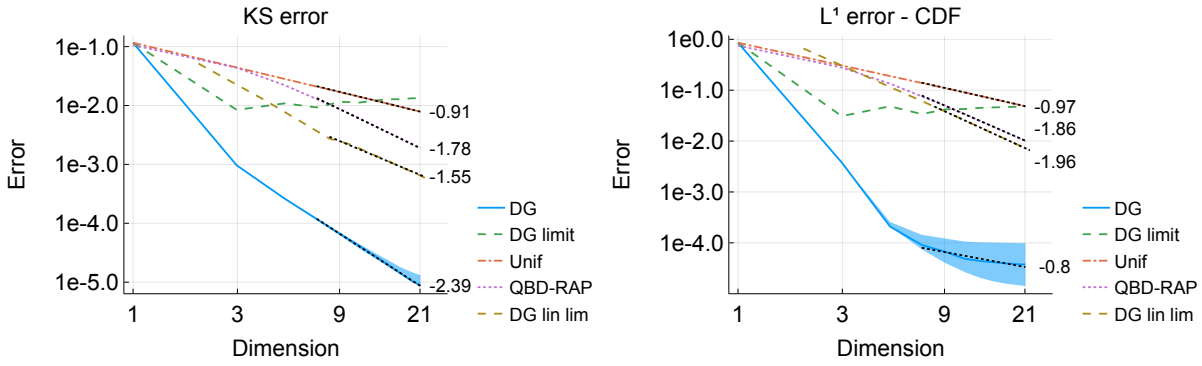

Figure 7.23: KS (left) and L^1 (right) errors between the true transient CDF at time $t = 2$ for Model 7.13 with the exponential initial condition, where the approximations were obtained via the DG (blue solid line), DG-lim (green dashed line), uniformisation (orange dashed line), QBD-RAP (purple dotted line) and DG-lin-lim (gold dashed line) schemes. Bootstrapped 90% confidence intervals are shown by the lighter coloured strip surrounding the lines. The black dotted lines are linear least-squares fits to the last 8 data points and the slopes of the least square lines are written next to the last point.

Model 7.13 with exponential initial condition Figure 7.23 shows the KS and L^1 metric on the CDF for the five different schemes (DG, DG-lim, DG-lin-lim, uniformisation and QBD-RAP schemes). For both error metrics the DG scheme converges rapidly until computational/simulation errors become significant. The uniformisation, DG-lin-lim and

QBD-RAP schemes converge at a slower rate than the DG scheme, with the QBD-RAP and DG-lin-lim schemes converging faster than the uniformisation scheme. For the KS metric, the DG-lin-lim scheme outperforms the QBD-RAP scheme at all dimensions, while for the L^1 metric, the two schemes perform similarly.

The DG-lim does not appear to be converging, which suggests there is at least one iteration during the numerical integration over time at which the numerical solution displays oscillations. However, when the DG approximations for the distribution at time $t = 2$ are plotted, they do not appear to display oscillations (not shown). This suggests that the oscillations which might occur with the DG scheme are transient, and have dissipated by time $t = 2$. The oscillations which the limiter detects in the DG scheme are likely from the reflecting boundary at $x = 0$. In the early stages of the evolution of the model the majority of the density in Phase 2 near zero will move to the left at rate 1 and hit the boundary at $x = 0$. Upon hitting the boundary, the density is reflected into Phase 1. Meanwhile, the initial density in Phase 1 moves to the right at rate 1 and the density reflected at the boundary from Phase 2 fills in to the region between $x = 0$ and $x = t$ in Phase 1. Intuitively, in the early evolution of the model, there will be a sharp peak in the transient density at $x = t$. The true transient density is continuous, but not differentiable, at the point $x = t$ for $t < 10$.

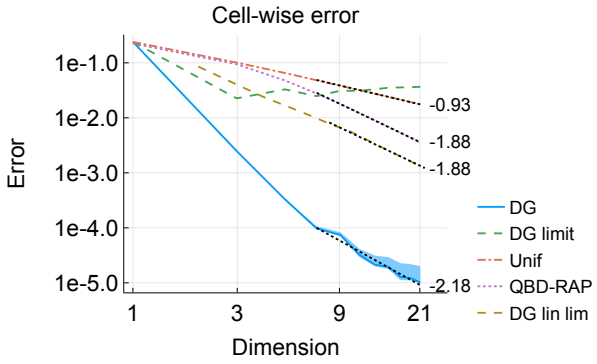

Figure 7.24: Cell-wise error metric from (7.10) for Model 7.13 with the exponential initial condition, where the approximations were obtained via the DG (blue solid line), DG-lim (green dashed line), uniformisation (orange dashed line), QBD-RAP (purple dotted line) and DG-lin-lim (gold dashed line) schemes. Bootstrapped 90% confidence intervals are shown by the lighter coloured bars surrounding the lines. The black dotted lines are linear least-squares fits to the last 8 data points and the slopes of the least square lines are written next to the last point.

Figure 7.24 plots the cell-wise error metric (7.10) for Model 7.13 with the exponential initial condition. Since the cell-wise errors do not require us to reconstruct the value of the solution within each cell, this metric allows us to observe the error characteristics of

the schemes without reconstruction. Figure 7.24 shows similar convergence characteristics to the KS error metric in Figure 7.23 (left).

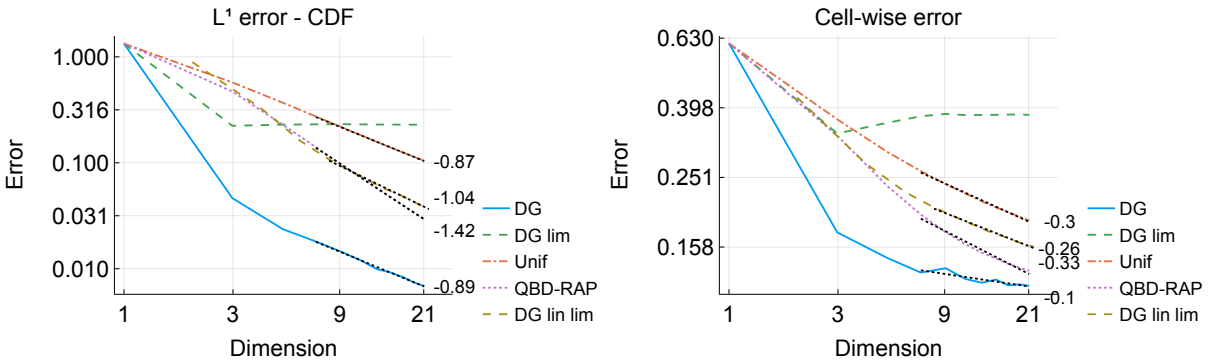

Figure 7.25: L^1 errors between the CDF, and cell-wise errors, at time $t = 2$ for Model 7.13 with the point-mass initial condition, where the approximations were obtained via the DG (blue solid line), DG-lim (green dashed line), uniformisation (orange dashed line), QBD-RAP (purple dotted line) and DG-lin-lim (gold dashed line) schemes. Bootstrapped 90% confidence intervals are shown by the lighter coloured bars surrounding the lines. The black dotted lines are linear least-squares fits to the last 8 data points and the slopes of the least square lines are written next to the last point.

Model 7.13 with a point-mass initial condition Figure 7.25 (left) shows L^1 error on the CDF for the five different schemes (DG, DG-lim, DG-lin-lim, uniformisation and QBD-RAP schemes). Comparing the L^1 error metrics in Figure 7.25 for the point mass initial condition with the ones in Figure 7.23 for the exponential initial condition all schemes perform worse for the point mass initial condition. Regarding comparative convergence of error for this problem, for both error metrics the DG scheme converges fastest, followed by the QBD-RAP and DG-lin-lim schemes which converge comparatively, then the uniformisation scheme. For the DG scheme, the rate at which the error decreases as the dimension of the scheme increases slows significantly after dimension 3. This is likely due to the DG scheme approximating smooth regions of the solution well when the dimension is 3 or greater, after which the majority of the error is due to the discontinuity in the solution, for which the rate of convergence is slower.

Figure 7.25 (right) plots the cell-wise error metric (7.10) for Model 7.13 with the point mass initial condition. Figure 7.25 (right) suggests that approximating the cell-wise probabilities seems to be a difficult problem. This is likely caused by the discontinuity at $x = 2$ in Phase 1, which lies exactly on a cell boundary. To investigate how the position of the discontinuity might affect the error metrics we evolved the model to time $t = 2.1$ and computed the L^1 error between the CDFs and the cell-wise error. Once again, we

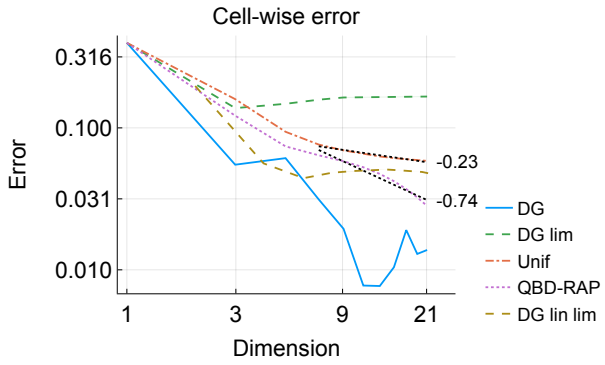

Figure 7.26: Cell-wise errors for Model 7.13 at time $t = 2.1$ with the point mass initial condition, where the approximations were obtained via the DG (blue solid line), DG-lim (green dashed line), uniformisation (orange dashed line), QBD-RAP (purple dotted line) and DG-lin-lim (gold dashed line) schemes. Bootstrapped 90% confidence intervals are shown by the lighter coloured bars surrounding the lines. The black dotted lines are linear least-squares fits to the last 8 data points and the slopes of the least square lines are written next to the last point.

use simulation and bootstrapping to approximate the true distribution. The plot (not shown) of the L^1 error between the CDFs for the model at times $t = 2$ and $t = 2.1$ is relatively similar to the one shown and not noteworthy. The plot of the cell-wise error metric at time $t = 2.1$ in Figure 7.26 is somewhat more interesting. Firstly, Figure 7.26 shows that all schemes improved from time $t = 2$ to time $t = 2.1$ with respect to this metric. Figure 7.26 also shows that the cell-wise error metric is somewhat volatile for the DG scheme; this is due to the oscillations in the DG approximation. In contrast, in Figure 7.26 the uniformisation and QBD-RAP schemes have monotonically decreasing error curves. Also interesting is the DG-lin-lim scheme, which does not improve past dimension 7. It seems that the discontinuity at $x = t$ is not easily resolved by the scheme given the presence of the limiter. Figure 7.27 plots the CDFs in Phase 1 around the discontinuity in the transient distribution at time $t = 2.1$ for the dimension 21 QBD-RAP and dimension 22 DG-lin-lim schemes. Neither scheme does a particularly good job of capturing the discontinuity, but the QBD-RAP scheme appears to perform slightly better.

Summary

For the exponential initial condition the DG scheme performed well, producing the lowest errors. This was followed by the DG-lin-lim scheme, then the QBD-RAP scheme, then the uniformisation scheme, while the DG-lim scheme did not perform well at all. The fact that the DG-lim scheme did not perform well suggests that there must have been

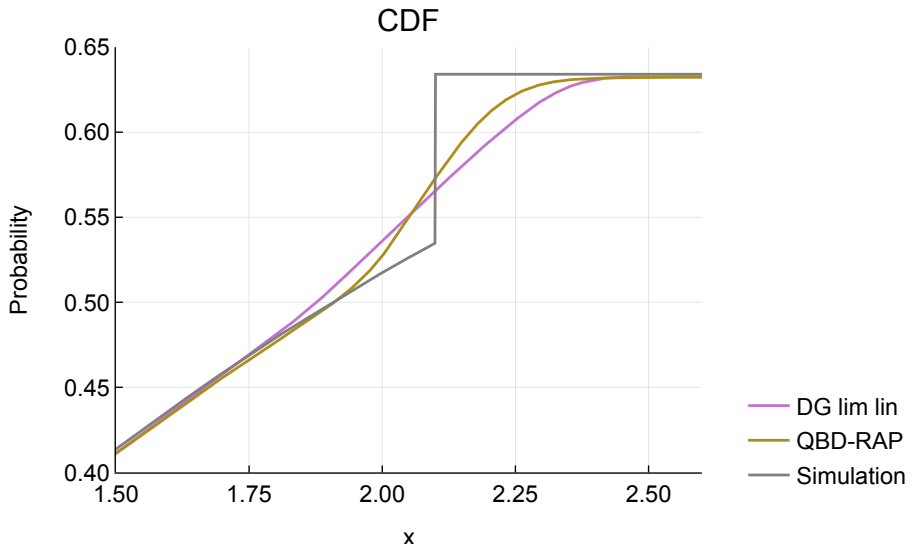

Figure 7.27: Dimension 21 QBD-RAP and dimension 22 DG-lin-lim approximations to the CDF in Phase 1 for Model 7.13 with a point-mass initial condition. An empirical estimate of the CDF obtained via simulation is plotted in grey.

oscillations in the DG solution at some point, although these oscillations are not obvious in the DG approximation at time $t = 2$. In general the methods all performed better for the exponential initial condition than for the point mass initial condition. For the point mass initial condition the DG scheme produced oscillatory solutions. Of the viable positivity preserving methods the uniformisation scheme appears to be converging but at a slower rate than the DG-lin-lim and QBD-RAP schemes. Of the two best performing positivity preserving schemes, the QBD-RAP scheme performed slightly better for the point mass initial condition, while the DG-lin-lim scheme performed slightly better for the exponential initial condition.

This section investigated the performance of the approximation schemes for a simple fluid queue. In the next section we analyse the approximation schemes in more detail by investigating their ability to approximate hitting times of fluid queues.

7.6 Hitting times

We can gain further insight into the approximation schemes by investigating how they perform approximating hitting times of fluid queues. This is also an important aspect to consider if we wish to apply the schemes to approximate operators from the analysis fluid-fluid queues. Recall, in the analysis of fluid-fluid queues in Section 2.4, we partition sample paths of the second fluid into periods where it is either increasing, decreasing,

or constant. The position of $(\varphi(t), X(t))$ determines the rate at which the second fluid moves, hence the hitting times on the boundaries of the sets $\mathcal{F}_i^+, \mathcal{F}_i^-$, and $\mathcal{F}_i^0$ determine the periods of time when the second fluid is either increasing, decreasing, or constant.

Once again we consider Model 7.13. Let $\zeta_X(\{0,1\}) = \{\inf t > 0 \mid X(t) = 0, \text{ or } X(t) = 1\}$, be the first hitting time of $\{X(t)\}$ on the set $\{0,1\}$. The distribution of the hitting time in phase $i \in \{1,2\}$ is

$$\mathbb{P}(\zeta_X(\{0,1\}) < t, \varphi(t) = i \mid \boldsymbol{X}(0) \sim \mu), \tag{7.11}$$

for some initial distribution μ. We look at two initial conditions; an exponential with equal mass in each phase,

$$\mathbb{P}(X(0) \in \mathrm{d}x, \varphi(0) = i) = \frac{1}{2}\frac{e^{-x}}{(1-e^{-1})}, \; i \in \{1,2\},$$

and a point mass at $X(0) = 0$ in phase $\varphi(0) = 1$.

We partition the interval $[0,1]$ into three intervals of width $1/3$. To capture the mass which has hit the set $\{0,1\}$, we suppose that when the process hits $\{0\}$ or $\{1\}$ it is absorbed forever and remains in the phase in which the process first hit $\{0\}$ or $\{1\}$. We integrate the schemes until time $t = 10$ using the SSPRK4 method with t-step size $0.005/3$.[‖] We use the DG scheme with and without a slope limiter during the time-integration and implement both the DG-lim and DG-lin-lim schemes. At each time-step of the numerical integration, we record the amount of mass at the absorbing boundaries in each phase, which gives us an approximation of the cumulative distribution function of the hitting time in each phase, up to time $t = 10$.

As a *ground truth* we simulate 5×10^{10} realisations and record the hitting time on the set $\{0,1\}$ and the phase at the time of hitting. We then compute the empirical CDF of the hitting probabilities

$$\mathbb{P}(\zeta_X(\{0,1\}) < t, \varphi(t) = i \mid \boldsymbol{X}(0) \sim \mu),$$

for $t = 0.005/3 \times k$, $k = 0, ..., 6000$.

To account for Monte-Carlo errors we take 1,000 bootstrap resamples of the original 5×10^{10} samples and compute the empirical CDF of the hitting times for each bootstrap sample. For each bootstrap sample, we resample 5×10^{10} points with replacement from the original 5×10^{10} realisations, then compute the error metrics between the empirical and approximate CDFs and, ultimately, estimate the 5th and 95th percentile of the distribution of the errors.

[‖] Since we use a smaller cell-width in this example than in previous examples we need to reduce the t-step size accordingly to ensure that numerical integration is stable for schemes up to dimension 21, adhering to a CFL-like condition (Hesthaven & Warburton 2007, Section 4.8).

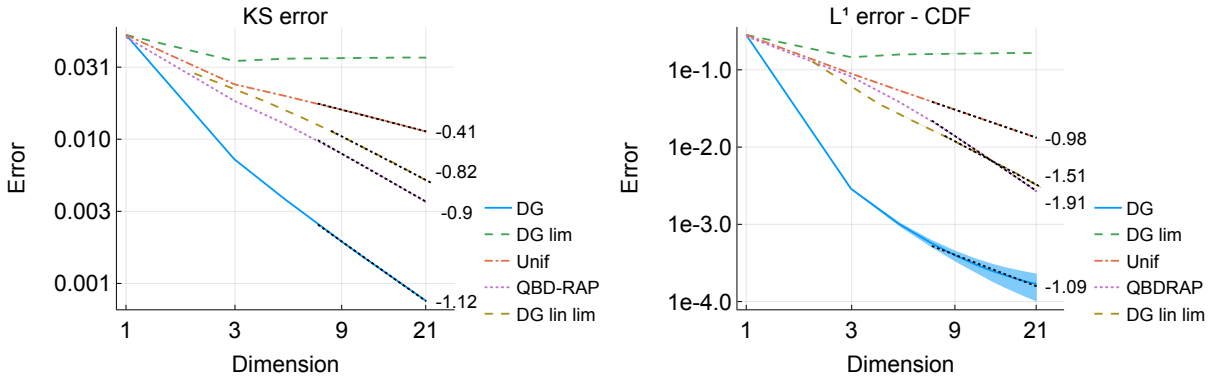

Figure 7.28: KS (left) and L^1 (right) errors between the simulated and approximate first hitting time CDFs in (7.11) for Model 7.13 with the exponential initial condition. The approximations were obtained via the DG (blue solid line), DG-lim (orange dashed line), uniformisation (green dashed line), QBD-RAP (purple dotted line) and DG-lin-lim (gold dashed line) schemes. Bootstrapped 90% confidence intervals are shown by the lighter coloured bars surrounding the lines. The black dotted lines are linear least-squares fits to the last 8 data points and the slopes of the least square lines are written next to the last point.

Exponential initial condition Figure 7.28 shows the error metrics recorded for the five numerical approximation schemes applied to the hitting time problem with the exponential initial condition. For both error metrics the DG scheme performs the best.

The DG-lim scheme does not appear to converge, suggesting oscillations in the DG approximation. With this initial condition the transient distribution of the fluid queue on the event that it remains in the interval $(0, 1)$ is discontinuous at $x = t$ in Phase 1 and $x = 1 - t$ in Phase 2 for $x, t < 1$. There is also a discontinuity in the hitting time PDF at time $t = 1$.

Regarding the other positivity preserving schemes, the uniformisation, DG-lin-lim and QBD-RAP schemes, all appear to converge, with the QBD-RAP and DG-lin-lim schemes performing similarly and better than the uniformisation scheme.

In Figure 7.29 we plot the PDFs of the hitting times $\zeta_X(\{0, 1\})$, in each phase, for the dimension 21 DG, dimension 21, QBD-RAP and dimension 22 DG-lin-lim schemes. Figure 7.29 demonstrates oscillations in the DG approximation, though the PDF is not negative in this case. All schemes seem to capture the discontinuity at $t = 1$ relatively well. There is an interesting artefact in the QBD-RAP approximation around $t = 0$ where the scheme has generated an oscillation. Let us investigate this artefact further.

Intuitively, most sample paths which exit near $t = 0$ start near a boundary, they see a change of phase shortly after $t = 0$, then remain in that phase until hitting $\{0, 1\}$. Consider such a sample path which starts at $x_0 = 0$ in Phase 1 and exits at time v with a

single change of phase at time $u \in (0, v)$. For the fluid queue, $u = v/2$, and we want the QBD-RAP scheme to approximate this. The QBD-RAP approximation to this sample paths has density**

$$\boldsymbol{\alpha} e^{(\boldsymbol{S}-\boldsymbol{I})u} \boldsymbol{D} e^{(\boldsymbol{S}-1.1\boldsymbol{I})(v-u)} \boldsymbol{s} = \boldsymbol{\alpha} e^{\boldsymbol{S}u} \boldsymbol{D} e^{\boldsymbol{S}(v-u)} \boldsymbol{s} e^{-1.1v+0.1u}. \tag{7.12}$$

We want (7.12) to approximate a point mass at $v = 2u$. To see that this is the case, recall that we can approximate a point mass at u by

$$\boldsymbol{k}(\Delta - u) \boldsymbol{e}^{\boldsymbol{S}x} \boldsymbol{s}, \tag{7.13}$$

where $\boldsymbol{k}(z) = \dfrac{\boldsymbol{\alpha} e^{\boldsymbol{S}z}}{\boldsymbol{\alpha} e^{\boldsymbol{S}z} \boldsymbol{e}}$, and recall that the matrix $\boldsymbol{D}$ is an approximation;

$$\begin{aligned} \boldsymbol{\alpha} e^{\boldsymbol{S}u} \boldsymbol{D} &= \mathbb{E}\left[\boldsymbol{k}(Z-u) 1(Z > u)\right] \\ &\approx \mathbb{P}(Z > u) \boldsymbol{k}(\Delta - u), \end{aligned} \tag{7.14}$$

where $Z \sim ME(\boldsymbol{\alpha}, \boldsymbol{S})$ and Z is concentrated around Δ.

Thus, (7.12) is approximately

$$e^{-1.1v+0.1u} \mathbb{P}(Z > u) \boldsymbol{k}(\Delta - u) \boldsymbol{e}^{\boldsymbol{S}(v-u)} \boldsymbol{s},$$

which approximates a point mass at $v - u = u$, or $2u = v$. Retracing our arguments, sources of error in this approximation come from the approximation in (7.14) and from the approximation of a point mass by (7.13), which we commented on in Section 7.2.1 where we conducted numerical experiments about the unnormalised closing operator given by $\boldsymbol{v}(x) = e^{\boldsymbol{S}x} \boldsymbol{s}$. Indeed, we saw in Example 7.4 that the unnormalised closing operator performed poorly for the exponential initial condition. The phenomenon discussed above is also related to Example 7.2, where the reconstructions for the QBD-RAP scheme failed to capture mass near the right-hand edge of the interval.

In Figure 7.29 there is another small oscillation present in the QBD-RAP approximation around $t = 0.35$ in both phases. The cause of this oscillation is likely related to the discussion above. Consider a sample path which starts at $x = 0.34$, just to the right of the cell-edge at $1/3$, in Phase 1, changes phase once at time u, then remains in that phase until the boundary at $x = 0$ is hit. For the fluid queue, the hitting time of this sample path is $0.34 + 2u$. The QBD-RAP approximation to this sample paths has density

$$\int_{v=u}^{\infty} \boldsymbol{\alpha} e^{(\boldsymbol{S}-\boldsymbol{I})u} \boldsymbol{D} e^{(\boldsymbol{S}-1.1\boldsymbol{I})(v-u)} \boldsymbol{s} \boldsymbol{\alpha} e^{(\boldsymbol{S}-1.1\boldsymbol{I})w} \boldsymbol{s} \, \mathrm{d}v. \tag{7.15}$$

The first term in (7.15) is the same as the first term in (7.12), and this is likely the source of the oscillation.

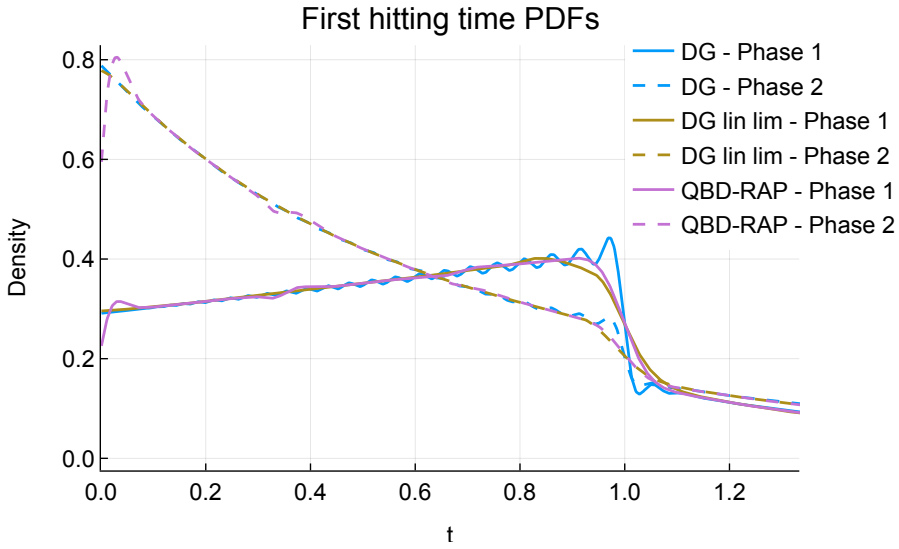

Figure 7.29: Approximations of the PDF of the first hitting time for Model 7.13 with the exponential initial condition. The blue lines were obtained from the dimension-21 DG scheme, the purple lines were obtained from the dimension-21 QBD-RAP scheme and the gold lines were obtained from the dimension 22 DG-lin-lim scheme. The DG scheme displays oscillations around the discontinuities at $t = 1$. The QBD-RAP scheme has oscillations near $t = 0$ and $t = 0.4$.

Point mass initial condition Figure 7.30 shows the L^1 error for the CDF for the five numerical approximation schemes applied to the hitting time problem with the point mass initial condition. Due to the discontinuity the KS error is not very insightful for this problem.

Observing Figure 7.30, the DG-lim scheme does not appear to converge, suggesting oscillations in the DG approximation as we might expect given the discontinuity. Of the other positivity preserving schemes, the uniformisation, DG-lin-lim and QBD-RAP schemes all appear to converge, with the QBD-RAP converging fastest and the DG-lin-lim and uniformisation schemes performing similarly.

The error for the DG scheme reduces significantly between dimension 1 and dimension 3, then reduces at a slower rate after that. For dimension 3 and above the DG scheme approximates the smooth regions of the solution well and the majority of the error is from the region around the discontinuity, which reduces at a slower rate. The DG scheme appears to perform best, however, oscillations are present in the DG approximation. In Figure 7.31 we plot the CDFs of the hitting time for Phase 1 obtained via the DG and QBD-RAP schemes, each using a dimension 21 basis, and for the DG-lin-lim scheme with

**One significant advantage of the QBD-RAP and uniformisation approaches is that, from their stochastic interpretation, we can use sample-path arguments to analyse the scheme.

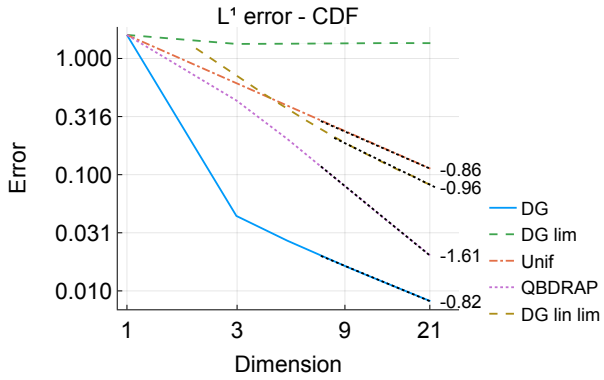

Figure 7.30: L^1 errors between the simulated and approximate first hitting time CDFs in (7.11) for Model 7.13 with the point mass initial condition. The approximations were obtained via the DG (blue solid line), DG-lim (orange dashed line), uniformisation (green dashed line), QBD-RAP (purple dotted line) and DG-lin-lim (gold dashed line) schemes. Bootstrapped 90% confidence intervals are shown by the lighter coloured bars surrounding the lines. The black dotted lines are linear least-squares fits to the last 8 data points and the slopes of the least square lines are written next to the last point.

dimension 22. Observing Figure 7.31 the DG approximation displays oscillations around the discontinuity at $t = 1$ while the other two schemes do not. Of the two positivity preserving schemes in Figure 7.31, the QBD-RAP scheme appears to approximate the discontinuity best.

Summary

The results of this experiment are somewhat similar to the last experiment on transient distributions; the DG-lim scheme performs poorly, the DG scheme exhibits oscillations, of the three other positivity preserving schemes, the uniformisation scheme performs poorest, while the QBD-RAP and DG-lin-lim schemes perform similarly, with the QBD-RAP scheme performing slightly better for the point mass initial condition and the DG-lin-lim scheme performing slightly better for the exponential initial condition.

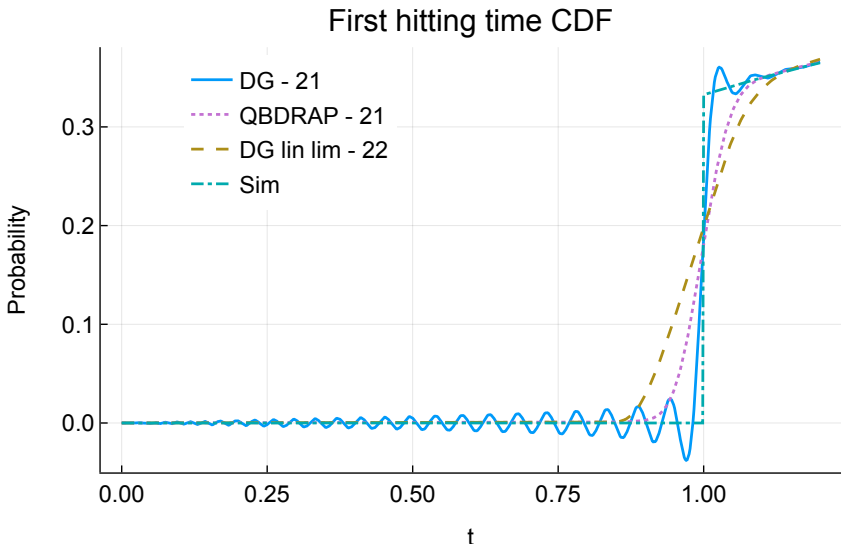

Figure 7.31: Approximations of the CDF of the first hitting time in Phase 1 for Model 7.13 with the point mass initial condition. The blue line was obtained from the dimension-21 DG scheme, the purple dotted line from the dimension-21 QBD-RAP scheme, and the gold dashed line from the dimension 22 DG-lin-lim scheme. The empirical CDF obtained via simulation is plotted as the teal dashed line. The DG scheme displays negative approximations to probabilities.

7.7 First-return distributions of fluid-fluid queues

Now consider the fluid-fluid queue first analysed in Bean et al. (2022) which is a modified version of a model first presented in Latouche et al. (2013). We refer the reader to Bean et al. (2022) for more on the DG scheme applied to this model.

Model 7.14. *Consider a stochastic fluid-fluid queue* $\{(Y(t), X(t), \varphi(t))\}_{t\geq 0}$, *where* $\{X(t)\}$ *and* $\{Y(t)\}$ *represent the workloads in Buffers 1 and 2 at time* $t \geq 0$, *respectively, both driven by the phase* $\{\varphi(t)\}$, *which is a Markov chain on the state space* $\mathcal{S} = \{11, 10, 01, 00\}$. *Both* $\{X(t)\}$ *and* $\{Y(t)\}$ *have a regulated boundary at 0. Here, the state* 11 *indicates inputs to both buffers being* ON, *the state* 00 *indicates both being* OFF, *the state* 10 *is when only the first input is* ON, *and the state* 01 *is when only the second is* ON. *The input of Buffer* k *is switched from* ON *to* OFF *with rate* γ_k, *and from* OFF *to* ON *with rate* β_k, *for* $k = 1, 2$. *Thus, the infinitesimal generator* T *for* $\varphi(t)$ *is given by*

$$T = \begin{bmatrix} -(\gamma_1+\gamma_2) & \gamma_2 & \gamma_1 & 0 \\ \beta_2 & -(\gamma_1+\beta_2) & 0 & \gamma_1 \\ \beta_1 & 0 & -(\gamma_2+\beta_1) & \gamma_2 \\ 0 & \beta_1 & \beta_2 & -(\beta_1+\beta_2) \end{bmatrix}.$$

The net rates of change for $X(t)$, denoted c_i, are given by

$$(c_{11}, c_{10}, c_{01}, c_{00}) = \ (\lambda_1 - \theta_1, \ \ \lambda_1 - \theta_1, \ \ -\theta_1, \ \ -\theta_1),$$

and the net rates of change for $Y(t)$, denoted r_i, are as follows

$$(r_{11}, r_{10}, r_{01}, r_{00}) = \begin{cases} (\lambda_2 - \kappa, & -\kappa, & \lambda_2 - \kappa, & -\kappa) & \text{if } X_t = 0, \\ (\lambda_2 - \theta_2, & -\theta_2, & \lambda_2 - \theta_2, & -\theta_2) & \text{if } X_t \in (0, x^*), \\ (\quad \lambda_2, & 0, & \lambda_2, & 0) & \text{if } X_t \geq x^*. \end{cases}$$

For our numerical experiments, we use the parameter choices given in (Latouche et al. 2013):

$$\gamma_1 = 11, \quad \beta_1 = 1, \quad \lambda_1 = 12.48, \quad \theta_1 = 1.6, \quad \kappa = 2.6, \tag{7.16}$$

$$\gamma_2 = 22, \quad \beta_2 = 1, \quad \lambda_2 = 16.25, \quad \theta_2 = 1.0, \quad x^* = 1.6. \tag{7.17}$$

We consider the initial distribution which is a point mass at $Y(0) = 0$, $X(0) = 5$, $\varphi(0) = 01$.

While the true problem has an unbounded domain $[0, \infty)$, the approximations require the domain of approximation to be a finite interval. Here we choose an upper bound of $b = 48$ and place a regulated boundary at the upper boundary. The effect of this truncation can be partly quantified by evaluating $\lim_{t\to\infty} \mathbb{P}(X(t) > 48) \approx 5.83 \times 10^{-9}$.

