UNIVERSITÀ DI PISA

DIPARTIMENTO DI MATEMATICA

Corso di Laurea in Matematica

Random walk in the quarter plane: numerical methods

Relatore:

Prof. Dario A. Bini

Controrelatore:

Prof.ssa Beatrice Meini

Candidato:

Luca Ferragina

Anno Accademico 2016-2017

Contents

Chapter 1

Introduction and Preliminaries

Random walks in the quarter-plane are frequently used to model queueing problems, they belong to the family of Quasi-Birth-Death processes and they are widely studied in literature, for example in [11] we can find a detailed analysis of the problem.

The main purpose of this thesis is to analyze and compare algorithms that allow us to calculate the invariant measure of the random walk when it is modelled as a discrete time Markov chain or as a continuous time Markov process. We do this with two approaches deeply different from each other.

The first, based on the works of Bini et al. ([5], [6], [7]), is an operator approach. The transition operator of this kind of process is semi-infinite, block-tridiagonal, almost block-Toeplitz with semi-infinite almost-Toeplitz blocks; the idea behind this approach is to adapt, to the infinite case, the available algorithms valid for blocks of finite size. To do this we need first to define the right space which these infinite blocks belong to and then to build an arithmetic in it. In particular we focus on the algorithm of Cyclic Reduction and on the matrix geometric approach of [13]. We prove that, similarly to the finite case, this algorithm converges to the minimal solution of certain quadratic operator equations from which we can build the invariant probability vector.

The second approach, named *compensation approach*, is based on the works of Adan ([2], [1]). This is a more practical approach that exploits the structure of the equilibrium equations in the interior of the quarter plane by imposing that linear combinations of product forms satisfy these equations. This leads to a kernel equation for the terms appearing in the product forms. Then, it is required that these linear combinations satisfy the equilibrium equations on the boundaries as well. As it turns out, this can be done by alternately compensating for the errors on the two boundaries, which eventually leads to infinite series of product forms. Convergence of these series is a crucial issue, some sufficient conditions for the convergence of

these series are provided.

After introducing the two approaches and relying on the work of [12, Kapodistria], we provide a comparison of the two approaches on both a theoretical and a computational basis. Some results are stated, in particular we show that eigenvalues and eigenvectors of the operators that we calculate in the first approach can be obtained in the construction of the infinite series of product forms in the compensation approach.

An accurate numerical simulation is carried out relying on test problems taken from the current literature. It turns out that the applicability of the compensation approach is more restricted than the operator approach. On the other hand the compensation approach shows a better performance in terms of accuracy. In fact it provides approximation with very small relative error, whereas the operator approach in the current implementation does not maintain a uniform bound to the relative error, even though the absolute error in the approximation is quite small.

The thesis is organized as follows. In Chapter 1 we describe the problem of the random walk in the quarter plane, we model it as a Markov chain and we recall some of the most important definitions and results about the subject. In Chapter 2 we extend the matrix geometric approach to the infinite case, we introduce the space of Quasi-Toeplitz operators and we describe a machine arithmetic for it. Chapter 3 concerns the Cyclic Reduction algorithm: we recall this algorithm together with its properties valid in the finite case, then we present its extension to the infinite case. In Chapter 4 we describe the compensation approach and we put it in relation with the operator approach. Finally, in Chapter 5 we exhibit the numerical results of our experimentation.

1.1 Short Summary on Markov Chain and Processes

Markov chains are used to model systems which evolve in time. They come under various guises but we only consider here discrete-state processes, meaning that the total number of states which the process may occupy is either finite or countably infinite. Time may either increase by discrete, constant amounts, as when the modeled system is controlled by a digital clock, or it may increase continuously.

A *stochastic process* is a family $\{\mathbf{X}_t : t \in T\}$ of random variables $\mathbf{X}_t$ indexed by some set T and with values in a common set S: $\mathbf{X}_t \in S$ for all $t \in T$. Here, S is a countable set and it is called the *state space*, and T is the *time space*. If T is countable, say $T = \mathbb{N}$, the process is said to be discrete, otherwise it is continuous.

Definition 1.1. The stochastic process $\{\mathbf{X}_t : t \in \mathbb{N}\}$ is a *Markov chain* if

$$\mathbb{P}\left[\mathbf{X}_{t+1} = j | \mathbf{X}_0, \mathbf{X}_1, \ldots, \mathbf{X}_{t-1}, \mathbf{X}_t\right] = \mathbb{P}\left[\mathbf{X}_{t+1} = j | \mathbf{X}_t\right],$$

for all states $j \in S$, and for all times $t \in \mathbb{N}$.

This means that if one knows the state $\mathbf{X}_t$ of the system at time t, then the past history $\mathbf{X}_0, \mathbf{X}_1, \ldots, \mathbf{X}_{t-1}$ does not help in determining which state might be occupied at time $t+1$. One also usually requires that the laws which govern the evolution of the system be time-invariant; this is formulated as follows.

Definition 1.2. A Markov chain $\{\mathbf{X}_t : t \in \mathbb{N}\}$ is *homogeneous* if

$$\mathbb{P}\left[\mathbf{X}_{t+1} = j | \mathbf{X}_t = i\right] = \mathbb{P}\left[\mathbf{X}_1 = j | \mathbf{X}_0 = i\right],$$

for all states $i, j \in S$, and for all times $t \in \mathbb{N}$.

In the sequel, we always assume that Markov chains are homogeneous. Define the matrix (operator, in the case when S is infinite) $P = (p_{ij})_{i,j \in S}$ with one row and one column for each state in S and such that

$$P_{ij} = \mathbb{P}\left[\mathbf{X}_1 = j | \mathbf{X}_0 = i\right],$$

for all $i, j \in S$. This is called the *transition matrix* of the Markov chain; it is a row-stochastic matrix, that is, its element are nonnegative and its row sums are all equal to 1.

The transition matrix plays a very important role in the dynamic behaviour of the Markov chains, as we can see in the following propositions.

Proposition 1.1. *For all times $n \geq 0$, all intervals of time $k \geq 0$ and all states $i, j \in S$, we have*

$$\mathbb{P}\left[\mathbf{X}_{t+k} = j | \mathbf{X}_t = i\right] = (P^k)_{i,j}.$$

Proposition 1.2. *For all times $t \geq 0$, all intervals of time $k \geq 1$ and all states $i, j_1, \ldots, j_k \in S$, we have*

$$\mathbb{P}\left[\mathbf{X}_{t+1} = j_1, \mathbf{X}_{t+2} = j_2, \ldots, \mathbf{X}_{t+k} = j_k | \mathbf{X}_t = i\right] = p_{ij_1} p_{j_1 j_2} \ldots p_{j_{k-1} j_k}.$$

Definition 1.3. Given a time-homogeneous Markov chain with transition matrix (operator) P, a row (infinite) vector $\pi = (\pi_i)_{i \in S}$ is said to be *invariant* if $\pi P = \pi$. If the invariant vector π is such that $\pi \geq 0$ and $\pi \mathbf{1} = 1$, then it is said to be an *invariant probability measure.*

Not all the Markov chains present an invariant probability measure. For that, additional hypotheses have to be assumed.

1.1.1 Classification of States and classes

The *transition graph* of a Markov chain with transition matrix P is the directed graph (S, E), where the vertices are all possible states, and there is an edge from

state i to state j if and only if $p_{ij} > 0$. We say that state i leads to j if there is a path from i to j in the transition graph. We say that states i and j communicate if each leads to the other. Communication forms an equivalence relation on the states, so we call *irreducible classes* its equivalence classes. We say that the Markov chain is *irreducible* if all its states communicate, else it is *reducible*.

Definition 1.4. Given a Markov chain, we say that a state $i \in S$ is *periodic* with period $\omega \geq 2$ if all closed paths through i in the transition graph have a length that is a multiple of ω. A state $i \in S$ is *aperiodic* if it is not periodic.

Definition 1.5. For any Markov chain $\mathbf{X}_t$ we define the *epochs of visits* to a subset S' of states as the sequence $\{\tau_t^{(S')}\}_{t=0,1,\ldots}$, where τ_t is the t-th time the Markov chain visits a state in S'. In formulas:

$$\tau_0^{(S')} = \inf\{i \geq 0 : \mathbf{X}_i \in S'\},$$
$$\tau_{t+1}^{(S')} = \inf\{i > \tau_t : \mathbf{X}_i \in S'\} \text{ for } t \geq 0,$$

when the subset S' is reduced to only a state, $S' = \{i\}$, we write $\tau_t^{(i)}$. We omit the superscript $\{(S')\}$ when there is no ambiguity. We also define the *restricted process* $\{\mathbf{X}_t^{(S')}\}_{t=0,1,\ldots}$ as follows

$$\mathbf{X}_t^{(S')} := \mathbf{X}_{\tau_t} \text{ for } t \geq 0,$$

so that $\mathbf{X}_t^{(S')}$ is the actual state visited when the Markov chain is in S' for the t-th time.

Let us consider the quantity $f_i := \mathbb{P}\left[\tau_0^{(i)} < \infty | \mathbf{X}_0 = i\right]$, that is the probability that, starting from i, the Markov chain returns to i in a finite time. We then classify state i of the Markov chain as:

- *transient* if $f_1 < 1$;
- *recurrent* if $f_i = 1$; in this case we also distinguish between:
  - *positive recurrent* if the expected return $\mathbb{E}\left[\tau_0^{(i)} | \mathbf{X}_0 = i\right]$ is finite;
  - *null recurrent* if the expected return $\mathbb{E}\left[\tau_0^{(i)} | \mathbf{X}_0 = i\right]$ is infinite;

It can be shown that all states of a single irreducible class of a Markov chain belong to the same classification, so irreducible classes in turn can be transient, positive recurrent or null recurrent. In particular, if a Markov chain is irreducible, then it is transient, positive recurrent or null recurrent, depending on the classification of its states. Similarly, it can be shown that periodicity is also a class property, that is all states of a single irreducible class are aperiodic, or are periodic with have the same period ω. We can then refer to the periodicity of each irreducible class, or even to the periodicity of an irreducible Markov chain.

Theorem 1.1. *An irreducible Markov chain has an invariant probability measure π if and only if it is positive recurrent. If the invariant probability measure exists, then it is unique.*

Theorem 1.2. *Consider an irreducible and aperiodic Markov chain $\{\mathbf{X}_t\}$ with state space S, and suppose that it has an invariant probability measure π. Then*

$$\lim_{t\to\infty} \mathbb{P}\left[\mathbf{X}_t = i | X_0 = j\right] = \pi_i$$

for all $i, j \in S$.

1.1.2 Continuous-time Markov Processes

A continuous-time Markov chain, or Markov process, is a stochastic process $\{\mathbf{X}_t : t \in T\}$, in which the index set is $T = [0, +\infty) \subseteq \mathbb{R}$, the set S of the possible values assumed by X_t is countable, and the following Markov property is satisfied:

$$\mathbb{P}\left[\mathbf{X}_{t_{n+1}} = i | \mathbf{X}_{t_0} = j_0, \ldots, \mathbf{X}_{t_n} = j_{n-1}\right] = \mathbb{P}\left[\mathbf{X}_{t_{n+1}} = i | \mathbf{X}_{t_{n-1}} = j_n\right]$$

for all $(n+2)$-tuple of states $i, j_0, \ldots, j_n \in S$ and for all $(n+2)$-tuple of times $0 \leq t_0 < \cdots < t_{n+1}$. We call transition probabilities the quantities

$$p_{ij}(s,t) := \mathbb{P}\left[\mathbf{X}_t = i | \mathbf{X}_s = j\right],$$

for $i, j \in S$ and $0 \leq s < t$.

Again, we will consider only a proper subclass of Markov processes:

Definition 1.6. A Markov process is *time-homogeneous* if it satisfies

$$p_{ij}(s, t+s) = p_{i,j}(0,t),$$

for all $i, j \in S$ and all $s, t \geq 0$, that is, if transition probabilities depend only on states and time difference. We will then indicate $p_{ij}(0,t)$ just by $p_{ij}(t)$.

From now on, all Markov processes that we consider are assumed to be time-homogeneous.

Definition 1.7. A Q-matrix on S is a matrix $Q = (q_{ij})_{i,j\in S}$ satisfying the following conditions:

1. $-\infty < q_{ii} \leq 0$ for all $i \in S$;
2. $q_{ij} \geq 0$ for all $ij \in S$ such that $i \neq j$;
3. $Q\mathbf{1} = \mathbf{0}$.

Definition 1.8. The transition matrix (operator, in the case when S is infinite) of a Markov process is $P(t) := (p_{ij}(t))_{i,j\in S}$. The transition rate matrix or generator matrix of a Markov process is

$$Q_{ij} := \begin{cases} \lim_{h\to 0} \frac{p_{ij}(h)}{h} & \text{for } i \neq j, \\ -\sum_{j\neq i} Q_{ij} & \text{for} i = j. \end{cases}$$

Under appropriate regularity hypotheses, it can be shown that, given a Markov chain with transition matrix $P(t)$ and generator matrix Q, they satisfy the differential equation $P'(t) = QP(t)$, and Q is a Q-matrix. In particular, the generator matrix is sufficient to define uniquely the behaviour of the Markov process. This suggests to give the following definition of invariant vector, in the continuous-time case:

Definition 1.9. Given a Markov process with generator matrix (operator) Q, a (infinite) row vector $\pi = (\pi_i)_{i\in S}$ is said to be invariant if $\pi Q = 0$. If the invariant vector π is such that $\pi \geq 0$ and $\pi\mathbf{1} = 1$, then it is said to be an invariant probability measure.

We can define the transition graph, the irreducible classes, transience and positive/null recurrence of Markov processes similarly as how we did for Markov chains. Then, it can be shown that most results about discrete-time Markov chains are still true for the continuous-time case. Most notably, the following theorems are valid:

Theorem 1.3. *An irreducible Markov process has an invariant probability measure π if and only if it is positive recurrent. If the invariant probability measure exists, then it is unique.*

Theorem 1.4. *Consider an irreducible and positive recurrent Markov process $\{\mathbf{X}_n\}$, with state space S and invariant probability measure π. Then*

$$\lim_{t\to\infty} \mathbb{P}\left[\mathbf{X}_t = j | \mathbf{X}_0 = i\right] = \pi_j,$$

for all $i, j \in S$.

For all not proven results in this section on Markov chains and processes and for a vast dissertation about this argument we refer to [17].

1.2 Reflecting Random Walk on an Orthant

The random walk in a quarter plane belongs to a more general class of problems whose full exposition can be found in [15], here we give a brief description of it.

We use some of standard notations for sets of numbers. Let $\mathbb{R}$ and $\mathbb{R}_+$ be the sets of all real and of all real nonnegative numbers, respectively. Similarly, let $\mathbb{Z}$ be

the set of all integers. Let d be a positive integer. Then, $S = \mathbb{Z}_+^d$ is referred to as a nonnegative orthant of $\mathbb{Z}^d$.

The reflecting random walk is defined on this orthant, that is, it has state space S. To describe a reflection mechanism, we partition S into disjoint subsets. Let $J = \{1, 2, \ldots, d\}$, for each subset $A \subset J$, we define

$$S_A = \{\mathbf{x} \in S; x_i \geq 1, i \in A, x_j = 0, j \notin A\}.$$

If $A \neq J$, then S_A is called a boundary face; S_J represents the interior part of S, and we also denote it by S_+. That is,

$$S_+ = S_J = \{\mathbf{x} = (x_1, \ldots, x_d) \in S; x_i > 0, i = 1, 2, \ldots, d\}.$$

The collection of all boundary faces is simply called the boundary, and denoted by ∂S. That is,

$$\partial S = \bigcup_{A \subsetneq J} S_A.$$

We now define the reflecting random walk. For each $A \subset J$, let $\{\mathbf{X}_t^A\}_{t=1,2,\ldots}$ be a sequence of independent identically distributed random variables; $\mathbf{X}_t^A$ represents a jump at time t when the random walk is in S_A. We denote its distribution by $\{p_{\mathbf{x}}^A; \mathbf{x} \in \mathbb{R}^d\}$, that is,

$$p_{\mathbf{x}}^A = \mathbb{P}(\mathbf{X}_t^A = \mathbf{x}), \quad \mathbf{x} \in \mathbb{Z}^d.$$

We omit the superscript A of $\mathbf{X}_t^A$ and $p_{\mathbf{x}}^A$ for $A = J$. We assume the following condition:

$$p_{\mathbf{x}}^A = 0 \text{ unless } x_i \geq -1 \ \forall i \in A \text{ and } x_j \geq 0 \ \forall j \notin A. \tag{1.1}$$

Let $\mathbf{Z}_0$ be a random vector taking values in S, and inductively define a discrete time process $\mathbf{Z}_t$ for $t = 0, 1, \ldots$ by

$$\mathbf{Z}_{t+1} = \mathbf{Z}_t + \sum_{A \subset J} \mathbf{X}_{t+1}^A \mathbb{1}_{\{\mathbf{Z} \in S_A\}}, \quad t = 0, 1, \ldots,$$

where we denote by $\mathbb{1}_{\mathcal{A}}$ the indicator function of an event $\mathcal{A}$.

By the assumption (1.1), $\mathbf{Z}_t$ remains in S for all $t \geq 0$. We refer to this process as a *reflecting random walk on a nonnegative orthant with downward skip-free transitions*, or simply as a *reflecting random walk*.

Clearly, $\mathbf{Z}_t$ is a discrete time Markov chain with state space S. Its transition probability $p(\mathbf{x}, \mathbf{y})$ becomes

$$p(\mathbf{x}, \mathbf{y}) = \mathbb{P}(\mathbf{Z}_{t+1} = \mathbf{y} | \mathbf{Z}_t = \mathbf{x}), \quad \mathbf{x}, \mathbf{y} \in S,$$

where the right side of this equation does not depend on $t \geq 0$ by the modeling assumption. Let P be the infinite-dimensional matrix whose $(\mathbf{x}, \mathbf{y})$th entry is $p(\mathbf{x}, \mathbf{y})$; P is a transition matrix, which is obviously stochastic.

We are interested in the stationary distribution of the reflecting random walk $\mathbf{Z}_t$. That is, we seek a distribution π on S such that

$$\lim_{t\to\infty} \mathbb{P}(\mathbf{Z}_t = \mathbf{x}) = \pi(\mathbf{x}), \quad \mathbf{x} \in S.$$

Let $\mathbf{Z}$ be a random variable subject to the distribution π, and $\mathbf{X}^A$ be a random variable with the same distribution of $\mathbf{X}_t^A$ for each $t = 0, 1, 2, \ldots$ and independent to them, then, from the definition of the Markov Chain, it follows that

$$\mathbf{Z} \simeq \mathbf{Z} + \sum_{A \subset J} \mathbf{X}^A(\mathbf{Z} \in S_A),$$

where " $\simeq$ " stands for the equality in distribution; we can view the distribution π as the row vector π whose $\mathbf{x}$th entry is $\pi(\mathbf{x})$ with $\mathbf{x} \in S$.

1.2.1 Jackson network

Let us consider a continuous time queueing network with d nodes, numbered as $1, 2, \ldots, d$. We assume that exogenous customers arrive at node i subject to a Poisson process with rate λ_i, and customers in node i have independent service times with an exponential distribution with mean $\frac{1}{\mu_i}$, and are served in first-in-first-out manner by a single server. A customer who completes service at node i goes to node j with probability r_{ij} or leaves the network with probability r_{i0}, where

$$\sum_{j=0}^{d} r_{ij} = 1, \quad i = 1, 2, \ldots, d.$$

We assume that all the movements are independent. This model, referred to as a *Jackson network*, is usually described by a continuous time Markov chain. For this, let $L_i(t)$ be the number of customers in node i at time t. The d-dimensional vector-valued process $\mathbf{L}(t) = (L_1(t), \ldots, L_d(t))$ is a continuous Markov chain, whose state space is the d-dimensional nonnegative integer orthant $S = \mathbb{Z}_+^d$. It is not hard to see that its transition rate matrix $Q = \{q(\mathbf{x}, \mathbf{y}),\ \mathbf{x}, \mathbf{y} \in S\}$ is given by, for $\mathbf{x} \neq \mathbf{y}$

$$q(\mathbf{x}, \mathbf{y}) = \begin{cases} \lambda_i & \text{if } \mathbf{y} = \mathbf{x} + \mathbf{e}_i,\ i \neq 0, \\ \mu_i r_{ij} & \text{if } \mathbf{y} = \mathbf{x} + \mathbf{e}_i + \mathbf{e}_j,\ x_i > 0,\ i, j \neq 0, \\ \mu_i r_{i0} & \text{if } \mathbf{y} = \mathbf{x} - \mathbf{e}_i,\ m_i > 0,\ i \neq 0, \\ 0 & \text{otherwise,} \end{cases}$$

and

$$q(\mathbf{x},\mathbf{x}) = -\sum_{\mathbf{y}\neq\mathbf{x}} q(\mathbf{x},\mathbf{y}).$$

For notation's convenience, we let

$$r_{00} = 0, \quad \mu_0 = \sum_{k=1}^{d} \lambda_k, \quad r_{0i} = \frac{\lambda_i}{\mu_i} \quad i = 1, 2, \ldots, d.$$

In this case the stationary distribution π is defined as

$$\lim_{t\to\infty} \mathbb{P}(\mathbf{L}(t) = \mathbf{x}) = \pi(\mathbf{x}), \quad \mathbf{x} \in S,$$

and it is obtained as a nonnegative summable solution of the stationary equation

$$\pi Q = \mathbf{0}.$$

The Jackson network can be described by the reflecting random walk in discrete time. We first note that if we change time from t to bt for a constant $b > 0$, that is, time scale is changed by b, then λ_i and μ_i are also increased b times. However, this does not change the stationary distribution. Hence, for studying the stationary distribution, we can assume without loss of generality that

$$\sum_{i=1}^{d} (\lambda_i + \mu_i) = 1.$$

By means of all previous notations, we define, for each $A \subset J$, the probability $\tilde{p}^A_{\mathbf{x}}$ of the reflecting random walk associated to the Jackson network as

$$\tilde{p}^A_{\mathbf{x}} = \sum_{i\in A\cup\{0\}} \sum_{j=0}^{d} \mathbb{1}_{\{\mathbf{x}=\mathbf{e}_j-\mathbf{e}_i\}} \mu_i r_{ij} + \mathbb{1}_{\{\mathbf{x}=\mathbf{e}_0\}} \sum_{i\notin A} \mu_i,$$

where $\mathbf{e}_0 = \mathbf{0}$. Moreover, the transition matrix $\tilde{P}$ of the reflecting random walk is defined by the following rules, for $\mathbf{x} \in S_A$,

$$\tilde{p}(\mathbf{x},\mathbf{y}) = \mathbb{1}_{\{\mathbf{x}\neq\mathbf{y}\}} q(\mathbf{x},\mathbf{y}) + \mathbb{1}_{\{\mathbf{x}=\mathbf{y}\}} \sum_{i\in J} \mu_i \mathbb{1}_{\{m_i=0\}}.$$

With these definitions it turns out that for the infinite row vector $\tilde{\pi}$ we have

$$\tilde{\pi}Q = 0 \Leftrightarrow \tilde{\pi}P = \tilde{\pi}.$$

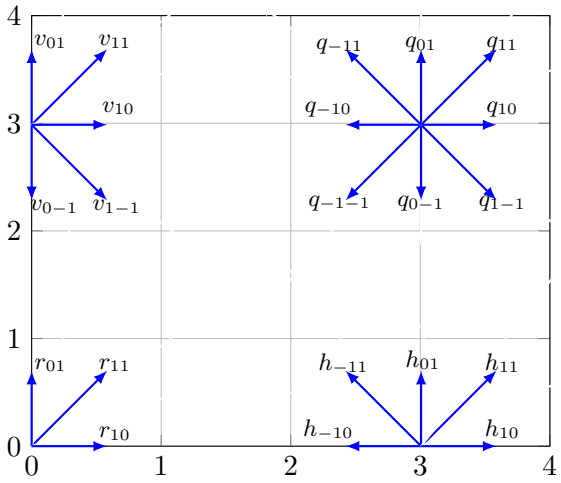


Figure 1.1: Transition rates of the random walk in the quarter plane.

1.2.2 Double QBD process

If all entries of $\mathbf{X}_t^A$ take values $0, 1$ or -1, then the process $\{\mathbf{Z}_t\}$ is called reflecting *skip-free* random walk, in queueing applications, it is also called *multidimensional quasi-birth-and-death* (QBD) process. This kind of process has simpler transitions, but still flexible for applications like Jackson network and some of its modifications.

The multiple QBD process for $d = 2$ is called *double* QBD process, we can see its transition diagram in Figure (1.1).

Remark 1.1. An effective way to visualize the double QBD process is to think of the space S as a semi-infinite chessboard and of the random variable $\mathbf{Z}_t$ as the movement of a King inside it.

The problem of finding the stationary distribution for a random walk in a quarter of plane is the main issue of this thesis. As we have seen in this chapter it can be modeled equivalently either as a discrete time Markov chain or as a continuous time Markov process. The approaches we present in the following chapters are based on this equivalence of models.

In particular we give the explicit representation of the operators both in the case of the Markov process and Markov chain, and also distinguishing between the case when the space $S = \mathbb{Z}_+^2$ is considered with the lexicographic order and the case when

it is considered with the anti-lexicographic order. First we define the quantities

$$r := r_{10} + r_{01} + r_{11}, \quad h := \sum_{i=-1}^{1}\sum_{j=0}^{1} h_{ij}, \quad v := \sum_{i=0}^{1}\sum_{j=-1}^{1} v_{ij}, \quad q := \sum_{i=-1}^{1}\sum_{j=-1}^{1} q_{ij}.$$

Markov process: anti-lexicographic order. In this case the generator is given by

$$Q^{(A)} = \begin{bmatrix} B_0^{(A)} & B_1^{(A)} & & & \\ A_{-1}^{(A)} & A_0^{(A)} & A_1^{(A)} & & \\ & A_{-1}^{(A)} & A_0^{(A)} & A_1^{(A)} & \\ & & \ddots & \ddots & \ddots \end{bmatrix}$$

where the blocks are of the form

$$B_0^{(A)} = \begin{bmatrix} -r & r_{10} & & & \\ h_{-10} & -h & h_{10} & & \\ & h_{-10} & -h & h_{10} & \\ & & \ddots & \ddots & \ddots \end{bmatrix} \quad B_1^{(A)} = \begin{bmatrix} r_{01} & r_{11} & & & \\ h_{-11} & h_{01} & h_{11} & & \\ & h_{-11} & h_{01} & h_{11} & \\ & & \ddots & \ddots & \ddots \end{bmatrix}$$

$$A_0^{(A)} = \begin{bmatrix} -v & v_{10} & & & \\ q_{-10} & -q & q_{10} & & \\ & q_{-10} & -q & q_{10} & \\ & & \ddots & \ddots & \ddots \end{bmatrix} \quad A_1^{(A)} = \begin{bmatrix} v_{01} & v_{11} & & & \\ q_{-11} & q_{01} & q_{11} & & \\ & q_{-11} & q_{01} & q_{11} & \\ & & \ddots & \ddots & \ddots \end{bmatrix}$$

$$A_{-1}^{(A)} = \begin{bmatrix} v_{0-1} & v_{1-1} & & & \\ q_{-1-1} & q_{0-1} & q_{1-1} & & \\ & q_{-1-1} & q_{0-1} & q_{1-1} & \\ & & \ddots & \ddots & \ddots \end{bmatrix}$$

Markov process: lexicographic order. In this case the generator is given by

$$Q^{(L)} = \begin{bmatrix} B_0^{(L)} & B_1^{(L)} & & & \\ A_{-1}^{(L)} & A_0^{(L)} & A_1^{(L)} & & \\ & A_{-1}^{(L)} & A_0^{(L)} & A_1^{(L)} & \\ & & \ddots & \ddots & \ddots \end{bmatrix}$$

where the blocks are of the form

$$B_0^{(L)} = \begin{bmatrix} -r & r_{01} & & & \\ v_{0-1} & -v & v_{01} & & \\ & v_{0-1} & -v & v_{01} & \\ & & \ddots & \ddots & \ddots \end{bmatrix} \quad B_1^{(L)} = \begin{bmatrix} r_{10} & r_{11} & & & \\ v_{1-1} & v_{10} & v_{11} & & \\ & v_{1-1} & v_{10} & v_{11} & \\ & & \ddots & \ddots & \ddots \end{bmatrix}$$

$$A_0^{(L)} = \begin{bmatrix} -h & h_{01} & & & \\ q_{0-1} & -q & q_{01} & & \\ & q_{0-1} & -q & q_{01} & \\ & & \ddots & \ddots & \ddots \end{bmatrix} \quad A_1^{(L)} = \begin{bmatrix} h_{10} & h_{11} & & & \\ q_{1-1} & q_{10} & q_{11} & & \\ & q_{1-1} & q_{10} & q_{11} & \\ & & \ddots & \ddots & \ddots \end{bmatrix}$$

$$A_{-1}^{(L)} = \begin{bmatrix} h_{-10} & h_{-11} & & & \\ q_{-1-1} & q_{-10} & q_{-11} & & \\ & q_{-1-1} & q_{-10} & q_{-11} & \\ & & \ddots & \ddots & \ddots \end{bmatrix}$$

Markov chain: anti-lexicographic order. In this case the generator is given by

$$P^{(A)} = \begin{bmatrix} \tilde{B}_0^{(A)} & \tilde{B}_1^{(A)} & & & \\ \tilde{A}_{-1}^{(A)} & \tilde{A}_0^{(A)} & \tilde{A}_1^{(A)} & & \\ & \tilde{A}_{-1}^{(A)} & \tilde{A}_0^{(A)} & \tilde{A}_1^{(A)} & \\ & & \ddots & \ddots & \ddots \end{bmatrix}$$

where the blocks are of the form

$$\tilde{B}_0^{(A)} = \begin{bmatrix} 0 & \frac{r_{10}}{r} & & & \\ \frac{h_{-10}}{h} & 0 & \frac{h_{10}}{h} & & \\ & \frac{h_{-10}}{h} & 0 & \frac{h_{10}}{h} & \\ & & \ddots & \ddots & \ddots \end{bmatrix} \quad \tilde{B}_1^{(A)} = \begin{bmatrix} \frac{r_{01}}{r} & \frac{r_{11}}{r} & & & \\ \frac{h_{-11}}{h} & \frac{h_{01}}{h} & \frac{h_{11}}{h} & & \\ & \frac{h_{-11}}{h} & \frac{h_{01}}{h} & \frac{h_{11}}{h} & \\ & & \ddots & \ddots & \ddots \end{bmatrix}$$

$$\tilde{A}_0^{(A)} = \begin{bmatrix} 0 & \frac{v_{10}}{v} & & & \\ \frac{q_{-10}}{q} & 0 & \frac{q_{10}}{q} & & \\ & \frac{q_{-10}}{q} & 0 & \frac{q_{10}}{q} & \\ & & \ddots & \ddots & \ddots \end{bmatrix} \quad \tilde{A}_1^{(A)} = \begin{bmatrix} \frac{v_{01}}{v} & \frac{v_{11}}{v} & & & \\ \frac{q_{-11}}{q} & \frac{q_{01}}{q} & \frac{q_{11}}{q} & & \\ & \frac{q_{-11}}{q} & \frac{q_{01}}{q} & \frac{q_{11}}{q} & \\ & & \ddots & \ddots & \ddots \end{bmatrix}$$

$$\tilde{A}_{-1}^{(A)} = \begin{bmatrix} \frac{v_{0-1}}{v} & \frac{v_{1-1}}{v} & & & \\ \frac{q_{-1-1}}{q} & \frac{q_{0-1}}{q} & \frac{q_{1-1}}{q} & & \\ & \frac{q_{-1-1}}{q} & \frac{q_{0-1}}{q} & \frac{q_{1-1}}{q} & \\ & & \ddots & \ddots & \ddots \end{bmatrix}$$