We obtain approximations to the generator of the fluid queue $\{(X(t), \varphi(t))\}$ via the DG, QBD-RAP and uniformisation schemes (as discussed at the start of this chapter, we do not apply slope limiting for this calculation). All the discretisations use a cell width of $\Delta = 0.4$. The approximate generator matrix is then used to approximate the first-return operator $\Psi(s)$ as discussed in Section 2.4.3. As in Section 2.4.3, let $\boldsymbol{B}^{m,n}$, $m, n \in \{+, -, 0\}$ be an approximation to $\mathbb{B}^{m,n}$, $m, n \in \{+, -, 0\}$ of dimension p, and recall that $\boldsymbol{D}^{n,m}$, $m, n \in \{+, -\}$ is given by

$$\boldsymbol{D}^{m,n} = \boldsymbol{R}^m(\boldsymbol{B}^{m,n} + \boldsymbol{B}^{m,0}(\boldsymbol{B}^{0,0})^{-1}\boldsymbol{B}^{0,n}), \ m, n \in \{+, -\},$$

where $\boldsymbol{R}^m$ is a diagonal matrix with diagonal blocks $1/r_i^k \boldsymbol{I}_p$, $i \in \mathcal{S}_k^m$, $k \in \mathcal{K}^m$, where r_i^k is the (constant) value of r_i on cell k. Then $\boldsymbol{\Psi}$, an approximation to Ψ, is given by the solution to the matrix-Ricatti equation

$$\boldsymbol{D}^{+-} + \boldsymbol{\Psi}\boldsymbol{D}^{-+}\boldsymbol{\Psi} + \boldsymbol{D}^{++}\boldsymbol{\Psi} + \boldsymbol{\Psi}\boldsymbol{D}^{--} = \boldsymbol{0}. \tag{7.18}$$

Due to the stochastic interpretation of the uniformisation and QBD-RAP schemes the approximations to $\Psi(s)$ have a stochastic interpretation as the first-return probabilities of a fluid queue driven by a CTMC and QBD-RAP, respectively (see (Peralta Gutierrez

2019, Chapter 7) and Bean et al. (2021) for details on the latter). For the DG scheme the approximation of $\Psi(s)$ is not as well-understood. Here, we assume that it is given by the solution to (7.18) and is a polynomial approximation.

Ultimately, we approximate the first-return distribution

$$\mathbb{P}(X(\zeta_Y(\{0\})) \leq x, \varphi(\zeta_Y(\{0\})) = i \mid \boldsymbol{X}(0) \sim \mu), \tag{7.19}$$

where we recall $\zeta_Y(E) = \inf\{t > 0 \mid Y(t) \in E\}$ is the first hitting time of $\{Y(t)\}$ on the set E. For Model 7.14, it is only possible for the process $Y(t)$ to return to 0 at time t when $(X(t), \varphi(t)) \in [0, 1.6) \times \{10, 00\}$, so we evaluate the approximations over this region only. We use a grid of 10,001 points on which we evaluate the approximations of the CDF in each phase.

For a 'ground truth' comparison, we simulated 5×10^{10} realisations of the fluid-fluid queue and recorded the value of $\{X(t)\}$ and $\{\varphi(t)\}$ at the time of first return of the second fluid, $\{Y(t)\}$. The empirical approximation of (7.19) was then constructed, and error metrics for the difference between the empirical CDF and the approximations were computed. To account for Monte-Carlo errors we used a bootstrap with 1,000 bootstrap samples to construct 1,000 bootstrap samples of the error estimates and ultimately estimated the 5th and 95th percentiles of the error distribution. Each of the 1,000 bootstrap samples was constructed by resampling the original 5×10^{10} realisations 5×10^{10} times with replacement.

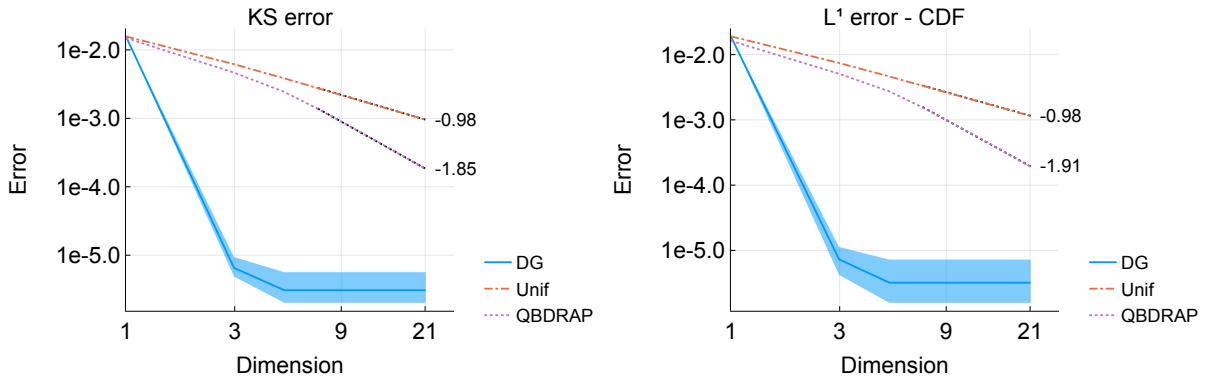

Figure 7.32: KS (left) and L^1 (right) errors between the simulated and approximate CDFs of $X(\zeta_Y(\{0\}))$ in (7.19) for Model 7.14. The approximations were obtained via the DG (blue solid line), uniformisation (green dashed line) and QBD-RAP (purple dotted line) schemes. Bootstrapped 90% confidence intervals are shown by the lighter coloured bars surrounding the lines. The black dotted lines are linear least-squares fits to the last 8 data points and the slopes of the least square lines are written next to the last point.

In Figures 7.32 and 7.33 we plot the error metrics for the approximations to the distribution (7.19). The DG scheme performs best converging rapidly until the error

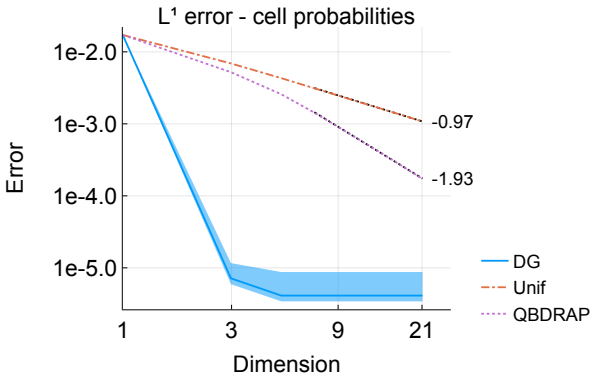

Figure 7.33: Cell-wise errors between the simulated and approximate probabilities of $X(\zeta_Y(\{0\}))$ residing on each cell $\mathcal{D}_k$ or at the boundary for Model 7.14. The approximations were obtained via the DG (blue solid line), uniformisation (green dashed line) and QBD-RAP (purple dotted line) schemes. Bootstrapped 90% confidence intervals are shown by the lighter coloured bars surrounding the lines. The black dotted lines are linear least-squares fits to the last 8 data points and the slopes of the least square lines are written next to the last point.

in the approximation scheme is swamped by other numerical (simulation) errors[††]. The QBD-RAP scheme is second best and the uniformisation scheme appears to be the slowest to converge. Here, the first return distribution appears to be smooth, hence we might expect the DG scheme to perform well. The initial condition, however, is not smooth and the DG approximation violates the axioms of probability for the initial distribition. Interestingly, however, there is no sign of this in the first return distribution.

By modifying Model 7.14 slightly, we can construct a first return distribution which is discontinuous.

Model 7.15. *Consider a fluid-fluid queue which is the same as Model 7.14 except*

$$r_{00}(X(t)) = \begin{cases} -\kappa, & \text{if } X(t) = 0, \\ -\theta_2, & \text{if } X(t) =\in (0, x^*), \\ \theta_2, & \text{if } X(t) \geq x^*, \end{cases} \tag{7.20}$$

with an initial distribution which is a point-mass at $Y(0) = 0$, $X(0) = 2, \varphi(0) = 00$.

[††]There are a significant number of other sources of error here; the largest contribution to error is simulation, but also machine precision errors, errors in solving the matrix Riccati equation to approximate $\Psi(s)$, errors from approximating error metrics (numerical integration/finding KS statistic), errors from approximating the initial condition and truncation errors. Furthermore, for the QBD-RAP, since the parameters $\boldsymbol{\alpha}$, $\boldsymbol{S}$, $\boldsymbol{s}$ and $\boldsymbol{D}$ are found numerically, then there is another source of error from this.

As before, we use the DG, uniformisation and QBD-RAP schemes to approximate the model, and compare the approximations to simulations.

Consider a sample path of Model 7.15 which begins at $X(0) = 2$ in Phase 00 and remains in Phase 00 until a time $t > 0.5$. On this sample path, at time $u \leq 0.5$ the position of the first fluid level is $X(u) = 2 - 1.6u$. Moreover, for $u \in [0, 0.25]$, $X(u) \geq x^*$ and $r_{00}(X(u)) = \theta_2 = 1 > 0$, and for $u \in (0.25, 0.5]$, $X(u) \in (0, x^*)$ and $r_{00}(X(u)) = -\theta_2 = -1 < 0$. Hence, for $u \leq 0.5$ the second fluid level is $Y(u) = u1(u \leq 0.25) + (0.25 - u)1(u > 0.25)$ and at time $u = 0.5$, the second fluid level first returns to 0, i.e. $\zeta_Y(\{0\}) = 0.5$ on this sample path. The probability of this sample path is $e^{-(\beta_1+\beta_2)\times 0.5}$. Thus, in the distribution of the first return time of the second fluid level there is a point mass at $u = 0.5$ with magnitude $e^{-(\beta_1+\beta_2)\times 0.5}$. Furthermore, on the same sample path $X(\zeta_Y(\{0\})) = X(0.5) = 1.2$, and hence there is a point mass with magnitude $e^{-(\beta_1+\beta_2)\times 0.5}$ in the distribution of $X(\zeta_Y(\{0\}))$ also.

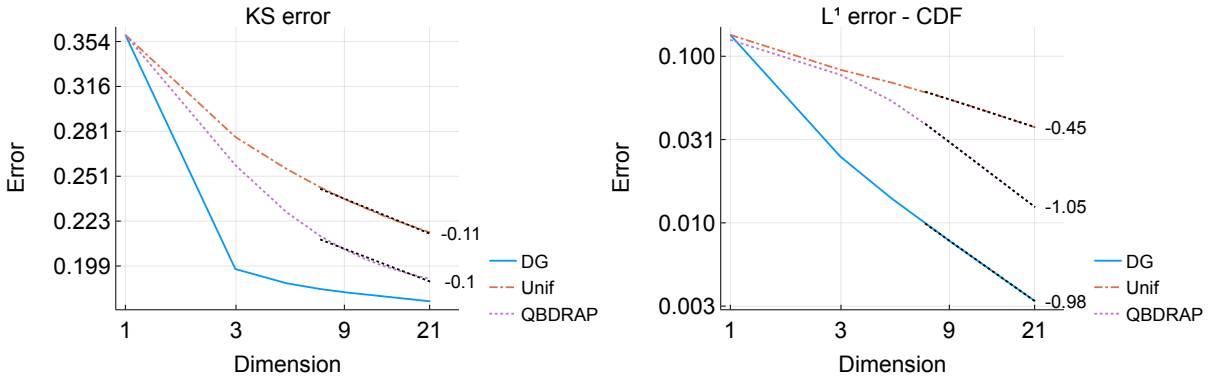

Figure 7.34: KS (left) and L^1 (right) errors between the simulated and approximate CDFs of $X(\zeta_Y(\{0\}))$ (Equation 7.19) for Model 7.15. The approximations were obtained via the DG (blue solid line), uniformisation (green dashed line) and QBD-RAP (purple dotted line) schemes. Bootstrapped 90% confidence intervals are shown by the lighter coloured bars surrounding the lines. The black dotted lines are linear least-squares fits to the last 8 data points and the slopes of the least square lines are written next to the last point.

Figure 7.34 plots error metrics for the first return distribution. Note that due to the discontinuity, the KS error is bounded below in how small it can be. The approximations all give a continuous approximation to the CDF, hence the smallest possible KS error is 1/2 the size of the discontinuity, which is ≈ 0.184 here. Observing Figure 7.34 we see that the DG scheme performs well, followed by the QBD-RAP scheme, then the uniformisation scheme performs worst. However, the DG scheme results in an oscillatory solution which does not represent a CDF, as shown in Figure 7.35.

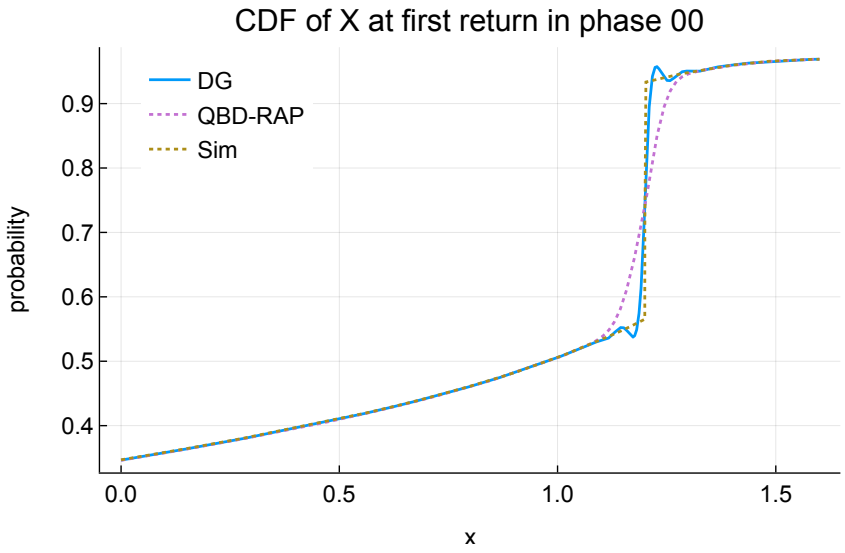

Figure 7.35: Approximations of the CDF of the distribution of X at the time $\zeta_Y(\{0\})$ in Phase 1 for Model 7.15. The blue line was obtained from the dimension-21 DG scheme, the purple dotted line from the dimension-21 QBD-RAP scheme, and the gold dotted line is the empirical CDF obtained via simulation. The DG scheme displays oscillations around the discontinuity, which mean that is does not represent a CDF.

7.7.1 Summary

Although the initial condition in the DG scheme violated the axioms of probability, for the smooth problem the DG scheme was highly effective, significantly outperforming the positivity preserving QBD-RAP and uniformisation methods. Interestingly the oscillations in the initial condition were not present in the first return distribution.

For the discontinuous problem, the DG scheme displayed oscillations (in both the initial condition and first-return distribution) and did not produce a legitimate CDF, but it did produce the lowest errors. Of the positivity preserving methods (QBD-RAP and uniformisation scheme), the QBD-RAP scheme performed the best for both problems.

7.8 Discussion

In this chapter we have numerically investigated some properties of the QBD-RAP approximation scheme and compared it to existing schemes; the uniformisation scheme of Bean & O'Reilly (2013*a*) and the discontinuous Galerkin scheme. In general, the numerical experiments show that the smoother the problem is, the better the performance of the DG scheme, and it emphatically outperforms the other two schemes for smooth problems. However, for problems with discontinuities the DG approximation can exhibit oscillations

and result in illegitimate approximations to probability distributions which violate the axioms of probability. The QBD-RAP and uniformisation approximations are guaranteed to produce legitimate distributions and, of the two, the QBD-RAP scheme almost always produces lower errors.

To avoid the problems of oscillations we can sometimes employ a *slope limiter* with the DG scheme which reduces the scheme to linear in the regions where oscillations are detected. We implemented two slope-limited DG schemes, the DG-lim scheme which takes a high order DG scheme and limits the solution as necessary, and the DG-lin-lim, which is a linear approximation on a finer grid designed to use approximately the same computational resources as the other schemes considered. The numerical experiments demonstrate a significant loss of accuracy in the approximation when a DG-lim scheme is used for discontinuous problems. We conclude that the DG-lim scheme is not a viable scheme for the discontinuous problems considered here. The numerical experiments suggest that the DG-lin-lim scheme can perform well, and is similar to the performance of the QBD-RAP scheme, in the presence of discontinuities. For highly discontinuous problems (i.e. problems with point masses) the QBD-RAP scheme can outperform the DG-lin-lim scheme.

As a first step, in Section 7.2, we examined the ability of each scheme to approximate various initial conditions. For the DG scheme, this is equivalent to a projection of the initial condition on to a set of basis polynomials, for the uniformisation scheme this is equivalent to projecting the initial condition on to a basis of constant functions, and for the QBD-RAP scheme the approximation of the initial condition is as described in Sections 4.6 and 4.7. Section 7.2.2 demonstrates that the DG scheme (projection) can result in oscillations and negative regions in the approximation when the initial condition is discontinuous. The uniformisation and QBD-RAP schemes avoid this, but appear to have higher errors and the QBD-RAP scheme appears to have the largest errors. For discontinuous initial conditions the rates of convergence are comparable for all three schemes. When the initial condition to be approximated is sufficiently smooth, then the DG approximation is superior. The QBD-RAP scheme performed the poorest for the examples considered.

Next, in Section 7.3, we investigated the performance of approximations for a simple travelling wave model with various initial conditions. The travelling wave model is useful as it allows us to investigate the ability of the schemes to capture the flow of probability without stochastic dynamics and, since the solution is known, there is no need for simulation. With the travelling wave model we demonstrate that, for problems with discontinuities, the DG scheme can display oscillations and violate the axioms of probability, while the other schemes (DG-lim, DG-lin-lim, uniformisation and QBD-RAP schemes) avoid oscillations. For discontinuous problems the rates of convergence of the QBD-RAP and DG-lin-lim schemes is similar. Interestingly, even though the QBD-RAP scheme performed worst for approximating initial conditions in the previous section, it outperformed

the uniformisation scheme for this model, demonstrating that the QBD-RAP scheme can capture the dynamics of the flow of probability better than the uniformisation scheme. For smooth problems the DG scheme is superior.

We then investigated the performance of the approximations on a simple fluid queue with two phases. First, in Section 7.4, we looked at the limiting distribution, which is known to be smooth. Since the problem is smooth, then the DG scheme was superior as expected. Of the uniformisation and QBD-RAP schemes, the QBD-RAP scheme gives more accurate solutions.

In Section 7.5 we turned our attention to approximating transient distributions for the same model and considered two different initial conditions, a point-mass and an exponential initial condition. The discontinuous initial condition results in a discontinuous transient distribution. As for the exponential initial condition, this example demonstrates that, even if the initial condition appears 'nice', it can still result in non-smooth (i.e. non-differetiable) solutions. The numerical evidence suggests that the lack of smoothness in the problem with the exponential initial condition is not problematic for the DG scheme, however, the DG scheme exhibits severe oscillations for the problem with the discontinuous initial condition. For both initial conditions, the DG-lim scheme detects oscillations and reduces the approximation to linear which renders the scheme not viable for these problems. Of the uniformisation, DG-lin-lim and QBD-RAP schemes, the latter two perform similarly, and better than the uniformisation scheme. Of the DG-lin-lim and QBD-RAP schemes the QBD-RAP scheme performed better in the presence of a point mass while the DG-lin-lim scheme performed better with the exponential initial condition.

Section 7.6 investigated the performance of the schemes for approximating the hitting time distribution for the same fluid queue with two initial conditions, an exponential initial condition and a point-mass. For this problem there is never any in-flow of mass at the boundaries of the interval and so, for a solution to be continuous, the initial condition needs to be chosen carefully, otherwise discontinuities in the transient distribution may result, as was the case for both initial conditions considered here. The numerical results suggest that, due to discontinuities in the problems, the DG scheme may display oscillations and violate the axioms of probability. Since the uniformisation, DG-lin-lim, and QBD-RAP schemes can handle discontinuities, they perform suitably, with the QBD-RAP and DG-lin-lim schemes performing similarly and better than the uniformisation.

Lastly, we applied the DG, uniformisation and QBD-RAP schemes to two simple fluid-fluid queues in Section 7.7. In the first fluid-fluid queue, which appears to have a smooth first return distribution, the DG scheme performs very well (however, the initial condition for the DG scheme displayed oscillations). Of the two positivity preserving schemes considered, the QBD-RAP scheme performs better than the uniformisation scheme. In the second example, which has a discontinuity in the first return distribution, the DG scheme produces the lowest errors, but exhibits oscillations in the solution. The QBD-RAP and uniformisation schemes do not produce oscillations and, of the two, the QBD-RAP scheme

performs best.

As we mentioned in the introduction to this chapter, in Sections 7.2, 7.2.2, 7.4, and 7.7 slope limiting is not used as it is not applicable: in these cases slope limiting is practically equivalent to post-processing the solution by detecting oscillations, then projecting oscillatory regions of the solution onto a basis of linear functions, and making some corrections to the solution if it happens to be negative.

Moreover, when computing approximations to the first-return operator of a fluid-fluid queue in Section 7.7, the QBD-RAP and uniformisation schemes lead to an algorithm that is theoretically justified. In contrast, to-date there no theory surrounding the DG approximation to the same operator (let alone any positivity preserving variants of the DG scheme).

In conclusion, when the problem is known to be smooth, the DG scheme is very likely to produce excellent results. However, for discontinuous problems, the scheme can show oscillations and may lead to illegitimate approximations to distribution functions which violate the axioms of probability. The slope limiter overcomes this, but reduces the accuracy of the DG scheme to linear near discontinuities, sometimes severely affecting the quality of the approximation. A piecewise-linear DG scheme with a limiter, such as the DG-lin-lim scheme considered here, can produce satisfactory results for some problems. The uniformisation and QBD-RAP schemes are other alternative approximation schemes which avoid oscillatory solutions and guarantee a legitimate approximation to distributions of fluid queues. Of the uniformisation and QBD-RAP schemes, the latter often produced lower errors and performed similarly to the DG-lin-lim scheme. Moreover, the uniformisation and QBD-RAP schemes have stochastic interpretations which can aid in their analysis. For the application to fluid-fluid queues in this chapter the stochastic interpretation is important as it justifies (and defines) an approximation to the first return operator. For the DG scheme there is no such theory (that the author is aware of) and this means that the approximation to the first return operator using the DG scheme is not a rigorously defined object. On a similar note, the uniformisation and QBD-RAP schemes result in operators which are linear, whereas the DG-lim and DG-lin-lim schemes do not result in linear operators. When approximating the Riccati equation for the first return operator this is necessary as the uniformisation and QBD-RAP schemes result in a matrix-Riccati equation which can be easily solved in software, whereas slope-limited DG schemes do not lead to matrix formulations as they are non-linear. In summary, the QBD-RAP and DG-lin-lim schemes perform similarly and guarantee solutions which obey the axioms of probability, however approximating operators of fluid-fluid queues is justified when using the QBD-RAP scheme, it is not justified for the DG-lin-lim (or DG schemes in general).

Chapter 8

Conclusion

A fluid-fluid queue is a stochastic fluid queue, where the driving process is a fluid queue itself. Fluid queues provide a model for a single continuous performance measure of a system in the presence of a random environment. Fluid queues have found a wide variety of applications including risk processes, telecommunications, and environmental modelling, among others. Given the success of fluid queues it is plausible that the extension to fluid-fluid queues, which enable us to track two continuous performance measures of a system, will also find success. Bean & O'Reilly (2014) provide an analysis of fluid-fluid queues and derive operator-analytic expressions for the first-return operator, and limiting distribution of a fluid-fluid queue.

Motivated by computation of the theoretical operators for stochastic fluid-fluid queues in Bean & O'Reilly (2014), this thesis investigates approximations method for the generator of stochastic fluid queues. Cell-based approximation methods are particularly useful as they allow flexible partitioning of the approximate operators, which is required in the analysis of fluid-fluid queues.

In Chapter 3 we introduced a discontinuous Galerkin (DG) scheme for the approximation of the generator of a fluid queue and show how to approximate the operators from Bean & O'Reilly (2014). High-order DG schemes are known to perform well for smooth problems, but problems, such as negative probability estimates, can occur in the presence of discontinuities. In some contexts, slope limiting can be used to rectify this, but they result in approximations to operators which are non-linear which complicates the computation of the operators in Bean & O'Reilly (2014). The uniformisation scheme of Bean & O'Reilly (2013*a*), which is equivalent to the simplest DG scheme, does not give erroneous approximations in the presence of discontinuities but can converge slowly compared to high-order DG schemes.

In Chapter 4, we introduce a new approximation to a fluid queue. Inspired by the uniformisation scheme (which is a QBD process) and its ability to handle discontinuous solutions, we construct a new discretisation of a fluid queue in the form of a QBD-RAP. Chapter 4 describes the construction of the approximation and the intuition behind it.

In Chapters 5 and 6 we prove that the QBD-RAP approximation scheme converges weakly (in the spatial and temporal variables) to the distribution of the fluid queue. Chapters 5 uses matrix-exponential-specific arguments in its proof and ultimately proves the convergence of the QBD-RAP scheme to the fluid queue when the latter remains within a given interval. Chapter 6 uses more traditional Markov process arguments to prove a global convergence result for the approximation scheme.

Chapter 7 investigates, via numerical experiments, the performance of the DG, QBD-RAP, and uniformisation schemes. In some contexts, we also implement two positivity preserving DG schemes which utilise a limiter. For smooth problems the DG scheme was superior, however, for discontinuous problems, the DG scheme exhibited oscillations and produced negative probability estimates. One of the slope limited DG schemes, in which a slope limiter is applied to a high-order DG scheme (which we title the DG-lim scheme), performed very poorly due to the limiter reducing the order of the scheme to linear near discontinuities. The QBD-RAP scheme performed similarly to the linear DG approximation on a finer grid with a slope limiter (the DG-lin-lim scheme). The QBD-RAP scheme performed slightly better on problems with point masses, and the DG-lin-lim scheme performed slightly better on the smoother problems. However, the QBD-RAP scheme constructs an approximation to the generator of the fluid queue which is linear and this is advantageous for the application to fluid-fluid queues. The uniformisation scheme performed better than the DG-lim scheme but worse than all the others.

In summary, this thesis investigated three main computational frameworks for the analysis of fluid queues and fluid-fluid queues, namely, the uniformisation, DG, and QBD-RAP schemes. The uniformisation scheme is from Bean & O'Reilly (2013*a*). The DG method is well known, but its application to fluid and fluid-fluid queues has not been well-studied. The QBD-RAP scheme is a novel approach which overcomes some known issues of the DG scheme, and converges faster than the uniformisation scheme. Analysis of the QBD-RAP scheme proves that it is convergent. Numerical experiments demonstrate the effectiveness of the approximation schemes for various problems.

Appendix A

DG applied to a toy example

This appendix has been taken from Appendix 2 of Bean et al. (2022) with only minor changes, such as notations, so that this chapter is consistent with the rest of the thesis. I am a co-author of the paper Bean et al. (2022).

Here we include a small toy example to show how we construct a DG approximation and to help clarify the notation.

Consider a process $\{(\dot{Y}(t), X(t), \varphi(t))\}_{t\geq 0}$ with two phases, $\varphi(t) \in \mathcal{S} = \{1, 2\}$ and generator matrix $\boldsymbol{T}$. Let $b = 1.8$, and the cell edges be $y_0 = 0$, $y_1 = 1$, $y_2 = 1.8$. We choose a basis of Lagrange polynomials of order 1 to define our approximation space. That is,

$$\begin{aligned} &\phi_0^1(x) = 1 - x, && \phi_0^2(x) = x, && x \in (0, 1), \\ &\phi_1^1(x) = \frac{1.8 - x}{0.8}, && \phi_1^2(x) = \frac{x - 1}{0.8}, && x \in (1, 1.8). \end{aligned}$$

The mesh and basis functions are shown in Figure 1. We can verify that the matrices $\boldsymbol{M}$ and $\boldsymbol{G}$ are given by

$$\boldsymbol{M} = \left[\begin{array}{c|c} \boldsymbol{M}_1 & \boldsymbol{0} \\ \hline \boldsymbol{0} & \boldsymbol{M}_2 \end{array}\right] = \left[\begin{array}{cc|cc} 1/3 & 1/6 & 0 & 0 \\ 1/6 & 1/3 & 0 & 0 \\ \hline 0 & 0 & 4/15 & 4/30 \\ 0 & 0 & 4/30 & 4/15 \end{array}\right],$$

$$\boldsymbol{G} = \left[\begin{array}{c|c} \boldsymbol{G}_1 & \boldsymbol{0} \\ \hline \boldsymbol{0} & \boldsymbol{G}_2 \end{array}\right] = \left[\begin{array}{cc|cc} -1/2 & 1/2 & 0 & 0 \\ -1/2 & 1/2 & 0 & 0 \\ \hline 0 & 0 & -1/2 & 1/2 \\ 0 & 0 & -1/2 & 1/2 \end{array}\right].$$

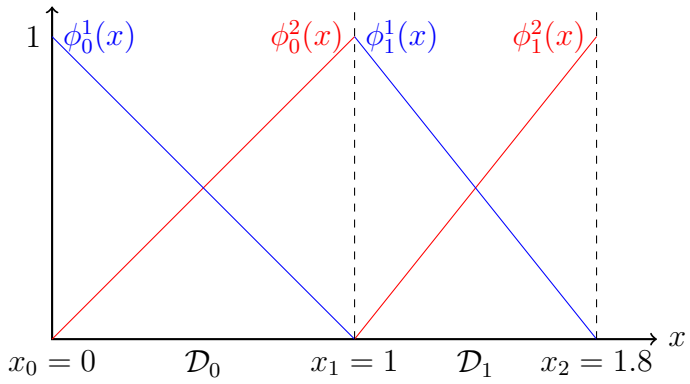

Figure 1: A mesh with nodes $y_0 = 0$, $y_1 = 1$ and $y_2 = 1.8$. There are two basis functions on each cell. Boundaries are located at $y_0 = 0$ and $y_2 = 1.8$.

Let $c_1 = 1$ and $c_2 = -2$. Then the cells are $\mathcal{D}_{0,1} = (0, 1)$, $\mathcal{D}_{1,1} = [1, 1.8)$, $\mathcal{D}_{0,2} = (0, 1]$, $\mathcal{D}_{1,2} = (1, 1.8)$ and the flux matrices are given by

$$\boldsymbol{F}_1 = \left[\begin{array}{c|c} \boldsymbol{F}_1^{00} & \boldsymbol{F}_1^{01} \\ \hline \boldsymbol{0} & \boldsymbol{F}_1^{11} \end{array}\right] = \left[\begin{array}{cc|cc} 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ \hline 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \end{array}\right],$$

$$\boldsymbol{F}_2 = \left[\begin{array}{c|c} \boldsymbol{F}_2^{00} & \boldsymbol{0} \\ \hline \boldsymbol{F}_2^{10} & \boldsymbol{F}_2^{11} \end{array}\right] = \left[\begin{array}{cc|cc} -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \hline 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right].$$

Suppose that $r_1(x) > 0$ on $\mathcal{D}_{0,1} = \mathcal{F}_1^+$ and $r_1(x) < 0$ on $\mathcal{D}_{1,1} \cup \{1.8\} = \mathcal{F}_1^-$, and further, that $r_2(x) < 0$ on $\{0\} \cup \mathcal{D}_{0,2} = \mathcal{F}_2^-$ and $r_2(x) > 0$ on $\mathcal{D}_{1,2} = \mathcal{F}_2^+$. Specifically, let

$$r_1(x) = \begin{cases} 1 & x \in (0, 1), \\ -1 & x \in [1, 1.8], \end{cases} \qquad r_2(x) = \begin{cases} -1 & x = 0, \\ -2 & x \in (0, 1], \\ 1 & x \in (1, 1.8). \end{cases}$$

Then, constructing $\boldsymbol{B}$ we get

-1	$\mathcal{D}_0$				$\mathcal{D}_2$				$K+1$
$\mathcal{F}_2^-$	$\mathcal{F}_1^+$		$\mathcal{F}_2^-$		$\mathcal{F}_1^-$		$\mathcal{F}_2^+$		$\mathcal{F}_1^-$
$q_{-1,0}$	$a_{0,1}^1$	$a_{0,1}^2$	$a_{0,2}^1$	$a_{0,2}^2$	$a_{1,1}^1$	$a_{1,1}^2$	$a_{1,2}^1$	$a_{1,2}^2$	$q_{K+1,1}$
T_{22}	$4T_{21}$	$-2T_{21}$	0	0	0	0	0	0	0
0	$T_{11}-3$	3	T_{12}	0	0	0	0	0	0
0	-1	$T_{11}-1$	0	T_{12}	5	-2.5	0	0	0
2	T_{21}	0	$T_{22}-2$	-2	0	0	0	0	0
0	0	T_{21}	6	$T_{22}-6$	0	0	0	0	0
0	0	0	0	0	$T_{11}-\frac{15}{4}$	$\frac{15}{4}$	T_{12}	0	0
0	0	0	0	0	$-\frac{5}{4}$	$T_{11}-\frac{5}{4}$	0	T_{12}	1
0	0	0	-4	8	T_{21}	0	$T_{22}-\frac{5}{2}$	$-\frac{5}{2}$	0
0	0	0	0	0	0	T_{21}	$\frac{15}{2}$	$T_{22}-\frac{15}{2}$	0
0	0	0	0	0	0	0	$-2T_{12}$	$4T_{12}$	T_{11}

We also have sub-matrices

$$\boldsymbol{B}_{11}^{++}=\begin{bmatrix} T_{11}-3 & 3 \\ -1 & T_{11}-1 \end{bmatrix},\ \boldsymbol{B}_{11}^{+-}=\left[\begin{array}{cc|c} 0 & 0 & 0 \\ 5 & -2.5 & 0 \end{array}\right],$$

$$\boldsymbol{B}_{11}^{--}=\left[\begin{array}{cc|c} T_{11}-\frac{15}{2} & \frac{15}{2} & 0 \\ -\frac{5}{2} & T_{11}-\frac{5}{2} & 1 \\ \hline 0 & 0 & T_{11} \end{array}\right],\ \boldsymbol{B}_{12}^{+-}=\left[\begin{array}{c|cc} 0 & T_{12} & 0 \\ 0 & 0 & T_{12} \end{array}\right],$$

$$\boldsymbol{B}_{12}^{-+}=\left[\begin{array}{cc} T_{12} & 0 \\ 0 & T_{12} \\ \hline -2T_{12} & 4T_{12} \end{array}\right],\ \boldsymbol{B}_{21}^{+-}=\left[\begin{array}{cc|c} T_{21} & 0 & 0 \\ 0 & T_{21} & 0 \end{array}\right],$$

$$\boldsymbol{B}_{21}^{-+}=\left[\begin{array}{cc} 4T_{21} & -2T_{21} \\ \hline T_{21} & 0 \\ 0 & T_{21} \end{array}\right],\ \boldsymbol{B}_{22}^{++}=\begin{bmatrix} T_{22}-\frac{5}{2} & -\frac{5}{2} \\ \frac{15}{2} & T_{22}-\frac{15}{2} \end{bmatrix},$$

$$\boldsymbol{B}_{22}^{+-}=\left[\begin{array}{c|cc} 0 & -4 & 8 \\ 0 & 0 & 0 \end{array}\right],\ \boldsymbol{B}_{22}^{--}=\left[\begin{array}{c|cc} T_{22} & 0 & 0 \\ \hline 2 & T_{22}-2 & -2 \\ 0 & 6 & T_{22}-6 \end{array}\right],$$

and $\boldsymbol{B}_{11}^{-+}=\mathbf{0}_{3\times 2}$, $\boldsymbol{B}_{12}^{++}=\mathbf{0}_{2\times 2}$, $\boldsymbol{B}_{12}^{--}=\mathbf{0}_{2\times 3}$, $\boldsymbol{B}_{21}^{++}=\mathbf{0}_{2\times 2}$, $\boldsymbol{B}_{21}^{--}=\mathbf{0}_{3\times 3}$, $\boldsymbol{B}_{22}^{-+}=\mathbf{0}_{3\times 2}$, where $\mathbf{0}_{n\times m}$ denotes an $n\times m$ matrix of zeros. Furthermore,

$$\boldsymbol{B}^{++}=\left[\begin{array}{cc|cc} T_{11}-3 & 3 & 0 & 0 \\ -1 & T_{11}-1 & 0 & 0 \\ \hline 0 & 0 & T_{22}-\frac{5}{2} & -\frac{5}{2} \\ 0 & 0 & \frac{15}{2} & -\frac{15}{2} \end{array}\right],$$

$$\boldsymbol{B}^{+-} = \left[\begin{array}{c|cc|cc|c} 0 & T_{12} & 0 & 0 & 0 & 0 \\ 0 & 0 & T_{12} & 5 & -2.5 & 0 \\ \hline 0 & -4 & 8 & T_{21} & 0 & 0 \\ 0 & 0 & 0 & 0 & T_{21} & 0 \end{array}\right],$$

$$\boldsymbol{B}^{-+} = \left[\begin{array}{cc|cc} 4T_{21} & -2T_{21} & 0 & 0 \\ \hline T_{21} & 0 & 0 & 0 \\ 0 & T_{21} & 0 & 0 \\ \hline 0 & 0 & T_{12} & 0 \\ 0 & 0 & 0 & T_{12} \\ \hline 0 & 0 & -2T_{12} & 4T_{12} \end{array}\right],$$

$$\boldsymbol{B}^{--} = \left[\begin{array}{c|cc|cc|c} T_{22} & 0 & 0 & 0 & 0 & 0 \\ \hline 2 & T_{22}-2 & -2 & 0 & 0 & 0 \\ 0 & 6 & T_{22}-6 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & T_{11}-\frac{15}{4} & -\frac{15}{4} & 0 \\ 0 & 0 & 0 & -\frac{5}{4} & T_{11}-\frac{5}{4} & 1 \\ \hline 0 & 0 & 0 & 0 & 0 & T_{11} \end{array}\right].$$

Since $r_1(x)$ and $r_2(x)$ are constant on each cell then $\boldsymbol{R}^+$ and $\boldsymbol{R}^-$ take a particularly simple form. We have

$$\boldsymbol{R}^+ = \left[\begin{array}{cc|cc} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \hline 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right], \quad \boldsymbol{R}^- = \left[\begin{array}{c|cc|cc|c} 1 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 1/2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right].$$

The DG approximations $\boldsymbol{D}^{mn}(s)$, $m, n \in \{+, -\}$ can now be constructed as

$$\boldsymbol{D}^{mn}(s) = \begin{cases} \boldsymbol{R}^m \left(\boldsymbol{B}^{mm} - s\boldsymbol{I}\right) & n = m, \\ \boldsymbol{R}^m \boldsymbol{B}^{mn} & n \neq m. \end{cases}$$

For a given value of s, we construct and solve the matrix Riccati equation,

$$\boldsymbol{D}^{+-}(s) + \boldsymbol{\Psi}(s)\boldsymbol{D}^{-+}(s)\Psi(s) + \boldsymbol{D}^{++}(s)\boldsymbol{\Psi}(s) + \boldsymbol{\Psi}(s)\boldsymbol{D}^{--}(s) = \boldsymbol{0},$$

for the matrix $\boldsymbol{\Psi}(s)$ using, for example, Newtons method (Bean et al. 2009*a*). To obtain the stationary distribution we require $\boldsymbol{\Psi}(0)$.