Markov chain: lexicographic order. In this case the generator is given by

$$P^{(L)} = \begin{bmatrix} \tilde{B}_0^{(L)} & \tilde{B}_1^{(L)} & & & \\ \tilde{A}_{-1}^{(L)} & \tilde{A}_0^{(L)} & \tilde{A}_1^{(L)} & & \\ & \tilde{A}_{-1}^{(L)} & \tilde{A}_0^{(L)} & \tilde{A}_1^{(L)} & \\ & & \ddots & \ddots & \ddots \end{bmatrix}$$

where the blocks are of the form

$$\tilde{B}_0^{(L)} = \begin{bmatrix} 0 & \frac{r_{01}}{r} & & & \\ \frac{v_{0-1}}{v} & 0 & \frac{v_{01}}{v} & & \\ & \frac{v_{0-1}}{v} & 0 & \frac{v_{01}}{v} & \\ & & \ddots & \ddots & \ddots \end{bmatrix} \quad \tilde{B}_1^{(L)} = \begin{bmatrix} \frac{r_{10}}{r} & \frac{r_{11}}{r} & & & \\ \frac{v_{1-1}}{v} & \frac{v_{10}}{v} & \frac{v_{11}}{v} & & \\ & \frac{v_{1-1}}{v} & \frac{v_{10}}{v} & \frac{v_{11}}{v} & \\ & & \ddots & \ddots & \ddots \end{bmatrix}$$

$$\tilde{A}_0^{(L)} = \begin{bmatrix} 0 & \frac{h_{01}}{h} & & & \\ \frac{q_{0-1}}{q} & 0 & \frac{q_{01}}{q} & & \\ & \frac{q_{0-1}}{q} & 0 & \frac{q_{01}}{q} & \\ & & \ddots & \ddots & \ddots \end{bmatrix} \quad \tilde{A}_1^{(L)} = \begin{bmatrix} \frac{h_{10}}{h} & \frac{h_{11}}{h} & & & \\ \frac{q_{1-1}}{q} & \frac{q_{10}}{q} & \frac{q_{11}}{q} & & \\ & \frac{q_{1-1}}{q} & \frac{q_{10}}{q} & \frac{q_{11}}{q} & \\ & & \ddots & \ddots & \ddots \end{bmatrix}$$

$$\tilde{A}_{-1}^{(L)} = \begin{bmatrix} \frac{h_{-10}}{h} & \frac{h_{-11}}{h} & & & \\ \frac{q_{-1-1}}{q} & \frac{q_{-10}}{q} & \frac{q_{-11}}{q} & & \\ & \frac{q_{-1-1}}{q} & \frac{q_{-10}}{q} & \frac{q_{-11}}{q} & \\ & & \ddots & \ddots & \ddots \end{bmatrix}$$

Remark 1.2. In the following we use both the model as Markov chain and the one as Markov process, but always with the lexicographic order. With these definitions, the invariant vectors π and $\tilde{\pi}$ such that

$$\pi Q^{(L)} = \mathbf{0}, \quad \tilde{\pi} P^{(L)} = \tilde{\pi},$$

are slightly different. It can be verified that, by partitioning them as

$$\pi = [\pi_0 \ \pi_1 \ \pi_2 \ \dots],$$
$$\tilde{\pi} = [\tilde{\pi}_0 \ \tilde{\pi}_1 \ \tilde{\pi}_2 \ \dots],$$

where, for $m \geq 0$,

$$\pi_m = [\pi_{m,0} \ \pi_{m,1} \ \pi_{m,2} \ \dots],$$
$$\tilde{\pi}_m = [\tilde{\pi}_{m,0} \ \tilde{\pi}_{m,1} \ \tilde{\pi}_{m,2} \ \dots],$$

we can transform them into each other with the following relations

$$\tilde{\pi}_{0,0} = r\pi_{0,0}, \quad \tilde{\pi}_{m,0} = h\pi_{m,0} \quad \tilde{\pi}_{0,n} = v\pi_{0,n}, \quad \tilde{\pi}_{m,n} = r\pi_{m,n} \quad m, n > 0.$$

An alternative approach consists into define the transition operator of the Markov chain approach starting from $Q^{(L)}$ in the following way

$$P^{(L)} := \frac{1}{\theta} Q^{(L)} + I, \quad \theta := \max\{q, v, h, r\}. \tag{1.2}$$

Let us observe that with this definition $P^{(L)}$ is always stochastic, besides $P^{(L)}$ and $Q^{(L)}$ share the same invariant vector π, indeed

$$\pi P^{(L)} = \frac{1}{\theta} \pi Q^{(L)} + \pi = \pi.$$

Chapter 2

Matrix geometric approach

In this chapter we consider our problem modeled as a discrete time Markov chain on the space $S = \mathbb{Z}_+^2$ which we partition as

$$S = \bigcup_{n \geq 0} \mathcal{L}(n),$$

where $\mathcal{L}(n) = \{(n,0),(n,1),(n,2),\dots\}$ represents the set of states at level n.

All the ideas and the proofs of this chapter are taken from [4] and [13] and they are adapted for the case of infinite states.

The stationary probability vector π is the unique solution of the system

$$\begin{cases} \pi = \pi P, \\ \pi \mathbf{1} = 1, \end{cases} \tag{2.1}$$

where P is the transition operator, in this chapter it doesn't matter whether it is the one based on the lexicographic order or the anti-lexicographic order or if it is the one defined in 1.2, indeed we could obtain the same results by partitioning the space of states defining $\mathcal{L}'(m) = \{(0,m),(1,m),(2,m),\dots\}$.

We partition the vector π by levels into subvectors π_n, for $n \geq 0$. Let us observe that all the subvectors π_n have infinite dimension. Because of this decomposition, the defining system may be written as

$$\begin{cases} \pi_0(B_0 - I) + \pi_1 A_{-1} = \mathbf{0}, \\ \pi_0 B_1 + \pi_1(A_0 - I) + \pi_2 A_1 = \mathbf{0}, \\ \pi_{n-1} A_{-1} + \pi_n(A_0 - I) + \pi_{n+1} A_1 = \mathbf{0} \quad \text{for} \quad n \geq 2, \\ \sum_{n \geq 0} \pi_n \mathbf{1} = 1, \end{cases}$$

where I represents the identity operator whose entries are equal to 1 on the diagonal and are equal to 0 on the off-diagonal.

Theorem 2.1. *If the QBD is positive recurrent, then there exists a nonnegative operator R such that*

$$\begin{cases} \pi_{n+1} & = \pi_n R \quad \textit{for} \quad n \geq 1, \\ \pi_1 & = \pi_0 B_1 A_1^{-1} R \end{cases} \tag{2.2}$$

Proof. Fix $n \geq 1$ and partition the state space as $T \cup T^C$, where $T = \mathcal{L}(0) \cup \cdots \cup \mathcal{L}(n)$ and $T^C = \mathcal{L}(n+1) \cup \mathcal{L}(n+2) \cup \ldots$. From this, the following partition of the matrix P holds

$$P = \begin{bmatrix} P_T & P_{TT^C} \\ P_{T^C T} & P_{T^C} \end{bmatrix},$$

where P_T and P_{T^C} are the submatrices of transition probabilities between states of T and T^C respectively. Let us observe that both P_{TT^C} and $P_{T^C T}$ have just one non zero block, in the lower left corner and in the upper right corner, respectively. From (2.1) we obtain the following relation

$$[\pi_{n+1}\ \pi_{n+2}\ \ldots] = [\pi_0 \ldots \pi_n]\, P_{TT^C} (I - P_{T^C})^{-1},$$

where $(I - P_{T^C})^{-1} = \sum_{i \geq 0} P_{T^C}^i$ converges since we have assumed that the QBD is irreducible.

We decompose the matrix $N_{T^C} := (I - P_{T^C})^{-1}$ into blocks $N_{kk'}$ with $k, k' \geq 1$ where each block represents the expected number of visits to the states in $\mathcal{L}(n+k')$, starting from a state in $\mathcal{L}(n+k)$, before the first visit to any of the states in T.

Because of the extremely sparse structure of the matrix P_{TT^C}, we have that

$$P_{TT^C}(I - P_{T^C})^{-1} = \begin{bmatrix} 0 & 0 & 0 & \ldots \\ \vdots & \vdots & \vdots & \\ 0 & 0 & 0 & \ldots \\ A_1 N_{11} & A_1 N_{12} & A_1 N_{13} & \ldots \end{bmatrix}$$

and that

$$[\pi_{n+1}\ \pi_{n+2}\ \ldots] = [\pi_n A_1 N_{11}\ \pi_n A_1 N_{12}\ \ldots].$$

In particular we obtain that $\pi_{n+1} = \pi_n A_1 N$, where we have defined $N := N_{11}$.

Because of the homogeneity of the process, independently from the value chosen for n, the matrix P_{T^C} is the same. Therefore, the matrix N is independent of n and it records the expected number of visits to $\mathcal{L}(n+1)$, starting from $\mathcal{L}(n+1)$, before the first visit to T, for all values of $n \geq 0$.

Furthermore, the structure of P_{TT^C} is independent of n, thus we have that $\pi_{n+1} = \pi_n A_1 N$ for all $n \geq 0$, which may be written as $\pi_{n+1} = \pi_0 R^n$ for all $n \geq 0$. We conclude the proof by observing that R records the expected number of visits to $\mathcal{L}(n+1)$ starting from $\mathcal{L}(n)$ and avoiding T.

For the case $n = 0$ with the same scheme of the other cases we obtain

$$[\pi_1 \ \pi_2 \ \ldots] = \pi_0 P_{TT^C} (I - P_{T^C})^{-1},$$

where $P_{T^C} = [B_1 \ 0 \ 0 \ \ldots]$ and P_{TT^C} is defined as before. Considering again the operator N, we get

$$[\pi_1 \ \pi_2 \ \ldots] = \pi_0 \left[B_1 N_{11} \ B_1 N_{12} \ \ldots\right],$$

from which in particular it follows that

$$\pi_1 = \pi_0 B_1 N_{11} = \pi_0 B_1 N = \pi_0 B_1 A_1^{-1} R.$$

□

Remark 2.1. The operator R is such that, for any $n \geq 0$, the entry R_{ij} is the expected number of visits to $(n+1, j)$ before a return to $\mathcal{L}(0) \cup \cdots \cup \mathcal{L}(n)$ given that the process starts in (n, i).

In a similar way, the operator N which appears in the proof of the theorem above, is such that, for any $n \geq 0$, the entry N_{ij} is the expected number of visits to (n, j), starting from (n, i), before the first visit to any of the states in $\mathcal{L}(n-1)$.

Let us assume that $\mathbf{X}_0$ is in $\mathcal{L}(n)$. Define τ as the first epoch of visit to the level $\mathcal{L}(n-1)$ and θ as the first epoch of return to level $\mathcal{L}(n)$. Let us define the operator U and G as follows,

$$\begin{aligned} U_{ij} &= \mathbb{P}\left[\theta < \tau \text{ and } \mathbf{X}_\theta = (n, j) | \mathbf{X}_0 = (n, i)\right], \\ G_{ij} &= \mathbb{P}\left[\tau < \infty \text{ and } \mathbf{X}_\tau = (n-1, j) | \mathbf{X}_0 = (n, i)\right]. \end{aligned}$$

The operator U records the probability, starting from $\mathcal{L}(n)$, of returning to $\mathcal{L}(n)$ before visiting $\mathcal{L}(n-1)$; the operator G records the probability, starting from $\mathcal{L}(n)$, of visiting $\mathcal{L}(n-1)$ in a finite time. In view of the homogeneity of the process the values of U and G do not depend on n.

The following theorems show that the operators R, G and U are connected by certain equation such that if one of the three operators is known, then it is possible to determine the other two.

Theorem 2.2. *With the previous definitions, the following equations hold:*

$$R = A_1 (I - U)^{-1}, \tag{2.3}$$
$$G = (I - U)^{-1} A_{-1}, \tag{2.4}$$
$$U = A_0 + A_1 G, \tag{2.5}$$
$$U = A_0 + R A_{-1}. \tag{2.6}$$

Proof. From the definition of the operator U it follows that its U^k for $k \geq 1$ records the probability starting from $\mathcal{L}(n)$, of returning to $\mathcal{L}(n)$ and remaining into this level for k units of time, before visiting $\mathcal{L}(n-1)$. Because of this we have that

$$N = \sum_{i \geq 0} U^i = (I - U)^{-1},$$

and therefore that

$$R = A_1(I - U)^{-1} \quad \text{and} \quad G = (I - U)^{-1} A_{-1}.$$

Moreover we have

$$U = A_0 + A_1 G,$$

indeed, in order to return to $\mathcal{L}(n)$ avoiding $\mathcal{L}(n-1)$, the process may either remain in $\mathcal{L}(n)$ at time 1, with probabilities recorded in A_0, or move up to $\mathcal{L}(n+1)$, with probabilities recorded in A_1; from $\mathcal{L}(n+1)$, the process returns to $\mathcal{L}(n)$ with probabilities recorded in G.

This last equation may also be written as

$$U = A_0 + A_1(I - U)A_{-1} = A_0 + RA_{-1}.$$

□

Remark 2.2. The equations we have obtained in the previous theorem allow us to rewrite the relation between π_0 and π_1 as

$$\pi_1 = \pi_0 B_1 (I - U)^{-1}.$$

Theorem 2.3. *The three operators U,G and R satisfy the following equations:*

$$U = A_0 + A_1(I - U)^{-1} A_{-1}, \tag{2.7}$$

$$G = A_{-1} + A_0 G + A_1 G^2, \tag{2.8}$$

$$R = A_1 + RA_0 + R^2 A_{-1}. \tag{2.9}$$

Proof. Equation (2.7) is simply obtained by inserting (2.4) into (2.5).

Starting from (2.4) and multiplying on the left by $I - U$ we obtain

$$G = A_{-1} + UG,$$

inserting (2.5) into this equation we obtain (2.8).

Starting from (2.3) and multiplying on the right by $I - U$ we obtain

$$R = A_1 + RU,$$

inserting (2.6) into this equation we obtain (2.9). □

2.1 Boundary Distribution

In order to completely specify the stationary distribution, we need determine the subvector π_0. We do this with the help of the following theorem, whose proof can be found in [13, pag. 119].

Theorem 2.4. *Let $\{\mathbf{X}_t\}_{t\geq 0}$ be an irreducible, homogeneous, positive recurrent Markov chain on a countable state space S with transition matrix P. Let S be partitioned into two arbitrary subsets $\mathcal{S}$ and $\mathcal{S}^C$.*

The restricted process $\{\mathbf{X}_t^{(T)}\}_{t\geq 0}$ is an irreducible, homogeneous, positive recurrent Markov chain on the state space $\mathcal{S}$. Its transition matrix $P(\mathcal{S})$ is given by

$$P(\mathcal{S}) = P_{\mathcal{S}} + P_{\mathcal{S}\mathcal{S}^C}\left(I - P_{\mathcal{S}^C}\right)^{-1} P_{\mathcal{S}^C\mathcal{S}}.$$

Its stationary probability vector $\pi_{\mathcal{S}}$ is given by

$$\pi_{\mathcal{S}} P(\mathcal{S}) = \pi_{\mathcal{S}}.$$

Consider the restricted process on $\mathcal{S} = \mathcal{L}(0)$. By Theorem 2.4, its transition matrix is given by

$$\begin{aligned}
P(\mathcal{S}) &= B_0 + [B_1\ 0\ 0\ \dots] \begin{bmatrix} I - A_0 & A_1 & 0 & & \\ A_{-1} & I - A_0 & A_1 & 0 & \\ & \ddots & \ddots & \ddots & \ddots \end{bmatrix}^{-1} \begin{bmatrix} A_{-1} \\ 0 \\ 0 \\ \vdots \end{bmatrix} \\
&= B_0 + [B_1\ 0\ 0\ \dots] \sum_{i\geq 0} \begin{bmatrix} A_0 & A_1 & 0 & & \\ A_{-1} & A_0 & A_1 & 0 & \\ & \ddots & \ddots & \ddots & \ddots \end{bmatrix}^{i} \begin{bmatrix} A_{-1} \\ 0 \\ 0 \\ \vdots \end{bmatrix} \\
&= B_0 + B_1 \sum_{i\geq 0} U^i A_{-1} \\
&= B_0 + B_1 (I - U)^{-1} A_{-1} = B_0 + B_1 G.
\end{aligned}$$

Moreover, the vector π_0 is such that $\pi_0 = \pi_0 P(\mathcal{S})$ and the normalization factor is determined by the constraint $\pi\mathbf{1} = 1$. Since

$$\pi\mathbf{1} = \sum_{i\geq 0} \pi_i \mathbf{1} = \pi_0\mathbf{1} + \pi_1 (I - R)^{-1}\mathbf{1} = \pi_0\mathbf{1} + \pi_0 B_1 A_1^{-1} R (I - R)^{-1}\mathbf{1},$$

we have thus proved the following

Lemma 2.1. *The stationary distribution π_0 of the boundary states in $\mathcal{L}(0)$ is the unique solution of the system*

$$\begin{cases} \pi_0 (B_0 + B_1 G) = \pi_0, \\ \pi_0\mathbf{1} + \pi_0 B_1 A_1^{-1} R (I - R)^{-1}\mathbf{1} = 1. \end{cases}$$

2.2 Minimal Solution of Quadratic Equation

The main purpose of this section is to prove that the operators R and G are completely characterized as minimal nonnegative solutions of (2.9) and (2.8) respectively.

We begin by defining $\gamma(i)$ as the first passage time at the level $\mathcal{L}(i)$:

$$\gamma(i) := \inf\{t \geq 1 : \mathbf{X}_t \in \mathcal{L}(i)\} \quad \text{for} \quad i \geq 0.$$

In terms of these quantities, we can review the definitions of the operators U and G as follows

$$\begin{aligned} G_{ij} &= \mathbb{P}\left[\gamma(0) < \infty, \mathbf{X}_{\gamma(0)} = (0,j) | \mathbf{X}_0 = (1,i)\right], \\ U_{ij} &= \mathbb{P}\left[\gamma(1) < \gamma(0), \gamma(1) < \infty, \mathbf{X}_{\gamma(1)} = (1,j) | \mathbf{X}_0 = (1,i)\right], \end{aligned}$$

for $i, j \geq 1$. In order to simplify the equations, we shall use the short-hand notations

$$\begin{aligned} G &= \mathbb{P}\left[\gamma(0) < \infty, \mathbf{X}_{\gamma(0)} | \mathbf{X}_0 \in \mathcal{L}(1)\right], \\ U &= \mathbb{P}\left[\gamma(1) < \gamma(0), \gamma(1) < \infty, \mathbf{X}_{\gamma(1)} | \mathbf{X}_0 \in \mathcal{L}(1)\right]. \end{aligned}$$

With the same notation let us define the sequences $\{G(k)\}_{k\geq 1}$ and $\{U(k)\}_{k\geq 1}$ as

$$\begin{aligned} G(k) &= \mathbb{P}\left[\gamma(0) < \gamma(k+1), \mathbf{X}_{\gamma(0)} | \mathbf{X}_0 \in \mathcal{L}(1)\right], \\ U(k) &= \mathbb{P}\left[\gamma(1) < \gamma(0), \gamma(1) < \gamma(k+1), \mathbf{X}_{\gamma(1)} | \mathbf{X}_0 \in \mathcal{L}(1)\right]. \end{aligned}$$

In this way $G_{ij}(k)$ is the probability that the process moves from the state $(1,i)$ at time 0 to the level $\mathcal{L}(0)$ in a finite amount of time by visiting the specific state $(0,j)$ under the taboo of the states in $\mathcal{L}(k+1)$ and higher level. Similarly $U_{ij}(k)$ is the probability that, starting from $(1,i)$, the process returns to the level $\mathcal{L}(1)$ by visiting the specific state $(1,j)$ under taboo of the states in $\mathcal{L}(0)$, $\mathcal{L}(k+1)$ and higher level.

Theorem 2.5. *The sequences $\{G(k)\}_{k\geq 1}$ and $\{U(k)\}_{k\geq 1}$ are connected by the iterations*

$$\begin{aligned} &G(0) = 0, \\ &U(k) = A_0 + A_1 G(k-1), && (2.10) \\ &G(k) = (I - U(k))^{-1} A_{-1} \quad \textit{for} \quad k \geq 1. && (2.11) \end{aligned}$$

Moreover they are monotonically increasing and converge to the operators G and U, respectively.

Proof. If $\mathbf{X}_0 \in \mathcal{L}(1)$, then necessarily $0 < \gamma(2) < \gamma(3) < \dots$ because the process may not increase by more than one level at a time. For the same reason, we have that $\gamma(k) \geq k-1$, so that $\lim_{k\to\infty} \gamma(k) = \infty$. Thus the last statement is obvious, and we only need to verify that the two sequences satisfy (2.10) and (2.11). Starting from $\mathbf{X}_0 \in \mathcal{L}(1)$, the process can make a passage into $\mathcal{L}(0)$, avoiding $\mathcal{L}(k+1)$, in two mutually exclusive ways. One way is to visit $\mathcal{L}(0)$ at the very first step with probabilities recorded in A_{-1}; the other way is to return to $\mathcal{L}(1)$ avoiding $\mathcal{L}(0)$ and $\mathcal{L}(k+1)$ with probabilities recorded in $U(k)$ and to start all over again from $\mathbf{X}_{\gamma(1)}$. This leads to the equation

$$G(k) = A_{-1} + U(k)G(k);$$

since $U(k) \leq U$, then $I - U$ is invertible and we immediately obtain (2.10).

To prove (2.11) we observe that for $k = 1$ the following equivalence of events holds

$$[\gamma(1) < \gamma(0), \gamma(1) < \gamma(2)] \equiv [\gamma(1) = 1] \equiv [\mathbf{X}_1 \in \mathcal{L}(1)]\,,$$

so that $U(1) = A_0$. For $k \geq 2$, in order to have $\gamma(1) < \gamma(0)$, it is necessary either that the process remains in $\mathcal{L}(1)$ at the first step with probabilities recorded in A_0, or that it moves up to $\mathcal{L}(2)$, with probabilities recorded in A_1, from which level the process will eventually have to return to $\mathcal{L}(1)$ avoiding $\mathcal{L}(k+1)$. Thus, we have that

$$U(k) = A_0 + A_1\tilde{G}(k),$$

where

$$\tilde{G}(k) = \mathbb{P}\left[\gamma(1) < \gamma(k+1), \mathbf{X}_{\gamma(1)} | \mathbf{X}_0 \in \mathcal{L}(2)\right].$$

We conclude the proof by observing that $\tilde{G}(k) = G(k-1)$, because of the homogeneity of the process. □

Now we have all the instruments we need to prove the main theorem of this section about operators G, R and U being the minimal solution of certain quadratic equation.

Theorem 2.6. *The operators U and G are the minimal nonnegative solutions of the system*

$$X = A_0 + A_1 Y, \quad Y = (I - X)^{-1} A_{-1}. \tag{2.12}$$

The operators U and R are the minimal nonnegative solutions of the system

$$X = A_0 + A_{-1} Z, \quad Z = A_1 (I - X)^{-1}. \tag{2.13}$$

The operator U is the minimal nonnegative solution of the equation

$$X = A_0 + A_1 (I - X)^{-1}. \tag{2.14}$$

The operator G is the minimal nonnegative solution of the equation

$$Y = A_{-1} + A_0 Y + A_1 Y^2. \tag{2.15}$$

The operator R is the minimal nonnegative solution of the equation

$$Z = A_1 + Z A_0 + Z^2 A_{-1}. \tag{2.16}$$

Proof. As we have seen the operators U, G and R are solutions of the stated equations, now we have only to prove their minimality.

Assume that (X^*, Y^*) is another nonnegative solution of (2.12). Since $Y^* \geq 0$, we have that $U(1) = A_0 \leq A_0 + A_1 Y^* = X^*$. Now let us assume that $U(k) \leq X^*$ for some k. Then

$$G(k) = (I - U(k))^{-1} A_{-1} = \sum_{j \geq 0} U(k)^j A_{-1} \leq \sum_{j \geq 0} (X^*)^j A_{-1} = Y^*$$

and

$$U(k+1) = A_0 + A_1 G(k) \leq A_0 + A_1 Y^* = X^*.$$

By induction, we obtain that $U(k) \leq X^*$ and $G(k) \leq Y^*$ for al k, so the inequalities hold at the limit, $U = \lim_{k \to \infty} U(k) \leq X^*$ and $G = \lim_{k \to \infty} G(k) \leq Y^*$.

To prove the second statement, we define the following sequence

$$R(k) = A_1 \sum_{j \geq 0} U(k)^j;$$

since $\{U(k)\}$ monotonically converges to U, then $\{R(k)\}$ monotonically converges to $R = A_1 \sum_{j \geq 0} U^j$. Now we proceed in a manner exactly similar to that of the first statement.

In the same way we obtain also the third statement, while the fourth is obtained by contradiction. Let us assume that there exists a nonnegative solution Y^* of (2.15) such that $Y^* \leq G$, $Y^* \neq G$. By defining $X^* = A_0 + A_1 Y^*$, we have that

$$\begin{aligned} Y^* &= A_{-1} + (A_0 + A_1 Y^*) Y^* \\ &= A_{-1} + X^* Y^* = A_{-1} + X^* Y^* + (X^*)^2 Y^* = \dots \\ &= \sum_{j \geq 0} (X^*)^j A_{-1}. \end{aligned}$$

The series converges since $X^* = A_0 + A_1 Y^* \leq A_0 + A_1 G = U$. We thus find that (X^*, Y^*) is a solution of the system (2.12) and is smaller than (U, G), this is absurd because of what we have proved in the first statement.

For the last statement, we assume that Z^* is a nonnegative solution of (2.16) and we define $X^* = A_0 + Z^* A_{-1}$. Since $Z^* \geq 0$, we have

$$Z^* \geq A_1 + Z^* A_0 \geq A_1 \sum_{j \geq 0} A_0^j = A_1 (I - U(1))^{-1} = R(1).$$

Now let us assume that $R(k) \leq Z^*$ for some k, then

$$\begin{aligned} U(k+1) &= A_0 + A_1 G(k) \\ &= A_0 + A_1 (I - U(k))^{-1} A_{-1} \\ &= A_0 + R(k) A_{-1} \\ &\leq A_0 + Z^* A_{-1} = X^*, \end{aligned}$$

by the inductive hypothesis. We also have that

$$\begin{aligned} Z^* &= A_1 + Z^*(A_0 + Z^* A_{-1}) \\ &= A_1 + Z^* X^* \\ &\geq A_1 + Z^* U(k+1) \\ &\geq A_1 (I - U(k+1))^{-1} \\ &= R(k+1), \end{aligned}$$

which proves the induction step. Thus, $R = \lim_{k \to \infty} R(k) \leq Z^*$ which concludes the proof. □

2.3 Quasi-Toeplitz operators

The operators A_i and B_j for $i = -1, 0, 1$ and $j = 0, 1$ share the same special structure that is the topic we focus on in this section.

First we give a short list of well known results about linear operators, especially from the spectral point of view, for an extended dissertation on this argument, we refer to [10].

We shall denote by $\mathcal{B}(X)$ the space of all linear bounded operators $A : X \mapsto X$ and we recall that it is a Banach space if X is a Banach space too with the natural induced norm

$$\|A\| = \|A\|_{\mathcal{B}(X)} = \sup_{\|x\|_X = 1} \|Bx\|_X.$$

The spectrum Λ_A of $A \in \mathcal{B}(X)$ is given by the set

$$\Lambda_A = \{z \in \mathbb{C} : zI - A \text{ is not invertible}\}.$$

It is well known that Λ_A is a closed subset of $\mathbb{C}$. The real number $\delta(A) = \sup_{\lambda\in\Lambda_A}|\lambda|$, called the spectral radius of A, is such that

$$\lim_{k\to\infty}\|A^k\|^{\frac{1}{k}} = \delta(A),$$

moreover $\|A^k\|^{\frac{1}{k}} \leq \|A\|$, so that $\delta(A) \leq \|A\|$. We recall also that, if $\delta(A) < 1$ then $I - A$ is invertible and $(I - A)^{-1} = \sum_{i=0}^{\infty} A^i$.

The following remark will be useful in next chapter.

Remark 2.3. From the limit property of $\|A^k\|^{\frac{1}{k}}$ it follows that for any $\epsilon > 0$ there exists an integer $N > 0$ such that for any $k \geq N$ it holds $\|A^k\|^{\frac{1}{k}} < \delta(A) + \epsilon$. In particular, if $\delta(A) < 1$ and if $\epsilon > 0$ is such that $\delta(A) + \epsilon < 1$, then $\|A^k\| < (\delta(A) + \epsilon)^k < 1$. Thus, if $k \geq N$ and $k = qN + r$, where q and r are quotient and remainder of the division of k by N, one finds that $\|A^k\| \leq \|A^k\|\|A^N\|^q$. This implies that if $\delta(A) < 1$, then $\lim_{k\to\infty}\|A^k\| = 0$.

We write $\rho(A)$ for the set of all values $z \in \mathbb{C}$ such that $(zI - A)$ is invertible, so we have that $\rho(A) = \mathbb{C} \setminus \Lambda_A$ is an open subset of $\mathbb{C}$. The map $R_A : \rho(A) \mapsto \mathcal{B}(X)$ defined by

$$R_A(z) := (zI - A)^{-1}$$

is called the resolvent operator.

In the following of the section our main goal is to construct a space which these operator belong to and an approximate matrix arithmetic in it. In order to do this we recall the following

Definition 2.1. A *Banach Algebra* $\mathcal{B}$ is a normed space such that

- it is complete in the metric induced by the norm,
- the norm is submultiplicative,
- it is closed under product.

Toeplitz operators can be studied from a functional point of view, by considering the following sets

$$\mathcal{W} = \left\{ a(z) : \partial\mathbb{B} \to \mathbb{C} :\ a(z) = \sum_{i\in\mathbb{Z}} a_i z^i,\ \sum_{i\in\mathbb{Z}}|a_i| < \infty \right\},$$

$$\mathcal{W}_1 = \left\{ a(z) \in \mathcal{W} :\ a'(z) := \sum_{i\in\mathbb{Z}} i a_i z^{i-1} \in \mathcal{W} \right\},$$

in the following, we denote by $a^+(z)$ and by $a^-(z)$ the power series defined by the coefficients of $a(z)$ with positive and with negative powers, respectively, that is,

$a^+(z) = \sum_{i\in\mathbb{Z}_+} a_i z^i$ and $a^-(z) = \sum_{i\in\mathbb{Z}_-} a_i z^i$, so that $a(z) = a_0 + a^-(z) + a^+(z)$. We associate with the Laurent series $a(z)$, and with the power series $b(z) = \sum_{i\geq 0} b_i z^i$ the following operators

$$\begin{aligned} T(a) =&(t_{ij}) := a_{j-i}, \\ H(b) =&(h_{ij}) := b_{i+j-1}, \quad i,j \in \mathbb{Z}_+, \end{aligned}$$

we refer to $a(z)$ as the symbol of the operator $A = T(a)$ and with a slight abuse of notation we write $A \in \mathcal{W}$. It is well known that both $\mathcal{W}$ and $\mathcal{W}_1$ are Banach algebras with the norms $\|A\|_{\mathcal{W}} = \sum_{i\in\mathbb{Z}} |a_i|$ and $\|A\|_{\mathcal{W}_1} = \sum_{i\in\mathbb{Z}} |a_i| + \sum_{i\in\mathbb{Z}} |ia_i|$, moreover the following Theorem holds, whose proof can be found in [9] as well as the proofs of the facts mentioned above.

Theorem 2.7. *For $a(z), b(z) \in \mathcal{W}$ let $c(z) = a(z)b(z)$. Then we have $T(a)T(b) = T(c) - H(a^-)H(b^+)$. Moreover, for any $a(z) \in \mathcal{W}$ and for any $p \geq 1$, including $p = \infty$, we have*

$$\|T(a)\|_p \leq \|a\|_{\mathcal{W}}, \quad \|H(a^-)\|_p \leq \|a^-\|_{\mathcal{W}}, \quad \|H(a^+)\|_p \leq \|a^+\|_{\mathcal{W}}.$$

The operators A_i and B_j for $i = -1, 0, 1$ and $j = 0, 1$ in general does not belong to both $\mathcal{W}$ and $\mathcal{W}_1$, so we need to introduce the following operator set

$$\mathcal{F} = \left\{ F = (f_{ij})_{i,j\in\mathbb{Z}_+}, \sum_{i,j\in\mathbb{Z}_+} |f_{ij}| < \infty \right\}.$$

Let us observe that $\mathcal{F}$ correspond with the space ℓ^1 if we look at its operators as infinite vectors. We have the following

Lemma 2.2. *The space $\mathcal{F}$ equipped with matrix sum and multiplication and with the norm $\|F\|_{\mathcal{F}} = \sum_{i,j\in\mathbb{Z}_+} |f_{ij}|$ is a Banach algebra over $\mathbb{C}$.*

Proof. We need to show that given $E, F \in \mathcal{F}$ and $\alpha \in \mathbb{C}$ it holds

1. $\alpha E \in \mathcal{F}$,
2. $E + F \in \mathcal{F}$,
3. $EF \in \mathcal{F}$ and $\|EF\|_{\mathcal{F}} \leq \|E\|_{\mathcal{F}} \|F\|_{\mathcal{F}}$,
4. $(\mathcal{F}, \|\cdot\|_{\mathcal{F}})$ is a complete metric space.