Now, to find $\boldsymbol{\xi}$, we solve the linear system in Equations (2.28)-(2.29). The result is a vector which we denote,

$$\boldsymbol{\xi} = \left[\begin{array}{c|cc|cc|c} \xi_{-1,2} & \xi^1_{0,2} & \xi^2_{0,2} & \xi^1_{1,1} & \xi^2_{1,1} & \xi_{K+1,1} \end{array}\right],$$

where $\xi_{-1,2}$ is an approximation to $\lim_{n\to\infty} \mathbb{P}\left(X(\theta_n) = 0, \varphi(\theta_n) = 2\right)$ and $\xi_{K+1,1}$ is an approximation to the artificial point mass $\lim_{n\to\infty} \mathbb{P}\left(X(\theta_n) = 1.8, \varphi(\theta_n) = 1\right)$. For $x \in \mathcal{D}_{0,2}$ an approximation to the density of $\lim_{n\to\infty} \mathbb{P}\left(X(\theta_n) \in \mathrm{d}x, \varphi(\theta_n) = 2\right)$, is constructed as $\xi^0_{0,2}\phi^1_0(x) + \xi^0_{0,2}\phi^1_0(x)$. For $x \in \mathcal{D}_{1,1}$ an approximation to the density of $\lim_{n\to\infty} \mathbb{P}\left(X(\theta_n) \in \mathrm{d}x, \varphi(\theta_n) = 1\right)$, is constructed as $\xi^1_{1,1}\phi^1_1(x) + \xi^2_{1,1}\phi^2_1(x)$.

Next, given a value of y, we solve the system (3.11)-(3.15) to find $\boldsymbol{p} = \boldsymbol{p}^-$ and $\boldsymbol{\pi}(y)$.

For the point masses we have

$$\boldsymbol{p}^- = \left[\ p_{-1,2} \,\middle|\, p^1_{0,2} \quad p^2_{0,2} \,\middle|\, p^1_{1,1} \quad p^2_{1,1} \,\middle|\, p_{K+1,1} \ \right],$$

where $p_{-1,2}$ is an approximation to $\lim_{t\to\infty} \mathbb{P}\left(Y(t) = 0, X(t) = 0, \varphi_t = 2\right)$ and $p_{K+1,1}$ is an approximation to the artificial point mass $\lim_{t\to\infty} \mathbb{P}\left(Y(t) = 0, X(t) = 1.8, \varphi(t) = 1\right)$. For $x \in \mathcal{D}_{0,2}$, an approximation to the density of $\lim_{t\to\infty} \mathbb{P}\left(Y(t) = 0, X(t) \in \mathrm{d}x, \varphi(t) = 2\right)$, is constructed as $p^1_{0,2}\phi^1_0(x) + p^2_{0,2}\phi^2_0(x)$. For $x \in \mathcal{D}_{1,1}$, an approximation to the density of $\lim_{t\to\infty} \mathbb{P}\left(Y(t) = 0, X(t) \in \mathrm{d}x, \varphi(t) = 1\right)$, is constructed as $p^1_{1,1}\phi^1_1(x) + p^2_{1,1}\phi^2_1(x)$.

Similarly, for $\boldsymbol{\pi}^-(y)$, we have

$$\boldsymbol{\pi}^-(y) = \left[\ \pi_{-1,2}(y) \,\middle|\, \pi^1_{0,2}(y) \quad \pi^2_{0,2}(y) \,\middle|\, \pi^1_{1,1}(y) \quad \pi^2_{1,1}(y) \,\middle|\, \pi_{K+1,1}(y) \ \right],$$

where $\pi_{-1,2}(y)$ is an approximation to $\lim_{t\to\infty} \mathbb{P}\left(Y(t) \in \mathrm{d}y, X(t) = 0, \varphi(t) = 2\right)$ and $\pi_{K+1,1}(y)$ is an approximation to the artificial point mass $\lim_{t\to\infty} \mathbb{P}\left(Y(t) \in \mathrm{d}y, X(t) = 1.8, \varphi(t) = 1\right)$. For $x \in \mathcal{D}_{0,2}$ an approximation to the density of $\lim_{t\to\infty} \mathbb{P}\left(Y(t) \in \mathrm{d}y, X(t) \in \mathrm{d}x, \varphi(t) = 2\right)$, is constructed as $\pi^1_{0,2}(y)\phi^1_0(x) + \pi^2_{0,2}(y)\phi^2_0(x)$. For $x \in \mathcal{D}_{1,1}$ an approximation to the density of $\lim_{t\to\infty} \mathbb{P}\left(Y(t) \in \mathrm{d}y, X(t) \in \mathrm{d}x, \varphi(t) = 1\right)$, is constructed as $\pi^1_{1,1}(y)\phi^1_1(x) + \pi^2_{1,1}(y)\phi^2_1(x)$.

For $\boldsymbol{\pi}^+(y)$, we have

$$\boldsymbol{\pi}^+(y) = \left[\ \pi^1_{0,1}(y) \quad \pi^2_{0,1}(y) \,\middle|\, \pi^1_{1,2}(y) \quad \pi^2_{1,2}(y) \ \right].$$

For $x \in \mathcal{D}_{0,1}$ an approximation to the joint density of $\lim_{t\to\infty} \mathbb{P}\left(Y(t) \in \mathrm{d}y, X_t \in \mathrm{d}x, \varphi(t) = 1\right)$ is constructed as $\pi^1_{0,1}(y)\phi^1_0(x) + \pi^2_{0,1}(y)\phi^2_0(x)$. For $x \in \mathcal{D}_{1,2}$ an approximation to the density of $\lim_{t\to\infty} \mathbb{P}\left(Y(t) \in \mathrm{d}y, X(t) \in \mathrm{d}x, \varphi(t) = 2\right)$ is constructed as $\pi^1_{1,2}(y)\phi^1_1(x) + \pi^2_{1,2}(y)\phi^2_1(x)$.

In summary, for $i \in \mathcal{S}$, a global approximation of the joint stationary distribution is

$$\lim_{t\to\infty} \mathbb{P}\left(Y(t) \in \mathrm{d}y, X(t) \in \mathrm{d}x, \varphi(t) = i\right) \approx \sum_{r\in\{1,2\}, k\in\{0,1\}} \pi^r_{i,k}(y)\phi^r_k(x)\,\mathrm{d}x\,\mathrm{d}y,$$

for $x \in (0, 1.8)$, $y > 0$, *and*

$$\lim_{t\to\infty} \mathbb{P}\left(Y(t) = 0, X(t) \in \mathrm{d}x, \varphi(t) = i\right) \approx \sum_{r\in\{1,2\}, k\in\{0,1\}} p^r_{k,i}\phi^r_k(x)\,\mathrm{d}x, \quad x \in (0, 1.8),$$

$$\lim_{t\to\infty} \mathbb{P}\left(Y(t) \in \mathrm{d}y, X(t) = 0, \varphi(t) = i\right) \approx \pi_{-1,i}(y)\,\mathrm{d}y, \quad y > 0,$$
$$\lim_{t\to\infty} \mathbb{P}\left(Y(t) = 0, X(t) = 0, \varphi(t) = i\right) \approx p_{-1,i},$$
$$\lim_{t\to\infty} \mathbb{P}\left(Y(t) \in \mathrm{d}y, X(t) = 1.8, \varphi(t) = i\right) \approx \pi_{K+1,i}(y)\,\mathrm{d}y, \quad y > 0,$$
$$\lim_{t\to\infty} \mathbb{P}\left(Y(t) = 0, X(t) = 1.8, \varphi(t) = i\right) \approx p_{K+1,i}.$$

Appendix B

Properties of the DG operator $\boldsymbol{B}$

In this section we prove that the DG approximation $\boldsymbol{B}$ conserves probability.

Recall that the coefficients $\boldsymbol{a}_{k,i}(t)$ can be used to construct an approximate solution to a differential equation at time t as $u_{k,i}(x,t) = \boldsymbol{a}_{k,i}(t)\boldsymbol{\phi}_k(x)'$. For $i \in \mathcal{S}$, $k \in \mathcal{K}$, $r \in \{1, ..., p_k\}$, define $\alpha^r_{k,i}(t) := a^r_{k,i}(t) \int_{x=y_k}^{y_{k+1}} \phi^r_k(x)\,\mathrm{d}x$, and row-vectors $\boldsymbol{\alpha}_{k,i}(t) = (\alpha^r_{k,i}(t))_{r\in\{1,...,p_k\}}$. Motivated by the fact that we may be interested in approximations of the probabilities $\mathbb{P}(X(t) \in \mathcal{D}_{k,i}, \varphi(t) = i)$ rather than the function $u_{k,i}$ itself, we can pose the problem equivalently in terms of the integrals

$$\mathbb{P}\left(X(t) \in \mathcal{D}_{k,i}, \varphi(t) = i\right) \approx \boldsymbol{a}_{k,i}(t) \int_{x\in\mathcal{D}_{k,i}} \boldsymbol{\phi}_k(x)'\,\mathrm{d}x = \boldsymbol{\alpha}_{k,i}(t)\mathbf{1}.$$

Define

$$\boldsymbol{\alpha}_k(t) = (\boldsymbol{\alpha}_{k,i}(t))_{i\in\mathcal{S}}, \text{ and } \boldsymbol{\alpha}(t) = (\boldsymbol{\alpha}_k(t)))_{k\in\mathcal{K}},$$

and matrices

$$\boldsymbol{P}_k = \operatorname{diag}\left(\int_{x=y_k}^{y_{k+1}} \phi^r_k(x)\,\mathrm{d}x\right)_{r\in\{1,...,p_k\}}, \; k \in \mathcal{K}^\circ,$$

$$\boldsymbol{P} = \begin{bmatrix} \boldsymbol{I}_N \otimes \boldsymbol{P}_0 & & \\ & \ddots & \\ & & \boldsymbol{I}_N \otimes \boldsymbol{P}_K \end{bmatrix}.$$

By choosing the basis $\{\phi^r_k\}_{r\in\{1,...,p_k\},k\in\mathcal{K}^\circ}$ such that $\int_{x=y_k}^{y_{k+1}} \phi^r_k(x)\,\mathrm{d}x \neq 0$ for all r, k, then $\boldsymbol{P}$ is invertible. This is the case for the Lagrange polynomials, but not, for example, for the Legendre polynomials. We can (loosely) interpret the new coefficients $\alpha^r_{k,i}(t)$ as representing the amount of probability captured by the basis function $\phi^r_k(x)$ in phase i.

The differential equation (3.9) can be equivalently expressed as $\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{\alpha}(t) = \boldsymbol{\alpha}(t)\mathfrak{B}$, where

$$\mathfrak{B} = \begin{bmatrix} \boldsymbol{I}_{|\mathcal{S}_{-1}|} & & \\ & \boldsymbol{P}^{-1} & \\ & & \boldsymbol{I}_{|\mathcal{S}_{K+1}|} \end{bmatrix} \boldsymbol{B} \begin{bmatrix} \boldsymbol{I}_{|\mathcal{S}_{-1}|} & & \\ & \boldsymbol{P} & \\ & & \boldsymbol{I}_{|\mathcal{S}_{K+1}|} \end{bmatrix}.$$

Let

$$\begin{aligned}
\mathfrak{B}^{-10} &:= \boldsymbol{T}_{-1+} \otimes \left(\boldsymbol{\phi}_0(0)\boldsymbol{M}_0^{-1}\boldsymbol{P}_0\right), \\
\mathfrak{B}^{0-1} &:= -\left[c_i p_{ij}^{-1} 1(c_i < 0)\right]_{i\in\mathcal{S},j\in\mathcal{S}_{-1}} \otimes \boldsymbol{P}_0^{-1}\boldsymbol{\phi}_0(0)', \\
\mathfrak{B}^{K+1K} &:= \boldsymbol{T}_{K+1-} \otimes \left(\boldsymbol{\phi}_K(b)\boldsymbol{M}_K^{-1}\boldsymbol{P}_K\right) \\
\mathfrak{B}^{KK+1} &:= \left[c_i p_{ij}^{K+1} 1(c_i > 0)\right]_{i\in\mathcal{S},j\in\mathcal{S}_{K+1}} \otimes \boldsymbol{P}_K^{-1}\boldsymbol{\phi}_K(b)', \\
\mathfrak{B}_{ij}^{kk} &:= \begin{cases} T_{ii}\boldsymbol{I}_{p_k} + c_i\boldsymbol{P}_k^{-1}(\boldsymbol{F}_i^{kk} + \boldsymbol{G}_k)\boldsymbol{M}_k^{-1}\boldsymbol{P}_k & i = j, \\ T_{ij}\boldsymbol{I}_{p_k} & i \neq j, \end{cases} \quad \text{for } k = 1, 2, \dots K-1, \\
\mathfrak{B}_{ij}^{00} &:= \begin{cases} T_{ii}\boldsymbol{I}_{p_0} + c_i\boldsymbol{P}_0^{-1}(\boldsymbol{F}_i^{00} + \boldsymbol{G}_0)\boldsymbol{M}_0^{-1}\boldsymbol{P}_0 & i = j, \\ T_{ij}\boldsymbol{I}_{p_0} - c_i p_{ij}^{-1} 1(c_i < 0)\boldsymbol{F}_i^{00}\boldsymbol{M}_0^{-1}\boldsymbol{P}_0 & i \neq j, \end{cases} \\
\mathfrak{B}_{ij}^{KK} &:= \begin{cases} T_{ii}\boldsymbol{I}_{p_K} + c_i\boldsymbol{P}_K^{-1}(\boldsymbol{F}_i^{KK} + \boldsymbol{G}_K)\boldsymbol{M}_K^{-1}\boldsymbol{P}_K & i = j, \\ T_{ij}\boldsymbol{I}_{p_K} + c_i p_{ij}^{K+1} 1(c_i > 0)\boldsymbol{P}_K^{-1}\boldsymbol{F}_i^{KK}\boldsymbol{M}_K^{-1}\boldsymbol{P}_K & i \neq j, \end{cases} \\
\mathfrak{B}_{ij}^{k,k+1} &:= \begin{cases} c_i\boldsymbol{P}_k^{-1}\boldsymbol{F}_i^{k,k+1}\boldsymbol{M}_{k+1}^{-1}\boldsymbol{P}_{k+1} & i = j, \\ \boldsymbol{0}_{p_k} & i \neq j, \end{cases} \quad \text{for } k = 0, 1, \dots K-1, \\
\mathfrak{B}_{ij}^{k-1,k} &:= \begin{cases} c_i\boldsymbol{P}_k^{-1}\boldsymbol{F}_i^{k,k-1}\boldsymbol{M}_{k-1}^{-1}\boldsymbol{P}_{k-1} & i = j, \\ \boldsymbol{0}_{p_k} & i \neq j, \end{cases}, \quad \text{for } k = 2, \dots K, \\
\mathfrak{B}^{kk} &=: \begin{bmatrix} \mathfrak{B}_{11}^{kk} & \dots & \mathfrak{B}_{1N}^{kk} \\ \vdots & \ddots & \vdots \\ \mathfrak{B}_{N1}^{kk} & \dots & \mathfrak{B}_{N,N}^{kk} \end{bmatrix}, \text{ for } k \in \mathcal{K}^\circ, \\
\mathfrak{B}^{k,k+1} &=: \begin{bmatrix} \mathfrak{B}_{1,1}^{k,k+1} & \dots & \mathfrak{B}_{1,N}^{k,k+1} \\ \vdots & \ddots & \vdots \\ \mathfrak{B}_{N,1}^{k,k+1} & \dots & \mathfrak{B}_{N,N}^{k,k+1} \end{bmatrix}, \text{ for } k = 0, 1, \dots K-1, \\
\mathfrak{B}^{k,k-1} &=: \begin{bmatrix} \mathfrak{B}_{1,1}^{k,k-1} & \dots & \mathfrak{B}_{1,N}^{k,k-1} \\ \vdots & \ddots & \vdots \\ \mathfrak{B}_{N,1}^{k,k-1} & \dots & \mathfrak{B}_{N,N}^{k,k-1} \end{bmatrix}, \text{ for } k = 1, 2, \dots K.
\end{aligned}$$

Then

$$\mathfrak{B} = \begin{bmatrix} \boldsymbol{T}_{-1,-1} & \mathfrak{B}^{-10} & & & & & \\ \mathfrak{B}^{0-1} & \mathfrak{B}^{00} & \mathfrak{B}^{01} & & & & \\ & \mathfrak{B}^{10} & \mathfrak{B}^{11} & \mathfrak{B}^{12} & & & \\ & & \ddots & \ddots & \ddots & & \\ & & \mathfrak{B}^{K-1,K-2} & \mathfrak{B}^{K-1,K-1} & \mathfrak{B}^{K-1,K} & & \\ & & & \mathfrak{B}^{K,K-1} & \mathfrak{B}^{K,K} & \mathfrak{B}^{K,K+1} & \\ & & & & \mathfrak{B}^{K+1,K} & \boldsymbol{T}_{K+1,K+1} \end{bmatrix}.$$

Remark B.1. *One may recognise the structure of $\mathfrak{B}$ as the structure of a quasi-birth-and-death process (QBD), with levels $k \in \mathcal{K}^\circ$. This raises whether $\mathfrak{B}$ is indeed a representation of the generator matrix of a QBD, or QBD-like process. In the case of a constant basis function on each cell, i.e. $p_k = 1$ and $\phi_k^1(x) \propto 1$, $k \in \mathcal{K}^\circ$, then $\mathfrak{B}$ is the generator of a QBD: it has zero row-sums, negative diagonal entries, and non-negative off-diagonal entries, the QBD-phase variable is $\{\varphi(t)\}$ and the level is $k \in \mathcal{K}$. In fact, if Δ_k is the same for every $k \in \mathcal{K}^\circ$, then this is the same QBD discretisation of a stochastic fluid process analysed by Bean & O'Reilly (2013a). However, for higher-degree polynomials $\mathfrak{B}$ is not necessarily the generator of a QBD process. We conjecture that, using polynomial basis functions, then $p_k = 1$ and $\phi_k^1(x) \propto 1$, $k \in \mathcal{K}^\circ$ is the only DG approximation which has an interpretation as a QBD-like process – not even as a QBD-RAP (Bean & Nielsen 2010).*

For simplicity, define $\mathcal{D}_k = [y_k, y_{k+1}]$. In the following lemma, we use the following properties of the Lagrange interpolating polynomials defined by the Gauss-Lobatto quadrature nodes.

Property 1 $\displaystyle\sum_{s=1}^{p_k} \phi_k^s(x) = \begin{cases} 1 & x \in \mathcal{D}_k, \\ 0 & x \notin \mathcal{D}_k. \end{cases}$

For $k \in \mathcal{K}^\circ$, let $\boldsymbol{e}_n^k$ be a row-vector of length p_k with a 1 in the nth position and zeros elsewhere.

Property 2 At the cell edges, $\boldsymbol{\phi}_k(y_k) = \boldsymbol{e}_1^k$ and $\boldsymbol{\phi}_k(y_{k+1}) = \boldsymbol{e}_{p_k}^k$, $k \in \mathcal{K}^\circ$.

Lemma B.2. *If $\{\phi_k^r(x)\}_{r\in\{1,\ldots,p_k\}}$, are chosen as the Lagrange interpolating polynomials on $\mathcal{D}_k$, $k \in \mathcal{K}^\circ$, then the matrix $\mathfrak{B}$ has zero row-sums.*

Proof. First we make numerous algebraic observations. Let $\mathbf{1}$ and $\mathbf{0}$ be column vectors of ones and zeros, respectively, with an appropriate length depending on the context. Using Property 1, observe that

$$\boldsymbol{M}_k \mathbf{1} = \left(\sum_{s=1}^{p_k} \int_{x\in\mathcal{D}_k} \phi_k^r(x) \phi_s^k(x) \, \mathrm{d}x \right)'_{r\in\{1,\ldots,p_k\}}$$

$$
\begin{aligned}
&= \left(\int_{x\in\mathcal{D}_k} \phi_k^r(x) \sum_{s=1}^{p_k} \phi_s^k(x)\,\mathrm{d}x \right)'_{r\in\{1,\ldots,p_k\}} \\
&= \left(\int_{x\in\mathcal{D}_k} \phi_k^r(x)\,\mathrm{d}x \right)'_{r\in\{1,\ldots,p_k\}} \\
&= \boldsymbol{P}_k\mathbf{1},
\end{aligned}
$$

hence $\boldsymbol{M}_k^{-1}\boldsymbol{P}_k\mathbf{1} = \mathbf{1}$. Also,

$$
\begin{aligned}
\boldsymbol{G}_k\boldsymbol{M}_k^{-1}\boldsymbol{P}_k\mathbf{1} = \boldsymbol{G}_k\mathbf{1} &= \left(\sum_{s=1}^{p_k} \int_{x\in\mathcal{D}_k} \phi_k^r(x) \frac{\mathrm{d}}{\mathrm{d}x}\phi_k^s(x)\,\mathrm{d}x \right)'_{r\in\{1,\ldots,p_k\}} \\
&= \left(\int_{x\in\mathcal{D}_k} \phi_k^r(x) \frac{\mathrm{d}}{\mathrm{d}x} \sum_{s=1}^{p_k} \phi_k^s(x)\,\mathrm{d}x \right)'_{r\in\{1,\ldots,p_k\}} \\
&= \left(\int_{x\in\mathcal{D}_k} \phi_k^r(x) \frac{\mathrm{d}}{\mathrm{d}x} 1\,\mathrm{d}x \right)'_{r\in\{1,\ldots,p_k\}} \\
&= \mathbf{0},
\end{aligned}
$$

where we have again used Property 1. Consider $c_i > 0$ and let $\boldsymbol{b}$ and $\boldsymbol{d}$ be arbitrary row-vectors of length p_k and p_{k+1}, respectively. By Property 2, for $k \in \mathcal{K}^\circ$,

$$
\begin{aligned}
\boldsymbol{F}_i^{kk}\boldsymbol{b} &= -\boldsymbol{\phi}_k(y_{k+1})'\boldsymbol{\phi}_k(y_{k+1})\boldsymbol{b} \\
&= -(\boldsymbol{e}_{p_k}^k)'\boldsymbol{e}_{p_k}^k\boldsymbol{b} \\
&= -b_{p_k}(\boldsymbol{e}_{p_k}^k)', \\
\boldsymbol{F}_i^{k,k+1}\boldsymbol{d} &= \boldsymbol{\phi}_k(y_{k+1})'\boldsymbol{\phi}_{k+1}(y_{k+1})\boldsymbol{d} \\
&= (\boldsymbol{e}_{p_k}^k)'\boldsymbol{e}_1^{k+1}\boldsymbol{d} \\
&= d_1(\boldsymbol{e}_{p_k}^k)'.
\end{aligned}
$$

Hence, observe that

$$
\boldsymbol{P}_k^{-1}\boldsymbol{F}_i^{kk}\boldsymbol{M}_k^{-1}\boldsymbol{P}_k\mathbf{1} = \boldsymbol{P}_k^{-1}\boldsymbol{\phi}_k(y_k)'\mathbf{1} = -\boldsymbol{P}_k^{-1}\boldsymbol{F}_i^{k,k+1}\boldsymbol{M}_k^{-1}\boldsymbol{P}_k\mathbf{1} = -\boldsymbol{P}_k^{-1}(\boldsymbol{e}_{p_k}^k)',
$$

for $k = 0, 1, \ldots, K-1$. Similarly, for $c_i < 0$,

$$
\boldsymbol{P}_k^{-1}\boldsymbol{F}_i^{kk}\boldsymbol{M}_k^{-1}\boldsymbol{P}_k\mathbf{1} = \boldsymbol{P}_k^{-1}\boldsymbol{\phi}_k(y_{k+1})'\mathbf{1} = -\boldsymbol{P}_k^{-1}\boldsymbol{F}_i^{k,k-1}\boldsymbol{M}_k^{-1}\boldsymbol{P}_k\mathbf{1} = -\boldsymbol{P}_k^{-1}(\boldsymbol{e}_1^k)',
$$

for $k = 1, \ldots, K$.

Now, with the above observations made, we first claim that, for $c_i > 0$, and $k = 0, 1, ..., K-1$,

$$
\sum_{j\in\mathcal{S}} \mathfrak{B}_{ij}^{kk}\mathbf{1} + \mathfrak{B}_{ij}^{kk+1}\mathbf{1} = \sum_{j\in\mathcal{S}} T_{ij}\boldsymbol{I}_{p_k}\mathbf{1} + c_i\boldsymbol{P}_k^{-1}(\boldsymbol{F}_i^{kk} + \boldsymbol{G}_k)\boldsymbol{M}_k^{-1}\boldsymbol{P}_k\mathbf{1} + c_i\boldsymbol{P}_k^{-1}\boldsymbol{F}_i^{k,k+1}\boldsymbol{M}_k^{-1}\boldsymbol{P}_k\mathbf{1}
$$

$$= \mathbf{0}.$$

The first sum is zero since $\boldsymbol{T}$ is a generator of a continuous-time Markov chain. This leaves the other two terms, which, using our observations, we get

$$\begin{aligned}
&c_i \boldsymbol{P}_k^{-1}(\boldsymbol{F}_i^{kk} + \boldsymbol{G}_k)\boldsymbol{M}_k^{-1}\boldsymbol{P}_k\mathbf{1} + c_i \boldsymbol{P}_k^{-1}\boldsymbol{F}_i^{k,k+1}\boldsymbol{M}_{k+1}^{-1}\boldsymbol{P}_{k+1}\mathbf{1} \\
&= c_i \boldsymbol{P}_k^{-1}\boldsymbol{F}_i^{kk}\boldsymbol{M}_k^{-1}\boldsymbol{P}_k\mathbf{1} + c_i \boldsymbol{P}_k^{-1}\boldsymbol{G}_k\boldsymbol{M}_k^{-1}\boldsymbol{P}_k\mathbf{1} + c_i \boldsymbol{P}_k^{-1}\boldsymbol{F}_i^{k,k+1}\boldsymbol{M}_{k+1}^{-1}\boldsymbol{P}_{k+1}\mathbf{1} \\
&= \mathbf{0}.
\end{aligned}$$

For $c_i > 0$ and $k = K$,

$$\begin{aligned}
&\sum_{j\in\mathcal{S}} \mathfrak{B}_{ij}^{KK}\mathbf{1} + \mathfrak{B}_{ij}^{KK+1}\mathbf{1} \\
&= \sum_{j\in\mathcal{S}} T_{ij}\boldsymbol{I}_{p_K}\mathbf{1} - \sum_{j\in\mathcal{S}_-} c_i p_{ij}^{K+1}\boldsymbol{P}_K^{-1}\boldsymbol{F}_i^{KK}\boldsymbol{M}_K^{-1}\boldsymbol{P}_K\mathbf{1} - \sum_{j\in\mathcal{S}_{K+1}} c_i p_{ij}^{K+1}\boldsymbol{P}_K^{-1}\boldsymbol{\phi}_K(0)'\mathbf{1} \\
&\quad + c_i \boldsymbol{P}_K^{-1}(\boldsymbol{F}_i^{KK} + \boldsymbol{G}_K)\boldsymbol{M}_K^{-1}\boldsymbol{P}_K\mathbf{1} \\
&= -\sum_{j\in\mathcal{S}_-} c_i p_{ij}^{K+1}\boldsymbol{P}_K^{-1}\boldsymbol{F}_i^{KK}\boldsymbol{M}_K^{-1}\boldsymbol{P}_K\mathbf{1} \\
&\quad - \sum_{j\in\mathcal{S}_{K+1}} c_i p_{ij}^{K+1}\boldsymbol{P}_K^{-1}\boldsymbol{F}_i^{KK}\boldsymbol{M}_K^{-1}\boldsymbol{P}_K\mathbf{1} + c_i \boldsymbol{P}_K^{-1}\boldsymbol{F}_i^{KK}\boldsymbol{M}_K^{-1}\boldsymbol{P}_K\mathbf{1} \\
&\quad + c_i \boldsymbol{P}_K^{-1}\boldsymbol{G}_K\boldsymbol{M}_K^{-1}\boldsymbol{P}_K\mathbf{1} \\
&= -c_i \boldsymbol{P}_K^{-1}\boldsymbol{F}_i^{KK}\boldsymbol{M}_K^{-1}\boldsymbol{P}_K\mathbf{1} + c_i \boldsymbol{P}_K^{-1}\boldsymbol{F}_i^{KK}\boldsymbol{M}_K^{-1}\boldsymbol{P}_K\mathbf{1} + c_i \boldsymbol{P}_K^{-1}\boldsymbol{G}_K\mathbf{1} \\
&= \mathbf{0}.
\end{aligned}$$

Similarly, for $c_i < 0$ and $k = 1, ..., K$, using the same arguments as before we have

$$\begin{aligned}
\sum_{j\in\mathcal{S}} \mathfrak{B}_{ij}^{kk}\mathbf{1} + \mathfrak{B}_{ij}^{kk-1}\mathbf{1} &= \sum_{j\in\mathcal{S}} T_{ij}\boldsymbol{I}_{p_k}\mathbf{1} + c_i \boldsymbol{P}_k^{-1}(\boldsymbol{F}_i^{kk} + \boldsymbol{G}_k)\boldsymbol{M}_k^{-1}\boldsymbol{P}_k\mathbf{1} + c_i \boldsymbol{P}_k^{-1}\boldsymbol{F}_i^{k,k-1}\boldsymbol{M}_k^{-1}\boldsymbol{P}_k\mathbf{1} \\
&= \mathbf{0}.
\end{aligned}$$

For $c_i < 0$ and $k = K$,

$$\begin{aligned}
\sum_{j\in\mathcal{S}} \mathfrak{B}_{ij}^{00}\mathbf{1} + \mathfrak{B}_{ij}^{0-1}\mathbf{1} &= \sum_{j\in\mathcal{S}} T_{ij}\boldsymbol{I}_{p_0}\mathbf{1} - \sum_{j\in\mathcal{S}_+} c_i p_{ij}^{-1}\boldsymbol{P}_0^{-1}\boldsymbol{F}_i^{00}\boldsymbol{M}_0^{-1}\boldsymbol{P}_0\mathbf{1} - \sum_{j\in\mathcal{S}_{-1}} c_i p_{ij}^{-1}\boldsymbol{P}_0^{-1}\boldsymbol{\phi}_0(0)'\mathbf{1} \\
&\quad + c_i \boldsymbol{P}_0^{-1}(\boldsymbol{F}_i^{0,0} + \boldsymbol{G}_0)\boldsymbol{M}_0^{-1}\boldsymbol{P}_0\mathbf{1} \\
&= \mathbf{0}.
\end{aligned}$$

This just leaves the boundaries. For the lower boundary,

$$\mathfrak{B}^{-1,-1}\mathbf{1} + \mathfrak{B}^{-10}\mathbf{1} = \boldsymbol{T}_{-1,-1}\mathbf{1} + \left[\boldsymbol{T}_{-1+} \otimes \left(\boldsymbol{\phi}_0(0)\boldsymbol{M}_0^{-1}\boldsymbol{P}_0\right)\right]\mathbf{1}.$$

Swapping the order of summation and recalling $\boldsymbol{M}_k^{-1}\boldsymbol{P}_k\mathbf{1} = \mathbf{1}$ then this is equal to

$$\begin{aligned}\boldsymbol{T}_{-1,-1}\mathbf{1} + \left[\boldsymbol{T}_{-1+} \otimes \left(\boldsymbol{\phi}_0(0)\boldsymbol{M}_0^{-1}\boldsymbol{P}_0\right)\mathbf{1}\right]\mathbf{1} &= \boldsymbol{T}_{-1,-1}\mathbf{1} + \left[\boldsymbol{T}_{-1+} \otimes \boldsymbol{e}_1^0\mathbf{1}\right]\mathbf{1} \\ &= \boldsymbol{T}_{-1,-1}\mathbf{1} + \left[\boldsymbol{T}_{-1+} \otimes \mathbf{1}\right]\mathbf{1} \\ &= \boldsymbol{T}_{-1,-1}\mathbf{1} + \boldsymbol{T}_{-1+}\mathbf{1} \\ &= \mathbf{0}.\end{aligned}$$

For the upper boundary,

$$\mathfrak{B}^{K+1,K+1}\mathbf{1} + \mathfrak{B}^{K+1K}\mathbf{1} = \boldsymbol{T}_{K+1,K+1}\mathbf{1} + [\boldsymbol{T}_{K+1-} \otimes (\boldsymbol{\phi}_K(b)\boldsymbol{M}_K^{-1}\boldsymbol{P}_K)]\mathbf{1}.$$

Swapping the order of summation and recalling $\boldsymbol{M}_k^{-1}\boldsymbol{P}_k\mathbf{1} = \mathbf{1}$ then this is equal to

$$\begin{aligned}\boldsymbol{T}_{K+1,K+1}\mathbf{1} + [\boldsymbol{T}_{K+1-} \otimes (\boldsymbol{\phi}_K(b)\boldsymbol{M}_K^{-1}\boldsymbol{P}_K)\mathbf{1}]\mathbf{1} &= \boldsymbol{T}_{K+1,K+1}\mathbf{1} + [\boldsymbol{T}_{K+1-} \otimes \boldsymbol{e}_{p_K}^K\mathbf{1}]\mathbf{1} \\ &= \boldsymbol{T}_{K+1,K+1}\mathbf{1} + [\boldsymbol{T}_{K+1-} \otimes \boldsymbol{e}_{p_K}^K]\mathbf{1} \\ &= \boldsymbol{T}_{K+1,K+1}\mathbf{1} + \boldsymbol{T}_{K+1-}\mathbf{1} \\ &= \mathbf{0}.\end{aligned}$$

Combining all the above we have shown that the row sums of $\mathfrak{B}$ are zero. □

Corollary B.3. *The DG approximation to the generator $\boldsymbol{B}$ conserves probability. That is, for all $t \geq 0$,*

$$\begin{aligned}&\sum_{i\in\mathcal{S}_{-1}} q_{-1,i}(t) + \sum_{i\in\mathcal{S}_{K+1}} q_{K+1,i}(t) + \sum_{i\in\mathcal{S}} \int_{x\in(0,b)} u_i(x,t)\,\mathrm{d}x \\ &= \sum_{i\in\mathcal{S}_{-1}} q_{-1,i}(0) + \sum_{i\in\mathcal{S}_{K+1}} q_{K+1,i}(0) + \sum_{i\in\mathcal{S}} \int_{x\in(0,b)} u_i(x,0)\,\mathrm{d}x.\end{aligned}$$

Proof. Let $\{\psi_k^r(x)\}_{r\in\{1,\ldots,p_k\},k\in\mathcal{K}^\circ}$, be a basis for $span(\phi_k^r(x), r \in \{1,...,p_k\}, k \in \mathcal{K}^\circ)$, where $\{\phi_k^r(x)\}_{r\in\{1,\ldots,p_k\},k\in\mathcal{K}^\circ}$ are the Lagrange polynomials. Also define $\psi_1^{-1}(x) = \delta(x)$ and $\psi_1^{K+1}(x) = \delta(x-b)$ to capture the point masses at the boundaries. Let us use the same vector notation for the basis $\psi_k^r(x)$ as we do for $\phi_k^r(x)$. For $k \in \mathcal{K}^\circ$, since $\{\psi_k^r(x)\}_{r\in\{1,\ldots,p_k\}}$ and $\{\phi_k^r(x)\}_{r\in\{1,\ldots,p_k\}}$ have the same span, then there is a matrix $\boldsymbol{V}^k$ such that $\boldsymbol{\psi}_k(x)' = \boldsymbol{V}^k\boldsymbol{\phi}_k(x)'$. Trivially, this also holds for $k = -1, K+1$.

Let

$$\boldsymbol{W} = \begin{bmatrix} \boldsymbol{I}_{|\mathcal{S}_{-1}|} & & \\ & \boldsymbol{P} & \\ & & \boldsymbol{I}_{|\mathcal{S}_{K+1}|} \end{bmatrix}$$

and

$$\boldsymbol{V} = \begin{bmatrix} \boldsymbol{I}_{|\mathcal{S}_{-1}|} & & & & & \\ & \boldsymbol{V}^0 & & & & \\ & & \ddots & & & \\ & & & \boldsymbol{V}^K & & \\ & & & & \boldsymbol{I}_{|\mathcal{S}_{K+1}|} \end{bmatrix}.$$

For a DG approximation, $\boldsymbol{B}$, constructed from the basis $\{\psi_k^r\}_{r\in\{1,\ldots,p_k\},k\in\mathcal{K}}$, it can be shown that $\boldsymbol{B}$ is similar to $\mathfrak{B}$ with similarity matrix, $\boldsymbol{VW}$, such that

$$\boldsymbol{B}_{ij} = \boldsymbol{VW}\mathfrak{B}_{ij}\boldsymbol{W}^{-1}\boldsymbol{V}^{-1},\ i,j\in\mathcal{S}.$$

Therefore,

$$\begin{aligned}\int_{x\in(0,b)} \boldsymbol{B}_{ij}\boldsymbol{\psi}(x)'\,\mathrm{d}x &= \boldsymbol{VW}\mathfrak{B}_{ij}\boldsymbol{W}^{-1}\boldsymbol{V}^{-1}\int_{x\in(0,b)}\boldsymbol{V}\boldsymbol{\phi}(x)'\,\mathrm{d}x\\ &= \boldsymbol{VW}\mathfrak{B}_{ij}\boldsymbol{W}^{-1}\boldsymbol{W}\boldsymbol{1}\\ &= \boldsymbol{VW}\mathfrak{B}_{ij}\boldsymbol{1},\end{aligned}$$

since $\displaystyle\int_{x\in(0,b)}\boldsymbol{\phi}(x)'\,\mathrm{d}x = \boldsymbol{W}\boldsymbol{1}$. The row sums of $\mathfrak{B}$ are 0, hence

$$\int_{x\in(0,b)}\sum_{j\in\mathcal{S}}\boldsymbol{B}_{ij}\boldsymbol{\psi}(x)'\,\mathrm{d}x = \boldsymbol{VW}\sum_{j\in\mathcal{S}}\mathfrak{B}_{ij}\boldsymbol{1} \tag{B.1}$$

$$= \boldsymbol{VW}\boldsymbol{0} \tag{B.2}$$

$$= \boldsymbol{0}. \tag{B.3}$$

Let ${}^{\psi}\boldsymbol{a}_i(t)$, $i\in\mathcal{S}$ denote the coefficients related to the DG approximation constructed with the basis $\{\psi_k^r\}_{r\in\{1,\ldots,p_k\},k\in\mathcal{K}}$ (to distinguish them from $\boldsymbol{a}$ and $\boldsymbol{\alpha}$ used above). The DE constructed by the DG method is

$$\frac{\mathrm{d}}{\mathrm{d}t}\left({}^{\psi}\boldsymbol{a}_j(t)\right)\boldsymbol{\psi}(x)' = \sum_{i\in\mathcal{S}}\left({}^{\psi}\boldsymbol{a}_i(t)\right)\boldsymbol{B}_{ij}\boldsymbol{\psi}(x)'.$$

Integrating over $x\in(0,b)$ and summing over $j\in\mathcal{S}$ we get

$$\int_{x\in(0,b)}\sum_{j\in\mathcal{S}}\frac{\mathrm{d}}{\mathrm{d}t}\left({}^{\psi}\boldsymbol{a}_j(t)\right)\boldsymbol{\psi}(x)'\,\mathrm{d}x = \int_{x\in(0,b)}\sum_{j\in\mathcal{S}}\sum_{i\in\mathcal{S}}\left({}^{\psi}\boldsymbol{a}_i(t)\right)\boldsymbol{B}_{ij}\boldsymbol{\psi}(x)'.$$

Exchanging the order of operations gives

$$\frac{\mathrm{d}}{\mathrm{d}t}\sum_{j\in\mathcal{S}}\left({}^{\psi}\boldsymbol{a}_j(t)\right)\int_{x\in(0,b)}\boldsymbol{\psi}(x)'\,\mathrm{d}x=\sum_{i\in\mathcal{S}}\left({}^{\psi}\boldsymbol{a}_i(t)\right)\int_{x\in(0,b)}\sum_{j\in\mathcal{S}}\boldsymbol{B}_{ij}\boldsymbol{\psi}(x)'\,\mathrm{d}x=0, \tag{B.4}$$

where the right-hand side is 0 due to Equation (B.1). This holds for all $t \geq 0$. The left-hand side of (B.4) is the rate of change (with respect to time) of the total mass of the system. Since this is 0 for all $t \geq 0$, there is no change in the total mass of the system and thus probability is conserved. □

Appendix C

Properties of closing operators

This appendix is dedicated to proving that the closing operators in (4.25)-(4.27) have the Properties 5.2, which we recall below, for convenience.

Properties 5.2. *Let $\{\boldsymbol{v}^{(p)}(x)\}_{p\geq 1}$ be a sequence of closing operators such that they may be decomposed into $\boldsymbol{v}^{(p)}(x) = \boldsymbol{w}^{(p)}(x) + \widetilde{\boldsymbol{w}}^{(p)}(x)$, where;*
(i) for $x \in [0, \Delta), u, v \geq 0$,

$$\boldsymbol{\alpha}^{(p)} e^{\boldsymbol{S}^{(p)}(u+v)} (-\boldsymbol{S}^{(p)})^{-1} \widetilde{\boldsymbol{w}}^{(p)}(x) \leq \boldsymbol{\alpha}^{(p)} e^{\boldsymbol{S}^{(p)}u} (-\boldsymbol{S}^{(p)})^{-1} \widetilde{\boldsymbol{w}}^{(p)}(x).$$

(ii) for $x \in [0, \Delta), u \geq 0$,

$$\boldsymbol{\alpha}^{(p)} e^{\boldsymbol{S}^{(p)}u} (-\boldsymbol{S}^{(p)})^{-1} \widetilde{\boldsymbol{w}}^{(p)}(x) = \widetilde{G}_{\boldsymbol{v}}^{(p)} \to 0, \quad \text{as } p \to \infty.$$

(iii) for $x \in [0, \Delta), u \geq 0$,

$$\boldsymbol{\alpha}^{(p)} e^{\boldsymbol{S}^{(p)}u} (-\boldsymbol{S}^{(p)})^{-1} \boldsymbol{w}^{(p)}(x) \leq \boldsymbol{\alpha}^{(p)} e^{\boldsymbol{S}^{(p)}u} \boldsymbol{e} G_{\boldsymbol{v}},$$

for some $0 \leq G_{\boldsymbol{v}} < \infty$ independent of p for $p > p_0$ where $p_0 < \infty$.
(iv) for $\boldsymbol{a} \in \mathcal{A}$, $u \geq 0$,

$$\int_{x\in[0,\Delta)} \boldsymbol{a}^{(p)} e^{\boldsymbol{S}^{(p)}u} \boldsymbol{v}^{(p)}(x)\, \mathrm{d}x \leq \boldsymbol{a}^{(p)} e^{\boldsymbol{S}^{(p)}u} \boldsymbol{e}.$$

(v) Let g be a function satisfying the Assumptions 5.1. For $u \leq \Delta - \varepsilon^{(p)}$, $v \in [0, \Delta)$, then

$$\left| \int_{x=0}^{\infty} \frac{\boldsymbol{\alpha}^{(p)} e^{\boldsymbol{S}^{(p)}(u+x)}}{\boldsymbol{\alpha}^{(p)} e^{\boldsymbol{S}^{(p)}u} \boldsymbol{e}} \boldsymbol{v}^{(p)}(v) g(x)\, \mathrm{d}x - g(\Delta - u - v) 1(u + v \leq \Delta - \varepsilon^{(p)}) \right| = |r_{\boldsymbol{v}}^{(p)}(u, v)|,$$

where

$$\int_{u=0}^{\Delta} \left| r_{\boldsymbol{v}}^{(p)}(u, v) \right| \, \mathrm{d}u \leq R_{\boldsymbol{v},1}^{(p)} \to 0$$

and

$$\int_{v=0}^{\Delta} \left| r_{\boldsymbol{v}}^{(p)}(u,v) \right| \, \mathrm{d}v \leq R_{\boldsymbol{v},2}^{(p)} \to 0$$

as $\mathrm{Var}(Z^{(p)}) \to 0$.

B.1 The closing operator $\boldsymbol{v}(x) = e^{\boldsymbol{S}x}\boldsymbol{s}$

For the closing operator $\boldsymbol{v}(x) = e^{\boldsymbol{S}x}\boldsymbol{s}$ we may set $\widetilde{\boldsymbol{w}}(x) = \boldsymbol{0}$, so Properties 5.2($i$) and 5.2($ii$) hold trivially.

Lemma B.1. *The closing operator* $\boldsymbol{v}(x) = e^{\boldsymbol{S}x}\boldsymbol{s}$ *has Property 5.2(iii).*

For any valid orbit, $\boldsymbol{a} \in \mathcal{A}$, $x, u \geq 0$,

$$\int_{x_n=0}^{\infty} \boldsymbol{a} e^{\boldsymbol{S}(x_n+u)}\boldsymbol{s} = \boldsymbol{a} e^{\boldsymbol{S}(u)}\boldsymbol{e}.$$

Proof. For any valid orbit, $\boldsymbol{a} \in \mathcal{A}$,

$$\boldsymbol{a} e^{\boldsymbol{S}(x+u)}\boldsymbol{e} = \mathbb{P}(Z > x+u) \leq \mathbb{P}(Z > u) = \boldsymbol{a} e^{\boldsymbol{S}u}\boldsymbol{e}.$$

□

Corollary B.2. *The closing operator* $\boldsymbol{v}(x) = e^{\boldsymbol{S}x}\boldsymbol{s}$ *has Property 5.2(iv).*

For $\boldsymbol{a} \in \mathcal{A}$, $u \geq 0$,

$$\int_{x=0}^{\Delta} \boldsymbol{a} e^{\boldsymbol{S}u}\boldsymbol{v}(x) \, \mathrm{d}x = \int_{x=0}^{\Delta} \boldsymbol{a} e^{\boldsymbol{S}u} e^{\boldsymbol{S}x}\boldsymbol{s} \, \mathrm{d}x \leq \boldsymbol{a} e^{\boldsymbol{S}u}\boldsymbol{e}.$$

Proof.

$$\int_{x=0}^{\Delta} \boldsymbol{a} e^{\boldsymbol{S}u} e^{\boldsymbol{S}x}\boldsymbol{s} \, \mathrm{d}x = \boldsymbol{a} e^{\boldsymbol{S}u}\boldsymbol{e} - \boldsymbol{a} e^{\boldsymbol{S}(u+\Delta)}\boldsymbol{e} \leq \boldsymbol{a} e^{\boldsymbol{S}u}\boldsymbol{e},$$

since $0 \leq \boldsymbol{a} e^{\boldsymbol{S}(u+\Delta)}\boldsymbol{e}$ as it is a probability. □

Corollary B.3. *The closing operator* $\boldsymbol{v}(x) = e^{\boldsymbol{S}x}\boldsymbol{s}$ *has Property 5.2(v).*

Let g *be a function satisfying the Assumptions 5.1 and consider the closing operator* $\boldsymbol{v}(x) = e^{\boldsymbol{S}x}\boldsymbol{s}$. *For* $u \leq \Delta - \varepsilon$, $v \geq 0$,

$$\int_{x=0}^{\infty} \frac{\boldsymbol{\alpha} e^{\boldsymbol{S}(u+x)}}{\boldsymbol{\alpha} e^{\boldsymbol{S}u}\boldsymbol{e}} \boldsymbol{v}(v) g(x) \, \mathrm{d}x = g(\Delta - u - v) 1(u + v \leq \Delta - \varepsilon) + r_{\boldsymbol{v}}(u,v),$$

where

$$R_{\boldsymbol{v},1} = \int_{u=0}^{\Delta-\varepsilon} |r_{\boldsymbol{v}}(u,v)|,$$

$$R_{\boldsymbol{v},2} = \int_{u=0}^{\Delta} |r_{\boldsymbol{v}}(u,v)|\,\mathrm{d}u,$$

and

$$R_{\boldsymbol{v},1}, R_{\boldsymbol{v},2} \leq r_2\Delta + 2\varepsilon G + \Delta G \frac{\mathrm{Var}(Z)/\varepsilon^2}{1 - \mathrm{Var}(Z)/\varepsilon^2}.$$

Proof. By Corollary 5.7,

$$\int_{x=0}^{\infty} \frac{\boldsymbol{\alpha} e^{\boldsymbol{S}(u+x)}}{\boldsymbol{\alpha} e^{\boldsymbol{S}u}\boldsymbol{e}} \boldsymbol{v}(v) g(x)\,\mathrm{d}x = g(\Delta - u - v)1(u + v \leq \Delta - \varepsilon) + r_3(u+v),$$

so $r_{\boldsymbol{v}}(u,v) = r_3(u+v)$. All that remains to be shown are the bounds. To this end, observe

$$\begin{aligned}
R_{\boldsymbol{v},1} &= \int_{u=0}^{\Delta} |r_3(u+v)|\,\mathrm{d}u \\
&\leq \int_{u=0}^{\Delta-\varepsilon} |r_3(u)|\,\mathrm{d}u + \int_{u=\Delta-\varepsilon}^{\Delta+\varepsilon} |r_3(u)|\,\mathrm{d}u + \int_{u=\Delta+\varepsilon}^{2\Delta+\varepsilon} |r_3(u)|\,\mathrm{d}u \\
&\leq \int_{u=0}^{\Delta-\varepsilon} |r_2|\,\mathrm{d}u + \int_{u=\Delta-\varepsilon}^{\Delta+\varepsilon} G\,\mathrm{d}u + \int_{u=\Delta+\varepsilon}^{2\Delta+\varepsilon} G \frac{\mathrm{Var}(Z)/\varepsilon^2}{1 - \mathrm{Var}(Z)/\varepsilon^2}\,\mathrm{d}u \\
&\leq |r_2|\Delta + 2\varepsilon G + \Delta G \frac{\mathrm{Var}(Z)/\varepsilon^2}{1 - \mathrm{Var}(Z)/\varepsilon^2}.
\end{aligned}$$

Similarly,

$$R_{\boldsymbol{v},2} = \int_{v=0}^{\Delta} |r_3(u+v)|\,\mathrm{d}v \leq r_2\Delta + 2\varepsilon G + \Delta G \frac{\mathrm{Var}(Z)/\varepsilon^2}{1 - \mathrm{Var}(Z)/\varepsilon^2}.$$

□

B.2 The closing operator $\widehat{\boldsymbol{v}}(x) = \left(e^{\boldsymbol{S}x} + e^{\boldsymbol{S}(2\Delta - x)}\right)\boldsymbol{s}$

Let $\widehat{\boldsymbol{v}}(x)$ be the closing operator,

$$\widehat{\boldsymbol{v}}(x) = \left(e^{\boldsymbol{S}x}\boldsymbol{s} + e^{\boldsymbol{S}(2\Delta - x)}\boldsymbol{s}\right),$$

for $x \in [0, \Delta)$.

For the closing operator $\widehat{\boldsymbol{v}}(x)$ we may set $\widetilde{\boldsymbol{w}}(x) = \boldsymbol{0}$, so Properties 5.2($i$) and 5.2($ii$) hold trivially.

Lemma B.4. *The closing operator* $\widehat{\boldsymbol{v}}(x)$ *has the Property 5.2(iii).*

For $x \in [0, \Delta), u \geq 0$,

$$\boldsymbol{a} e^{\boldsymbol{S}u}(-\boldsymbol{S})^{-1}\boldsymbol{u}(x) \leq 2\boldsymbol{a} e^{\boldsymbol{S}u}\boldsymbol{e}.$$

Proof. Let $\boldsymbol{a} \in \mathcal{A}$ be arbitrary. By definition

$$\begin{aligned}\boldsymbol{a}e^{\boldsymbol{S}u}(-\boldsymbol{S})^{-1}\boldsymbol{u}(x) &= \boldsymbol{a}e^{\boldsymbol{S}u}(-\boldsymbol{S})^{-1}\left(e^{\boldsymbol{S}x}\boldsymbol{s} + e^{\boldsymbol{S}(2\Delta-x)}\boldsymbol{s}\right)\\ &= \boldsymbol{a}e^{\boldsymbol{S}u}\left(e^{\boldsymbol{S}x}\boldsymbol{e} + e^{\boldsymbol{S}(2\Delta-x)}\boldsymbol{e}\right)\end{aligned}$$

since $(-\boldsymbol{S})^{-1}$ and $e^{\boldsymbol{S}x}$ commute and $\boldsymbol{s} = -\boldsymbol{S}e$. Since $\boldsymbol{a}e^{\boldsymbol{S}x}\boldsymbol{e}$ is decreasing this is less than or equal to

$$\boldsymbol{a}e^{\boldsymbol{S}u}\left(\boldsymbol{e} + \boldsymbol{e}\right) = 2\boldsymbol{a}e^{\boldsymbol{S}u}\boldsymbol{e}. \tag{B.1}$$

□

Lemma B.5. *The closing operator $\widehat{\boldsymbol{v}}(x)$ has the property 5.2(iv).*

For $\boldsymbol{a} \in \mathcal{A}$, $u \geq 0$,

$$\int_{x=0}^{\Delta} \boldsymbol{a}e^{\boldsymbol{S}u}\widehat{\boldsymbol{v}}(x)\,\mathrm{d}x \leq \boldsymbol{a}e^{\boldsymbol{S}u}\boldsymbol{e}$$

Proof.

$$\int_{x=0}^{\Delta} \boldsymbol{a}e^{\boldsymbol{S}u}\widehat{\boldsymbol{v}}(x)\,\mathrm{d}x = \boldsymbol{a}e^{\boldsymbol{S}u}\boldsymbol{e} - \boldsymbol{a}e^{\boldsymbol{S}(u+2\Delta)}\boldsymbol{e} \leq \boldsymbol{a}e^{\boldsymbol{S}u}\boldsymbol{e}.$$

□

Corollary B.6. *The closing operator $\widehat{\boldsymbol{v}}(x)$ has the Property 5.2(v).*

Let g be a function satisfying the Assumptions 5.1. For $u \leq \Delta - \varepsilon$, $v \in [0, \Delta)$,

$$\int_{x=0}^{\infty} \frac{\boldsymbol{\alpha}e^{\boldsymbol{S}(u+x)}}{\boldsymbol{\alpha}e^{\boldsymbol{S}u}\boldsymbol{e}}\widehat{\boldsymbol{v}}(v)g(x)\,\mathrm{d}x = g(\Delta - u - v)1(u + v \leq \Delta - \varepsilon) + r_{\widehat{\boldsymbol{v}}}(u, v),$$

where

$$|r_{\widehat{\boldsymbol{v}}}(u, v)| \leq r_3(u + v) + r_3(u + 2\Delta - v).$$

Furthermore,

$$\begin{aligned}R_{\widehat{\boldsymbol{v}},1} &= \int_{u=0}^{\Delta} |r_{\widehat{\boldsymbol{v}}}(u, v)|\,\mathrm{d}u,\\ R_{\widehat{\boldsymbol{v}},2} &= \int_{v=0}^{\Delta} |r_{\widehat{\boldsymbol{v}}}(u, v)|\,\mathrm{d}u,\end{aligned}$$

where

$$R_{\widehat{\boldsymbol{v}},1},\ R_{\widehat{\boldsymbol{v}},2} \leq 2\left(\Delta r_2 + 2\varepsilon G + \Delta\frac{\mathrm{Var}(Z)/\varepsilon^2}{1 - \mathrm{Var}(Z)/\varepsilon^2}\right).$$

Proof. By the definition of the operator $\widehat{\boldsymbol{v}}(x)$,

$$\int_{x=0}^{\infty} \frac{\boldsymbol{\alpha} e^{\boldsymbol{S}(u+x)}}{\boldsymbol{\alpha} e^{\boldsymbol{S}u}\boldsymbol{e}} \widehat{\boldsymbol{v}}(v) g(x)\,\mathrm{d}x = \int_{x=0}^{\infty} \frac{\boldsymbol{\alpha} e^{\boldsymbol{S}(u+x)}}{\boldsymbol{\alpha} e^{Su}\boldsymbol{e}} e^{\boldsymbol{S}v}\boldsymbol{s} g(x) + \frac{\boldsymbol{\alpha} e^{\boldsymbol{S}(u+x)}}{\boldsymbol{\alpha} e^{Su}\boldsymbol{e}} e^{\boldsymbol{S}(2\Delta - v)}\boldsymbol{s} g(x)\,\mathrm{d}x. \quad \text{(B.2)}$$

By Corollary 5.7

$$\int_{x=0}^{\infty} \frac{\boldsymbol{\alpha} e^{\boldsymbol{S}(u+x)}}{\boldsymbol{\alpha} e^{Su}\boldsymbol{e}} e^{\boldsymbol{S}v}\boldsymbol{s} g(x)\,\mathrm{d}x = g(\Delta - u - v)1(u + v \le \Delta - \varepsilon) + r_3(u+v), \quad \text{(B.3)}$$

$$\int_{x=0}^{\infty} \frac{\boldsymbol{\alpha} e^{\boldsymbol{S}(u+x)}}{\boldsymbol{\alpha} e^{Su}\boldsymbol{e}} e^{\boldsymbol{S}(2\Delta - v)}\boldsymbol{s} g(x)\,\mathrm{d}x = r_3(u + 2\Delta - v). \quad \text{(B.4)}$$

Therefore, (B.2) is,

$$g(\Delta - u - v)1(u + v \le \Delta - \varepsilon) + r_3(u+v) + r_3(u + 2\Delta - v). \quad \text{(B.5)}$$

Now,

$$\begin{aligned} R_{\boldsymbol{u},1} &\le \int_{u=0}^{\Delta} |r_{\boldsymbol{u}}(u,v)|\,\mathrm{d}u \\ &\le \int_{u=0}^{\Delta} r_3(u+v) + r_3(u + 2\Delta - v)\,\mathrm{d}u \\ &\le \int_{u=0}^{2\Delta} 2r_3(u+v)\,\mathrm{d}u \\ &\le 2\left(\Delta r_2 + 2\varepsilon G + \Delta \frac{\mathrm{Var}(Z)/\varepsilon^2}{1 - \mathrm{Var}(Z)/\varepsilon^2}\right). \end{aligned}$$

Similarly,

$$\begin{aligned} R_{\boldsymbol{u},2} &\le \int_{v=0}^{\Delta} |r_{\boldsymbol{u}}(u,v)|\,\mathrm{d}v \\ &= \int_{v=0}^{\Delta} r_3(u+v) + r_3(u + 2\Delta - v) \\ &\le 2\left(\Delta r_2 + 2\varepsilon G + \Delta \frac{\mathrm{Var}(Z)/\varepsilon^2}{1 - \mathrm{Var}(Z)/\varepsilon^2}\right). \end{aligned}$$

□

B.3 The closing operator $\overline{\boldsymbol{v}}(x)$

Let $\overline{\boldsymbol{v}}(x)$ be the closing operator

$$\overline{\boldsymbol{v}}(x) = \left(e^{\boldsymbol{S}x} + e^{\boldsymbol{S}(2\Delta - x)}\right)\left[I - e^{\boldsymbol{S}2\Delta}\right]^{-1}\boldsymbol{s},$$

for $x \in [0, \Delta)$. Notice that

$$\boldsymbol{a}\overline{\boldsymbol{v}}(x) = \boldsymbol{a}\left(e^{\boldsymbol{S}x} + e^{\boldsymbol{S}(2\Delta - x)}\right)\sum_{n=0}^{\infty} e^{\boldsymbol{S}2n\Delta}\boldsymbol{s}.$$

We decompose the closing operator $\overline{\boldsymbol{v}}(x) = \boldsymbol{w}(x) + \widetilde{\boldsymbol{w}}(x)$, where $\boldsymbol{w}(x) = \widehat{\boldsymbol{v}}(x)$ and

$$\widetilde{\boldsymbol{w}}(x) = \left(e^{\boldsymbol{S}x} + e^{\boldsymbol{S}(2\Delta - x)}\right)\sum_{n=1}^{\infty} e^{\boldsymbol{S}2n\Delta}\boldsymbol{s}.$$

Lemma B.7. *The closing operator $\overline{\boldsymbol{v}}(x)$ has Property 5.2(i).*
For $\boldsymbol{a} \in \mathcal{A}$, $u \geq 0$,

$$\boldsymbol{a}e^{\boldsymbol{S}(u+v)}(-\boldsymbol{S})^{-1}\overline{\boldsymbol{v}}(x) \leq \boldsymbol{a}e^{\boldsymbol{S}u}(-\boldsymbol{S})^{-1}\overline{\boldsymbol{v}}(x).$$

Proof.

$$\begin{aligned}
\boldsymbol{a}e^{\boldsymbol{S}(u+v)}(-\boldsymbol{S})^{-1}\overline{\boldsymbol{v}}(x)\,\mathrm{d}x &= \sum_{n=0}^{\infty} \boldsymbol{a}e^{\boldsymbol{S}(x+u+v+2n\Delta)}\boldsymbol{e} + \boldsymbol{a}e^{\boldsymbol{S}(2\Delta-x+u+v+2n\Delta)}\boldsymbol{e} \\
&\leq \sum_{n=0}^{\infty} \boldsymbol{a}e^{\boldsymbol{S}(x+u+2n\Delta)}\boldsymbol{e} + \boldsymbol{a}e^{\boldsymbol{S}(2\Delta-x+u+2n\Delta)}\boldsymbol{e} \\
&= \boldsymbol{a}e^{\boldsymbol{S}u}(-\boldsymbol{S})^{-1}\overline{\boldsymbol{v}}(x)\,\mathrm{d}x,
\end{aligned}$$

where the inequality holds since $\boldsymbol{a}e^{\boldsymbol{S}x}\boldsymbol{e}$ is decreasing. □

Lemma B.8. *The closing operator $\overline{\boldsymbol{v}}(x)$ has Property 5.2(ii).*
For $x \in [0, \Delta), u \geq 0$,

$$\boldsymbol{\alpha}e^{\boldsymbol{S}u}(-\boldsymbol{S})^{-1}\widetilde{\boldsymbol{w}}(x) \leq \frac{\mathrm{Var}(Z)}{\Delta^2}\frac{\pi^2}{4}. \tag{B.6}$$

Proof. The expression on the left-hand side of (B.6) is

$$\boldsymbol{\alpha}e^{\boldsymbol{S}u}(-\boldsymbol{S})^{-1}\left(e^{\boldsymbol{S}x} + e^{\boldsymbol{S}(2\Delta - x)}\right)\sum_{n=1}^{\infty} e^{\boldsymbol{S}2n\Delta}\boldsymbol{s}$$

$$
\begin{aligned}
&= \boldsymbol{\alpha} e^{\boldsymbol{S}u} \left(e^{\boldsymbol{S}x} + e^{\boldsymbol{S}(2\Delta - x)} \right) \sum_{n=1}^{\infty} e^{\boldsymbol{S}2n\Delta} \boldsymbol{e} \\
&= \sum_{n=1}^{\infty} \mathbb{P}(Z > u + x + 2n\Delta) + \mathbb{P}(Z > u + 2\Delta - x + 2n\Delta) \\
&\leq 2 \sum_{n=1}^{\infty} \mathbb{P}(Z > 2n\Delta).
\end{aligned}
$$

By Chebyshev's inequality, $\mathbb{P}(Z > 2n\Delta) \leq \dfrac{\mathrm{Var}(Z)}{\Delta^2(1 + 2(n-1))^2}$. Therefore,

$$
2 \sum_{n=1}^{\infty} \mathbb{P}(Z > 2n\Delta) \leq 2 \frac{\mathrm{Var}(Z)}{\Delta^2} \sum_{n=0}^{\infty} \frac{1}{(1+2n)^2}. \tag{B.7}
$$

Now, consider the sum

$$
\sum_{n=1}^{\infty} \frac{1}{n^2} = \sum_{n=1}^{\infty} \frac{1}{(2n)^2} + \sum_{n=0}^{\infty} \frac{1}{(1+2n)^2} = \frac{1}{4} \sum_{n=1}^{\infty} \frac{1}{n^2} + \sum_{n=0}^{\infty} \frac{1}{(1+2n)^2}. \tag{B.8}
$$

The solution to the *Basel problem* states that $\displaystyle\sum_{n=1}^{\infty} 1/n^2 = \pi^2/6$. Hence,

$$
\frac{\pi^2}{6} = \frac{1}{4} \frac{\pi^2}{6} + \sum_{n=0}^{\infty} \frac{1}{(1+2n)^2}
$$

and therefore

$$
\sum_{n=0}^{\infty} \frac{1}{(1+2n)^2} = \frac{\pi^2}{8}.
$$

Thus, (B.7) is less than or equal to

$$
\frac{\mathrm{Var}(Z)}{\Delta^2} \frac{\pi^2}{4}.
$$

□

Since $\boldsymbol{w}(x) = \widehat{\boldsymbol{v}}(x)$ then, from the results of the previous section, $\overline{\boldsymbol{v}}(x)$ has Property 5.2(*iii*).