Clearly, $\sum_{i,j\in\mathbb{Z}_+} |\alpha e_{ij}| = |\alpha| \sum_{i,j\in\mathbb{Z}_+} |e_{ij}| < \infty$ which proves 1. By the triangular inequality one obtains that $\sum_{i,j\in\mathbb{Z}_+} |e_{ij} + f_{ij}| \leq \sum_{i,j\in\mathbb{Z}_+} |e_{ij}| + \sum_{i,j\in\mathbb{Z}_+} |f_{ij}| < \infty$

which implies 2. If $H = EF = (h_{ij})$ then $h_{ij} = \sum_{r\in\mathbb{Z}_+}|e_{ir}f_{rj}|$ so that, defining $\alpha_r = \sum_{i\in\mathbb{Z}_+}|e_{ir}|$, and $\beta_r = \sum_{j\in\mathbb{Z}_+}|f_{rj}|$, we have

$$\|EF\|_{\mathcal{F}} \leq \sum_{i,j,r\in\mathbb{Z}_+} |e_{ir}||f_{rj}| = \sum_{r\in\mathbb{Z}_+} \alpha_r\beta_r \leq \left(\sum_{r\in\mathbb{Z}_+} \alpha_r\right)\left(\sum_{r\in\mathbb{Z}_+} \beta_r\right) = \|E\|_{\mathcal{F}}\|F\|_{\mathcal{F}},$$

which proves 3. Finally, we observe that any operator $E \in \mathcal{F}$ can be viewed as a vector $v = (v_k)_{k\in\mathbb{Z}_+}$ obtained by suitably ordering the entries e_{ij}. Moreover, the norm $\|\cdot\|_{\mathcal{F}}$ corresponds to the ℓ^1 norm in the space of infinite vectors having finite sum of their moduli. This way, the space $\mathcal{F}$ actually coincides with ℓ^1, which is a Banach space. Thus, we get 4. □

Thanks to $\mathcal{F}$ we are now able to build the right spaces for our operators:

$$\mathcal{QT} := \{T(a) + F,\ a \in \mathcal{W}, F \in \mathcal{F}\},$$
$$\mathcal{QT}_1 := \{T(a) + F,\ a \in \mathcal{W}_1, F \in \mathcal{F}\}.$$

Observe that given $A \in \mathcal{QT}$ there is a unique way to decompose it. In fact, suppose by contradiction that there exist $a_1(z), a_2(z) \in W$ and $E_1, E_2 \in \mathcal{F}$ with $a_1 \neq a_2$ and $E_1 \neq E_2$ such that $A = T(a_1) + E_1 = T(a_2) + E_2$. Then we should have $E_1 - E_2 = T(a_2) - T(a_1) = T(a_2 - a_1)$, hence $\|E_1 - E_2\|_{\mathcal{F}} = \|T(a_2 - a_1)\|_{\mathcal{F}}$. On the other hand, since $T(a_2 - a_1) \neq 0$ we have $\|T(a_2 - a_1)\|_{\mathcal{F}} = \infty$, which contradicts the fact that $E_1 - E_2 \in \mathcal{F}$. Obviously the same result holds for $A \in \mathcal{QT}_1$, since $\mathcal{QT}_1 \subset \mathcal{QT}$.

Lemma 2.3. *The set $\mathcal{QT}$ endowed with the norm $\|T(a) + E\|_{\mathcal{QT}} = \|a\|_{\mathcal{W}} + \|E\|_{\mathcal{F}}$ is a Banach space.*

Proof. It clearly holds the isomorphism $\mathcal{QT} \simeq \mathcal{W} \oplus \mathcal{F}$. Since both $\mathcal{W}$ and $\mathcal{F}$ are Banach spaces, the composition of the 1-norm of $\mathbb{R}^2$ with the vector valued function $T(a) + E \mapsto (\|a\|_{\mathcal{W}}, \|E\|_{\mathcal{F}})$ makes $\mathcal{W} \oplus \mathcal{F}$ a complete metric space. □

The symbols associated with the operators in the process we are dealing with are composed by a finite sum, so they belong to both $\mathcal{QT}$ and $\mathcal{QT}_1$. As we will see in the following, $\mathcal{QT}_1$ has more structure properties than $\mathcal{QT}$ and so it will be the space in which we will develop our theory. The next statements have the common aim to prove that $\mathcal{QT}_1$ is a Banach algebra.

Lemma 2.4. *Let $a(z), b(z) \in \mathcal{W}_1$ and set $c(z) = a(z)b(z)$. Then $T(a)T(b) = T(c) + E$, where $E \in \mathcal{F}$. Moreover,*

$$\|E\|_{\mathcal{F}} \leq \|H(a^-)\|_{\mathcal{F}}\|H(b^+)\|_{\mathcal{F}} = \|(a^-)'\|_{\mathcal{W}}\|(b^+)'\|_{\mathcal{W}} \leq \|a'\|_{\mathcal{W}}\|b'\|_{\mathcal{W}}.$$

Proof. From Theorem 2.7 we deduce that $T(a)T(b) = T(c) + E$ where we set $E = -H(a^-)H(b^+)$.

Let us prove that $H(a^-), H(b^+) \in \mathcal{F}$. We have $\|H(b^+)\|_{\mathcal{F}} = \sum_{i,j\in\mathbb{Z}_+} |b_{i+j-1}|$. Setting $k = i + j - 1$ we may write $\|H(b^+)\|_{\mathcal{F}} = \sum_{k\in\mathbb{Z}_+} k|b_k|$ which is finite since $b(z) \in \mathcal{W}_1$. The same argument applies to $H(a^-)$, $\mathcal{F}$ is a normed matrix algebra therefore $\|E\|_{\mathcal{F}} \leq \|H(a^-)\|_{\mathcal{F}}\|H(b^+)\|_{\mathcal{F}} < \infty$. We conclude the proof by observing that the quantities $\sum_{i\in\mathbb{Z}_+} i|a_{-i}|$ and $\sum_{i\in\mathbb{Z}_+} i|b_i|$ coincide with the $\mathcal{W}$-norms of the first derivatives of the functions $a^-(z)$ and $b^+(z)$, respectively, and that hold the inequalities $\|(a^-)'\|_{\mathcal{W}} \leq \|a'\|_{\mathcal{W}}$ and $\|(b^+)'\|_{\mathcal{W}} \leq \|b'\|_{\mathcal{W}}$. □

Theorem 2.8. *Let $A, B \in \mathcal{QT}_1$, where $A = T(a) + E_a$ and $B = T(b) + E_b$. Then we have $C = AB = T(c) + E_c \in \mathcal{QT}_1$ with $c(z) = a(z)b(z)$ and*

$$\|E_c\| \leq \|H(a^-)\|_{\mathcal{W}}\|H(b^+)\|_{\mathcal{W}} + \|a\|_{\mathcal{W}}\|E_b\|_{\mathcal{F}} + \|b\|_{\mathcal{W}}\|E_a\|_{\mathcal{F}} + \|E_a\|_{\mathcal{F}}\|E_b\|_{\mathcal{F}}.$$

Proof. Applying Theorem 2.7 yields $C = T(c) + E_c$, where

$$E_c := -H(a^-)H(b^+) + T(a)E_b + E_aT(b) + E_aE_b. \tag{2.17}$$

Therefore it is sufficient to prove that $\|E_c\|$ is finite. From Lemmas 2.2 and 2.4 it follows that both $\|H(a^-)H(b^+)\|_{\mathcal{F}}$ and $\|E_aE_b\|_{\mathcal{F}}$ are finite. It remains to show that $\|E_aT(b)\|_{\mathcal{F}}$ and $\|T(a)E_b\|_{\mathcal{F}}$ are finite. We prove this property only for the first since the boundedness of the other matrix norm follows by transposition, in fact, for any $F \in \mathcal{F}$ one has $\|F\|_{\mathcal{F}} = \|F^T\|_{\mathcal{F}}$ and $T(a)^T = T(\widehat{a})$ where $\widehat{a}(z) = a(z^{-1})$ and $\|a\|_{\mathcal{W}} = \|\widehat{a}\|_{\mathcal{W}}$.

Denote $H = T(a)E_b = (h_{ij})$ and $E_b = (e_{ij})$. We have $h_{ij} = \sum_{r\geq 1} a_{r-i}e_{rj}$ so that

$$\|H\|_{\mathcal{F}} = \sum_{i,j\in\mathbb{Z}_+} |h_{ij}| \leq \sum_{i,j\in\mathbb{Z}_+} \sum_{r\geq 1} |a_{r-i}e_{rj}|,$$

substituting $k = r - i$ yields

$$\|H\|_{\mathcal{F}} \leq \sum_{k\in\mathbb{Z}_+} |a_k| \sum_{j\geq 1} \sum_{i\geq -k+1} |e_{k+ij}|.$$

Since $\sum_{j\geq 1}\sum_{i\geq -k+1}|e_{k+ij}| = \sum_{j\geq 1}\sum_{i\geq 1}|e_{ij}| = \|E_b\|_{\mathcal{F}}$ for any k, we have

$$\|H\|_{\mathcal{F}} \leq \sum_{k\in\mathbb{Z}} |a_k|\|E_b\|_{\mathcal{F}} = \|a\|_{\mathcal{W}}\|E_b\|_{\mathcal{F}} < \infty.$$

Thus taking norms in (2.17) yields

$$\|E_c\| \leq \|H(a^-)\|_{\mathcal{F}}\|H(b^+)\|_{\mathcal{F}} + \|a\|_{\mathcal{W}}\|E_b\|_{\mathcal{F}} + \|E_a\|_{\mathcal{F}}\|b\|_{\mathcal{W}} + \|E_a\|_{\mathcal{F}}\|E_b\|_{\mathcal{F}}$$

which completes the proof. □

Observe that we may rewrite the inequality of the previous Theorem as

$$\|E_c\| \leq \|a'\|_{\mathcal{W}}\|b'\|_{\mathcal{W}} + \|a\|_{\mathcal{W}}\|E_b\|_{\mathcal{F}} + \|b\|_{\mathcal{W}}\|E_a\|_{\mathcal{F}} + \|E_a\|_{\mathcal{F}}\|E_b\|_{\mathcal{F}}. \tag{2.18}$$

Theorem 2.9. *The set $\mathcal{QT}_1$ endowed with the norm $\|T(a)+E\|_{\mathcal{QT}_1} = \|a\|_{\mathcal{W}_1} + \|E\|_{\mathcal{F}}$ is a Banach algebra.*

Proof. Theorem 2.8 ensures the closure of $\mathcal{QT}_1$ under multiplication. To prove the submultiplicative property of the norm we observe that for any $A, B \in \mathcal{QT}_1$, with $A = T(a) + E_a$ and $B = T(b) + E_b$, we have

$$\|ab\|_{\mathcal{W}_1} = \|ab\|_{\mathcal{W}} + \|a'b + ab'\|_{\mathcal{W}} \leq \|a\|_{\mathcal{W}}\|b\|_{\mathcal{W}} + \|a'\|_{\mathcal{W}}\|b\|_{\mathcal{W}} + \|a\|_{\mathcal{W}}\|b'\|_{\mathcal{W}}. \tag{2.19}$$

Since $\|AB\|_{\mathcal{QT}_1} = \|ab\|_{\mathcal{W}_1} + \|E_c\|_{\mathcal{F}}$, for $c(z) = a(z)b(z)$, and where E_c is defined as in Theorem 2.8, by applying (2.18) and (2.19) we obtain

$$\begin{aligned}
\|AB\|_{\mathcal{QT}_1} \leq & \|ab\|_{\mathcal{W}_1} + \|a'\|_{\mathcal{W}}\|b'\|_{\mathcal{W}} + \|a\|_{\mathcal{W}}\|E_b\|_{\mathcal{F}} + \|b\|_{\mathcal{W}}\|E_a\|_{\mathcal{F}} + \|E_a\|_{\mathcal{F}}\|E_b\|_{\mathcal{F}} \\
\leq & \|a\|_{\mathcal{W}}\|b\|_{\mathcal{W}} + \|a'\|_{\mathcal{W}}\|b\|_{\mathcal{W}} + \|a\|_{\mathcal{W}}\|b'\|_{\mathcal{W}} \\
& + \|a'\|_{\mathcal{W}}\|b'\|_{\mathcal{W}} + \|a\|_{\mathcal{W}}\|E_b\|_{\mathcal{F}} + \|b\|_{\mathcal{W}}\|E_a\|_{\mathcal{F}} + \|E_a\|_{\mathcal{F}}\|E_b\|_{\mathcal{F}} \\
= & \big(\|a\|_{\mathcal{W}} + \|a'\|_{\mathcal{W}}\big) + \big(\|b\|_{\mathcal{W}} + \|b'\|_{\mathcal{W}}\big) \\
& + \|a\|_{\mathcal{W}}\|E_b\|_{\mathcal{F}} + \|b\|_{\mathcal{W}}\|E_a\|_{\mathcal{F}} + \|E_a\|_{\mathcal{F}}\|E_b\|_{\mathcal{F}} \\
\leq & (\|a\|_{\mathcal{W}_1} + \|E_a\|_{\mathcal{F}}) + (\|b\|_{\mathcal{W}_1} + \|E_b\|_{\mathcal{F}}) \\
= & \|A\|_{\mathcal{QT}_1}\|B\|_{\mathcal{QT}_1}.
\end{aligned}$$

Concerning the completeness, observe that the isomorphism $\mathcal{QT}_1 \simeq \mathcal{W}_1 \oplus \mathcal{F}$ holds. Since both $\mathcal{W}_1$ and $\mathcal{F}$ are Banach spaces, the composition of the 1-norm of R^2 with the vector valued function $T(a) + E \mapsto (\|a\|_{\mathcal{W}_1}, \|E\|_{\mathcal{F}})$ makes $\mathcal{W}_1 \oplus \mathcal{F}$ a complete metric space. □

2.3.1 $\mathcal{QT}_1$ arithmetic

The properties that we have described in the previous sections imply that any finite computation which takes as input a set of $\mathcal{QT}_1$ operators and that performs additions, multiplications, inversions, and multiplications by a scalar, generates results that belong to $\mathcal{QT}_1$. If the computation can be carried out with no breakdown, say caused by singularity, then the output still belongs to $\mathcal{QT}_1$.

In order to manipulate $\mathcal{QT}_1$ operators effectively we have to provide a simple and effective way of representing them, up to an arbitrarily small error, by means of a finite number of parameters.

Given $A = T(a) + E_a$, an element in $\mathcal{QT}_1$, since the symbol $a(z)$ belongs to $\mathcal{W}_1$, and since the correction matrix E_a has entries with finite sum of their moduli, we

may write A through its truncated form $\widetilde{A}$. That is, for any $\epsilon > 0$ there exist integers n_-, n_+, k_-, k_+ such that

$$A = \widetilde{A} + \mathcal{E}_a, \quad \|\mathcal{E}_a\|_{\mathcal{QT}_1} \leq \epsilon, \quad \widetilde{A} = T(\widetilde{a}) + \widetilde{E}_a, \quad \widetilde{a}(z) = \sum_{i=-n_-}^{n_+} a_i z^i,$$

where $\widetilde{E}_a = (\widetilde{e}_{ij})$, is such that $\widetilde{e}_{ij} = e_{ij}$ for $i = 1, \ldots, k_-$ and $j = 1, \ldots, k_+$, while $\widetilde{e}_{ij} = 0$ elsewhere.

In this way, we can approximate any given $\mathcal{QT}_1$ operator A, to any desired precision, with a $\mathcal{QT}_1$ operator $\widetilde{A}$ where the Toeplitz part is banded and the correction $\widetilde{E}_a$ has a finite dimensional nonzero part.

Remark 2.4. In the context of our process the operators we are concerning of are already in a truncated form, since the Toeplitz part has only three bands and the correction has at most two nonzero elements, so they can be represented with $\epsilon = 0$.

From the computational point of view, it is convenient to express the matrix $\widetilde{E}_a$ by means of a factorization of the kind $\widetilde{E}_a = F_a G_a^T$, where matrices F_a and G_a have a number of columns given by the rank of $\widetilde{E}_a$ and infinitely many rows. In this way, in presence of low-rank corrections, the storage is reduced together with the computational cost for performing matrix arithmetic. This representation in product form can be obtained by means of SVD up to some error which can be controlled at run time and which can be included in $\mathcal{E}_a$.