Lemma B.9. *The closing operator $\overline{\boldsymbol{v}}(x)$ has Property 5.2(iv).*

For $\boldsymbol{a} \in \mathcal{A}$, $u \geq 0$,

$$
\int_{x=0}^{\Delta} \boldsymbol{a} e^{\boldsymbol{S}u} \overline{\boldsymbol{v}}(x) \, \mathrm{d}x = \boldsymbol{a} e^{\boldsymbol{S}u} \boldsymbol{e}.
$$

Proof.

$$\begin{aligned}\int_{x=0}^{\Delta} \boldsymbol{a}e^{\boldsymbol{S}u}\overline{\boldsymbol{v}}(x)\,\mathrm{d}x &= \boldsymbol{a}e^{\boldsymbol{S}u}(\boldsymbol{S})^{-1}\left(e^{\boldsymbol{S}\Delta} - I + e^{\boldsymbol{S}2\Delta} - e^{\boldsymbol{S}\Delta}\right)\left[I - e^{\boldsymbol{S}2\Delta}\right]^{-1}\boldsymbol{s} \\ &= \boldsymbol{a}e^{\boldsymbol{S}u}(-\boldsymbol{S})^{-1}\boldsymbol{s} \\ &= \boldsymbol{a}e^{\boldsymbol{S}u}\boldsymbol{e}.\end{aligned}$$

□

Corollary B.10. *The closing operator $\overline{\boldsymbol{v}}(x)$ has Property 5.2(v).*

Let g be a function satisfying the Assumptions 5.1. For $u \le \Delta - \varepsilon$, $v \in [0, \Delta)$,

$$\int_{x=0}^{\infty} \frac{\boldsymbol{\alpha}e^{\boldsymbol{S}(u+x)}}{\boldsymbol{\alpha}e^{\boldsymbol{S}u}\boldsymbol{e}}\overline{\boldsymbol{v}}(v)g(x)\,\mathrm{d}x = g(\Delta - u - v)1(u + v \le \Delta - \varepsilon) + r_{\overline{\boldsymbol{v}}}(u, v),$$

where

$$|r_{\overline{\boldsymbol{v}}}(u, v)| \le r_{\widehat{\boldsymbol{v}}}(u, v) + \frac{G\varepsilon^2\pi^2}{4\Delta^2}.$$

Furthermore,

$$R_{\overline{\boldsymbol{v}},1} = \int_{u=0}^{\Delta} |r_{\overline{\boldsymbol{v}}}(u, v)|\,\mathrm{d}u \le R_{\widehat{\boldsymbol{v}},1} + \frac{G\varepsilon^2\pi^2}{4\Delta},$$

and

$$R_{\overline{\boldsymbol{v}},2} = \int_{v=0}^{\Delta} |r_{\overline{\boldsymbol{v}}}(u, v)|\,\mathrm{d}u \le R_{\widehat{\boldsymbol{v}},2} + \frac{G\varepsilon^2\pi^2}{4\Delta}.$$

Proof. By the definition of the operator $\overline{\boldsymbol{v}}(v)$,

$$\begin{aligned}&\int_{x=0}^{\infty} \frac{\boldsymbol{\alpha}e^{\boldsymbol{S}(u+x)}}{\boldsymbol{\alpha}e^{\boldsymbol{S}u}\boldsymbol{e}}\overline{\boldsymbol{v}}(v)g(x)\,\mathrm{d}x \\ &= \int_{x=0}^{\infty} \frac{\boldsymbol{\alpha}e^{\boldsymbol{S}(u+x)}}{\boldsymbol{\alpha}e^{\boldsymbol{S}u}\boldsymbol{e}}\widehat{\boldsymbol{v}}(v)g(x)\,\mathrm{d}x + \int_{x=0}^{\infty} \frac{\boldsymbol{\alpha}e^{\boldsymbol{S}(u+x)}}{\boldsymbol{\alpha}e^{\boldsymbol{S}u}\boldsymbol{e}}\left(e^{\boldsymbol{S}v} + e^{\boldsymbol{S}(2\Delta - v)}\right)\sum_{n=1}^{\infty} e^{\boldsymbol{S}2n\Delta}\boldsymbol{s}g(x)\,\mathrm{d}x. \quad \text{(B.9)}\end{aligned}$$

By Lemma B.6 the first term is $g(\Delta - u - v)1(u + v \le \Delta - \varepsilon) + r_{\widehat{\boldsymbol{v}}}(u, v)$, where $|r_{\widehat{\boldsymbol{v}}}(u, v)| \le r_3(u + v) + r_3(u + 2\Delta - v)$.

Since $g \le G$, the second term in (B.9) is less than or equal to

$$G\int_{x=0}^{\infty} \frac{\boldsymbol{\alpha}e^{\boldsymbol{S}(u+x)}}{\boldsymbol{\alpha}e^{\boldsymbol{S}u}\boldsymbol{e}}\left(e^{\boldsymbol{S}v} + e^{\boldsymbol{S}(2\Delta - v)}\right)\sum_{n=1}^{\infty} e^{\boldsymbol{S}2n\Delta}\boldsymbol{s}\,\mathrm{d}x = G\frac{\boldsymbol{\alpha}e^{\boldsymbol{S}u}}{\boldsymbol{\alpha}e^{\boldsymbol{S}u}\boldsymbol{e}}\left(e^{\boldsymbol{S}v} + e^{\boldsymbol{S}(2\Delta - v)}\right)\sum_{n=1}^{\infty} e^{\boldsymbol{S}2n\Delta}\boldsymbol{e}.$$

By similar arguments to those used in the proof of Lemma B.8 we can show that this is less than or equal to

$$\frac{G}{\boldsymbol{\alpha}e^{\boldsymbol{S}u}\boldsymbol{e}}\frac{\mathrm{Var}(Z)}{\Delta^2}\frac{\pi^2}{4}.$$

Now, as $u \leq \Delta - \varepsilon$, then $\boldsymbol{\alpha}e^{\boldsymbol{S}u}\boldsymbol{e} \geq \mathrm{Var}(Z)/\varepsilon^2$ by Chebyshev's inequality, hence

$$\frac{G}{\boldsymbol{\alpha}e^{\boldsymbol{S}u}\boldsymbol{e}}\frac{\mathrm{Var}(Z)}{\Delta^2}\frac{\pi^2}{4} \leq \frac{G}{\mathrm{Var}(Z)/\varepsilon^2}\frac{\mathrm{Var}(Z)}{\Delta^2}\frac{\pi^2}{4} = \frac{G\varepsilon^2\pi^2}{4\Delta^2}.$$

Putting it all together, we have shown

$$\int_{x=0}^{\infty} \frac{\boldsymbol{\alpha}e^{\boldsymbol{S}(u+x)}}{\boldsymbol{\alpha}e^{\boldsymbol{S}u}\boldsymbol{e}}\overline{\boldsymbol{v}}(v)g(x)\,\mathrm{d}x = g(\Delta - u - v)1(u + v \leq \Delta - \varepsilon) + r_{\overline{\boldsymbol{v}}}(u, v) \tag{B.10}$$

where

$$|r_{\overline{\boldsymbol{v}}}(u, v)| \leq \left| r_{\widehat{\boldsymbol{v}}}(u, v) + \frac{G\varepsilon^2\pi^2}{4\Delta^2} \right|.$$

Lastly, observe

$$\begin{aligned} R_{\overline{\boldsymbol{v}},1} = \int_{u=0}^{\Delta} |r_{\overline{\boldsymbol{v}}}(u, v)|\,\mathrm{d}u &\leq \int_{u=0}^{\Delta} |r_{\widehat{\boldsymbol{v}}}(u, v)| + \left| \frac{G\varepsilon^2\pi^2}{4\Delta^2} \right| \mathrm{d}u \\ &= R_{\widehat{\boldsymbol{v}},1} + \frac{G\varepsilon^2\pi^2}{4\Delta} \end{aligned}$$

and similarly,

$$R_{\overline{\boldsymbol{v}},2} = \int_{v=0}^{\Delta} |r_{\overline{\boldsymbol{v}}}(u, v)|\,\mathrm{d}v \leq R_{\widehat{\boldsymbol{v}},2} + \frac{G\varepsilon^2\pi^2}{4\Delta}$$

where we have used Lemma B.6. □

Appendix C

Convergence without ephemeral phases

For completeness, we include here results needed to prove a convergence of the QBD-RAP scheme to the fluid queue without the need for the ephemeral initial states. Note that we only need to prove convergence up to, and at, the time of the first change of level of the QBD-RAP, then we can use the results from Chapter 6 to obtain global convergence results.

For $\varphi(0) = k \in \mathcal{S}_{-0}$ (or $\varphi(0) = k \in \mathcal{S}_{+0}$) the added complexity comes from the fact, upon the phase process first leaving $\mathcal{S}_{-0}$ ($\mathcal{S}_{+0}$), that it is possible the phase transitions to a state in $\mathcal{S}_+$ ($\mathcal{S}_-$). Since the orbit of the QBD-RAP is constant on $\varphi(t) \in \mathcal{S}_{-0}$ ($\varphi(t) \in \mathcal{S}_{+0}$), then upon a first transition out of $\mathcal{S}_{-0}$ ($\mathcal{S}_{+0}$) and into $\mathcal{S}_+$ ($\mathcal{S}_-$) the orbit jumps to $\boldsymbol{a}^{(p)}_{\ell_0,i}(x_0)\boldsymbol{D}^{(p)}$. For $k \in \mathcal{S}_{-0}$, $m \geq 0$, the corresponding Laplace transform of the QBD-RAP is

$$\begin{aligned}
&\widehat{f}^{\ell_0,(p)}_{m,-0,+}(\lambda)(x, j; x_0, k)\,\mathrm{d}x \\
&:= \int_{x_1=0}^{\infty} \cdots \int_{x_{2m+1}=0}^{\infty} \boldsymbol{e}_k[\boldsymbol{I} - \boldsymbol{T}_{00}]^{-1}\boldsymbol{T}_{0+}\boldsymbol{M}^m_{++}(\lambda, x_1, \ldots, x_{2m+1})\boldsymbol{e}'_j \\
&\boldsymbol{a}^{(p)}_{\ell_0,k}(x_0)\boldsymbol{D}^{(p)}\boldsymbol{N}^{2m+1,(p)}(\lambda, x_1, \ldots, x_{2m+1})\boldsymbol{v}^{(p)}_{\ell_0,j}(x)\,\mathrm{d}x_{2m+1}\ldots\,\mathrm{d}x_2\,\mathrm{d}x_1\,\mathrm{d}x \\
&+ \int_{x_1=0}^{\infty} \cdots \int_{x_{2m+2}=0}^{\infty} \boldsymbol{e}_k[\boldsymbol{I} - \boldsymbol{T}_{00}]^{-1}\boldsymbol{T}_{0-}\boldsymbol{M}^{m+1}_{-+}(\lambda, x_1, \ldots, x_{2m+2})\boldsymbol{e}'_j \\
&\boldsymbol{a}^{(p)}_{\ell_0,k}(x_0)\boldsymbol{N}^{2m+2,(p)}(\lambda, x_1, \ldots, x_{2m+2})\boldsymbol{v}^{(p)}_{\ell_0,j}(x)\,\mathrm{d}x_{2m+2}\ldots\,\mathrm{d}x_2\,\mathrm{d}x_1\,\mathrm{d}x. \qquad \text{(C.1)}
\end{aligned}$$

The Laplace transform of the fluid queue corresponding to (C.1) is

$$\begin{aligned}
\widehat{\mu}^{\ell_0}_{m,-0,+}(\lambda)(\,\mathrm{d}x, j; x_0, k) &:= \sum_{i\in\mathcal{S}_+} \boldsymbol{e}_k\left[\lambda\boldsymbol{I} - \boldsymbol{T}_{00}\right]^{-1}\boldsymbol{T}_{0i}\widehat{\mu}^{\ell_0}_{m,+,+}(\lambda)(\,\mathrm{d}x, j; x_0, i) \\
&\quad + \sum_{i\in\mathcal{S}_-} \boldsymbol{e}_k\left[\lambda\boldsymbol{I} - \boldsymbol{T}_{00}\right]^{-1}\boldsymbol{T}_{0i}\widehat{\mu}^{\ell_0}_{m+1,-,+}(\lambda)(\,\mathrm{d}x, j; x_0, i), \qquad \text{(C.2)}
\end{aligned}$$

$m \geq 0$.

The second term of (C.1) is a linear combination of $\widehat{f}^{\ell_0,(p)}_{m+1,-,+}(\lambda)(x,j;x_0,k)\,\mathrm{d}x$ which converges to $\widehat{\mu}^{\ell_0}_{m,-,+}(\lambda)(\,\mathrm{d}x,j;x_0,i)$, so there are no issues here. The first term of (C.1) creates significantly more work. Essentially, we need to derive more bounds, analogous to the results from Chapter 5, but with the initial vector $\boldsymbol{a}_{\ell_0,i}(x_0)\boldsymbol{D}$.

Analogously, for $k \in \mathcal{S}_{-0}$, $m \geq 0$, we also have

$$\begin{aligned}\widehat{f}^{\ell_0,(p)}_{m,-0,-}(\lambda)(x,j;x_0,k) := &\int_{x_1=0}^{\infty}\cdots\int_{x_{2m+2}=0}^{\infty}\boldsymbol{e}_k[\boldsymbol{I}-\boldsymbol{T}_{00}]^{-1}\boldsymbol{T}_{0+}\boldsymbol{M}^{m+1}_{+-}(\lambda,x_1,\ldots,x_{2m+2})\boldsymbol{e}'_j\\ &\boldsymbol{a}^{(p)}_{\ell_0,k}(x_0)\boldsymbol{D}^{(p)}\boldsymbol{N}^{2m+2,(p)}(\lambda,x_1,\ldots,x_{2m+2})\boldsymbol{v}^{(p)}_{\ell_0,j}(x)\,\mathrm{d}x_{2m+2}\ldots\,\mathrm{d}x_1\\ &+\int_{x_1=0}^{\infty}\cdots\int_{x_{2m+1}=0}^{\infty}\boldsymbol{e}_k[\boldsymbol{I}-\boldsymbol{T}_{00}]^{-1}\boldsymbol{T}_{0-}\boldsymbol{M}^{m}_{--}(\lambda,x_1,\ldots,x_{2m+1})\boldsymbol{e}'_j\\ &\boldsymbol{a}^{(p)}_{\ell_0,k}(x_0)\boldsymbol{N}^{2m+1,(p)}(\lambda,x_1,\ldots,x_{2m+1})\boldsymbol{v}^{(p)}_{\ell_0,j}(x)\,\mathrm{d}x_{2m+1}\ldots\,\mathrm{d}x_1.\end{aligned}$$

For $k \in \mathcal{S}_{+0}$, $m \geq 0$, we have

$$\begin{aligned}\widehat{f}^{\ell_0,(p)}_{m,+0,+}(\lambda)(x,j;x_0,k) := &\int_{x_1=0}^{\infty}\cdots\int_{x_{2m+2}=0}^{\infty}\boldsymbol{e}_k[\boldsymbol{I}-\boldsymbol{T}_{00}]^{-1}\boldsymbol{T}_{0-}\boldsymbol{M}^{m+1}_{-+}(\lambda,x_1,\ldots,x_{2m+2})\boldsymbol{e}'_j\\ &\boldsymbol{a}^{(p)}_{\ell_0,k}(x_0)\boldsymbol{D}^{(p)}\boldsymbol{N}^{2m+2,(p)}(\lambda,x_1,\ldots,x_{2m+2})\boldsymbol{v}^{(p)}_{\ell_0,j}(x)\,\mathrm{d}x_{2m+2}\ldots\,\mathrm{d}x_1\\ &+\int_{x_1=0}^{\infty}\cdots\int_{x_{2m+1}=0}^{\infty}\boldsymbol{e}_k[\boldsymbol{I}-\boldsymbol{T}_{00}]^{-1}\boldsymbol{T}_{0+}\boldsymbol{M}^{m}_{++}(\lambda,x_1,\ldots,x_{2m+1})\boldsymbol{e}'_j\\ &\boldsymbol{a}^{(p)}_{\ell_0,k}(x_0)\boldsymbol{N}^{2m+1,(p)}(\lambda,x_1,\ldots,x_{2m+1})\boldsymbol{v}^{(p)}_{\ell_0,j}(x)\,\mathrm{d}x_{2m+1}\ldots\,\mathrm{d}x_1,\end{aligned}$$

and

$$\begin{aligned}\widehat{f}^{\ell_0,(p)}_{m,+0,-}(\lambda)(x,j;x_0,k) := &\int_{x_1=0}^{\infty}\cdots\int_{x_{2m+1}=0}^{\infty}\boldsymbol{e}_k[\boldsymbol{I}-\boldsymbol{T}_{00}]^{-1}\boldsymbol{T}_{0-}\boldsymbol{M}^{m}_{--}(\lambda,x_1,\ldots,x_{2m+1})\boldsymbol{e}'_j\\ &\boldsymbol{a}^{(p)}_{\ell_0,k}(x_0)\boldsymbol{D}^{(p)}\boldsymbol{N}^{2m+1,(p)}(\lambda,x_1,\ldots,x_{2m+1})\boldsymbol{v}^{(p)}_{\ell_0,j}(x)\,\mathrm{d}x_{2m+1}\ldots\,\mathrm{d}x_1\\ &+\int_{x_1=0}^{\infty}\cdots\int_{x_{2m+2}=0}^{\infty}\boldsymbol{e}_k[\boldsymbol{I}-\boldsymbol{T}_{00}]^{-1}\boldsymbol{T}_{0+}\boldsymbol{M}^{m+1}_{+-}(\lambda,x_1,\ldots,x_{2m+2})\boldsymbol{e}'_j\\ &\boldsymbol{a}^{(p)}_{\ell_0,k}(x_0)\boldsymbol{N}^{2m+2,(p)}(\lambda,x_1,\ldots,x_{2m+2})\boldsymbol{v}^{(p)}_{\ell_0,j}(x)\,\mathrm{d}x_{2m+2}\ldots\,\mathrm{d}x_1.\end{aligned}$$

In general, for $q \in \{+,-\}$, $q' \in \{+,-\}$, $m \geq 0$,

$$\widehat{\mu}^{\ell_0}_{m,q0,q'}(\lambda)(\,\mathrm{d}x,j;x_0,k) := \sum_{r\in\{+,-\}}\sum_{i\in\mathcal{S}_r}\boldsymbol{e}_k\left[\lambda\boldsymbol{I}-\boldsymbol{T}_{00}\right]^{-1}\boldsymbol{T}_{0i}\widehat{\mu}^{\ell_0}_{m+1(r\neq q'),r,q'}(\lambda)(\,\mathrm{d}x,j;x_0,i).$$

Remark C.1. *For technical reasons we should not have point masses at $x_0 \in y_{\ell_0}, y_{\ell_0+1}$ when $\varphi(0) \in \mathcal{S}_{+0} \cup \mathcal{S}_{-0}$. Intuitively, if $\varphi(0) = k \in \mathcal{S}_{+0}$ and $x_0 = y_{\ell_0}$ then, upon exiting*

$\mathcal{S}_{+0}$, if the phase process transitions to $\mathcal{S}_-$ then the fluid queue will instantaneously leave the interval $\mathcal{D}_{\ell_0}$ upon this transition. On the same event, the orbit of the QBD-RAP will be $\boldsymbol{\alpha}^{(p)}\boldsymbol{D}^{(p)}$ at the instant of the transition to $\mathcal{S}_-$. Roughly speaking $\boldsymbol{D}^{(p)}$ maps $\boldsymbol{\alpha}^{(p)}$ to approximately $\dfrac{\boldsymbol{\alpha}^{(p)}e^{\boldsymbol{S}^{(p)}\Delta}}{\boldsymbol{\alpha}^{(p)}e^{\boldsymbol{S}^{(p)}\Delta}\boldsymbol{e}}$ (asymptotically). Our asymptotic arguments rely on Chebyshev's inequality, in the form of a bound in terms of the distance of the random variable $Z^{(p)} \sim ME(\boldsymbol{\alpha}^{(p)}, \boldsymbol{S}^{(p)})$ from its mean Δ. However, we cannot use such a technique to bound terms such as $\dfrac{\boldsymbol{\alpha}^{(p)}e^{\boldsymbol{S}^{(p)}\Delta}}{\boldsymbol{\alpha}^{(p)}e^{\boldsymbol{S}^{(p)}\Delta}\boldsymbol{e}}e^{\boldsymbol{S}^{(p)}z}\boldsymbol{s}^{(p)}$.

In practice, it may be possible to avoid this issue by choosing the intervals $\{\mathcal{D}_\ell\}$ so that the boundaries do not align with any point masses. Another option is to append an ephemeral class of phases to the fluid queue as previously stated.

Theorem C.2 below is analogous to Theorem 5.3 and proves a certain convergence of the QBD-RAP scheme to the fluid queue in the case that $\phi(0) \in \mathcal{S}_{+0} \cup \mathcal{S}_{-0}$. Like the proof of Theorem 5.3, the proof of Theorem C.2 relies on technical bounds which we establish with the remainder of this Appendix.

Theorem C.2. *As $p \to \infty$, for $(q, r) \in \{(+0, -), (-0, +)\}$, $r \in \{+, -\}$ and $m = 0$,*

$$\int_{x\in\mathcal{D}_{\ell_0}} \widehat{f}^{\ell_0,(p)}_{0,q,r}(\lambda)(x, j; x_0, k)\psi(x)\,\mathrm{d}x \to \int_{x\in\mathcal{D}_{\ell_0}} \widehat{\mu}^{\ell_0}_{0,q,r}(\lambda)(x, j; x_0, k)\psi(x)\,\mathrm{d}x. \tag{C.3}$$

For $q \in \{+0, -0\}, r \in \{+, -\}$ and $m \geq 1$,

$$\int_{x\in\mathcal{D}_{\ell_0}} \widehat{f}^{\ell_0,(p)}_{m,q,r}(\lambda)(x, j; x_0, k)\psi(x)\,\mathrm{d}x \to \int_{x\in\mathcal{D}_{\ell_0}} \widehat{\mu}^{\ell_0}_{m,q,r}(\lambda)(x, j; x_0, k)\psi(x)\,\mathrm{d}x. \tag{C.4}$$

Proof. Cases $(q, r) \in \{(+0, -), (-0, +)\}$, and $m = 0$. First, take $q = -0$ and $r = +$, then

$$\begin{aligned}
&\int_{x\in\mathcal{D}_k} \widehat{f}^{\ell_0,(p)}_{0,-0,+}(\lambda)(x, j; x_0, k)\psi(x)\,\mathrm{d}x \\
&:= \int_{x_1=0}^{\infty}\int_{x\in\mathcal{D}_k} \boldsymbol{e}_k[\boldsymbol{I} - \boldsymbol{T}_{00}]^{-1}\boldsymbol{T}_{0+}\boldsymbol{M}^0_{++}(\lambda, x_1)\boldsymbol{e}_j'\boldsymbol{a}^{(p)}_{\ell_0,k}(x_0)\boldsymbol{D}^{(p)}\boldsymbol{N}^{1,(p)}(\lambda, x_1) \\
&\boldsymbol{v}^{(p)}_{\ell_0,j}(x)\psi(x)\,\mathrm{d}x\,\mathrm{d}x_1 \\
&+ \int_{x_1=0}^{\infty}\int_{x_2=0}^{\infty}\int_{x\in\mathcal{D}_k} \boldsymbol{e}_k[\boldsymbol{I} - \boldsymbol{T}_{00}]^{-1}\boldsymbol{T}_{0-}\boldsymbol{M}^1_{-+}(\lambda, x_1, x_2)\boldsymbol{e}_j'\boldsymbol{a}^{(p)}_{\ell_0,k}(x_0)\boldsymbol{N}^{2,(p)}(\lambda, x_1, x_2) \\
&\boldsymbol{v}^{(p)}_{\ell_0,j}(x)\psi(x)\,\mathrm{d}x\,\mathrm{d}x_2\,\mathrm{d}x_1.
\end{aligned} \tag{C.5}$$

The second term is a linear combination of $\int_{x\in\mathcal{D}_k} \widehat{f}^{\ell_0,(p)}_{1,-,+}(\lambda)(x, j; x_0, i)\psi(x)\,\mathrm{d}x$ terms, each of which converge to $\int_{x\in\mathcal{D}_k} \widehat{\mu}^{\ell_0,(p)}_{1,-,+}(\lambda)(\,\mathrm{d}x, j; x_0, i)\psi(x)$, by Theorem 5.3. As for the first

term, it is a linear combination of terms

$$\int_{x_1=0}^{\infty}\int_{x\in\mathcal{D}_k} \boldsymbol{e}_i\boldsymbol{H}^{++}(\lambda,x_1)\boldsymbol{e}_j'\boldsymbol{a}_{\ell_0,k}^{(p)}(x_0)\boldsymbol{D}^{(p)}e^{\boldsymbol{S}^{(p)}x_1}\boldsymbol{v}_{\ell_0,j}^{(p)}(x)\psi(x)\,\mathrm{d}x\,\mathrm{d}x_1.$$

Lemma C.3, below, proves that such terms converge to $\int_{x\in\mathcal{D}_k}\widehat{\mu}_{0,+,+}^{\ell_0}(\lambda)(\,\mathrm{d}x,j;x_0,i)\psi(x)$. Therefore, (C.5) is a finite linear combination of terms, each of which converge, hence (C.5) converges and converges to

$$\begin{aligned}&\int_{x_1=0}^{\infty}\int_{x\in\mathcal{D}_k}\boldsymbol{e}_k\sum_{i\in\mathcal{S}_+}[\boldsymbol{I}-\boldsymbol{T}_{00}]^{-1}\boldsymbol{T}_{0i}\int_{x\in\mathcal{D}_k}\widehat{\mu}_{0,+,+}^{\ell_0}(\lambda)(\,\mathrm{d}x,j;x_0,i)\psi(x)\\&+\int_{x_1=0}^{\infty}\int_{x_2=0}^{\infty}\int_{x\in\mathcal{D}_k}\sum_{i\in\mathcal{S}_-}\boldsymbol{e}_k[\boldsymbol{I}-\boldsymbol{T}_{00}]^{-1}\boldsymbol{T}_{0i}\int_{x\in\mathcal{D}_k}\widehat{\mu}_{1,-,+}^{\ell_0}(\lambda)(\,\mathrm{d}x,j;x_0,i)\psi(x),\end{aligned}\tag{C.6}$$

which is $\widehat{\mu}_{0,-0,+}^{\ell_0}(\lambda)(\,\mathrm{d}x,j;x_0,k)$. This proves the result for $r=+$ and $q=-0$. Analogous arguments prove the result for $r=-$ and $q=+0$.

Cases $q\in\{+0,-0\}$, $r\in\{+,-\}$ and $m\geq 1$. First, take $q=+0$ and $r=+$, then

$$\begin{aligned}&\int_{x\in\mathcal{D}_k}\widehat{f}_{m,-0,+}^{\ell_0,(p)}(\lambda)(x,j;x_0,k)\psi(x)\,\mathrm{d}x\\&:=\int_{x\in\mathcal{D}_k}\int_{x_1=0}^{\infty}\cdots\int_{x_{2m+1}=0}^{\infty}\boldsymbol{e}_k[\boldsymbol{I}-\boldsymbol{T}_{00}]^{-1}\boldsymbol{T}_{0+}\boldsymbol{M}_{++}^{m}(\lambda,x_1,\ldots,x_{2m+1})\boldsymbol{e}_j'\\&\boldsymbol{a}_{\ell_0,k}^{(p)}(x_0)\boldsymbol{D}^{(p)}\boldsymbol{N}^{2m+1,(p)}(\lambda,x_1,\ldots,x_{2m+1})\boldsymbol{v}_{\ell_0,j}^{(p)}(x)\psi(x)\,\mathrm{d}x_{2m+1}\ldots\,\mathrm{d}x_2\,\mathrm{d}x_1\,\mathrm{d}x\\&+\int_{x\in\mathcal{D}_k}\int_{x_1=0}^{\infty}\cdots\int_{x_{2m+2}=0}^{\infty}\boldsymbol{e}_k[\boldsymbol{I}-\boldsymbol{T}_{00}]^{-1}\boldsymbol{T}_{0-}\boldsymbol{M}_{-+}^{m+1}(\lambda,x_1,\ldots,x_{2m+2})\boldsymbol{e}_j'\\&\boldsymbol{a}_{\ell_0,k}^{(p)}(x_0)\boldsymbol{N}^{2m+2,(p)}(\lambda,x_1,\ldots,x_{2m+2})\boldsymbol{v}_{\ell_0,j}^{(p)}(x)\psi(x)\,\mathrm{d}x_{2m+2}\ldots\,\mathrm{d}x_2\,\mathrm{d}x_1\,\mathrm{d}x.\end{aligned}\tag{C.7}$$

The second term is a linear combination of $\int_{x\in\mathcal{D}_k}\widehat{f}_{m+1,-,+}^{\ell_0,(p)}(\lambda)(x,j;x_0,i)\psi(x)\,\mathrm{d}x$ terms, each of which converge to $\int_{x\in\mathcal{D}_k}\widehat{\mu}_{m+1,-,+}^{\ell_0,(p)}(\lambda)(\,\mathrm{d}x,j;x_0,i)\psi(x)$. As for the first term, it is a linear combination of terms

$$\begin{aligned}&\int_{x\in\mathcal{D}_k}\int_{x_1=0}^{\infty}\cdots\int_{x_{2m+1}=0}^{\infty}\boldsymbol{e}_i\boldsymbol{M}_{++}^{m}(\lambda,x_1,\ldots,x_{2m+1})\boldsymbol{e}_j'\\&\boldsymbol{a}_{\ell_0,k}^{(p)}(x_0)\boldsymbol{D}^{(p)}\boldsymbol{N}^{2m+1,(p)}(\lambda,x_1,\ldots,x_{2m+1})\boldsymbol{v}_{\ell_0,j}^{(p)}(x)\psi(x)\,\mathrm{d}x_{2m+1}\ldots\,\mathrm{d}x_2\,\mathrm{d}x_1\,\mathrm{d}x\\&=\int_{x\in\mathcal{D}_k}\int_{x_1=0}^{\infty}\cdots\int_{x_{2m+1}=0}^{\infty}\boldsymbol{e}_i\boldsymbol{H}^{+-}(\lambda,x_1)\prod_{r=1}^{m-1}\boldsymbol{H}^{-+}(\lambda,x_{2r})\boldsymbol{H}^{+-}(\lambda,x_{2r+1})\\&\boldsymbol{H}^{-+}(\lambda,x_{2m})\boldsymbol{H}^{++}(\lambda,x_{2m+1})\boldsymbol{e}_j'\boldsymbol{a}_{\ell_0,k}^{(p)}(x_0)\prod_{r=1}^{2m}e^{\boldsymbol{S}^{(p)}x_r}\boldsymbol{D}^{(p)}e^{\boldsymbol{S}^{(p)}x_{2m+1}}\end{aligned}$$

$$\boldsymbol{v}_{\ell_0,j}^{(p)}(x)\psi(x)\,\mathrm{d}x_{2m+1}\ldots\,\mathrm{d}x_2\,\mathrm{d}x_1\,\mathrm{d}x,$$

which satisfies the assumptions of Lemma C.8. To see this take $n = 2m+1$, $\boldsymbol{G}_1(x_1) = \boldsymbol{e}_i\boldsymbol{H}^{+-}(\lambda, x_1)$, $\boldsymbol{G}_{2k}(x_{2k}) = \boldsymbol{H}^{-+}(\lambda, x_{2k})$, $\boldsymbol{G}_{2k+1}(x_{2k}) = \boldsymbol{H}^{+-}(\lambda, x_{2k+1})$, $k = 1,\ldots,m-1$, $\boldsymbol{G}_{2m}(x_{2m}) = \boldsymbol{H}^{-+}(x_{2m})$ and $\boldsymbol{G}_{2m+1} = \boldsymbol{H}^{++}(\lambda, x_{2m+1})\boldsymbol{e}_j'$ in Equation (C.45). By the remarks following Lemma C.8, this gives the required convergence for this case. Analogous arguments prove the result for the remaining combinations of (q, r). □

C.1 Technical results

As we did in Chapter 5, we treat the cases of $m = 0$, and $m \geq 1$, up-down/down-up transitions separately. We start with the $m = 0$ case. The following result is analogous to Lemma 5.4, though the proof is somewhat more tedious.

Lemma C.3. *Let $\psi : \mathcal{D}_{\ell_0} \to \mathbb{R}$ be bounded $|\psi(x)| \leq F$ and let $\lambda > 0$. For $i \in \mathcal{S}_-, j \in \mathcal{S}_- \cup \mathcal{S}_{-0}$, $x_0 \in (0, \Delta)$,*

$$\int_{x_1=0}^{\infty}\int_{x=0}^{\Delta}\boldsymbol{k}^{(p)}(x_0)\boldsymbol{D}^{(p)}e^{\boldsymbol{S}x_1}\boldsymbol{v}_{\ell_0,j}^{(p)}(x)h_{ij}^{--}(\lambda, x_1)\psi(x)\,\mathrm{d}x\,\mathrm{d}x_1 \to \int_{x=0}^{x_0}h_{ij}^{--}(\lambda, x_0 - x)\psi(x)\,\mathrm{d}x, \tag{C.8}$$

as $p \to \infty$. Similarly, for $i \in \mathcal{S}_+, j \in \mathcal{S}_+ \cup \mathcal{S}_{+0}$

$$\begin{aligned}&\int_{x_1=0}^{\infty}\int_{x=0}^{\Delta}\boldsymbol{k}^{(p)}(x_0)\boldsymbol{D}^{(p)}e^{\boldsymbol{S}^{(p)}x_1}\boldsymbol{v}_{\ell_0,j}^{(p)}(x)h_{ij}^{++}(\lambda, x_1)\psi(\Delta - x)\,\mathrm{d}x\,\mathrm{d}x_1\\ &\to \int_{x=\Delta-x_0}^{\Delta}h_{ij}^{++}(\lambda, x - x_0)\psi(x)\,\mathrm{d}x,\end{aligned} \tag{C.9}$$

Proof. Assume, for simplicity and without loss of generality, that $\ell_0 = 0$ so $\mathcal{D}_{\ell_0} = [0, \Delta]$. In the following we choose $\varepsilon^{(p)} = \mathrm{Var}(Z^{(p)})^{1/3}$ which tends to 0 as $p \to \infty$. Therefore, assume p is sufficiently large so that $x_0 \in (2\varepsilon, \Delta - \varepsilon)$. Such a $p < \infty$ always exists since $x_0 \in (0, \Delta)$, by assumption.

Now, upon substituting the definition of $\boldsymbol{D}$ and exchanging the order of integration (justified by the Fubini-Tonelli Theorem), the difference between the left and right-hand sides of (C.8) is

$$\begin{aligned}\Bigg|&\int_{u=0}^{\infty}\boldsymbol{k}(x_0)e^{\boldsymbol{S}u}\boldsymbol{s}\int_{x=0}^{\Delta}\int_{x_1=0}^{\infty}\boldsymbol{k}(u)e^{\boldsymbol{S}x_1}\boldsymbol{v}_{\ell_0,j}(x)h_{ij}^{--}(\lambda, x_1)\psi(x)\,\mathrm{d}x_1\,\mathrm{d}x\,\mathrm{d}u\\ &- \int_{x=0}^{x_0}h_{ij}^{--}(\lambda, x_0 - x)\psi(x)\,\mathrm{d}x\Bigg|.\end{aligned} \tag{C.10}$$

We wish to apply Property 5.2(v) to the integral over x_1, however, to do so requires that $u \le \Delta - \varepsilon$. Breaking up the integral with respect to u, then (C.10) is equal to

$$\left| \int_{u=0}^{\Delta-\varepsilon} \boldsymbol{k}(x_0) e^{\boldsymbol{S}u} \boldsymbol{s} \int_{x=0}^{\Delta} \int_{x_1=0}^{\infty} \boldsymbol{k}(u) e^{\boldsymbol{S}x_1} \boldsymbol{v}_{\ell_0,j}(x) h_{ij}^{--}(\lambda, x_1) \psi(x) \,\mathrm{d}x_1 \,\mathrm{d}x \,\mathrm{d}u + d_1 \right.$$
$$\left. - \int_{x=0}^{x_0} h_{ij}^{--}(\lambda, x_0 - x) \psi(x) \,\mathrm{d}x \right| \tag{C.11}$$
$$\le \left| \int_{u=0}^{\Delta-\varepsilon} \boldsymbol{k}(x_0) e^{\boldsymbol{S}u} \boldsymbol{s} \int_{x=0}^{\Delta} \int_{x_1=0}^{\infty} \boldsymbol{k}(u) e^{\boldsymbol{S}x_1} \boldsymbol{v}_{\ell_0,j}(x) h_{ij}^{--}(\lambda, x_1) \psi(x) \,\mathrm{d}x_1 \,\mathrm{d}x \,\mathrm{d}u \right.$$
$$\left. - \int_{x=0}^{x_0} h_{ij}^{--}(\lambda, x_0 - x) \psi(x) \,\mathrm{d}x \right| + |d_1| \tag{C.12}$$

where

$$|d_1| = \left| \int_{u=\Delta-\varepsilon}^{\infty} \boldsymbol{k}(x_0) e^{\boldsymbol{S}u} \boldsymbol{s} \int_{x=0}^{\Delta} \int_{x_1=0}^{\infty} \boldsymbol{k}(u) e^{\boldsymbol{S}x_1} \boldsymbol{v}_{\ell_0,j}(x) h_{ij}^{--}(\lambda, x_1) \psi(x) \,\mathrm{d}x_1 \,\mathrm{d}x \,\mathrm{d}u \right|.$$

We show later that $|d_1|$ can be made arbitrarily small by choosing Z with sufficiently small variance. For now, let us focus on the first absolute value in (C.12). By Property 5.2(v) and swapping the order of integration (justified by the Fubini-Tonelli Theorem) the first absolute value in (C.12) is

$$\left| \int_{x=0}^{\Delta} \int_{u=0}^{\Delta-\varepsilon} \boldsymbol{k}(x_0) e^{\boldsymbol{S}u} \boldsymbol{s} \left[h_{ij}^{--}(\lambda, \Delta - u - x) 1(u + x \le \Delta - \varepsilon) + r_{\boldsymbol{v}}(u, x) \right] \psi(x) \,\mathrm{d}u \,\mathrm{d}x \right.$$
$$\left. - \int_{x=0}^{x_0} h_{ij}^{--}(\lambda, x_0 - x) \psi(x) \,\mathrm{d}x \right|$$
$$\le \left| \int_{x=0}^{\Delta} \int_{u=0}^{\Delta-\varepsilon} \boldsymbol{k}(x_0) e^{\boldsymbol{S}u} \boldsymbol{s} h_{ij}^{--}(\lambda, \Delta - u - x) 1(u + x \le \Delta - \varepsilon) \psi(x) \,\mathrm{d}u \,\mathrm{d}x \right.$$
$$\left. - \int_{x=0}^{x_0} h_{ij}^{--}(\lambda, x_0 - x) \psi(x) \,\mathrm{d}x \right| + |d_2|, \tag{C.13}$$

where

$$|d_2| = \left| \int_{x=0}^{\Delta} \int_{u=0}^{\Delta-\varepsilon} \boldsymbol{k}(x_0) e^{\boldsymbol{S}u} \boldsymbol{s} r_{\boldsymbol{v}}(u, x) \psi(x) \,\mathrm{d}u \,\mathrm{d}x \right|.$$

We show later that $|d_2|$ can be made arbitrarily small by choosing Z with sufficiently

small variance. The remaining term in (C.13) can be written as

$$\left| \int_{x=0}^{\Delta-\varepsilon} \int_{u=0}^{\Delta-x-\varepsilon} \boldsymbol{k}(x_0) e^{\boldsymbol{S}u} \boldsymbol{s} h_{ij}^{--}(\lambda, \Delta - u - x)\psi(x)\,\mathrm{d}u\,\mathrm{d}x - \int_{x=0}^{x_0} h_{ij}^{--}(\lambda, x_0 - x)\psi(x)\,\mathrm{d}x \right|. \tag{C.14}$$

Intuitively, when the variance of Z is low, we expect a significant contribution to the integral with respect to u in (C.14) to come from the portion of the integral over the interval $(\Delta - x_0 - \varepsilon, \Delta - x_0 + \varepsilon)$. Although, the integral with respect to u only contains this interval when x is sufficiently small. Breaking up the integral with respect to u in (C.14) into an integral over the interval $(\Delta - x_0 - \varepsilon, \Delta - x_0 + \varepsilon)$ and integrals over the rest, then applying the triangle inequality, (C.14) is less than or equal to

$$\begin{aligned} &\left| \int_{x=0}^{\Delta-\varepsilon} \int_{u=\Delta-x_0-\varepsilon}^{\Delta-x_0+\varepsilon} \boldsymbol{k}(x_0) e^{\boldsymbol{S}u} \boldsymbol{s} h_{ij}^{--}(\lambda, \Delta - u - x)\psi(x)\,\mathrm{d}u 1(x \leq x_0 - 2\varepsilon)\,\mathrm{d}x \right. \\ &\qquad \left. - \int_{x=0}^{x_0} h_{ij}^{--}(\lambda, x_0 - x)\psi(x)\,\mathrm{d}x \right| + |d_3| + |d_4| + |d_5|, \end{aligned} \tag{C.15}$$

where

$$\begin{aligned} |d_3| &= \left| \int_{x=0}^{\Delta-\varepsilon} \int_{u=0}^{\Delta-x-\varepsilon} \boldsymbol{k}(x_0) e^{\boldsymbol{S}u} \boldsymbol{s} h_{ij}^{--}(\lambda, \Delta - u - x)\psi(x)\,\mathrm{d}u 1(x \geq x_0)\,\mathrm{d}x \right|, \\ |d_4| &= \left| \int_{x=0}^{\Delta-\varepsilon} \int_{u=\Delta-x_0+\varepsilon}^{\Delta-x-\varepsilon} \boldsymbol{k}(x_0) e^{\boldsymbol{S}u} \boldsymbol{s} h_{ij}^{--}(\lambda, \Delta - u - x)\psi(x)\,\mathrm{d}u 1(x \leq x_0 - 2\varepsilon)\,\mathrm{d}x \right|, \\ |d_5| &= \left| \int_{x=0}^{\Delta-\varepsilon} \int_{u=\Delta-x_0-\varepsilon}^{\Delta-x-\varepsilon} \boldsymbol{k}(x_0) e^{\boldsymbol{S}u} \boldsymbol{s} h_{ij}^{--}(\lambda, \Delta - u - x)\psi(x)\,\mathrm{d}u \times 1(x \in [x_0 - 2\varepsilon, x_0))\,\mathrm{d}x \right|. \end{aligned}$$

We show later that $|d_3|$, $|d_4|$ and $|d_5|$ can be made arbitrarily small by choosing Z with sufficiently small variance.

In the first integral with respect to x in (C.15), since $x_0 \in (2\varepsilon, \Delta - \varepsilon)$, then we can absorb the indicator function into the limits of the integral which results in

$$\int_{x=0}^{x_0-2\varepsilon} \int_{u=\Delta-x_0-\varepsilon}^{\Delta-x_0+\varepsilon} \boldsymbol{k}(x_0) e^{\boldsymbol{S}u} \boldsymbol{s} h_{ij}^{--}(\lambda, \Delta - u - x)\psi(x)\,\mathrm{d}u\,\mathrm{d}x. \tag{C.16}$$

With this, and breaking up the integral over h_{ij}^{--}, we can write the first absolute value in (C.15) as

$$\left| \int_{x=0}^{x_0-2\varepsilon} \int_{u=\Delta-x_0-\varepsilon}^{\Delta-x_0+\varepsilon} \boldsymbol{k}(x_0) e^{\boldsymbol{S}u} \boldsymbol{s} h_{ij}^{--}(\lambda, \Delta - u - x)\psi(x)\,\mathrm{d}u\,\mathrm{d}x \right.$$

$$
\begin{aligned}
&\quad - \int_{x=0}^{x_0-2\varepsilon} h_{ij}^{--}(\lambda, x_0 - x)\psi(x)\,\mathrm{d}x - \int_{x=x_0-2\varepsilon}^{x_0} h_{ij}^{--}(\lambda, x_0 - x)\psi(x)\,\mathrm{d}x \Bigg| \\
&\leq \Bigg| \int_{x=0}^{x_0-2\varepsilon} \int_{u=\Delta-x_0-\varepsilon}^{\Delta-x_0+\varepsilon} \boldsymbol{k}(x_0) e^{\boldsymbol{S}u} \boldsymbol{s} h_{ij}^{--}(\lambda, \Delta - u - x)\psi(x)\,\mathrm{d}u\,\mathrm{d}x \\
&\quad - \int_{x=0}^{x_0-2\varepsilon} h_{ij}^{--}(\lambda, x_0 - x)\psi(x)\,\mathrm{d}x \Bigg| + |d_6|, \qquad \text{(C.17)}
\end{aligned}
$$

where

$$
|d_6| = \left| \int_{x=x_0-2\varepsilon}^{x_0} h_{ij}^{--}(\lambda, x_0 - x)\psi(x)\,\mathrm{d}x \right|.
$$

We show later that $|d_6|$ can be made arbitrarily small by choosing Z with sufficiently small variance.

Now, since probability densities integrate to 1, then we can write

$$
\begin{aligned}
\int_{x=0}^{x_0-2\varepsilon} h_{ij}^{--}(\lambda, x_0 - x)\psi(x)\,\mathrm{d}x &= \int_{x=0}^{x_0-2\varepsilon} h_{ij}^{--}(\lambda, x_0 - x)\psi(x)\,\mathrm{d}x \mathbb{P}(|Z - \Delta| \leq \varepsilon \mid Z > x_0) \\
&\quad + \int_{x=0}^{x_0-2\varepsilon} h_{ij}^{--}(\lambda, x_0 - x)\psi(x)\,\mathrm{d}x \mathbb{P}(|Z - \Delta| > \varepsilon \mid Z > x_0) \\
&= \int_{x=0}^{x_0-2\varepsilon} \int_{u=\Delta-x_0-\varepsilon}^{\Delta-x_0+\varepsilon} \boldsymbol{k}(x_0) e^{\boldsymbol{S}u} \boldsymbol{s} h_{ij}^{--}(\lambda, x_0 - x)\psi(x)\,\mathrm{d}u\,\mathrm{d}x \\
&\quad + \int_{x=0}^{x_0-2\varepsilon} h_{ij}^{--}(\lambda, x_0 - x)\psi(x)\,\mathrm{d}x \mathbb{P}(|Z - \Delta| > \varepsilon \mid Z > x_0).
\end{aligned}
$$

Therefore, the first absolute value in (C.17) can be written as

$$
\begin{aligned}
&\Bigg| \int_{x=0}^{x_0-2\varepsilon} \int_{u=\Delta-x_0-\varepsilon}^{\Delta-x_0+\varepsilon} \boldsymbol{k}(x_0) e^{\boldsymbol{S}u} \boldsymbol{s} h_{ij}^{--}(\lambda, \Delta - u - x)\psi(x)\,\mathrm{d}u\,\mathrm{d}x \\
&\quad - \int_{x=0}^{x_0-2\varepsilon} \int_{u=\Delta-x_0-\varepsilon}^{\Delta-x_0+\varepsilon} \boldsymbol{k}(x_0) e^{\boldsymbol{S}u} \boldsymbol{s} h_{ij}^{--}(\lambda, x_0 - x)\psi(x)\,\mathrm{d}u\,\mathrm{d}x \\
&\quad - \int_{x=0}^{x_0-2\varepsilon} h_{ij}^{--}(\lambda, x_0 - x)\psi(x)\,\mathrm{d}x \mathbb{P}(|Z - \Delta| > \varepsilon \mid Z > x_0) \Bigg| \\
&= \Bigg| \int_{x=0}^{x_0-2\varepsilon} \int_{u=\Delta-x_0-\varepsilon}^{\Delta-x_0+\varepsilon} \boldsymbol{k}(x_0) e^{\boldsymbol{S}u} \boldsymbol{s} \left[h_{ij}^{--}(\lambda, \Delta - u - x) - h_{ij}^{--}(\lambda, x_0 - x)\psi(x) \right] \mathrm{d}u\,\mathrm{d}x \\
&\quad - \int_{x=0}^{x_0-2\varepsilon} h_{ij}^{--}(\lambda, x_0 - x)\psi(x)\,\mathrm{d}x \mathbb{P}(|Z - \Delta| > \varepsilon \mid Z > x_0) \Bigg|
\end{aligned}
$$

$$\leq \left| \int_{x=0}^{x_0-2\varepsilon} \int_{u=\Delta-x_0-\varepsilon}^{\Delta-x_0+\varepsilon} \boldsymbol{k}(x_0) e^{\boldsymbol{S}u} \boldsymbol{s} \left[h_{ij}^{--}(\lambda, \Delta - u - x) - h_{ij}^{--}(\lambda, x_0 - x)\psi(x) \right] \mathrm{d}u \, \mathrm{d}x \right| + |d_7|$$
$$\leq \int_{x=0}^{x_0-2\varepsilon} \int_{u=\Delta-x_0-\varepsilon}^{\Delta-x_0+\varepsilon} \boldsymbol{k}(x_0) e^{\boldsymbol{S}u} \boldsymbol{s} \left| h_{ij}^{--}(\lambda, \Delta - u - x) - h_{ij}^{--}(\lambda, x_0 - x) \, || \, \psi(x) \right| \mathrm{d}u \, \mathrm{d}x + |d_7|, \tag{C.18}$$

where

$$|d_7| = \left| \int_{x=0}^{x_0-2\varepsilon} h_{ij}^{--}(\lambda, x_0 - x)\psi(x) \, \mathrm{d}x \mathbb{P}(|Z - \Delta| > \varepsilon \mid Z > x_0) \right|.$$

We show later that $|d_7|$ can be made arbitrarily small by choosing Z with sufficiently small variance.

Since h_{ij}^{--} is Lipschitz and $|(\Delta - u - x) - (x_0 - x)| \leq 2\varepsilon$ for all $u \in (\Delta - x_0 - \varepsilon, \Delta - x_0 + \varepsilon)$, then the first absolute value of (C.18) is less than or equal to

$$\int_{u=\Delta-x_0-\varepsilon}^{\Delta-x_0+\varepsilon} \boldsymbol{k}(x_0) e^{\boldsymbol{S}u} \boldsymbol{s} \, \mathrm{d}u 2L\varepsilon \int_{x=0}^{x_0-2\varepsilon} |\psi(x)| \, \mathrm{d}x. \tag{C.19}$$

Now, $\int_{u=\Delta-x_0-\varepsilon}^{\Delta-x_0+\varepsilon} \boldsymbol{k}(x_0) e^{\boldsymbol{S}u} \boldsymbol{s} \, \mathrm{d}u \leq 1$ as it is a probability, and $\int_{x=0}^{x_0-2\varepsilon} |\psi(x)| \, \mathrm{d}x \leq F\Delta$ as $|\psi| \leq F$ and $x_0 \in (2\varepsilon, \Delta - \varepsilon)$. Therefore, (C.19) is less than or equal to $2\varepsilon LF\Delta$.

What remains is to bound the terms $|d_\ell|$, $\ell = 1, \ldots, 7$.

Since $|\psi| \leq F$ then

$$|d_1| \leq \left| \int_{u=\Delta-\varepsilon}^{\infty} \boldsymbol{k}(x_0) e^{\boldsymbol{S}u} \boldsymbol{s} \int_{x=0}^{\Delta} \int_{x_1=0}^{\infty} \boldsymbol{k}(u) e^{\boldsymbol{S}x_1} \boldsymbol{v}(x) h_{ij}^{--}(\lambda, x_1) \, \mathrm{d}x_1 \, \mathrm{d}x \, \mathrm{d}u \right| F. \tag{C.20}$$

From Property 5.2(iv), (C.20) is less than or equal to

$$\begin{aligned} &\left| \int_{u=\Delta-\varepsilon}^{\infty} \boldsymbol{k}(x_0) e^{\boldsymbol{S}u} \boldsymbol{s} \int_{x_1=0}^{\infty} \boldsymbol{k}(u) e^{\boldsymbol{S}x_1} \boldsymbol{e} h_{ij}^{--}(\lambda, x_1) \, \mathrm{d}x_1 \, \mathrm{d}u \right| F \\ &\leq \left| \int_{u=\Delta-\varepsilon}^{\infty} \boldsymbol{k}(x_0) e^{\boldsymbol{S}u} \boldsymbol{s} \int_{x_1=0}^{\infty} \boldsymbol{k}(u) \boldsymbol{e} h_{ij}^{--}(\lambda, x_1) \, \mathrm{d}x_1 \, \mathrm{d}u \right| F \\ &= \left| \int_{u=\Delta-\varepsilon}^{\infty} \boldsymbol{k}(x_0) e^{\boldsymbol{S}u} \boldsymbol{s} \int_{x_1=0}^{\infty} h_{ij}^{--}(\lambda, x_1) \, \mathrm{d}x_1 \, \mathrm{d}u \right| F, \end{aligned} \tag{C.21}$$

since $\boldsymbol{k}(u) e^{\boldsymbol{S}x_1} \boldsymbol{e}$ is decreasing in x_1 and $\boldsymbol{k}(u)\boldsymbol{e} = 1$. Now, as $h_{ij}^{--}(\lambda, x_1)$ is integrable with $\int_{x_1=0}^{\infty} h_{ij}^{--}(\lambda, x_1) \, \mathrm{d}x_1 \leq \widehat{G}$, then (C.21) is less than or equal to

$$\left| \int_{u=\Delta-\varepsilon}^{\infty} \boldsymbol{k}(x_0) e^{\boldsymbol{S}u} \boldsymbol{s} \, \mathrm{d}u \right| \widehat{G}F = \mathbb{P}(Z > x_0 + \Delta - \varepsilon \mid Z > x_0) \widehat{G}F$$

$$\leq \frac{\mathrm{Var}(Z)/\varepsilon^2}{1-\mathrm{Var}(Z)/\varepsilon^2}\widehat{G}F,$$

by Chebyshev's inequality, since $x_0 \in (2\varepsilon, \Delta - \varepsilon)$.

Since $|\psi(x)| \leq F$, then

$$|d_2| \leq \left| \int_{x=0}^{\Delta} \int_{u=0}^{\Delta-\varepsilon} \boldsymbol{k}(x_0) e^{\boldsymbol{S}u} \boldsymbol{s} r_{\boldsymbol{v}}(u,x)\,\mathrm{d}u\,\mathrm{d}x \right| F \tag{C.22}$$

$$\leq \left| \int_{u=0}^{\Delta-\varepsilon} \boldsymbol{k}(x_0) e^{\boldsymbol{S}u} \boldsymbol{s}\,\mathrm{d}u \right| R_{\boldsymbol{v},2} F \tag{C.23}$$

$$\leq R_{\boldsymbol{v},2} F, \tag{C.24}$$

where the first inequality holds from Property 5.2(v), and the last inequality holds since $\int_{u=0}^{\Delta-\varepsilon} \boldsymbol{k}(x_0) e^{\boldsymbol{S}u} \boldsymbol{s}\,\mathrm{d}u = \mathbb{P}(Z \in (x_0, x_0+\Delta-\varepsilon) \mid Z > x_0) \leq 1$.

Since $|\psi(x)| \leq F$ and $h_{ij}^{--}(\lambda, \Delta - u - x) \leq G$, then

$$\begin{aligned} |d_3| &\leq \int_{u=0}^{\min(\Delta-x_0-\varepsilon, \Delta-x-\varepsilon)} \boldsymbol{k}(x_0) e^{\boldsymbol{S}u} \boldsymbol{s}\,\mathrm{d}u \Delta G F \\ &\leq \mathbb{P}(Z \leq \Delta - \varepsilon \mid Z > x_0)\,\mathrm{d}u \Delta G F \\ &\leq \frac{\mathrm{Var}(Z)/\varepsilon^2}{1-\mathrm{Var}(Z)/\varepsilon^2} \Delta G F, \end{aligned} \tag{C.25}$$

where the last inequality holds from Chebyshev's inequality.

Similarly, since $|\psi(x)| \leq F$ and $h_{ij}^{--}(\lambda, \Delta - u - x) \leq G$, then

$$\begin{aligned} |d_4| &\leq \int_{x=0}^{x_0-2\varepsilon} \int_{u=\Delta-x_0+\varepsilon}^{\Delta-x-\varepsilon} \boldsymbol{k}(x_0) e^{\boldsymbol{S}u} \boldsymbol{s}\,\mathrm{d}u\,\mathrm{d}x G F \\ &= \int_{x=0}^{x_0-2\varepsilon} \mathbb{P}(Z \in [\Delta+\varepsilon, \Delta - x - \varepsilon + x_0] \mid Z > x_0)\,\mathrm{d}x G F \\ &\leq \int_{x=0}^{x_0-2\varepsilon} \mathbb{P}(Z \in [\Delta+\varepsilon, \Delta - \varepsilon + x_0] \mid Z > x_0)\,\mathrm{d}x G F \\ &\leq \mathbb{P}(Z \in [\Delta+\varepsilon, \Delta - \varepsilon + x_0] \mid Z > x_0) \Delta G F \\ &\leq \frac{\mathrm{Var}(Z)/\varepsilon^2}{1-\mathrm{Var}(Z)/\varepsilon^2} \Delta G F. \end{aligned}$$

Since $|\psi(x)| \leq F$ and $h_{ij}^{--}(\lambda, \Delta - u - x) \leq G$, then

$$|d_5| \leq \int_{x=0}^{\Delta-\varepsilon} \int_{u=\Delta-x_0-\varepsilon}^{\Delta-x-\varepsilon} \boldsymbol{k}(x_0) e^{\boldsymbol{S}u} \boldsymbol{s}\,\mathrm{d}u 1(x \in [x_0 - 2\varepsilon, x_0))\,\mathrm{d}x G F \tag{C.26}$$

$$\leq \int_{x=0}^{\Delta-\varepsilon} 1(x \in [x_0 - 2\varepsilon, x_0))\,\mathrm{d}x GF$$
$$= 2\varepsilon GF,$$

where the second inequality holds since $\int_{u=\Delta-x_0-\varepsilon}^{\Delta-x-\varepsilon} \boldsymbol{k}(x_0)e^{\boldsymbol{S}u}\boldsymbol{s}\,\mathrm{d}u \leq 1$.

Since $|\psi(x)| \leq F$ and $h_{ij}^{--}(\lambda, \Delta - u - x) \leq G$, then

$$|d_6| \leq \int_{x=x_0-2\varepsilon}^{x_0} \mathrm{d}x GF$$
$$= 2\varepsilon GF. \tag{C.27}$$

Since $|\psi(x)| \leq F$ and $h_{ij}^{--}(\lambda, \Delta - u - x) \leq G$ then

$$|d_7| \leq \mathbb{P}(|Z - \Delta| > \varepsilon \mid Z > x_0)\Delta GF \tag{C.28}$$
$$\leq \frac{\mathrm{Var}(Z)/\varepsilon^2}{1 - \mathrm{Var}(Z)/\varepsilon^2}\Delta GF \tag{C.29}$$

Convergence follows after setting $\varepsilon^{(p)} = \mathrm{Var}(Z^{(p)})^{1/3}$ and observing that all the bounds $|d_1|, \ldots, |d_7|$ tend to 0, as does the bound on (C.19), given by $2\varepsilon^{(p)}LF\Delta$, as $p \to \infty$. □

Now we show bounds for certain Laplace transform expressions which arise when the QBD-RAP starts in phases in $\mathcal{S}_{+0} \cup \mathcal{S}_{-0}$ and there is more than one up-down or down-up transition before the first change of level. These expressions have the form

$$\int_{x_1=0}^{\infty} g_1(x_1)\boldsymbol{k}(x_0)\boldsymbol{D}e^{\boldsymbol{S}x_1}\,\mathrm{d}x_1\boldsymbol{D}\left[\prod_{n=2}^{k-1}\int_{x_n=0}^{\infty} g_n(x_n)e^{\boldsymbol{S}x_n}\,\mathrm{d}x_n\boldsymbol{D}\right]\int_{x_n=0}^{\infty} g_n(x_n)e^{\boldsymbol{S}x_n}\,\mathrm{d}x_n\boldsymbol{v}(x). \tag{C.30}$$

Here, we ultimately wish to show that (C.30) converges to $g_{1,n}^*(\Delta - x_0, x)$. We do not do this directly, instead, we show that (C.30) is 'close' to $w_n(\Delta - x_0, x)$, then rely on the results from Chapter 5 to get the desired convergence.

Observe that by substituting the first matrix $\boldsymbol{D}$ in the expression above for its integral expression, then (C.30) is equal to

$$\int_{x_1=0}^{\infty} g_1(x_1)\boldsymbol{k}(x_0)\int_{z_0=0}^{\infty} e^{\boldsymbol{S}z_0}\boldsymbol{s}\frac{\boldsymbol{\alpha}e^{\boldsymbol{S}z_0}}{\boldsymbol{\alpha}e^{\boldsymbol{S}z_0}\boldsymbol{e}}\mathrm{d}z_0 e^{\boldsymbol{S}x_1}\,\mathrm{d}x_1\boldsymbol{D}\left[\prod_{n=2}^{k-1}\int_{x_n=0}^{\infty} g_n(x_n)e^{\boldsymbol{S}x_n}\,\mathrm{d}x_n\boldsymbol{D}\right] \tag{C.31}$$
$$\times \int_{x_n=0}^{\infty} g_n(x_n)e^{\boldsymbol{S}x_n}\,\mathrm{d}x_n\boldsymbol{v}(x)$$
$$= \boldsymbol{k}(x_0)\int_{z_0=0}^{\infty} e^{\boldsymbol{S}z_0}\boldsymbol{s}w_n(z_0, x)\,\mathrm{d}z_0. \tag{C.32}$$

Intuitively, when the variance of Z is low, we expect that the integral in (C.32) above will be approximately equal to $w_n(\Delta - x_0, x)$. Indeed, we proved in Lemma 5.6 that this is the case for functions g satisfying the Assumptions 5.1. However, here we do not immediately have that $w_n(x_0, x)$ is Lipschitz in x_0, which we would need for it to satisfy Assumptions 5.1. Instead, we can show a Lipschitz-like condition in x_0 for $w_n(x_0, x)$, which suffices.

For later use, observe that

$$\begin{aligned} g^*_{2,n}(u_1, x) &= \int_{u_2=0}^{\Delta-u_1} g_2(\Delta - u_2 - u_1)\,\mathrm{d}u_1 \ldots \int_{u_{n-1}=0}^{\Delta-u_{n-2}} g_{n-1}(\Delta - u_{n-1} - u_{n-2})\,\mathrm{d}u_{n-2} \\ &\qquad g_n(\Delta - x - u_{n-1})1(\Delta - x - u_{n-1} \geq 0)\,\mathrm{d}u_{n-1} \\ &\leq G^{n-1} \int_{u_2=0}^{\Delta-u_1} \mathrm{d}u_1 \ldots \int_{u_{n-1}=0}^{\Delta-u_{n-2}} \mathrm{d}u_{n-1} \\ &\leq G^{n-1}\Delta^{n-2} := G^*_n. \end{aligned} \tag{C.33}$$

Corollary C.4. *For $x_0, x \in [0, \Delta)$, $n \geq 2$,*

$$|w_n(x_0, x) - w_n(z_0, x)| \leq 2|r_5(n)| + 2|r_6(n)| + 2(n-1)|r_4(n)| + |x_0 - z_0|G^*_n(G + L\Delta). \tag{C.34}$$

Proof. Assume, without loss of generality $x_0 < z_0$. By adding and subtracting both $\int_{u_1=0}^{\Delta-x_0} g_1(\Delta - u_1 - x_0)g^*_{2,n}(u_1, x)\,\mathrm{d}u_1$ and $\int_{u_1=0}^{\Delta-z_0} g_1(\Delta - u_1 - z_0)g^*_{2,n}(u_1, x)\,\mathrm{d}u_1$, we can write the left-hand side of (C.34) as

$$\begin{aligned} \Bigg| & w_n(x_0, x) - \int_{u_1=0}^{\Delta-x_0} g_1(\Delta - u_1 - x_0)g^*_{2,n}(u_1, x)\,\mathrm{d}u_1 \\ & - w_n(z_0, x) + \int_{u_1=0}^{\Delta-z_0} g_1(\Delta - u_1 - z_0)g^*_{2,n}(u_1, x)\,\mathrm{d}u_1 \\ & + \int_{u_1=0}^{\Delta-x_0} g_1(\Delta - u_1 - x_0)g^*_{2,n}(u_1, x)\,\mathrm{d}u_1 - \int_{u_1=0}^{\Delta-z_0} g_1(\Delta - u_1 - z_0)g^*_{2,n}(u_1, x)\,\mathrm{d}u_1 \Bigg| \end{aligned}$$

which, by the triangle inequality, is less than or equal to

$$\begin{aligned} & \left| w_n(x_0, x) - \int_{u_1=0}^{\Delta-x_0} g_1(\Delta - u_1 - x_0)g^*_{2,n}(u_1, x)\,\mathrm{d}u_1 \right| \\ & + \left| w_n(z_0, x) - \int_{u_1=0}^{\Delta-z_0} g_1(\Delta - u_1 - z_0)g^*_{2,n}(u_1, x)\,\mathrm{d}u_1 \right| \end{aligned}$$

$$+\left|\int_{u_1=0}^{\Delta-x_0} g_1(\Delta-u_1-x_0)g_{2,n}^*(u_1,x)\,\mathrm{d}u_1-\int_{u_1=0}^{\Delta-z_0} g_1(\Delta-u_1-z_0)g_{2,n}^*(u_1,x)\,\mathrm{d}u_1\right|. \tag{C.35}$$

By Corollary 5.9, the first two terms of (C.35) are less than or equal to

$$|r_5(n)|+|r_6(n)|+(n-1)|r_4(n)|.$$

As for the last term, adding and subtracting $\int_{u_1=0}^{\Delta-z_0} g_1(\Delta-u_1-x_0)g_{2,n}^*(u_1,x)\,\mathrm{d}u_1$ gives

$$\begin{aligned}
&= \left|\int_{u_1=0}^{\Delta-x_0} g_1(\Delta-u_1-x_0)g_{2,n}^*(u_1,x)\,\mathrm{d}u_1-\int_{u_1=0}^{\Delta-z_0} g_1(\Delta-u_1-x_0)g_{2,n}^*(u_1,x)\,\mathrm{d}u_1\right.\\
&\qquad\left.-\int_{u_1=0}^{\Delta-z_0}(g_1(\Delta-u_1-z_0)-g_1(\Delta-u_1-x_0))g_{2,n}^*(u_1,x)\,\mathrm{d}u_1\right|\\
&\le \left|\int_{u_1=\Delta-z_0}^{\Delta-x_0} g_1(\Delta-u_1-x_0)g_{2,n}^*(u_1,x)\,\mathrm{d}u_1\right|\\
&\qquad+\int_{u_1=0}^{\Delta-z_0}|g_1(\Delta-u_1-z_0)-g_1(\Delta-u_1-x_0)|g_{2,n}^*(u_1,x)\,\mathrm{d}u_1\\
&\le GG_n^*|x_0-z_0|+\int_{u_1=0}^{\Delta-z_0} L|x_0-z_0|G_n^*\,\mathrm{d}u_1,
\end{aligned} \tag{C.36}$$

since g_1 is Lipschitz by Assumption 5.1(iv), and $g_{2,n}^*\le G_n^*$. Bounding the integral over u_1 by Δ, then (C.36) is less than or equal to

$$GG_n^*|x_0-z_0|+\Delta L|x_0-z_0|G_n^*. \tag{C.37}$$

□

Corollary C.5. *Let $g_1,g_2,\ldots,$ be functions satisfying Assumptions 5.1 and let $\boldsymbol{v}(x)$, $x\in[0,\Delta)$, be a closing operator with Properties 5.2. For $x_0,x\in[0,\Delta)$, $n\ge 2$,*

$$\left|\boldsymbol{k}(x_0)\int_{z_0=0}^{\infty} e^{\boldsymbol{S}z_0}\boldsymbol{s}w_n(z_0,x)\,\mathrm{d}z_0-w_n(\Delta-x_0,x)\right|=r_8(n),$$

where

$$\begin{aligned}
|r_8(n)|\le{}&(2|r_5(n)|+2|r_6(n)|+2(n-1)|r_4(n)|+\varepsilon G_n^*(G+L\Delta))\\
&+2\widehat{G}^{n-2}GG_{\boldsymbol{v}}\frac{\mathrm{Var}(Z)/\varepsilon^2}{1-\mathrm{Var}(Z)/(\Delta-x_0)^2}.
\end{aligned}$$

Proof. Consider

$$\begin{aligned}
&\left|\boldsymbol{k}(x_0)\int_{z_0=0}^{\infty} e^{\boldsymbol{S}z_0}\boldsymbol{s}w_n(z_0,x)\,\mathrm{d}z_0 - w_n(\Delta-x_0,x)\right| \\
&= \left|\boldsymbol{k}(x_0)\int_{z_0=0}^{\infty} e^{\boldsymbol{S}z_0}\boldsymbol{s}(w_n(z_0,x)-w_n(\Delta-x_0,x))\,\mathrm{d}z_0\right| \\
&\leq \boldsymbol{k}(x_0)\int_{z_0=0}^{\infty} e^{\boldsymbol{S}z_0}\boldsymbol{s}\,|w_n(z_0,x)-w_n(\Delta-x_0,x)|\,\mathrm{d}z_0 \\
&= \boldsymbol{k}(x_0)\int_{z_0=0}^{\Delta-\varepsilon-x_0} e^{\boldsymbol{S}z_0}\boldsymbol{s}\,|w_n(z_0,x)-w_n(\Delta-x_0,x)|\,\mathrm{d}z_0 \\
&\quad + \boldsymbol{k}(x_0)\int_{z_0=\Delta+\varepsilon-x_0}^{\infty} e^{\boldsymbol{S}z_0}\boldsymbol{s}\,|w_n(z_0,x)-w_n(\Delta-x_0,x)|\,\mathrm{d}z_0 \\
&\quad + \boldsymbol{k}(x_0)\int_{z_0=\Delta-\varepsilon-x_0}^{\Delta+\varepsilon-x_0} e^{\boldsymbol{S}z_0}\boldsymbol{s}\,|w_n(z_0,x)-w_n(\Delta-x_0,x)|\,\mathrm{d}z_0. \qquad \text{(C.38)}
\end{aligned}$$