In the following, we represent the truncation of a $\mathcal{QT}_1$ operator A with $\widetilde{E}_a = F_a G_a^T$ where F_a has f_a nonzero rows and k_a columns, G_a has g_a nonzero rows and k_a columns, and the error $\mathcal{E}_a$ has a sufficiently small norm. This way, $\widetilde{E}_a$ has f_a nonzero rows, g_a nonzero columns and rank at most k_a.

With this notation we may easily implement the operations of addition, multiplication and inversion of two $\mathcal{QT}_1$ operator $\widetilde{A}$ and $\widetilde{B}$, which are the truncated representations of two $\mathcal{QT}$ operator A and B, that is

$$A = \widetilde{A} + \mathcal{E}_a, \ \widetilde{A} = \mathrm{tr}(A) = T(\widetilde{a}) + \widetilde{E}_a, \quad B = \widetilde{B} + \mathcal{E}_b, \ \widetilde{B} = \mathrm{tr}(B) = T(\widetilde{b}) + \widetilde{E}_b.$$

Denoting by $\star$ any arithmetic operation, let us define $C = A \star B$, $\widehat{C} = \widetilde{A} \star \widetilde{B}$ and $\widetilde{C} = \mathrm{tr}(\widehat{C})$. Moreover we define the *total error* in the operation $\star$ as $\mathcal{E}_c^{tot} = C - \widetilde{C}$, the *local error* as $\mathcal{E}_c^{loc} = \widehat{C} - \widetilde{C}$ and the *inherent error* as $\mathcal{E}_c^{in} = C - \widehat{C}$ so that $\mathcal{E}_c^{tot} = \mathcal{E}_c^{in} + \mathcal{E}_c^{loc}$.

Addition. Let $\widetilde{a}(z)$ and $\widetilde{b}(z)$ be the Laurent polynomials of degrees $n_a^{\pm}$ and $n_b^{\pm}$ respectively, associated to the truncation of A and B and $\widehat{E}_a = F_a G_a^T$, $\widehat{E}_b = F_b G_b^T$. For the operator $C = A + B$ we have the representation

$$C = \widetilde{A} + \widetilde{B} + \mathcal{E}_a + \mathcal{E}_b,$$

from which we deduce that the inherent error is $\mathcal{E}_c^{in} = \mathcal{E}_a + \mathcal{E}_b$. On the other hand, concerning $\widehat{C} = \widetilde{A} + \widetilde{B}$ we have

$$\widehat{C} = T(\widetilde{a} + \widetilde{b}) + \widetilde{E}_a + \widetilde{E}_b,$$

where $\widetilde{a}(z) + \widetilde{b}(z)$ is a Laurent polynomial of degrees $n_c^- = \max(n_a^-, n_b^-)$ and $n_c^+ = \max(n_a^+, n_b^+)$, while

$$E_c = \widetilde{E}_a + \widetilde{E}_b = F_c G_c^T, \quad F_c = [F_a, F_b], \quad G_c = [G_a, G_b],$$

where $f_c = \max(f_a, f_b)$ and $g_c = \max(g_a, g_b)$ are the number of nonzero rows of F_c and G_c, respectively, and $k_c = k_a + k_b$ is the number of columns of F_c and G_c.

The Laurent polynomial $\widetilde{a}(z) + \widetilde{b}(z)$ can be truncated and replaced by a Laurent polynomial $\widetilde{c}(z)$ of possibly less degree. Also the value of k_c, can be reduced and the operators F_c, G_c can be compressed, by using a technique, explained in next sections, which guarantees a local error with norm bounded by a given ϵ. Denoting by $\widetilde{F}_c$ and $\widetilde{G}_c$ the operators obtained after compressing F_c and G_c, we have

$$\widetilde{C} = \mathrm{tr}(\widehat{C}) = T(\widetilde{c}) + \widetilde{E}_c + \mathcal{E}_c^{loc}, \quad \widetilde{E}_c = \widetilde{F}_c \widetilde{G}_c^T,$$

where $\mathcal{E}_c^{loc} = \widetilde{A} + \widetilde{B} - \mathrm{tr}(\widetilde{A} + \widetilde{B})$. This way we have

$$A + B = T(\widetilde{c}) + \widetilde{E}_c + \mathcal{E}_c^{loc} + \mathcal{E}_c^{in}.$$

Multiplication. For the product $C = AB$ we have the equation

$$AB = \widetilde{A}\widetilde{B} + \widetilde{A}\mathcal{E}_b + \mathcal{E}_a\widetilde{B} + \mathcal{E}_a\mathcal{E}_b$$

from which we deduce that the inherent error is $\mathcal{E}_c^{in} = \widetilde{A}\mathcal{E}_b + \mathcal{E}_a\widetilde{B} + \mathcal{E}_a\mathcal{E}_b$. Moreover we have

$$\begin{aligned}
\widehat{C} = \widetilde{A}\widetilde{B} &= T(\widetilde{a})T(\widetilde{b}) + T(\widetilde{a})\mathcal{E}_b + \mathcal{E}_a T(\widetilde{b}) + \mathcal{E}_a\mathcal{E}_b \\
&= T(\widetilde{a}\widetilde{b}) - H(\widetilde{a}^-)H(\widetilde{b}^+) + T(\widetilde{a})\mathcal{E}_b + \mathcal{E}_a T(\widetilde{b}) + \mathcal{E}_a\mathcal{E}_b \\
&= T(\widetilde{a}\widetilde{b}) + E_c,
\end{aligned}$$

where $E_c := -H(\widetilde{a}^-)H(\widetilde{b}^+) + T(\widetilde{a})\mathcal{E}_b + \mathcal{E}_a T(\widetilde{b}) + \mathcal{E}_a\mathcal{E}_b$. Since $\widetilde{a}^-(z)$ and $\widetilde{b}^+(z)$ are polynomials, the operators $H(\widetilde{a}^-)$ and $H(\widetilde{b}^+)$ have a finite number of nonzero entries. Therefore, we may factorize the product $H(\widetilde{a}^-)H(\widetilde{b}^+)$ in the form FG^T. Thus, the operator E_c can be written as $E_c = F_c G_c^T$ where

$$F_c = [F, T(\widetilde{a})F_b, F_b], \quad G_c = \left[G, G_b, T(\widetilde{b})^T G_a + G_b(F_b^T G_a)\right].$$

This provides the finite representation of the product $\widehat{C} = \widetilde{A}\widetilde{B}$ with $n_c^- = n_a^- + n_b^-$, $n_c^+ = n_a^+ + n_b^+$, $f_c = \max(f_b + n_a^-, f_a)$, $f_c = \max(n_b^+, g_a + n_b^-, g_b)$ and $k_c = k_a + k_b + n_b^+$.

Also in this case we may apply a compression technique for reducing the degree of the Laurent polynomial $\widetilde{a}(z)\widetilde{b}(z)$. We introduce a local error $\mathcal{E}_c^{loc} = \widetilde{A}\widetilde{B} - \mathrm{tr}(\widetilde{A}\widetilde{B})$, denoting by $\widetilde{c}(z)$ the truncation of the Laurent polynomial $\widetilde{a}(z)\widetilde{b}(z)$ and with $\widetilde{F}_c\widetilde{G}_c^T$ the compression of $F_c G_c^T$, we have

$$\widehat{C} = \widetilde{A}\widetilde{B} = T(\widetilde{c}) + \widetilde{F}_c\widetilde{G}_c^T + \mathcal{E}_c^{loc}.$$

This way we have

$$C = AB = T(\widetilde{c}) + \widetilde{F}_c\widetilde{G}_c^T + \mathcal{E}_c^{loc} + \mathcal{E}_c^{in}.$$

which expresses the result C of the multiplication in terms of the approximated value $\widehat{C} = T(\widehat{c}) + \widehat{E}_c$, the local error $\mathcal{E}_c^{loc}$ and the inherent error $\mathcal{E}_c^{in}$. The overall error is given by $\mathcal{E}_c = \mathcal{E}_c^{loc} + \mathcal{E}_c^{in}$.

Inversion. First, we consider the problem of inverting $A = T(a)$, that is the special case in which $E_a = 0$. For this, let us recall the following well known theorems, the first relates the invertibility of the operator $T(a)$ to the winding number of $a(z)$, that is, the (integer) number of times that the complex number $a(\cos\theta + i\sin\theta)$, where $i^2 = -1$, winds around the origin as θ moves from 0 to 2π; the second is about Wiener-Hopf factorization of $a(z)$.

Theorem 2.10. *Let $a(z)$ be a continuous function from $\partial\mathbb{B}$ in $\mathbb{C}$. Then the linear operator $T(a)$ is invertible if and only if the winding number of $a(z)$ is zero and $a(z)$ does not vanish on $\partial\mathbb{B}$.*

Theorem 2.11. *Let $a(z) \in \mathcal{W}$ be a function which does not vanish for $z \in \partial\mathbb{B}$ and such that its winding number is κ. Then $a(z)$ admits the Wiener-Hopf factorization*

$$a(z) = u(z)z^\kappa l(z),$$

where $u(z) = \sum_{i\geq 0} u_i z^i$, $l(z) = \sum_{i\geq 0} l_i z^{-i}$ belong to $\mathcal{W}$ and $u(z)$, $l(z^{-1})$ do not vanish in the closed unit disk. If $\kappa = 0$ the factorization is said canonical.

Assume that $a(z) \in \mathcal{W}_1$ does not vanishes on the unit circle and its winding number is zero, from theorem 2.11 we deduce the following operator factorization

$$T(a) = T(u)T(l),$$

where $T(l)$ is lower triangular and $T(u)$ is upper triangular. Since $u(z)$ and $l(z^{-1})$ do not vanish in the unit disk, the functions $u(z)$ and $l(z)$ have inverse in $\mathcal{W}_1$, such that $T(u)T(u^{-1}) = T(u^{-1})T(u) = I$, and $T(l)T(l^{-1}) = T(l^{-1})T(l) = I$, so we have

$$T(a)^{-1} = T(l)^{-1}T(u)^{-1} = T(l^{-1})T(u^{-1}),$$

and from what we have seen until now we get

$$T(a)^{-1} = T(a^{-1}) - H((l^{-1})^{-})H((u^{-1})^{+}) = T(a^{-1}) - H(l^{-1})H(u^{-1}) \in \mathcal{QT}_1. \quad (2.20)$$

Thus, a finite representation of A^{-1} is obtained by truncating the Laurent series of $\frac{1}{a(z)}$ to a Laurent polynomial and by approximating the operators $H((l^{-1})^{-})$ and $H((u^{-1})^{+})$ by means of operators having a finite number of nonzero entries, an infinite number of rows and the same finite number of columns. The latter operation can be achieved by truncating the power series $l^{-1}(z)$ and $u^{-1}(z)$ to polynomials and by numerically compressing the product of the Hankel operators obtained this way.

Now consider the more general case of the matrix $A = T(a) + F_a G_a^T$ which we assume already in its truncated form. Assume $T(a)$ invertible and write $A = T(a)(I + T(a)^{-1}F_a G_a^T)$. Denoting for simplicity $U = T(u)$, $L = T(l)$ we have

$$(T(a) + F_a G_a^T)^{-1} = T(a)^{-1} - L^{-1}(U^{-1}F_a)Y^{-1}(G_a^T L^{-1})U^{-1},$$

where $Y = I + G_a^T L^{-1} U^{-1} F_a$ is a finite matrix which is invertible if and only if A is invertible. This way, the algorithm for computing A^{-1} in its finite representation is given by the following steps:

1. compute the spectral factorization $a(z) = u(z)l(z)$;
2. compute the coefficients of the power series $\widetilde{u}(z) = \frac{1}{u(z)}$ and $\widetilde{l}(z) = \frac{1}{l(z)}$ so that $U^{-1} = T(\widetilde{u})$ and $L^{-1} = T(\widetilde{l})$;
3. represent the operator $H = L^{-1}U^{-1}$ as $T(c) + F_h G_h^T$ where $c(z) = \widetilde{l}\widetilde{u}$;
4. compute the products $G_1 = T(\widetilde{l})G_a$ and $F_1 = T(\widetilde{u})F_a$;
5. compute $Y = I + G_1^T F_1$, $F_2 = F_1 Y^{-1}$, $F_3 = T(\widetilde{l})F_2$ and $G_2 = T(\widetilde{u})G_1$;
6. output the coefficients of $c(z)$ and the operators $F_c = [F_h, F_3]$ and $G_c = [G_h, G_2]$.

Compression Given the matrix E in the form $E = FG^T$ where F and G are matrices of size $m \times k$ and $n \times k$, respectively, we aim to reduce the size k and to approximate E in the form $\widetilde{F}\widetilde{G}^T$ where $\widetilde{F}$ and $\widetilde{G}$ are matrices of size $m \times \widetilde{k}$ and $n \times \widetilde{k}$ with $\widetilde{k} < k$.

We use the following procedure. Compute the pivoted (rank-revealing) QR factorizations $F = Q_f R_f P_f$ and $G = Q_g R_g P_g$, where P_f and P_g are permutation matrices, Q_f and Q_g are orthogonal and R_f and R_g are upper triangular; remove the last negligible rows from the matrices R_f and R_g, remove the corresponding columns of Q_f and Q_g. In this way we obtain matrices $\widehat{R}_f$, $\widehat{R}_g$, $\widehat{Q}_f$, $\widehat{Q}_g$ such that, up

to within a small error, satisfy the equations $F = \widehat{R}_f \widehat{Q}_f P_f$, $F = \widehat{R}_g \widehat{Q}_g P_g$. Then, in the factorization $FG^T = F = \widehat{Q}_f (\widehat{R}_f P_f P_g^T \widehat{R}_g^T) \widehat{Q}_g^T$, compute the SVD of the matrix in the middle $\widehat{R}_f P_f P_g^T \widehat{R}_g^T = U \Sigma V^T$, and replace U, Σ and V with matrices $\widehat{U}$, $\widehat{\Sigma}$, $\widehat{V}$, obtained by removing the singular values σ_i and the corresponding singular vectors if $\sigma_i < \epsilon \sigma_1$, where ϵ is a given tolerance. In output, the matrices $\widetilde{F} = \widehat{Q}_f \widehat{U} \widehat{\Sigma}^{\frac{1}{2}}$ and $\widetilde{F} = \widehat{Q}_g \widehat{V} \widehat{\Sigma}^{\frac{1}{2}}$ are delivered.

All the operations cited above are fully explained in [6] and are implemented in a Matlab toolbox which is introduced in [7].