Using Equations (5.52), (5.60) and (5.73) we can claim

$$\begin{aligned}
|w_n(x_0,x)| &\leq \frac{1}{\boldsymbol{\alpha}e^{\boldsymbol{S}x_0}\boldsymbol{e}}G^2\widehat{G}^{n-2}\int_{u_k=0}^{\infty}\boldsymbol{\alpha}e^{\boldsymbol{S}u_k}\boldsymbol{e}\,\mathrm{d}u_k G_{\boldsymbol{v}} + \frac{1}{\boldsymbol{\alpha}e^{\boldsymbol{S}x_0}\boldsymbol{e}}G\widehat{G}^n\widetilde{G}_{\boldsymbol{v}} \\
&= \frac{1}{\boldsymbol{\alpha}e^{\boldsymbol{S}x_0}\boldsymbol{e}}G^2\widehat{G}^{n-2}G_{\boldsymbol{v}} + \frac{1}{\boldsymbol{\alpha}e^{\boldsymbol{S}x_0}\boldsymbol{e}}G\widehat{G}^n\widetilde{G}_{\boldsymbol{v}} \\
&=: W_n. \qquad \text{(C.39)}
\end{aligned}$$

Therefore, the sum of the first two terms in (C.38) is less than or equal to

$$\begin{aligned}
&2W_n\left(\int_{z_0=0}^{\Delta-\varepsilon-x_0}\boldsymbol{k}(x_0)e^{\boldsymbol{S}z_0}\boldsymbol{s}\,\mathrm{d}z_0 + \int_{z_0=\Delta+\varepsilon-x_0}^{\infty}\boldsymbol{k}(x_0)e^{\boldsymbol{S}z_0}\boldsymbol{s}\,\mathrm{d}z_0\right) \\
&= 2W\frac{\mathbb{P}(|Z-\Delta|>\varepsilon)}{\mathbb{P}(Z>x_0)} \\
&\leq 2W_n\frac{\mathrm{Var}(Z)/\varepsilon^2}{1-\mathrm{Var}(Z)/(\Delta-x_0)^2} \qquad \text{(C.40)}
\end{aligned}$$

by Chebyshev's inequality. As for the last term in (C.38), we can use Corollary C.4 to bound the integrand so that the last term is less than or equal to

$$\begin{aligned}
&\boldsymbol{k}(x_0)\int_{z_0=\Delta-\varepsilon-x_0}^{\Delta+\varepsilon-x_0} e^{\boldsymbol{S}z_0}\boldsymbol{s}\left(2|r_5(n)|+2|r_6(n)|+2(n-1)|r_4(n)|+\varepsilon G^{n-1}\Delta^{n-2}(G+L\Delta)\right)\mathrm{d}z_0 \\
&\leq \left(2|r_5(n)|+2|r_6(n)|+2(n-1)|r_4(n)|+\varepsilon G^{n-1}\Delta^{n-2}(G+L\Delta)\right), \qquad \text{(C.41)}
\end{aligned}$$

since $\boldsymbol{k}(x_0)\displaystyle\int_{z_0=\Delta-\varepsilon-x_0}^{\Delta+\varepsilon-x_0} e^{\boldsymbol{S}z_0}\boldsymbol{s}\,\mathrm{d}z_0 \leq 1$. Thus, (C.38) is bounded by (C.41). □

Corollary C.6. *Let $g_1, g_2, \ldots,$ be functions satisfying Assumptions 5.1 and let $\boldsymbol{v}(x)$, $x \in (0, \Delta)$, be a closing operator with Properties 5.2. For $x_0, x \in (0, \Delta)$, $n \geq 2$*

$$\begin{aligned}&\left|\boldsymbol{k}(x_0)\int_{z_0=0}^{\infty} e^{\boldsymbol{S}z_0}\boldsymbol{s}w_n(z_0,x)\,\mathrm{d}z_0 - g^*_{1,n}(\Delta - x_0, x)\right| \\ &\leq |r_8(n)| + |r_5(n)| + |r_6(n)| + (n-1)|r_4(n)|. \end{aligned} \tag{C.42}$$

Proof. Adding and subtracting $w_n(\Delta - x_0, x)$ within the absolute value on the left-hand side of (C.42)

$$\begin{aligned}&\left|\boldsymbol{k}(x_0)\int_{z_0=0}^{\infty} e^{\boldsymbol{S}z_0}\boldsymbol{s}w_n(z_0,x)\,\mathrm{d}z_0 - w_n(\Delta - x_0, x) + w_n(\Delta - x_0, x) - g^*_{1,n}(\Delta - x_0, x)\right| \\ &\leq \left|\boldsymbol{k}(x_0)\int_{z_0=0}^{\infty} e^{\boldsymbol{S}z_0}\boldsymbol{s}w_n(z_0,x)\,\mathrm{d}z_0 - w_n(\Delta - x_0, x)\right| + \left|w_n(\Delta - x_0, x) - g^*_{1,n}(\Delta - x_0, x)\right|\end{aligned}$$

where the first absolute value is less than or equal to $|r_8(n)|$ by Corollary C.5 and the second absolute value is less than or equal to $|r_5(n)| + |r_6(n)| + (n-1)|r_4(n)|$ by Corollary 5.9. □

Corollary C.7. *Let ψ be bounded and Lipschitz, let $g_1, g_2, \ldots,$ be functions satisfying Assumptions 5.1 and let $\boldsymbol{v}(x)$, $x \in (0, \Delta)$, be a closing operator with Properties 5.2. For $x_0, x \in (0, \Delta)$, $n \geq 2$*

$$\begin{aligned}&\left|\int_{x\in[0,\Delta)} \boldsymbol{k}(x_0)\int_{z_0=0}^{\infty} e^{\boldsymbol{S}z_0}\boldsymbol{s}w_n(z_0,x)\,\mathrm{d}z_0\psi(x)\,\mathrm{d}x - \int_{x\in[0,\Delta)} g^*_{1,n}(\Delta - x_0, x)\psi(x)\,\mathrm{d}x\right| \\ &\leq \left(|r_8(n)| + |r_5(n)| + |r_6(n)| + (n-1)|r_4(n)|\right) F\Delta. \end{aligned} \tag{C.43}$$

Proof. The left-hand side of (C.43) is less than or equal to

$$\int_{x\in[0,\Delta)} \left|\boldsymbol{k}(x_0)\int_{z_0=0}^{\infty} e^{\boldsymbol{S}z_0}\boldsymbol{s}w_n(z_0,x)\,\mathrm{d}z_0 - g^*_{1,n}(\Delta - x_0, x)\right| |\psi(x)|\,\mathrm{d}x. \tag{C.44}$$

Now, using $|\psi(x)| \leq F$ and Corollary C.6 then (C.44) is less than or equal to

$$\left(|r_8(n)| + |r_5(n)| + |r_6(n)| + (n-1)|r_4(n)|\right)\Delta F.$$

□

We now extend the previous results to the matrix case.

Lemma C.8. *Let $\boldsymbol{G}_k(x)$, $k \in \{1, 2, ...\}$, be matrix functions with dimensions $N_k \times N_{k+1}$. Further, suppose the scalar functions $[\boldsymbol{G}_k(x)]_{ij}$, $k \in \{1, 2, ...\}$ satisfy Assumptions 5.1. Then,*

$$\left|\int_{x\in[0,\Delta)}\int_{x_1=0}^{\infty} \boldsymbol{G}_1(x_1)\otimes\boldsymbol{k}(x_0)\boldsymbol{D}e^{\boldsymbol{S}x_1}\,\mathrm{d}x_1\boldsymbol{D}\left[\prod_{k=2}^{n-1}\int_{x_k=0}^{\infty}\boldsymbol{G}_k(x_k)\otimes e^{\boldsymbol{S}x_k}\,\mathrm{d}x_k\boldsymbol{D}\right]\right.$$

$$\times \int_{x_n=0}^{\infty} \boldsymbol{G}_n(x_n) \otimes e^{\boldsymbol{S}x_n}\,\mathrm{d}x_n \boldsymbol{v}(x)\psi(x)\,\mathrm{d}x$$
$$- \int_{x\in[0,\Delta)} \int_{u_1=0}^{x_0} \boldsymbol{G}_1(x_0-u_1) \left[\prod_{k=2}^{n-1} \int_{u_k=0}^{\Delta-u_{k-1}} \boldsymbol{G}_k(\Delta-u_k-u_{k-1})\,\mathrm{d}u_{k-1}\right]$$
$$\boldsymbol{G}_n(\Delta-x-u_{n-1}) \times 1(\Delta-x-u_{n-1}\geq 0)\,\mathrm{d}u_{n-1}\psi(x)\,\mathrm{d}x\Bigg|$$
$$\leq (|r_8(n)| + |r_5(n)| + |r_6(n)| + (n-1)|r_4(n)|)\, F\Delta \prod_{k=2}^{n} N_k. \tag{C.45}$$

Moreover, choosing $\varepsilon = \mathrm{Var}(Z)$, then, for fixed n, the bound is $\mathcal{O}(\mathrm{Var}(Z)^{1/3})$.

Proof. The proof is the same as the proof of Lemma 5.11, with Corollary C.7 replacing Corollary 5.10. □

Lemma C.8 effectively shows that, as $p\to\infty$, then

$$\int_{x\in[0,\Delta)} \int_{x_1=0}^{\infty} \boldsymbol{G}_1(x_1)\otimes \boldsymbol{k}^{(p)}(x_0)\boldsymbol{D}^{(p)} e^{\boldsymbol{S}^{(p)}x_1}\,\mathrm{d}x_1 \boldsymbol{D}^{(p)} \left[\prod_{k=2}^{n-1}\int_{x_k=0}^{\infty} \boldsymbol{G}_k(x_k)\otimes e^{\boldsymbol{S}^{(p)}x_k}\,\mathrm{d}x_k \boldsymbol{D}^{(p)}\right]$$
$$\times \int_{x_n=0}^{\infty} \boldsymbol{G}_n(x_n)\otimes e^{\boldsymbol{S}^{(p)}x_n}\,\mathrm{d}x_n \boldsymbol{v}^{(p)}(x)\psi(x)\,\mathrm{d}x$$
$$\to \int_{x\in[0,\Delta)} \int_{u_1=0}^{x_0} \boldsymbol{G}_1(x_0-u_1)\left[\prod_{k=2}^{n-1}\int_{u_k=0}^{\Delta-u_{k-1}} \boldsymbol{G}_k(\Delta-u_k-u_{k-1})\,\mathrm{d}u_{k-1}\right]$$
$$\boldsymbol{G}_n(\Delta-x-u_{n-1})\times 1(\Delta-x-u_{n-1}\geq 0)\,\mathrm{d}u_{n-1}\psi(x)\,\mathrm{d}x.$$

The technical results in this section are enough to prove Theorem C.2.

C.1.1 More results

We are not quite done yet. If we want to use Theorem C.2 to prove convergence before the first orbit restart epoch, we need a domination condition like that in Lemma 5.12.

Lemma C.9. *For all $M\geq 0$, $x\in\mathcal{D}_{\ell_0,j}$, $x_0\in\mathcal{D}_{\ell_0,i}$, $\ell_0\in\mathcal{K}$, $\lambda>0$, $q\in\{+0,-0\}$, $r\in\{+,-\}$, $i\in\mathcal{S}_q$, $j\in\mathcal{S}_r\cup\mathcal{S}_{r0}$, and for any bounded function ψ, $|\psi|<F$,*

$$\sum_{m=M+1}^{\infty} \left| \int_{x\in\mathcal{D}_{\ell_0}} \widehat{f}^{\ell_0,(p)}_{m,q,r}(\lambda)(x,j;x_0,i)\psi(x)\,\mathrm{d}x - \int_{x\in\mathcal{D}_{\ell_0}} \widehat{\mu}^{\ell_0}_{m,q,r}(\lambda)(\,\mathrm{d}x,j;x_0,i)\psi(x)\right| \leq r_6^M \tag{C.46}$$

where

$$r_6^M = F(G\Delta + \widehat{G})\left(\frac{\gamma}{\gamma+\lambda}\right)^{2M+2}\left(1-\left(\frac{\gamma}{\gamma+\lambda}\right)^2\right)^{-1}.$$

Proof. The proof follows the same arguments as the proof for the case $q = 0$, $i \in \mathcal{S}_0^*$ in the proof of Lemma 5.12.

□

The last thing we need to prove is convergence at the first change of level. Since the result in Corollary C.6 is pointwise in x, choosing the closing operator as $e^{\boldsymbol{S}x}\boldsymbol{s}$ and setting $x = 0$, then we get convergence at the first change of level, on the event that are $m > 0$ up-down or down-up transitions. The only things that remain are to show convergence at the first change of level on the event that there is no up-down or down-up transitions, and a domination condition so that we may sum over the number of up-down and down-up transitions ad prove convergence at the time of the first orbit restart epoch (analogous to the domination condition in the proof of Lemma 5.15). Regarding the former, we have the following lemma.

Lemma C.10. *Let g satisfy the Assumptions 5.1 and $x_0 \in (2\varepsilon, \Delta - \varepsilon)$. Then*

$$\left|\int_{x=0}^{\infty} \boldsymbol{k}(x_0)\boldsymbol{D}e^{\boldsymbol{S}x}g(x)\boldsymbol{s}\,\mathrm{d}x - g(x_0)\right| \leq \frac{\mathrm{Var}(Z)/\varepsilon^2}{1-\mathrm{Var}(Z)/(\Delta-x_0)^2}4G + 3L\varepsilon + 6G\frac{\mathrm{Var}(Z)}{\varepsilon^2}. \tag{C.47}$$

Proof. First rewrite the left-hand side as

$$\left|\int_{x=0}^{\infty} \boldsymbol{k}(x_0)\boldsymbol{D}e^{\boldsymbol{S}x}(g(x)-g(x_0))\boldsymbol{s}\,\mathrm{d}x\right| \leq \int_{x=0}^{\infty} \boldsymbol{k}(x_0)\boldsymbol{D}e^{\boldsymbol{S}x}|g(x)-g(x_0)|\boldsymbol{s}\,\mathrm{d}x. \tag{C.48}$$

Substituting in the expression for $\boldsymbol{D}$ gives,

$$\begin{aligned}
&\int_{x=0}^{\infty} \boldsymbol{k}(x_0)\int_{u=0}^{\infty} e^{\boldsymbol{S}u}\boldsymbol{s}\frac{\boldsymbol{\alpha}e^{\boldsymbol{S}u}}{\boldsymbol{\alpha}e^{\boldsymbol{S}u}\boldsymbol{e}}\mathrm{d}u e^{\boldsymbol{S}x}|g(x)-g(x_0)|\boldsymbol{s}\,\mathrm{d}x \\
&= \int_{x=0}^{\infty} \boldsymbol{k}(x_0)\int_{u=0}^{\Delta-\varepsilon} e^{\boldsymbol{S}u}\boldsymbol{s}\frac{\boldsymbol{\alpha}e^{\boldsymbol{S}u}}{\boldsymbol{\alpha}e^{\boldsymbol{S}u}\boldsymbol{e}}\mathrm{d}u e^{\boldsymbol{S}x}|g(x)-g(x_0)|\boldsymbol{s}\,\mathrm{d}x \\
&\quad + \int_{x=0}^{\infty} \boldsymbol{k}(x_0)\int_{u=\Delta-\varepsilon}^{\infty} e^{\boldsymbol{S}u}\boldsymbol{s}\frac{\boldsymbol{\alpha}e^{\boldsymbol{S}u}}{\boldsymbol{\alpha}e^{\boldsymbol{S}u}\boldsymbol{e}}\mathrm{d}u e^{\boldsymbol{S}x}|g(x)-g(x_0)|\boldsymbol{s}\,\mathrm{d}x.
\end{aligned} \tag{C.49}$$

Since g is bounded, the second term is less than or equal to

$$\int_{x=0}^{\infty} \boldsymbol{k}(x_0)\int_{u=\Delta-\varepsilon}^{\infty} e^{\boldsymbol{S}u}\boldsymbol{s}\frac{\boldsymbol{\alpha}e^{\boldsymbol{S}u}}{\boldsymbol{\alpha}e^{\boldsymbol{S}u}\boldsymbol{e}}\mathrm{d}u e^{\boldsymbol{S}x}\boldsymbol{s}\,\mathrm{d}x 2G = \boldsymbol{k}(x_0)\int_{u=\Delta-\varepsilon}^{\infty} e^{\boldsymbol{S}u}\boldsymbol{s}\frac{\boldsymbol{\alpha}e^{\boldsymbol{S}u}}{\boldsymbol{\alpha}e^{\boldsymbol{S}u}\boldsymbol{e}}\mathrm{d}u\boldsymbol{e}2G$$

$$= \boldsymbol{k}(x_0) \int_{u=\Delta-\varepsilon}^{\infty} e^{\boldsymbol{S}u} \boldsymbol{s} \, \mathrm{d}u 2G$$
$$= \frac{\mathbb{P}(Z \geq x_0 + \Delta - \varepsilon)}{\mathbb{P}(Z > x_0)} 2G. \tag{C.50}$$

For $x_0 \in (2\varepsilon, \Delta - \varepsilon)$, then (C.50) is less than or equal to

$$\frac{\mathrm{Var}(Z)/\varepsilon^2}{1 - \mathrm{Var}(Z)/(\Delta - x_0)^2} 2G.$$

As for the first term in (C.49), it can be written as

$$\begin{aligned}
&\int_{x=0}^{\infty} \boldsymbol{k}(x_0) \int_{u=\Delta-x_0-\varepsilon}^{u=\Delta-x_0+\varepsilon} e^{\boldsymbol{S}u} \boldsymbol{s} \frac{\boldsymbol{\alpha} e^{\boldsymbol{S}u}}{\boldsymbol{\alpha} e^{\boldsymbol{S}u} \boldsymbol{e}} \mathrm{d}u e^{\boldsymbol{S}x} |g(x) - g(x_0)| \boldsymbol{s} \, \mathrm{d}x \\
&\quad + \int_{x=0}^{\infty} \boldsymbol{k}(x_0) \int_{u=0}^{\Delta-x_0-\varepsilon} e^{\boldsymbol{S}u} \boldsymbol{s} \frac{\boldsymbol{\alpha} e^{\boldsymbol{S}u}}{\boldsymbol{\alpha} e^{\boldsymbol{S}u} \boldsymbol{e}} \mathrm{d}u e^{\boldsymbol{S}x} |g(x) - g(x_0)| \boldsymbol{s} \, \mathrm{d}x \\
&\quad + \int_{x=0}^{\infty} \boldsymbol{k}(x_0) \int_{u=\Delta-x_0+\varepsilon}^{\Delta-\varepsilon} e^{\boldsymbol{S}u} \boldsymbol{s} \frac{\boldsymbol{\alpha} e^{\boldsymbol{S}u}}{\boldsymbol{\alpha} e^{\boldsymbol{S}u} \boldsymbol{e}} \mathrm{d}u e^{\boldsymbol{S}x} |g(x) - g(x_0)| \boldsymbol{s} \, \mathrm{d}x.
\end{aligned} \tag{C.51}$$

Since g is bounded, then the last two terms in (C.51) are

$$\begin{aligned}
&2G \Bigg(\int_{x=0}^{\infty} \boldsymbol{k}(x_0) \int_{u=0}^{\Delta-x_0-\varepsilon} e^{\boldsymbol{S}u} \boldsymbol{s} \frac{\boldsymbol{\alpha} e^{\boldsymbol{S}u}}{\boldsymbol{\alpha} e^{\boldsymbol{S}u} \boldsymbol{e}} \mathrm{d}u e^{\boldsymbol{S}x} \boldsymbol{s} \, \mathrm{d}x \\
&\qquad + \int_{x=0}^{\infty} \boldsymbol{k}(x_0) \int_{u=\Delta-x_0+\varepsilon}^{\Delta-\varepsilon} e^{\boldsymbol{S}u} \boldsymbol{s} \frac{\boldsymbol{\alpha} e^{\boldsymbol{S}u}}{\boldsymbol{\alpha} e^{\boldsymbol{S}u} \boldsymbol{e}} \mathrm{d}u e^{\boldsymbol{S}x} \boldsymbol{s} \, \mathrm{d}x \Bigg) \\
&= 2G \Bigg(\boldsymbol{k}(x_0) \int_{u=0}^{\Delta-x_0-\varepsilon} e^{\boldsymbol{S}u} \boldsymbol{s} \frac{\boldsymbol{\alpha} e^{\boldsymbol{S}u}}{\boldsymbol{\alpha} e^{\boldsymbol{S}u} \boldsymbol{e}} \boldsymbol{e} \, \mathrm{d}u + \boldsymbol{k}(x_0) \int_{u=\Delta-x_0+\varepsilon}^{\Delta-\varepsilon} e^{\boldsymbol{S}u} \boldsymbol{s} \frac{\boldsymbol{\alpha} e^{\boldsymbol{S}u}}{\boldsymbol{\alpha} e^{\boldsymbol{S}u} \boldsymbol{e}} \boldsymbol{e} \, \mathrm{d}u \Bigg) \\
&= 2G \frac{\mathbb{P}(Z > x_0, Z \notin (\Delta - \varepsilon, \Delta + \varepsilon))}{\mathbb{P}(Z > x_0)} \\
&\leq 2G \frac{\mathrm{Var}(Z)/\varepsilon^2}{1 - \mathrm{Var}(Z)/(\Delta - x_0)^2}.
\end{aligned} \tag{C.52}$$

Exchanging the order of integration for the first term in (C.51) (justified by the Fubini-Tonelli Theorem)

$$\int_{u=\Delta-x_0-\varepsilon}^{u=\Delta-x_0+\varepsilon} \boldsymbol{k}(x_0) e^{\boldsymbol{S}u} \boldsymbol{s} \int_{x=0}^{\infty} \frac{\boldsymbol{\alpha} e^{\boldsymbol{S}u}}{\boldsymbol{\alpha} e^{\boldsymbol{S}u} \boldsymbol{e}} e^{\boldsymbol{S}x} |g(x) - g(x_0)| \boldsymbol{s} \, \mathrm{d}x \, \mathrm{d}u, \tag{C.53}$$

from which we see that we can apply Corollary 5.7 with $v = 0$ to the integral over x, implying that (C.53) is less than or equal to

$$\int_{u=\Delta-x_0-\varepsilon}^{u=\Delta-x_0+\varepsilon} \boldsymbol{k}(x_0)e^{\boldsymbol{S}u}\boldsymbol{s}\left(|g(\Delta-u)-g(x_0)|+r_3(u)\right)\,\mathrm{d}u. \tag{C.54}$$

Noting that $\sup|r_3(u)| \leq |r_2|$ for $u \leq \Delta - \varepsilon$, and since g is Lipschitz, then (C.54) is less than or equal to

$$\int_{u=\Delta-x_0-\varepsilon}^{u=\Delta-x_0+\varepsilon} \boldsymbol{k}(x_0)e^{\boldsymbol{S}u}\boldsymbol{s}\left(L\varepsilon+|r_2|\right)\,\mathrm{d}u \leq L\varepsilon+|r_2|. \tag{C.55}$$

Putting all the bounds together proves the result. □

We conclude this appendix but stating the convergence results formally.

Lemma C.11. *For all $x \in \mathcal{D}_{\ell_0,j}$, $x_0 \in \mathcal{D}_{\ell_0,i}$, $i \in \mathcal{S}_{0+}\cup\mathcal{S}_{0-}$, $j \in \mathcal{S}$, $\ell_0 \in \mathcal{K}$, $\lambda > 0$,*

$$\left|\int_{x\in\mathcal{D}_{\ell_0}} \widehat{f}^{\ell_0,(p)}(\lambda)(x,j;x_0,i)\psi(x)\,\mathrm{d}x - \int_{x\in\mathcal{D}_{\ell_0}} \widehat{\mu}^{\ell_0}(\lambda)(\,\mathrm{d}x,j;x_0,i)\psi(x)\right| \to 0 \tag{C.56}$$

as $p \to \infty$.

Proof. The convergence in Theorem C.2, the domination condition in Lemma C.9, and the Dominated Convergence Theorem can be used to obtain the result. □

Corollary C.12. *Recall $\boldsymbol{y}_0^{(p)} = (\ell_0, \boldsymbol{a}_{\ell_0,j}^{(p)}(x_0), i)$. For $\ell_0 \in \mathcal{K}$ $x_0 \in \mathcal{D}_{\ell_0,i}$, $i \in \mathcal{S}_{+0}\cup\mathcal{S}_{0-}$, $j \in \mathcal{S}_+\cup\mathcal{S}_-$,*

$$\begin{aligned}&\mathbb{P}(L^{(p)}(\tau_1^{(p)}) = \ell(\ell_0,j), \varphi(\tau_1^{(p)}) = j, \tau_1^{(p)} \leq E^\lambda \mid \boldsymbol{Y}^{(p)}(0) = \boldsymbol{y}_0^{(p)})\\ &\to \mathbb{P}(\boldsymbol{X}(\tau_1^X) = (y_{\ell(\ell_0,j)+1(j\in\mathcal{S}_-)}, j), \tau_1^X \leq E^\lambda \mid \boldsymbol{X}(0) = (x_0,i))\end{aligned} \tag{C.57}$$

where $\ell(\ell_0,j)$ can take values

$$\ell(\ell_0,j) = \begin{cases} \ell_0 - 1, & \text{if } \ell_0 \in \{0,1,\ldots,K+1\},\ j \in \mathcal{S}_- \\ \ell_0, & \text{if } \ell_0 = 0, j \in \mathcal{S}_+,\ \text{or } \ell_0 = K, j \in \mathcal{S}_-, \\ \ell_0 + 1, & \text{if } \ell_0 \in \{-1,0,1,\ldots,K\},\ j \in \mathcal{S}_+. \end{cases}$$

Proof. An analogue of the domination condition required in the proof of Lemma 5.15 can be established by extending Lemma C.9 in the same way we extended Lemma 5.12 in the proof of Lemma 5.15.

With the aforementioned domination condition, the point-wise convergence in Corollary C.6, the convergence in Lemma C.10, and the Dominated Convergence Theorem we can prove the result. □

Appendix D

Kronecker properties

Here we detail some properties of Kronecker sums, products, and exponential (see (Bladt & Nielsen 2017), Appendix A.4).

Let

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & \dots & a_{1m} \\ \dots & & \dots \\ a_{n1} & \dots & a_{nm} \end{bmatrix} \qquad \boldsymbol{B} = \begin{bmatrix} b_{11} & \dots & b_{1m'} \\ \dots & & \dots \\ b_{n'1} & \dots & b_{n'm'} \end{bmatrix}$$

be matrices with dimensions $n \times m$ and $n' \times m'$, respectively. The operator $\otimes$ is the Kronecker product of two matrices;

$$\boldsymbol{A} \otimes \boldsymbol{B} = \begin{bmatrix} a_{11}\boldsymbol{B} & \dots & a_{1m}\boldsymbol{B} \\ \dots & & \dots \\ a_{n1}\boldsymbol{B} & \dots & a_{nm}\boldsymbol{B} \end{bmatrix},$$

which is an $nn' \times mm'$ matrix.

Let $\boldsymbol{C}, \boldsymbol{D}$ be matrices with dimensions $m \times k$ and $m' \times k'$. A property of the Kronecker Product is

$$(\boldsymbol{A} \otimes \boldsymbol{B})(\boldsymbol{C} \otimes \boldsymbol{D}) = \boldsymbol{AC} \otimes \boldsymbol{BD}. \qquad \text{(Mixed Product Rule)}$$

Proof. The proof follows from

$$\begin{bmatrix} a_{i1}\boldsymbol{B} & a_{i2}\boldsymbol{B} & \dots & a_{in}\boldsymbol{B} \end{bmatrix} \begin{bmatrix} c_{1j}\boldsymbol{D} \\ c_{2j} \\ \vdots \\ c_{nj}\boldsymbol{D} \end{bmatrix} = \left(\sum_{\ell} a_{i\ell} c_{\ell j} \right) \boldsymbol{BD}$$
$$= (\boldsymbol{AC})_{ij}\, \boldsymbol{BD}.$$

□

If $\boldsymbol{A}$ and $\boldsymbol{B}$ are invertible matrices, then

$$(\boldsymbol{A} \otimes \boldsymbol{B})^{-1} = \boldsymbol{A}^{-1} \otimes \boldsymbol{B}^{-1}. \tag{D.1}$$

Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be $n \times n$ and $m \times m$ matrices, respectively. The Kronecker sum of $\boldsymbol{A}$ and $\boldsymbol{B}$ is denoted by $\oplus$ and defined as

$$\boldsymbol{A} \oplus \boldsymbol{B} := \boldsymbol{A} \otimes \boldsymbol{I}_m + \boldsymbol{I}_n \otimes \boldsymbol{B}.$$

The exponential of a square matrix $\boldsymbol{B}$ is

$$e^{\boldsymbol{B}} := \sum_{n=0}^{\infty} \frac{1}{n!} \boldsymbol{B}^n.$$

A property of the Kronecker sum is

$$e^{\boldsymbol{A} \oplus \boldsymbol{B}} = e^{\boldsymbol{A}} \otimes e^{\boldsymbol{B}}. \tag{D.2}$$

Proof. First, the matrices $\boldsymbol{A} \otimes \boldsymbol{I}_m$ and $\boldsymbol{I}_n \otimes \boldsymbol{B}$ commute; from the mixed product rule their product is $\boldsymbol{A} \otimes \boldsymbol{B}$. Hence,

$$e^{\boldsymbol{A} \oplus \boldsymbol{B}} = e^{\boldsymbol{A} \otimes \boldsymbol{I}_m} e^{\boldsymbol{I}_n \otimes \boldsymbol{B}}.$$

We now show that $e^{\boldsymbol{A} \otimes \boldsymbol{I}_m} = e^{\boldsymbol{A}} \otimes \boldsymbol{I}_m$ and $e^{\boldsymbol{I}_n \otimes \boldsymbol{B}} = \boldsymbol{I}_n \otimes e^{\boldsymbol{B}}$. The latter follows from the fact that $\boldsymbol{I}_n \otimes \boldsymbol{B}$ is a block diagonal matrix with blocks $\boldsymbol{B}$, hence its exponential is also block diagonal with blocks equal to the exponential of $\boldsymbol{B}$. The former follows from

$$\begin{aligned} e^{\boldsymbol{A} \otimes \boldsymbol{I}_m} &= \sum_{n=0}^{\infty} \frac{1}{n!} (\boldsymbol{A} \otimes \boldsymbol{I}_m)^n \\ &= \sum_{n=0}^{\infty} \frac{1}{n!} (\boldsymbol{A}^n \otimes \boldsymbol{I}_m) \\ &= \left(\sum_{n=0}^{\infty} \frac{1}{n!} \boldsymbol{A}^n \otimes \boldsymbol{I}_m \right) \\ &= e^{\boldsymbol{A}} \otimes \boldsymbol{I}_m. \end{aligned} \tag{D.3}$$

Therefore

$$e^{\boldsymbol{A} \oplus \boldsymbol{B}} = \left(e^{\boldsymbol{A}} \otimes \boldsymbol{I}_m \right) \left(\boldsymbol{I}_n \otimes e^{\boldsymbol{B}} \right),$$

and the result follows by the mixed product rule. □

Lemma D.1. *Let $\boldsymbol{T}$ and $\boldsymbol{C}$ be $n \times n$, square matrices with $\boldsymbol{C}$ diagonal and invertible; let $\boldsymbol{S}$ be a $p \times p$ matrix. Further, suppose $[\boldsymbol{T} \otimes \boldsymbol{I} + \boldsymbol{C} \otimes \boldsymbol{S} - \lambda \boldsymbol{I}]$ is invertible for $\lambda > 0$. Then*

$$\int_{t=0}^{\infty} e^{-\lambda t} e^{(\boldsymbol{T} \otimes \boldsymbol{I} + \boldsymbol{C} \otimes \boldsymbol{S})t} \, \mathrm{d}t = \int_{x=0}^{\infty} e^{\boldsymbol{C}^{-1}(\boldsymbol{T} - \lambda \boldsymbol{I})x} \otimes e^{\boldsymbol{S}x} \, \mathrm{d}x \, (\boldsymbol{C} \otimes \boldsymbol{I})^{-1}. \tag{D.4}$$

Proof. Computing the integral on the left-hand side and then factorising the result and using the Mixed Product Rule multiple times gives

$$\begin{aligned}\int_{t=0}^{\infty} e^{-\lambda t} e^{(\boldsymbol{T}\otimes\boldsymbol{I}+\boldsymbol{C}\otimes\boldsymbol{S})t}\,\mathrm{d}t &= -\left[\boldsymbol{T}\otimes\boldsymbol{I}+\boldsymbol{C}\otimes\boldsymbol{S}-\lambda\boldsymbol{I}\right]^{-1}\\ &= -\left[\boldsymbol{T}\otimes\boldsymbol{I}+(\boldsymbol{C}\otimes\boldsymbol{I})(\boldsymbol{I}\otimes\boldsymbol{S})-\lambda\boldsymbol{I}\right]^{-1}\\ &= -\left[(\boldsymbol{C}\otimes\boldsymbol{I})\left((\boldsymbol{C}\otimes\boldsymbol{I})^{-1}(\boldsymbol{T}\otimes\boldsymbol{I})+\boldsymbol{I}\otimes\boldsymbol{S}-(\boldsymbol{C}\otimes\boldsymbol{I})^{-1}\lambda\boldsymbol{I}\right)\right]^{-1}.\end{aligned} \tag{D.5}$$

By Equation (D.1) and since $\boldsymbol{C}$ is invertible, (D.5) is equal to

$$-\left[(\boldsymbol{C}\otimes\boldsymbol{I})\left(\left(\boldsymbol{C}^{-1}\otimes\boldsymbol{I}\right)(\boldsymbol{T}\otimes\boldsymbol{I})+\boldsymbol{I}\otimes\boldsymbol{S}-\left(\boldsymbol{C}^{-1}\otimes\boldsymbol{I}\right)\lambda\boldsymbol{I}\right)\right]^{-1}. \tag{D.6}$$

Using the Mixed Product Rule and algebraic manipulation, (D.6) is equal to

$$\begin{aligned}&-\left[(\boldsymbol{C}\otimes\boldsymbol{I})\left(\left(\boldsymbol{C}^{-1}\boldsymbol{T}\right)\otimes\boldsymbol{I}+\boldsymbol{I}\otimes\boldsymbol{S}-\left(\boldsymbol{C}^{-1}\lambda\boldsymbol{I}\right)\otimes\boldsymbol{I}\right)\right]^{-1}\\ &=-\left[(\boldsymbol{C}\otimes\boldsymbol{I})\left(\left(\boldsymbol{C}^{-1}(\boldsymbol{T}-\lambda\boldsymbol{I})\right)\otimes\boldsymbol{I}+\boldsymbol{I}\otimes\boldsymbol{S}\right)\right]^{-1}\\ &=-\left[\left(\boldsymbol{C}^{-1}(\boldsymbol{T}-\lambda\boldsymbol{I})\right)\otimes\boldsymbol{I}+\boldsymbol{I}\otimes\boldsymbol{S}\right]^{-1}(\boldsymbol{C}\otimes\boldsymbol{I})^{-1}\\ &=-\left[\left(\boldsymbol{C}^{-1}(\boldsymbol{T}-\lambda\boldsymbol{I})\right)\oplus\boldsymbol{S}\right]^{-1}(\boldsymbol{C}\otimes\boldsymbol{I})^{-1},\end{aligned} \tag{D.7}$$

by definition of the Kronecker sum.