Chapter 3

Cyclic Reduction in a Banach Algebra

3.1 Cyclic Reduction in the Finite Case

The Cyclic Reduction algorithm is one of the most powerful methods to calculate the invariant vector of a QBD with finite size blocks, for a vast dissertation about the properties of this algorithm we refer to [8]. Despite its success into the Markov chain field, the original formulation of the Cyclic Reduction appears in the context of solving the Poisson equation over a rectangle. Discretizing this problem with finite-differences formulas leads to a block linear system of the form

$$\begin{bmatrix} A & C & & 0 \\ B & A & \ddots & \\ & \ddots & \ddots & C \\ 0 & & B & A \end{bmatrix} \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \vdots \\ \mathbf{u}_n \end{bmatrix} = \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \vdots \\ \mathbf{b}_n \end{bmatrix}$$

where $A, B, C \in \mathbb{R}^{m\times m}$ and $\mathbf{u}_i, \mathbf{b}_i \in \mathbb{R}^m$. Assuming that $n = 2^s - 1$, we apply an odd-even permutation to both block-columns and block-rows in the above system and get

$$\begin{bmatrix} A & & & 0 & C & & 0 \\ & \ddots & & & B & \ddots & \\ & & \ddots & & & \ddots & C \\ 0 & & & A & 0 & & B \\ B & C & & 0 & A & & 0 \\ & \ddots & \ddots & & & \ddots & \\ 0 & & B & C & 0 & & A \end{bmatrix} \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_3 \\ \vdots \\ \mathbf{u}_{2^s-1} \\ \mathbf{u}_2 \\ \mathbf{u}_4 \\ \vdots \\ \mathbf{u}_{2^s-2} \end{bmatrix} = \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_3 \\ \vdots \\ \mathbf{b}_{2^s-1} \\ \mathbf{b}_2 \\ \mathbf{b}_4 \\ \vdots \\ \mathbf{b}_{2^s-2} \end{bmatrix}.$$

Now, rewrite the above system as

$$\begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix} \begin{bmatrix} \mathbf{u}_{\text{odd}} \\ \mathbf{u}_{\text{even}} \end{bmatrix} = \begin{bmatrix} \mathbf{b}_{\text{odd}} \\ \mathbf{b}_{\text{even}} \end{bmatrix},$$

assume A nonsingular, eliminate the odd block components by means of block Gaussian elimination, that is compute the Schur complement of H_{11} and obtain the smaller system of block size $2^{s-1}-1$:

$$\left[H_{22} - H_{21}H_{11}^{-1}H_{12}\right]\mathbf{u}_{\text{even}} = \mathbf{b}^{(1)}, \quad \mathbf{b}^{(1)} = \mathbf{b}_{\text{even}} - H_{21}H_{11}^{-1}\mathbf{b}_{\text{odd}}.$$

Surprisingly, the Schur complement has the same structure as the original matrix and the above system takes the form

$$\begin{bmatrix} A^{(1)} & C^{(1)} & & 0 \\ B^{(1)} & A^{(1)} & \ddots & \\ & \ddots & \ddots & C^{(1)} \\ 0 & & B^{(1)} & A^{(1)} \end{bmatrix} \begin{bmatrix} \mathbf{u}_2 \\ \mathbf{u}_4 \\ \vdots \\ \mathbf{u}_{2^s-2} \end{bmatrix} = \begin{bmatrix} \mathbf{b}_1^{(1)} \\ \mathbf{b}_2^{(1)} \\ \vdots \\ \mathbf{b}_{2^{s-1}-1}^{(1)} \end{bmatrix}$$

where

$$\mathbf{b}_i^{(1)} = \mathbf{b}_{2i} - BA^{-1}\mathbf{b}_{2i-1} - CA^{-1}\mathbf{b}_{2i+1}, \quad i = 1, \ldots, 2^{s-1}-1$$

and

$$\begin{aligned} A^{(1)} &= A - BA^{-1}C - CA^{-1}B, \\ B^{(1)} &= -BA^{-1}B, \\ C^{(1)} &= -CA^{-1}C, \end{aligned}$$

while for the odd indexed block components one has

$$A\mathbf{u}_{2i-1} = \mathbf{b}_{2i-1} - B\mathbf{u}_{2i-2} - C\mathbf{u}_{2i+2}, \quad i = 1, \ldots, 2^{s-1},$$

where we set $\mathbf{u}_0 = \mathbf{u}_{n+1} = 0$.

This process, can be cyclically repeated and generates the sequence of systems of block size $2^{s-k}-1$:

$$\begin{bmatrix} A^{(k)} & C^{(k)} & & 0 \\ B^{(k)} & A^{(k)} & \ddots & \\ & \ddots & \ddots & C^{(k)} \\ 0 & & B^{(k)} & A^{(k)} \end{bmatrix} \begin{bmatrix} \mathbf{u}_{2^k} \\ \mathbf{u}_{2^{k+1}} \\ \vdots \\ \mathbf{u}_{2^s-2^k} \end{bmatrix} = \begin{bmatrix} \mathbf{b}_1^{(k)} \\ \mathbf{b}_2^{(k)} \\ \vdots \\ \mathbf{b}_{2^{s-k}-1}^{(k)} \end{bmatrix},$$

for $k = 0, 1, \ldots, s-1$. The block matrices are defined by

$$\begin{aligned} A^{(k+1)} &= A^{(k)} - B^{(k)} \left(A^{(k)}\right)^{-1} C^{(k)} - C^{(k)} \left(A^{(k)}\right)^{-1} B^{(k)}, \\ B^{(k+1)} &= -B^{(k)} \left(A^{(k)}\right)^{-1} B^{(k)}, \\ C^{(k+1)} &= -C^{(k)} \left(A^{(k)}\right)^{-1} C^{(k)}, \end{aligned}$$

while for the vectors the following recursion holds

$$\mathbf{b}_i^{(k+1)} = \mathbf{b}_{2i}^{(k)} - B^{(k)} \left(A^{(k)}\right)^{-1} \mathbf{b}_{2i-1}^{(k)} - C^{(k)} \left(A^{(k)}\right)^{-1} \mathbf{b}_{2i+1}^{(k)}, \quad i = 1, \ldots, 2^{s-1} - 1$$

for $k = 0, 1, \ldots, q-2$ and $A^{(0)} = A$, $B^{(0)} = B$, $C^{(0)} = C$, provided that $\det A^{(k)} \neq 0$ for any k. At the end of the process one recovers $\mathbf{u}_{2^{s-1}}$ by solving the following system of size m:

$$A^{(s-1)} \mathbf{u}_{2^{s-1}} = \mathbf{u}_1^{s-1},$$

then, back substitution, performed by solving the system

$$A^{(k)} \mathbf{u}_{(2i-1)2^k} = \mathbf{b}_{2i-1}^{(k)} - B^{(k)} \mathbf{u}_{(2i-2)2^k} - C^{(k)} \mathbf{u}_{(2i)2^k},$$

for $i = 1, 2, \ldots, 2^{s-k-1}$ and $k = s-2, \ldots, 0$ allows one to compute the remaining unknowns.

3.2 Factorization and quadratic matrix equations

As we have seen in the previous chapter, the context of a Banach algebra is the most natural space to handle with the operators which arise in our model. Because of this, in this section, we assume that the operators A_{-1}, A_0 and A_1 belong to a Banach algebra $\mathcal{B}$ with norm $\|\cdot\|$ and identity I.

In the finite case Cyclic Reduction is one of the most powerful algorithms known in literature to solve the quadratic matrix equations. Now we want to apply Cyclic Reduction iterations substituting all the matrices with Banach algebra elements, this idea leads to the following sequences

$$\begin{aligned} S^{(h)} &= (I - A_0^{(h)})^{-1}, \\ A_0^{(h+1)} &= A_0^{(h)} + A_1^{(h)} S^{(h)} A_{-1}^{(h)} + A_{-1}^{(h)} S^{(h)} A_1^{(h)}, \\ A_1^{(h+1)} &= A_1^{(h)} S^{(h)} A_1^{(h)}, \\ A_{-1}^{(h+1)} &= A_{-1}^{(h)} S^{(h)} A_{-1}^{(h)} \\ \widehat{A}^{(h+1)} &= \widehat{A}^{(h)} + A_{-1}^{(h)} S^{(h)} A_1^{(h)}, \\ \widetilde{A}^{(h+1)} &= \widetilde{A}^{(h)} + A_1^{(h)} S^{(h)} A_{-1}^{(h)}, \end{aligned}$$

for $h = 0, 1, \ldots$ and with $A_0^{(0)} = \widetilde{A}^{(0)} = \widehat{A}^{(0)} = A_0 \in \mathcal{B}$, $A_1^{(0)} = A_1 \in \mathcal{B}$ and $A_{-1}^{(0)} = A_{-1} \in \mathcal{B}$.

In the following of the chapter we report some of the most important results about Cyclic Reduction in a Banach algebra that can be found in [5], [6] and [8].

Let us introduce the Laurent operator polynomials

$$\varphi(z) = A_{-1}z^{-1} + A_0 - I + A_1 z, \quad \varphi^{(h)}(z) = A_{-1}^{(h)}z^{-1} + A_0^{(h)} - I + A_1^{(h)}z,$$

we have the following Lemma about their structure.

Lemma 3.1. *Assume that $\varphi(z)$ is invertible for $z \in \mathbb{A}(t^{-1}, t)$ for a given $t > 1$. Then $\psi(z) = \varphi(z)^{-1} = \sum_{j\in\mathbb{Z}} z^j H_j$ with $H_j \in \mathcal{B}$. Moreover for any $t^{-1} < s < t$ we have*

$$\|H_j\| \leq M(s)s^{-j}$$

with $M(s) := \max_{z\in\partial\mathbb{B}}\|\varphi(zs)\|$.

Proof. Let s such that $t^{-1} < s < t$. Since $\varphi(z)$ is invertible for $|z| = s$, then $\|\psi(z)\|$ depends continuously on z, so that the value $M(s)$ is well defined and finite. The coefficients of a Laurent series can be represented in the following integral form

$$H_j = \frac{1}{2\pi\mathbf{i}}\int_{|z|=s} z^{-(j+1)}\phi(z)dz,$$

where the integral is meant componentwise. Applying the norm on both sides yields

$$\|H_j\| \leq \frac{1}{2\pi}\int_{|z|=s} s^{-(j+1)}\|\phi(z)\|dz \leq M(s)s^{-j}.$$

□

There exist a special factorization of $\varphi(z)$ under certain condition on the operators R and G, as we state in the following Lemma whose proof can be found in [14].

Lemma 3.2. *The equations* (2.16) *and* (2.15) *have solutions respectively R and G with disjoint spectra if and only if there exists the following factorization*

$$\varphi(z) = (I - zR)W(I - z^{-1}G), \tag{3.1}$$

for some invertible $W \in \mathcal{B}$.

Theorem 3.1. *Assume that there exists $R, G \in \mathcal{B}$ with disjoint spectra which solve* (2.16) *and* (2.15) *respectively, or equivalently that*

$$\varphi(z) = (I - zR)W(I - z^{-1}G),$$

for some invertible $W \in \mathcal{B}$. Moreover, assume that $\Sigma_R, \Sigma_G \in t^{-1}\mathbb{B}$ for some $t > 1$.

Then $\varphi(z)$ is invertible for $z \in \mathbb{A}(t^{-1}, t)$, the operator

$$H_0 = \sum_{j=0}^{\infty} G^j W^{-1} R^j,$$

belongs to $\mathcal{B}$, and setting $H_i = H_0 R^i$ for $i > 0$ and $H_i = G^{-i} H_0$ for $i < 0$, it follows that

$$\psi(z) = \varphi(z)^{-1} = \sum_{i \in \mathbb{Z}} z^i H_i, \quad z \in \mathbb{A}(t^{-1}, t),$$

with $H_i \in \mathcal{B}$. Finally, if H_0 is invertible, then $G = H_{-1} H_0^{-1}$ and $R = H_0^{-1} H_1$.

Proof. Since $\Lambda_G, \Lambda_R \subset t^{-1}\mathbb{B}$, then $\delta(zG), \delta(zR) < 1$ for $z \in \mathbb{A}(t^{-1}, t)$, so that $I - zG$ and $I - z^{-1}R$ are invertible. Thus, from (3.1) we obtain

$$\psi(z) = (I - z^{-1}G)^{-1} W^{-1} (I - zR)^{-1} = \left(\sum_{j \geq 0} z^{-j} G^j \right) W^{-1} \left(\sum_{j \geq 0} z^j R^j \right).$$

This shows that $\varphi(z)$ is invertible for any $z \in \mathbb{A}(t^{-1}, t)$. In view of Lemma 3.1 we can write $\psi = \sum_{i \in \mathbb{Z}} z^i H_i$. By equating the terms in the same power of z in these two expressions of $\psi(z)$, we find that the coefficients H_j satisfy the following equations

$$H_0 = \sum_{j \geq 0} G^j W^{-1} R^j, \quad H_i = \begin{cases} H_0 R^i & i > 0, \\ G^{-i} H_0 & i < 0. \end{cases}$$

It remains to prove that $H_0^{(n)} = \sum_{j=0}^{n} G^j W^{-1} R^j$ forms a Cauchy sequence, so that there exists $\lim_{n \to \infty} H_0^{(n)} = H_0 \in \mathcal{B}$. In order to do this, for $n > m$ we consider

$$H_0^{(n)} - H_0^{(m)} = \sum_{j=m+1}^{n} G^j W^{-1} R^j = G^{m+1} \left(\sum_{j=0}^{n-m-1} G^j W^{-1} R^j \right) R^{m+1}.$$

by taking the norm on both sides we get

$$\|H_0^{(n)} - H_0^{(m)}\| \leq \|G^{m+1}\| \|S\| \|R^{m+1}\|, \quad S = \sum_{j=0}^{n-m-1} G^j W^{-1} R^j.$$

From Remark (2.3) for any $\epsilon > 0$ such that $\lambda_G := \delta(G) + \epsilon < 1$ and $\lambda_R := \delta(R) + \epsilon < 1$, there exists $N > 0$ such that for $k \geq N$ we have

$$\|G^k\| \leq \|G^r\| \|G^N\|^q \leq \|G^r\| \lambda_G^q,$$
$$\|R^k\| \leq \|R^r\| \|R^N\|^q \leq \|R^r\| \lambda_R^q,$$

where q and r are quotient and remainder of the division of k by N, that is $k = Nq+r$. Thus,

$$\|H_0^{(n)} - H_0^{(m)}\| \leq \|G^r\|\|S\|\|R^r\|\lambda_G^q\lambda_R^q,$$

and since $\lambda_G, \lambda_R < 1$, if $\|S\|$ is bounded from above by a constant independent of n and m, it follows that $H_0^{(n)}$ is a Cauchy sequence. In order to prove the boundedness of $\|S\|$, it is sufficient to consider the division of $n-m+1$ by N, that is $n-m+1 = \widehat{q}N + \widehat{r}$, so that we have

$$S = \sum_{k=0}^{\widehat{q}-1} G^{kN}TR^{kN} + G^{\widehat{q}N}T_{\widehat{r}}R^{\widehat{q}N}, \quad T = \sum_{j=0}^{N-1} G^jW^{-1}R^j, \quad T_{\widehat{r}} = \sum_{j=0}^{\widehat{r}} G^jW^{-1}R^j.$$

From this we obtain the bound

$$\|S\| \leq \eta \sum_{k=0}^{\widehat{q}} \lambda_G^k\lambda_R^k,$$

where $\eta = \sum_{j=0}^{N-1}\|G^j\|\|W^{-1}\|\|R^j\|$ is a constant independent of n and m. This completes the proof. □

3.3 Convergence of the Cyclic Reduction

Cyclic Reduction algorithm can be expressed in functional form by means of the polynomials $\varphi^{(h)}(z)$, indeed let us consider the quantity $\varphi^{(h+1)}(z^2)$, by using Cyclic Reduction iteration formulas we can turn it into the following form

$$\begin{aligned}
\varphi^{(h+1)}(z^2) &= z^{-2}A_{-1}^{(h+1)} + A_0^{(h+1)} - I + z^2A_1^{(h+1)} \\
&= z^{-2}A_{-1}^{(h)}S^{(h)}A_{-1}^{(h)} + A_0^{(h)} + A_1^{(h)}S^{(h)}A_{-1}^{(h)} + A_{-1}^{(h)}S^{(h)}A_1^{(h)} - I \\
&+ z^2A_1^{(h)}S^{(h)}A_1^{(h)} \\
&= \left[z^{-1}A_{-1}^{(h)} + A_0^{(h)} - I + zA_1^{(h)}\right] S^{(h)} \left[z^{-1}A_{-1}^{(h+1)} - A_0^{(h+1)} + I + zA_1^{(h+1)}\right] \\
&= -\varphi^{(h)}(z)S^{(h)}\varphi^{(h)}(-z). \qquad (3.2)
\end{aligned}$$

Moreover, let us consider the sequence $\{\psi^{(h)}(z)\}$ recursively defined by

$$\begin{cases} \psi^{(0)}(z) & = \psi(z), \\ \psi^{(h+1)}(z) & = \frac{1}{2}(\psi^{(h)}(z) + \psi^{(h)}(-z)), \end{cases}$$

we can observe that

$$\psi^{(h)}(z) = \sum_{j\in\mathbb{Z}} z^jH_{j2^h}.$$

Indeed, the function $\psi^{(h)}(z)$ is defined on $\mathbb{A}(t^{-1},t)$, but we can prove that analyticity domain is much wider.