Now, for an invertible matrix $\boldsymbol{A}$ we can write $-\boldsymbol{A}^{-1}=\int_{x=0}^{\infty} e^{\boldsymbol{A}x}\,\mathrm{d}x$. Therefore, (D.7) is

$$-\left[\left(\boldsymbol{C}^{-1}(\boldsymbol{T}-\lambda\boldsymbol{I})\right)\oplus\boldsymbol{S}\right]^{-1}(\boldsymbol{C}\otimes\boldsymbol{I})^{-1}=\int_{x=0}^{\infty} e^{\left(\boldsymbol{C}^{-1}(\boldsymbol{T}-\lambda\boldsymbol{I})x\right)\oplus\boldsymbol{S}x}\,\mathrm{d}x\,(\boldsymbol{C}\otimes\boldsymbol{I})^{-1}.$$

Using the rule in Equation (D.2) gives

$$\int_{x=0}^{\infty} e^{\left(\boldsymbol{C}^{-1}(\boldsymbol{T}-\lambda\boldsymbol{I})\right)x}\otimes e^{\boldsymbol{S}x}\,\mathrm{d}x\,(\boldsymbol{C}\otimes\boldsymbol{I})^{-1},$$

which is the result. □

Lemma D.2 (Latouche & Nguyen (2015))**.** *Let $\boldsymbol{B}$ be the block-partitioned matrix*

$$\boldsymbol{B}=\begin{bmatrix}\boldsymbol{B}_{11} & \boldsymbol{B}_{12}\\ \boldsymbol{B}_{21} & \boldsymbol{B}_{22}\end{bmatrix}$$

where $\boldsymbol{B}_{11}$ and $\boldsymbol{B}_{22}$ are matrices of order m_1 and m_2, respectively. Denote by $\boldsymbol{H}_{11}(t)$ the top-left quadrant of order m_1 of $e^{\boldsymbol{B}t}$:

$$\boldsymbol{H}_{11}(t)=\begin{bmatrix}\boldsymbol{I}_{m_1\times m_1} & \boldsymbol{0}\end{bmatrix}e^{\boldsymbol{B}t}\begin{bmatrix}\boldsymbol{I}_{m_1\times m_1}\\ \boldsymbol{0}\end{bmatrix}.$$

The matrix $\boldsymbol{H}_{11}(t)$ *is the solution of*

$$\boldsymbol{H}_{11}(t) = e^{\boldsymbol{B}_{11}t} + \int_{v=0}^{t}\int_{u=v}^{t} e^{\boldsymbol{B}_{11}(t-u)}\boldsymbol{B}_{12}e^{\boldsymbol{B}_{22}(u-v)}\boldsymbol{B}_{21}\boldsymbol{H}_{11}(v)\,\mathrm{d}u\,\mathrm{d}v. \tag{D.8}$$

Proof. See Latouche & Nguyen (2015). □

Let $\boldsymbol{H}_{12}(t)$ be the top-right quadrant of $e^{\boldsymbol{B}t}$ of size $m_1 \times m_2$, i.e.

$$\boldsymbol{H}_{12}(t) = \begin{bmatrix}\boldsymbol{I}_{m_1\times m_1} & \boldsymbol{0}\end{bmatrix} e^{\boldsymbol{B}t}\begin{bmatrix}\boldsymbol{0}\\ \boldsymbol{I}_{m_2\times m_2}\end{bmatrix}. \tag{D.9}$$

Denote by $\widehat{\boldsymbol{H}}_{11}(\lambda) := \int_{t=0}^{\infty} e^{-\lambda t}\boldsymbol{H}_{11}(t)\,\mathrm{d}t$ and by $\widehat{\boldsymbol{H}}_{12}(\lambda) := \int_{t=0}^{\infty} e^{-\lambda t}\boldsymbol{H}_{12}(t)\,\mathrm{d}t$, the Laplace transforms of $\boldsymbol{H}_{11}(t)$ and $\boldsymbol{H}_{12}(t)$, respectively. Using Lemma D.2 we can show the following result.

Lemma D.3. *Assuming* $\lambda > 0$ *and* $\lambda\boldsymbol{I} - \boldsymbol{B}_{22}$ *is invertible, then*

$$\widehat{\boldsymbol{H}}_{11}(\lambda) = \int_{x=0}^{\infty} e^{\left(\boldsymbol{B}_{11}-\lambda\boldsymbol{I}_{m_1\times m_1}+\boldsymbol{B}_{12}(\lambda\boldsymbol{I}_{m_2\times m_2}-\boldsymbol{B}_{22})^{-1}\boldsymbol{B}_{21}\right)x}\,\mathrm{d}x, \tag{D.10}$$

$$\widehat{\boldsymbol{H}}_{12}(\lambda) = \int_{x=0}^{\infty} e^{\left(\boldsymbol{B}_{11}-\lambda\boldsymbol{I}_{m_1\times m_1}+\boldsymbol{B}_{12}(\lambda\boldsymbol{I}_{m_2\times m_2}-\boldsymbol{B}_{22})^{-1}\boldsymbol{B}_{21}\right)x}\boldsymbol{B}_{12}(\lambda\boldsymbol{I}_{m_2\times m_2}-\boldsymbol{B}_{22})^{-1}\,\mathrm{d}x. \tag{D.11}$$

Proof. First we show the result for $\widehat{\boldsymbol{H}}_{11}(\lambda)$. Taking the Laplace transform of (D.8) shows that $\widehat{\boldsymbol{H}}_{11}(\lambda)$ is equal to

$$\begin{aligned}&\int_{t=0}^{\infty}\int_{v=0}^{t}\int_{u=v}^{t} e^{-\lambda(t-u)}e^{\boldsymbol{B}_{11}(t-u)}\boldsymbol{B}_{12}e^{-\lambda(u-v)}e^{\boldsymbol{B}_{22}(u-v)}\boldsymbol{B}_{21}e^{-\lambda v}\boldsymbol{H}_{11}(v)\,\mathrm{d}u\,\mathrm{d}v\\ &\qquad + (\lambda\boldsymbol{I}_{m_1\times m_1}-\boldsymbol{B}_{11})^{-1}\\ &= (\lambda\boldsymbol{I}_{m_1\times m_1}-\boldsymbol{B}_{11})^{-1} + (\lambda\boldsymbol{I}_{m_1\times m_1}-\boldsymbol{B}_{11})^{-1}\boldsymbol{B}_{12}(\lambda\boldsymbol{I}_{m_2\times m_2}-\boldsymbol{B}_{22})^{-1}\boldsymbol{B}_{21}\widehat{\boldsymbol{H}}_{11}(\lambda),\end{aligned} \tag{D.12}$$

by the convolution theorem for Laplace transforms. This implies

$$\begin{aligned}&\left[\boldsymbol{I}_{m_1\times m_1} - (\lambda\boldsymbol{I}_{m_1\times m_1}-\boldsymbol{B}_{11})^{-1}\boldsymbol{B}_{12}(\lambda\boldsymbol{I}_{m_2\times m_2}-\boldsymbol{B}_{22})^{-1}\boldsymbol{B}_{21}\right]\widehat{\boldsymbol{H}}_{11}(\lambda)\\ &= (\lambda\boldsymbol{I}_{m_1\times m_1}-\boldsymbol{B}_{11})^{-1},\end{aligned}$$

and therefore

$$\begin{aligned}&\widehat{\boldsymbol{H}}_{11}(\lambda)\\ &= \left[\boldsymbol{I}_{m_1\times m_1} - (\lambda\boldsymbol{I}_{m_1\times m_1}-\boldsymbol{B}_{11})^{-1}\boldsymbol{B}_{12}(\lambda\boldsymbol{I}_{m_2\times m_2}-\boldsymbol{B}_{22})^{-1}\boldsymbol{B}_{21}\right]^{-1}(\lambda\boldsymbol{I}_{m_1\times m_1}-\boldsymbol{B}_{11})^{-1}\\ &= \left[(\lambda\boldsymbol{I}_{m_1\times m_1}-\boldsymbol{B}_{11})\left(\boldsymbol{I}_{m_1\times m_1} - (\lambda\boldsymbol{I}_{m_1\times m_1}-\boldsymbol{B}_{11})^{-1}\boldsymbol{B}_{12}(\lambda\boldsymbol{I}_{m_2\times m_2}-\boldsymbol{B}_{22})^{-1}\boldsymbol{B}_{21}\right)\right]^{-1}\end{aligned}$$

$$= \left[\lambda \boldsymbol{I}_{m_1\times m_1} - \boldsymbol{B}_{11} - \boldsymbol{B}_{12}(\lambda \boldsymbol{I}_{m_2\times m_2} - \boldsymbol{B}_{22})^{-1}\boldsymbol{B}_{21}\right]^{-1}$$
$$= \int_{t=0}^{\infty} e^{\left(\boldsymbol{B}_{11}-\lambda \boldsymbol{I}_{m_1\times m_1}+\boldsymbol{B}_{12}(\lambda \boldsymbol{I}_{m_2\times m_2}-\boldsymbol{B}_{22})^{-1}\boldsymbol{B}_{21}\right)t}\,\mathrm{d}t,$$

which is (D.10).

Now, to show (D.11), differentiate (D.9)

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{H}_{12}(t) &= \begin{bmatrix}\boldsymbol{I}_{m_1\times m_1} & \boldsymbol{0}\end{bmatrix} e^{\boldsymbol{B}t}\begin{bmatrix}\boldsymbol{B}_{11} & \boldsymbol{B}_{12}\\ \boldsymbol{B}_{21} & \boldsymbol{B}_{22}\end{bmatrix}\begin{bmatrix}\boldsymbol{0}\\ \boldsymbol{I}_{m_2\times m_2}\end{bmatrix}\\ &= \begin{bmatrix}\boldsymbol{I}_{m_1\times m_1} & \boldsymbol{0}\end{bmatrix} e^{\boldsymbol{B}t}\begin{bmatrix}\boldsymbol{B}_{12}\\ \boldsymbol{B}_{22}\end{bmatrix}\\ &= \boldsymbol{H}_{11}(t)\boldsymbol{B}_{12} + \boldsymbol{H}_{12}(t)\boldsymbol{B}_{22}. \end{aligned} \tag{D.13}$$

Now take the Laplace transform

$$\lambda\widehat{\boldsymbol{H}}_{12}(\lambda) - \boldsymbol{H}_{12}(0) = \widehat{\boldsymbol{H}}_{11}(\lambda)\boldsymbol{B}_{12} + \widehat{\boldsymbol{H}}_{12}(\lambda)\boldsymbol{B}_{22}. \tag{D.14}$$

Since $\boldsymbol{H}_{12}(0) = \boldsymbol{0}$ and after rearranging we get

$$\widehat{\boldsymbol{H}}_{12}(\lambda) = \widehat{\boldsymbol{H}}_{11}(\lambda)\boldsymbol{B}_{12}(\lambda \boldsymbol{I}_{m_2\times m_2} - \boldsymbol{B}_{22})^{-1}, \tag{D.15}$$

which gives (D.11) upon substituting (D.10). □

Now, recall the matrix-functions

$$\begin{aligned}\boldsymbol{Q}_{++}(\lambda) &= \boldsymbol{C}_{+}^{-1}\left(\boldsymbol{T}_{++} - \lambda\boldsymbol{I} + \boldsymbol{T}_{+0}\left[\lambda\boldsymbol{I} - \boldsymbol{T}_{00}\right]^{-1}\boldsymbol{T}_{0+}\right),\\ \boldsymbol{Q}_{+-}(\lambda) &= \boldsymbol{C}_{+}^{-1}\left(\boldsymbol{T}_{+-} + \boldsymbol{T}_{+0}\left[\lambda\boldsymbol{I} - \boldsymbol{T}_{00}\right]^{-1}\boldsymbol{T}_{0-}\right),\\ \boldsymbol{Q}_{--}(\lambda) &= \boldsymbol{C}_{-}^{-1}\left(\boldsymbol{T}_{--} - \lambda\boldsymbol{I} + \boldsymbol{T}_{-0}\left[\lambda\boldsymbol{I} - \boldsymbol{T}_{00}\right]^{-1}\boldsymbol{T}_{0-}\right),\\ \boldsymbol{Q}_{-+}(\lambda) &= \boldsymbol{C}_{-}^{-1}\left(\boldsymbol{T}_{-+} + \boldsymbol{T}_{-0}\left[\lambda\boldsymbol{I} - \boldsymbol{T}_{00}\right]^{-1}\boldsymbol{T}_{0+}\right),\end{aligned}$$

from Chapter 5.

Corollary D.4. *For $m \in \{+,-\}$ the top-left quadrant of size $m_1 \times m_1 = |\mathcal{S}_m|p \times |\mathcal{S}_m|p$ of $e^{\boldsymbol{B}_{mm}t}$,*

$$\begin{aligned}&\begin{bmatrix}\boldsymbol{I}_{|\mathcal{S}_m|p\times|\mathcal{S}_m|p} & \boldsymbol{0}_{|\mathcal{S}_m|p\times|\mathcal{S}_0|p}\end{bmatrix}\int_{t=0}^{\infty} e^{-\lambda t}\exp\left\{\begin{bmatrix}\boldsymbol{T}_{mm}\otimes\boldsymbol{I} + \boldsymbol{C}_m\otimes\boldsymbol{S} & \boldsymbol{T}_{m0}\otimes\boldsymbol{I}\\ \boldsymbol{T}_{0m}\otimes\boldsymbol{I} & \boldsymbol{T}_{00}\otimes\boldsymbol{I}\end{bmatrix}t\right\}\mathrm{d}t\\ &\quad\times\begin{bmatrix}\boldsymbol{I}_{|\mathcal{S}_m|p\times|\mathcal{S}_m|p}\\ \boldsymbol{0}_{|\mathcal{S}_0|p\times|\mathcal{S}_m|p}\end{bmatrix},\end{aligned} \tag{D.16}$$

is given by

$$\int_{x=0}^{\infty} e^{\boldsymbol{Q}_{mm}(\lambda)x} \otimes e^{\boldsymbol{S}x}\,\mathrm{d}x(\boldsymbol{C}_m^{-1} \otimes \boldsymbol{I}). \tag{D.17}$$

For $m \in \{+,-\}$ the top-right quadrant of size $m_1 \times m_2 = |\mathcal{S}_m|p \times |\mathcal{S}_0|p$ of $e^{\boldsymbol{B}_{mm}t}$,

$$\begin{aligned}\begin{bmatrix}\boldsymbol{I}_{|\mathcal{S}_m|p\times|\mathcal{S}_m|p} & \boldsymbol{0}_{|\mathcal{S}_m|p\times|\mathcal{S}_0|p}\end{bmatrix} \int_{t=0}^{\infty} e^{-\lambda t} \exp\left\{ \begin{bmatrix} \boldsymbol{T}_{mm}\otimes\boldsymbol{I} + \boldsymbol{C}_m\otimes\boldsymbol{S} & \boldsymbol{T}_{m0}\otimes\boldsymbol{I} \\ \boldsymbol{T}_{0m}\otimes\boldsymbol{I} & \boldsymbol{T}_{00}\otimes\boldsymbol{I} \end{bmatrix} t \right\} \mathrm{d}t \\ \times \begin{bmatrix} \boldsymbol{0}_{|\mathcal{S}_m|p\times|\mathcal{S}_m|p} \\ \boldsymbol{I}_{|\mathcal{S}_0|p\times|\mathcal{S}_0|p} \end{bmatrix},\end{aligned} \tag{D.18}$$

is given by

$$\int_{x=0}^{\infty} e^{\boldsymbol{Q}_{mm}(\lambda)x} \otimes e^{\boldsymbol{S}x}\,\mathrm{d}x((\boldsymbol{C}_m^{-1}\boldsymbol{T}_{m0}(\lambda\boldsymbol{I} - \boldsymbol{T}_{00})^{-1}) \otimes \boldsymbol{I}). \tag{D.19}$$

Also,

$$\begin{bmatrix}\boldsymbol{I}_{|\mathcal{S}_m|p\times|\mathcal{S}_m|p} & \boldsymbol{0}_{|\mathcal{S}_m|p\times|\mathcal{S}_0|p}\end{bmatrix} \int_{t=0}^{\infty} e^{-\lambda t} e^{\boldsymbol{B}_{mm}t}\,\mathrm{d}t\boldsymbol{B}_{mn} \tag{D.20}$$

$$= \int_{x=0}^{\infty} \boldsymbol{H}^{mn}(\lambda, x) \otimes e^{\boldsymbol{S}x}\boldsymbol{D}\,\mathrm{d}x \left(\begin{bmatrix}\boldsymbol{I}_{|\mathcal{S}_n|p} & \boldsymbol{0}_{|\mathcal{S}_n|p\times|\mathcal{S}_0|p}\end{bmatrix}\right), \tag{D.21}$$

for $m, n \in \{+,-\}$, $m \neq n$.

Proof. From Lemma D.2 the top-left quadrant of size $m_1 \times m_1 = |\mathcal{S}_m|p \times |\mathcal{S}_m|p$ of the integral with respect to t on the left-hand side of (D.17) is

$$\int_{t=0}^{\infty} e^{\left(\boldsymbol{T}_{mm}\otimes\boldsymbol{I} + \boldsymbol{C}_m\otimes\boldsymbol{S} - \lambda\boldsymbol{I} + (\boldsymbol{T}_{m0}\otimes\boldsymbol{I})(\lambda\boldsymbol{I} - \boldsymbol{T}_{00}\otimes\boldsymbol{I})^{-1}(\boldsymbol{T}_{0m}\otimes\boldsymbol{I})\right)t}\,\mathrm{d}t. \tag{D.22}$$

By Lemma D.1, (D.22) is equal to

$$\begin{aligned}&\int_{x=0}^{\infty} e^{\boldsymbol{C}_m^{-1}\left(\boldsymbol{T}_{mm} - \lambda\boldsymbol{I} + \boldsymbol{T}_{m0}(\lambda\boldsymbol{I} - \boldsymbol{T}_{00})^{-1}\boldsymbol{T}_{0m}\right)x} \otimes e^{\boldsymbol{S}x}\,\mathrm{d}x(\boldsymbol{C}_m \otimes \boldsymbol{I})^{-1} \\ &= \int_{x=0}^{\infty} e^{\boldsymbol{Q}_{mm}(\lambda)x} \otimes e^{\boldsymbol{S}x}\,\mathrm{d}x(\boldsymbol{C}_m \otimes \boldsymbol{I})^{-1},\end{aligned} \tag{D.23}$$

from the definition of $\boldsymbol{Q}_{mm}(\lambda)$. This proves (D.17).

Now, from Lemma D.2 the top-right quadrant of size $m_1 \times m_2 = |\mathcal{S}_m|p \times |\mathcal{S}_0|p$ of the integral with respect to t on the left-hand side of (D.19) is

$$\int_{t=0}^{\infty} e^{\left(\boldsymbol{T}_{mm}\otimes\boldsymbol{I} + \boldsymbol{C}_m\otimes\boldsymbol{S} - \lambda\boldsymbol{I} + (\boldsymbol{T}_{m0}\otimes\boldsymbol{I})(\lambda\boldsymbol{I} - \boldsymbol{T}_{00}\otimes\boldsymbol{I})^{-1}(\boldsymbol{T}_{0m}\otimes\boldsymbol{I})\right)t} (\boldsymbol{T}_{m0}\otimes\boldsymbol{I})(\lambda\boldsymbol{I} - \boldsymbol{T}_{00}\otimes\boldsymbol{I})^{-1}$$

$$\times (\boldsymbol{T}_{0m} \otimes \boldsymbol{I})\,\mathrm{d}t. \tag{D.24}$$

By Lemma D.1, (D.24) is equal to

$$\int_{x=0}^{\infty} e^{\boldsymbol{Q}_{mm}(\lambda)x} \otimes e^{\boldsymbol{S}x}\,\mathrm{d}x(\boldsymbol{C}_m \otimes \boldsymbol{I})^{-1}(\boldsymbol{T}_{m0} \otimes \boldsymbol{I})(\lambda \boldsymbol{I} - \boldsymbol{T}_{00} \otimes \boldsymbol{I})^{-1}(\boldsymbol{T}_{0m} \otimes \boldsymbol{I}). \tag{D.25}$$

Now,

$$\begin{aligned}(\lambda \boldsymbol{I} - \boldsymbol{T}_{00} \otimes \boldsymbol{I})^{-1} &= \int_{u=0}^{\infty} e^{-(\lambda \boldsymbol{I} - \boldsymbol{T}_{00} \otimes \boldsymbol{I})u}\,\mathrm{d}u \\ &= \int_{u=0}^{\infty} e^{-\lambda u} e^{(\boldsymbol{T}_{00} \otimes \boldsymbol{I})u}\,\mathrm{d}u \\ &= \int_{u=0}^{\infty} e^{-\lambda u} e^{\boldsymbol{T}_{00}u} \otimes \boldsymbol{I}\,\mathrm{d}u,\end{aligned}$$

by (D.3). Using this and the Mixed Product Rule we can write

$$\begin{aligned}&(\boldsymbol{C}_m \otimes \boldsymbol{I})^{-1}(\boldsymbol{T}_{m0} \otimes \boldsymbol{I})(\lambda \boldsymbol{I} - \boldsymbol{T}_{00} \otimes \boldsymbol{I})^{-1}(\boldsymbol{T}_{0m} \otimes \boldsymbol{I}) \\ &= (\boldsymbol{C}_m^{-1} \otimes \boldsymbol{I})(\boldsymbol{T}_{m0} \otimes \boldsymbol{I}) \int_{u=0}^{\infty} e^{-\lambda u} e^{\boldsymbol{T}_{00}u} \otimes \boldsymbol{I}\,\mathrm{d}u(\boldsymbol{T}_{0m} \otimes \boldsymbol{I}) \\ &= \left(\boldsymbol{C}_m^{-1}\boldsymbol{T}_{m0}(\lambda \boldsymbol{I} - \boldsymbol{T}_{00}u)^{-1}\boldsymbol{T}_{0m}\right) \otimes \boldsymbol{I}).\end{aligned} \tag{D.26}$$

Substituting (D.26) into (D.25) completes the proof of (D.19).

Now, using (D.17) and (D.19) we can write

$$\begin{aligned}&\left[\boldsymbol{I}_{|\mathcal{S}_m|p \times |\mathcal{S}_m|p} \quad \boldsymbol{0}_{|\mathcal{S}_m|p \times |\mathcal{S}_0|p}\right] \int_{t=0}^{\infty} e^{-\lambda t} e^{\boldsymbol{B}_{mm}t}\,\mathrm{d}t \boldsymbol{B}_{mn} \\ &= \left[\boldsymbol{I}_{|\mathcal{S}_m|p \times |\mathcal{S}_m|p} \quad \boldsymbol{0}_{|\mathcal{S}_m|p \times |\mathcal{S}_0|p}\right] \int_{t=0}^{\infty} e^{-\lambda t} e^{\boldsymbol{B}_{mm}t}\,\mathrm{d}t \begin{bmatrix} \boldsymbol{T}_{mn} \otimes \boldsymbol{D} & \boldsymbol{0} \\ \boldsymbol{T}_{0n} \otimes \boldsymbol{D} & \boldsymbol{0} \end{bmatrix} \\ &= \int_{x=0}^{\infty} e^{\boldsymbol{Q}_{mm}(\lambda)x} \otimes e^{\boldsymbol{S}x}\,\mathrm{d}x(\boldsymbol{C}_m^{-1}(\boldsymbol{T}_{mn} + \boldsymbol{T}_{m0}(\lambda \boldsymbol{I} - \boldsymbol{T}_{00})^{-1}\boldsymbol{T}_{0n}) \left[\boldsymbol{I}_{|\mathcal{S}_n| \times |\mathcal{S}_n|} \quad \boldsymbol{0}_{|\mathcal{S}_n| \times |\mathcal{S}_0|}\right] \otimes \boldsymbol{D}) \\ &= \int_{x=0}^{\infty} e^{\boldsymbol{Q}_{mm}(\lambda)x} \otimes e^{\boldsymbol{S}x}\,\mathrm{d}x \left(\left(\boldsymbol{Q}_{mn}(\lambda) \left[\boldsymbol{I}_{|\mathcal{S}_n|} \quad \boldsymbol{0}_{|\mathcal{S}_n| \times |\mathcal{S}_0|}\right]\right) \otimes \boldsymbol{D}\right) \\ &= \int_{x=0}^{\infty} \left(\boldsymbol{H}^{mn}(\lambda, x) \left[\boldsymbol{I}_{|\mathcal{S}_n|} \quad \boldsymbol{0}_{|\mathcal{S}_n| \times |\mathcal{S}_0|}\right]\right) \otimes e^{\boldsymbol{S}x}\boldsymbol{D}\,\mathrm{d}x, \\ &= \int_{x=0}^{\infty} \boldsymbol{H}^{mn}(\lambda, x) \otimes e^{\boldsymbol{S}x}\boldsymbol{D}\,\mathrm{d}x, \left[\boldsymbol{I}_{|\mathcal{S}_m|p \times |\mathcal{S}_n|p} \quad \boldsymbol{0}_{|\mathcal{S}_n|p \times |\mathcal{S}_0|p}\right],\end{aligned} \tag{D.27}$$

for $m, n \in \{+, -\}$, $m \neq n$ which is (D.21), where the last line holds from the Mixed Product Rule. □

Bibliography

Abate, J. & Whitt, W. (2006), 'A unified framework for numerically inverting Laplace transforms', *INFORMS Journal on Computing* **18**(4), 408–421.

Ahn, S., Jeon, J. & Ramaswami, V. (2005), 'Steady state analysis of finite fluid flow models using finite QBDs', *Queueing Syst.* **49**, 223–259.

Ahn, S. & Ramaswami, V. (2003), 'Fluid flow models and queues—a connection by stochastic coupling', *Stochastic Models* **19**(3), 325–348.

Ahn, S. & Ramaswami, V. (2004), 'Transient analysis of fluid flow models via stochastic coupling to a queue', *Stochastic Models* **20**(1), 71–101.

Aldous, D. & Shepp, L. (1987), 'The least variable phase type distribution is Erlang', *Stochastic Models* **3**(3), 467–473.

Anick, D., Mitra, D. & Sondhi, M. (1982), 'Stochastic theory of a data-handling system with multiple sources', *The Bell System Technical Journal* **61**(8), 1871–1894.

Asmussen, S. (2008), *Applied probability and queues*, Vol. 51, Springer Science & Business Media.

Badescu, A., Breuer, L., Soares, A. D. S., Latouche, G., Remiche, M.-A. & Stanford, D. (2005), 'Risk processes analyzed as fluid queues', *Scandinavian Actuarial Journal* **2005**(2), 127–141.

Bean, N. G., Lewis, A., Nguyen, G., O'Reilly, M. M. & Sunkara, V. (2022), 'A discontinuous Galerkin method for approximating the stationary distribution of stochastic fluid-fluid processes', *arXiv preprint ARXIV TICKER* .

Bean, N. G., Nguyen, G. T., Nielsen, B. F. & Peralta Gutierrez, O. (2021), 'RAP-modulated fluid processes: First passages and the stationary distribution'.

Bean, N. G. & Nielsen, B. F. (2010), 'Quasi-birth-and-death processes with rational arrival process components', *Stochastic Models* **26**(3), 309–334.

Bean, N. G. & O'Reilly, M. M. (2008), 'Performance measures of a multi-layer Markovian fluid model', *Annals of Operations Research* **160**(1), 99–120.

Bean, N. G. & O'Reilly, M. M. (2013*a*), 'Spatially-coherent uniformization of a stochastic fluid model to a quasi-birth-and-death process', *Performance Evaluation* **70**(9), 578 – 592.

Bean, N. G. & O'Reilly, M. M. (2013*b*), 'A stochastic two-dimensional fluid model', *Stochastic Models* **29**(1), 31–63.

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 itself', *Stochastic Processes and their Applications* **124**, 1741–1772.

Bean, N. G., O'Reilly, M. M. & Palmowski, Z. (2020), 'Matrix-analytic methods for the analysis of stochastic fluid-fluid models'.

Bean, N. G., O'Reilly, M. M. & Sargison, J. (2010), 'A stochastic fluid flow model of the operation and maintenance of power generation systems', *IEEE Transactions on Power Systems* **25**(3), 1361–1374.

Bean, N. G., O'Reilly, M. M. & Taylor, P. G. (2005*a*), 'Algorithms for return probabilities for stochastic fluid flows', *Stochastic Models* **21**(1), 149–184.

Bean, N. G., O'Reilly, M. M. & Taylor, P. G. (2005*b*), 'Hitting probabilities and hitting times for stochastic fluid flows', *Stochastic Processes and their Applications* **115**(9), 1530–1556.

Bean, N. G., O'Reilly, M. M. & Taylor, P. G. (2009*a*), 'Algorithms for the Laplace-Stieltjes transforms of first return times for stochastic fluid flows', *Methodology and Computing in Applied Probability* **10**, 381–408.

Bean, N. G., O'Reilly, M. M. & Taylor, P. G. (2009*b*), 'Hitting probabilities and hitting times for stochastic fluid flows: the bounded model', *Probability in the Engineering and Informational Sciences* **23**(1), 121–147.

Bean, N., Nguyen, G. T. & Poloni, F. (2018), 'Doubling algorithms for stationary distributions of fluid queues: A probabilistic interpretation', *Performance Evaluation* **125**, 1–20.

Bhatia, R. & Rosenthal, P. (1997), 'How and why to solve the operator equation AX-XB = Y', *Bulletin of the London Mathematical Society* **29**(1), 1–21.

Bladt, M. & Nielsen, B. (2017), *Matrix-Exponential Distributions in Applied Probability*, Springer.

Borthwick, D. (n.d.), *Introduction to Partial Differential Equations*, Universitext, Springer International Publishing, Cham.

Cockburn, B. (1999), Discontinous Garlekin methods for convection-dominated problems, *in* T. Barth & H. Deconink, eds, 'Higher-Order Methods for Computational Physics', Vol. 9 of *Lecture Notes in Computational Science and Engineering*, Springer Verlag, Berlin, pp. 69–224.

da Silva Soares, A. (2005), Fluid queues: building upon the analogy with QBD processes, PhD thesis, Université Libre de Bruxelles.

Élteto, T., Ŕacz, S. & Telek, M. (2006), Minimal coefficient of variation of matrix exponential distributions, *in* '2nd Madrid Conference on Queueing Theory'.

Ethier, S. N. (1986), *Markov processes : characterization and convergence*, Wiley series in probability and mathematical statistics, Wiley, New York.

Feller, W. (1957), *An introduction to probability theory and its applications*, Wiley series in probability and mathematical statistics, 2nd ed. edn, Wiley, New York.

Hautphenne, S., Massaro, M. & Taylor, P. (2017), 'How old is this bird? The age distribution under some phase sampling schemes', *Journal of Mathematical Biology* **75**(6), 1319–1347.

Hesthaven, J. S. & Warburton, T. (2007), *Nodal discontinuous Galerkin methods: algorithms, analysis, and applications*, Springer Science & Business Media.

Horváth, I., Sáfár, O., Telek, M. & Zambó, B. (2016), Concentrated matrix exponential distributions, *in* 'In proceedings of European Performance Evaluation Workshop'.

Horváth, G., Horváth, I., Almousa, S. A.-D. & Telek, M. (2020), 'Numerical inverse Laplace transformation using concentrated matrix exponential distributions', *Performance Evaluation* **137**, 102067.

Horváth, G., Horváth, I. & Telek, M. (2020), 'High order concentrated matrix-exponential distributions', *Stochastic Models* **36**(2), 176–192.

Kallenberg, O. (2021), *Foundations of modern probability*, Probability Theory and Stochastic Modelling, 99, third edition. edn, Springer, Cham, Switzerland.

Karandikar, R. L. & Kulkarni, V. G. (1995), 'Second-order fluid flow models: Reflected Brownian motion in a random environment', *Operations Research* **43**(1), 77–88.

Klenke, A. (2014), *Probability Theory*, Springer, London.

Koltai, P. (2011), *Efficient approximation methods for the global long-term behavior of dynamical systems: theory, algorithms and examples*, Logos Verlag Berlin GmbH.

Latouche, G. & Nguyen, G. T. (2015), 'The morphing of fluid queues into Markov-modulated Brownian motion', *Stochastic Systems* **5**(1), 62–86.

Latouche, G. & Nguyen, G. T. (2019), 'Analysis of fluid flow models', *arXiv preprint arXiv:1901.10635* .

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. Squillante & D. Yao, eds, 'Matrix-Analytic Methods in Stochastic Models', Vol. 27 of *Springer Proceedings in Mathematics & Statistics*, Springer Science, New York, NY, chapter 9, pp. 187–207.

Loynes, R. (1962), 'A continuous-time treatment of certain queues and infinite dams', *Journal of the Australian Mathematical Society* **2**(4), 484–498.

Miyazawa, M. & Zwart, B. (2012), 'Wiener-Hopf factorizations for a multidimensional Markov additive process and their applications to reflected processes', *Stochastic Systems* **2**, 67–114.

Mészáros, A. & Telek, M. (2021), 'Concentrated matrix exponential distributions with real eigenvalues', *Probability in the Engineering and Informational Sciences* pp. 1–17.

Peralta Gutierrez, O. (2019), Advances of matrix-analytic methods in risk modelling, PhD thesis, DTU Compute.

Rabehasaina, L. & Sericola, B. (2003), Transient analysis of a Markov modulated fluid queue with linear service rate, *in* 'Proceedings of the 10th Conference in Analytical Stochastic Modelling Techniques and Applications', ASMTA'03, pp. 234–239.

Reed, W. H. & Hill, T. R. (1973), Triangular mesh methods for the neutron transport equation, Technical report, Los Alamos Scientific Lab., N. Mex.(USA).

Ross, S. M. (1996), *Stochastic processes / Sheldon M. Ross.*, Wiley series in probability and statistics, 2nd ed. edn, Wiley, New York.

Sericola, B. (1998), 'Transient analysis of stochastic fluid models', *Performance evaluation* **32**(4), 245–263.

Sericola, B. (2001), 'A finite buffer fluid queue driven by a markovian queue', *Queueing Systems* **38**(2), 213–220.

Sericola, B. & Tuffin, B. (1999), 'A fluid queue driven by a markovian queue', *Queueing Systems* **31**(3), 253–264.

Shalizi, C. (2010), 'Almost none of the theory of stochastic processes'.
URL: *https://www.stat.cmu.edu/ cshalizi/almost-none/v0.1.1/almost-none.pdf*

Sonenberg, N. (2017), Networks of interacting stochastic fluid models, PhD thesis, School of Mathematics and Statistics, University of Melbourne.

Spiteri, R. J. & Ruuth, S. J. (2002), 'A new class of optimal high-order strong-stability-preserving time discretization methods', *SIAM Journal on Numerical Analysis* **40**(2), 469–23.

Stein, E. M. & Shakarchi, R. (2009), *Real analysis*, Princeton University Press.

Wurm, M. (2020), Stochastic modelling of coral-algal symbiosis on the Great Barrier Reef, Master's thesis, School of Mathematical Sciences, The University of Adelaide.