Theorem 3.2. *The operator function $\psi^{(h)}(z)$ is analytic in the annulus $\mathbb{A}(t^{-2^h},t^{2^h})$. Moreover, if H_0 is invertible, then the sequence $\varphi^{(h)}(z)$ converges to H_0^{-1} uniformly over all the compact sets $K\subset\mathbb{A}(t^{-2^h},t^{2^h})$ and there exists $h_0>0$ such that $\psi^{(h)}(z)$ is invertible for $h\geq h_0$.*

Proof. From Lemma 3.1, we have that $\|H_{j2^h}\|\leq M(s)(t^{-1}+\epsilon)^{j2^h}$ for $j>0$ and $\|H_{j2^h}\|\leq M(s)(t-\epsilon)^{j2^h}$ for $j<0$. Since ϵ is arbitrary, it follows that $\psi^{(h)}(z)$ is analytic in $\mathbb{A}(t^{-2^h},t^{2^h})$. Let K be a compact set in $\mathbb{A}(t^{-1},t)$, then

$$\sup_{z\in K}|\psi^{(h)}(z)-H_0|=\sup_{z\in K}|\sum_{j\neq 0}z^{j2^h}H_{j2^h}|,$$

since $z\in K\subset\mathbb{A}(t^{-1},t)$, there exists $\delta>0$ such that $t^{-1}+\delta<|z|<t-\delta$ so that

$$\sup_{z\in K}|\psi^{(h)}(z)-H_0|\leq\sum_{j>0}(t-\delta)^{j2^h}\|H_{j2^h}\|+\sum_{j<0}(t^{-1}-\delta)^{j2^h}\|H_{j2^h}\|.$$

By choosing $\epsilon<\delta$, using Lemma 3.1, we obtain that $\sup_{z\in K}|\psi^{(h)}(z)-H_0|$ converges to 0, which means that the sequence $\psi^{(h)}$ uniformly converges to H_0 over K. □

Notice that we can rewrite (3.2) as

$$\begin{aligned}\varphi^{(h+1)}(z^2)&=-\varphi^{(h)}(z)\left(\frac{\varphi^{(h)}(z)+\varphi^{(h)}(-z)}{2}\right)^{-1}\varphi^{(h)}(-z)\\&=-\left(\varphi^{(h)}(z)^{-1}\right)^{-1}\left(\frac{\varphi^{(h)}(z)+\varphi^{(h)}(-z)}{2}\right)^{-1}\left(\varphi^{(h)}(-z)^{-1}\right)^{-1}\\&=-\left(\varphi^{(h)}(-z)^{-1}\frac{\varphi^{(h)}(z)+\varphi^{(h)}(-z)}{2}\varphi^{(h)}(z)^{-1}\right)^{-1}\\&=\left(\frac{\varphi^{(h)}(z)^{-1}+\varphi^{(h)}(-z)^{-1}}{2}\right)^{-1}.\end{aligned}$$

This equation is the basis to prove the following

Theorem 3.3. *Let $\varphi(z)$ be invertible for $z\in\mathbb{A}(t^{-1},t)$. If $I-A_0^{(i)}$ is invertible for $i=0,\dots,h-1$, then $\varphi^{(i)}(z)$ is well defined and invertible for $z\in\mathbb{A}(t^{-2^i},t^{2^i})$ and $i=0,\dots,h$. Moreover, if $\psi^{(i)}$ is invertible for $z\in\mathbb{A}(t^{-2^i},t^{2^i})$ and $i=0,\dots,h$, then $\varphi^{(i)}(z)$ exists for $i=0,\dots,h$. In both cases it holds*

$$\varphi^{(i)}(z)^{-1}=\psi^{(i)}(z),\quad i=0,\dots,h. \tag{3.3}$$

Proof. Since $I - A_0^{(i)}$ is invertible for $i = 0, \dots, h-1$, then $\varphi^{(i)}(z)$ is well defined for $i = 0, \dots, h$ in view of (3.2). Moreover, from the same equation we observe that $\varphi^{(i)}(z^{2^i})$ is invertible if and only if $\varphi(z)$ is invertible, so that $\varphi^{(i)}(z)$ is invertible for $z \in \mathbb{A}(t^{-2^i}, t^{2^i})$.

Finally, by definition we have $\psi^{(0)}(z) = \varphi(z)^{-1} = \varphi^{(0)}(z)^{-1}$, if we suppose that (3.3) is true for a certain $h > 0$, then, using (3.2), by inductive hypothesis we have

$$\varphi^{(h+1)}(z^2) = \left(\frac{\psi^{(h)}(z) + \psi^{(h)}(-z)}{2} \right)^{-1} = \psi^{(h+1)}(z^2)^{-1},$$

which completes the proof. □

Now we are ready to prove the main convergence results of the Cyclic Reduction iterations, which are stated in the following theorems.

Theorem 3.4. *Let us assume that H_0 is invertible and $\varphi(z)$ is invertible for $z \in \mathbb{A}(t^{-1}, t)$. If the Cyclic Reduction algorithm can be carried out without breakdown, then $A_i^{(h)}$ are Cauchy sequences for $i = -1, 0, 1$ and*

$$\lim_{h\to\infty} A_{-1}^{(h)} = \lim_{h\to\infty} A_1^{(h)} = 0, \quad \lim_{h\to\infty} A_0^{(h)} = I + H_0^{-1}.$$

Moreover, for any $1 < s < t$ there exists $\gamma > 0$ such that

$$\|A_{-1}^{(h)}\| \le \gamma s^{-2^h}, \quad \|A_1^{(h)}\| \le \gamma s^{-2^h}, \quad \|A_0^{(h)} - I - H_0^{-1}\| \le \gamma s^{-2^{h+1}}.$$

Proof. Equating the coefficients of the same degree in z in equation

$$\psi^{(h)}(z) \left(z^{-1} A_{-1}^{(h)} + A_0^{(h)} - I + z A_1^{(h)} \right) = I,$$

yields

$$\begin{cases} H_0 A_{-1}^{(h)} + H_{-2^h}(A_0^{(h)} - I) + H_{-2^{h+1}} A_1^{(h)} & = 0, \\ H_{2^h} A_{-1}^{(h)} + H_0(A_0^{(h)} - I) + H_{-2^h} A_1^{(h)} & = I, \\ H_{2^{h+1}} A_{-1}^{(h)} + H_{2^h}(A_0^{(h)} - I) + H_0 A_1^{(h)} & = 0, \end{cases}$$

whence, multiplying all the equations on the left by H_0^{-1} and adding the quantity $-H_0^{-1} H_{-2^h} H_0^{-1}$ to each side of the first and the quantity $-H_0^{-1} H_{2^h} H_0^{-1}$ to each side of the third equation, we obtain

$$\begin{bmatrix} I & H_0^{-1}H_{-2^h} & H_0^{-1}H_{-2^{h+1}} \\ H_0^{-1}H_{2^h} & I & H_0^{-1}H_{-2^h} \\ H_0^{-1}H_{2^{h+1}} & H_0^{-1}H_{2^h} & I \end{bmatrix} \begin{bmatrix} A_{-1}^{(h)} \\ A_0^{(h)} - I - H_0^{-1} \\ A_1^{(h)} \end{bmatrix} = \begin{bmatrix} H_0^{-1}H_{2^h}H_0^{-1} \\ 0 \\ H_0^{-1}H_{2^h}H_0^{-1} \end{bmatrix}$$

Since the inverse of the above matrix can be written as

$$\begin{bmatrix} I & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{bmatrix} + \sum_{i\geq 0}(-1)^i \begin{bmatrix} 0 & H_0^{-1}H_{-2^h} & H_0^{-1}H_{-2^{h+1}} \\ H_0^{-1}H_{2^h} & 0 & H_0^{-1}H_{-2^h} \\ H_0^{-1}H_{2^{h+1}} & H_0^{-1}H_{2^h} & 0 \end{bmatrix}^i,$$

it follows that

$$\begin{bmatrix} A_{-1}^{(h)} \\ A_0^{(h)} - I - H_0^{-1} \\ A_1^{(h)} \end{bmatrix} = \begin{bmatrix} -H_0^{-1}H_{-2^h}H_0^{-1} \\ H_0^{-1}H_{-2^h}H_0^{-1}H_{2^h}H_0^{-1} + H_0^{-1}H_{2^h}H_0^{-1}H_{-2^h}H_0^{-1} \\ -H_0^{-1}H_{2^h}H_0^{-1} \end{bmatrix}.$$

By taking the norm, we get

$$\begin{cases} \|A_{-1}^{(h)}\| \leq \|H_0^{-1}\|^2\|H_{-2^h}\|, \\ \|A_0^{(h)} - I - H_0^{-1}\| \leq 2\|H_0^{-1}\|^3\|H_{-2^h}\|\|H_{2^h}\|, \\ \|A_1^{(h)}\| \leq \|H_0^{-1}\|^2\|H_{2^h}\|, \end{cases}$$

this completes the proof in view of Lemma 3.1. □

Concerning convergence to the operators G and R, we have the following result.

Theorem 3.5. *Assume that the hypotheses of Theorem 3.4 hold and that there exists $R, G \in \mathcal{B}$ with $\Lambda_R, \Lambda_G \subset t^{-1}\mathbb{B}$ and $t > 1$, such that (3.1) holds. If the sequences $(I - \widehat{A}^{(h)})^{-1}$ and $(I - \widetilde{A}^{(h)})^{-1}$ are uniformly bounded in norm, then*

$$G = \lim_{h\to\infty}(I - \widehat{A}^{(h)})^{-1}A_{-1}, \quad R = \lim_{h\to\infty} A_1(I - \widetilde{A}^{(h)})^{-1}.$$

Moreover, for any $1 < s < t$ there exists $\gamma > 0$ such that

$$\|G - (I - \widehat{A}^{(h)})^{-1}A_{-1}\| \leq \gamma s^{-2^{h+1}}, \quad \|R - A_1(I - \widetilde{A}^{(h)})^{-1}\| \leq \gamma s^{-2^{h+1}}.$$

Proof. Since G satisfies (2.8), then the following system of equations hold

$$\begin{bmatrix} I - A_0 & -A_1 & 0 & & \\ -A_{-1} & I - A_0 & -A_1 & 0 & \\ 0 & \ddots & \ddots & \ddots & \ddots \end{bmatrix} \begin{bmatrix} G \\ G^2 \\ \vdots \end{bmatrix} = \begin{bmatrix} A_{-1} \\ 0 \\ \vdots \end{bmatrix},$$

by applying to it the even-odd permutation, eliminating the odd block components by means of block Gaussian elimination and iterating the process in very similar way as the finite case, we obtain the new system

$$\begin{bmatrix} I - \widehat{A}^{(h)} & -A_1^{(h)} & 0 & & \\ -A_{-1}^{(h)} & I - A_0^{(h)} & -A_1^{(h)} & 0 & \\ 0 & \ddots & \ddots & \ddots & \ddots \end{bmatrix} \begin{bmatrix} G \\ G^{2^h+1} \\ \vdots \end{bmatrix} = \begin{bmatrix} A_{-1} \\ 0 \\ \vdots \end{bmatrix}.$$

Let us consider the first equation of the above system, that is

$$-A_{-1} + (I - \widehat{A}^{(h)})G = A_1^{(h)} G^{2^h+1},$$

by taking the norm, we get

$$\|G - (I - \widehat{A}^{(h)})^{-1} A_{-1}\| \leq \|(I - \widehat{A}^{(h)})^{-1}\| \|A_1^{(h)} G^{2^h+1}\|.$$

Since $\delta(G) < 1$ and $\lim_{h\to\infty} \|A_1^{(h)}\| = 0$, then $\lim_{h\to\infty} \|A_1^{(h)} G^{2^h+1}\| = 0$. Moreover, by hypothesis, we have that $(I - \widehat{A}^{(h)})^{-1}$ is uniformly bounded so that the sequence converges to G. The bound on the speed of convergence follows directly from Theorem 3.4. A similar argument applies for R. □

3.4 Cyclic Reduction in $\mathcal{QT}_1$

We are now able to consider the case we are interested, that is the one with the operators A_i, with $i = -1, 0, 1$, $\widetilde{A}_0$ and $\widetilde{A}_1$ that all belong to the Banach algebra $\mathcal{QT}_1$. Since it is an algebra, all the operators generated by Cyclic Reduction belong to $\mathcal{QT}_1$. Moreover, the Toeplitz part of these matrices have associated symbols $a_{-1}^{(h)}(z)$, $a_0^{(h)}(z)$, $a_1^{(h)}(z)$, $\widetilde{a}^{(h)}(z)$ and $\widehat{a}^{(h)}(z)$ which satisfy the same recurrence equations as their operator counterpart. More precisely we have the scalar functional relations

$$\begin{aligned}
a_0^{(h+1)}(z) &= a_0^{(h)}(z) + 2\frac{a_1^{(h)}(z) a_{-1}^{(h)}(z)}{(1 - a_0^{(h)}(z))}, \\
a_1^{(h+1)}(z) &= \frac{a_1^{(h)}(z)^2}{(1 - a_0^{(h)}(z))}, \\
a_{-1}^{(h+1)}(z) &= \frac{a_{-1}^{(h)}(z)^2}{(1 - a_0^{(h)}(z))}, \\
\widetilde{a}^{(h+1)}(z) &= \widetilde{a}^{(h)}(z) + \frac{a_1^{(h)}(z) a_{-1}^{(h)}(z)}{(1 - a_0^{(h)}(z))},
\end{aligned}$$

where $a_i^{(0)}(z) = a_i(z)$, for $i = -1, 0, 1$ and $\widetilde{a}^{(0)}(z) = a_0(z)$. Observe that since all the quantities are scalar functions, they commute so that $\widehat{a}^{(h)}(z)$ coincides with $\widetilde{a}^{(h)}(z)$. Moreover, it is easy to verify that $\widetilde{a}(z) = \frac{1}{2}\left(a_0^{(h)}(z) + a_0^{(0)}(z)\right)$ for any $h \geq 0$.

From this functional point of view the Cyclic Reduction algorithm reduces to the process known in literature as *Graeffe iteration*. About this, it is known (see [8],[3],[18] for more details about the subject) that if, for a given $z \in \partial\mathbb{B}$, the polynomial

$$p_z(x) := a_1(z)x^2 + (a_0(z) - 1)x + a_{-1}(z)$$

has one root $g(z)$ inside the unit disk and one root $r(z)^{-1}$ outside, then the sequences $a_{-1}(z)(1-\widetilde{a}^{(h)}(z))^{-1}$ and $a_1(z)(1-\widetilde{a}^{(h)}(z))^{-1}$ tend to $g(z)$ and $r(z)$. Thus, for what we have proved, the functions $g(z)$ and $r(z)$ correspond to the symbols of the operators G and R which are the minimal nonnegative solutions of the quadratic equations.

About the roots of the polynomial $p_z(x)$ we have the following

Theorem 3.6. *Let $a_i(z) = a_{i,-1}z^{-1} + a_{i,0} + a_{i,1}z$, for $i = -1, 1$, and $a_0(z) = a_{0,-1}z^{-1} + (a_{0,0} - 1) + a_{0,1}z$ be such that the sum of all their coefficients is equal to zero, $a_{0,0} - 1 < 0$, and all their other coefficients are nonnegative. If*

1. *$a_{-1,0} > 0$ or $a_{1,0} > 0$,*
2. *$a_{i,j} \neq 0$ for at least a pair (i, j), with $j \neq 0$,*

then for any $z \in \partial\mathbb{B}$, the polynomial $p_z(x)$ has a root of modulus less than 1 and a root of modulus larger than 1.

Proof. Without loss of generality we may assume that the coefficients of $a_i(z)$ belong to the interval $[-1, 1]$. If not, we may scale equations (2.8) and (2.9) by a suitable constant and reduce it to this case.

As a first step we show that there are no roots of modulus 1. Assume by contradiction that x is a root of modulus 1. Obviously, we have $p_z(x) = 0$ and $p_z(x) + x = x$. Observe that, if $z \in \partial\mathbb{B}$, the left hand-side of the previous equation is a convex combination of the points in the discrete set

$$\mathcal{C}_{x,z} := \left\{x^i z^j \; i = 0, 1, 2 \; j = -1, 0, 1\right\} \subset \partial\mathbb{B}.$$

If $z \neq 1$, condition 1. and the fact that $0 \leq a_{0,0} < 1$ ensure that the convex combination involves at least two different points of the unit circle, either x and 1 or x and x^2. Therefore, this convex combination $p_z(x)+x$ is equal to a point which belongs to the interior of the unit disc. This contradicts the fact that $|p_z(x) + x| = |x| = 1$. This argument excludes roots on $\partial\mathbb{B}$ for $z \in \partial\mathbb{B}$, $z \neq 1$.

We conclude by showing that there is exactly one root of modulus less than 1. In order to prove this, we first show that $|a_0(z)| > |a_{-1}(z) + a_1(z)|$ holds for any $z \in \partial\mathbb{B}$, $z \neq 1$. Therefore, by applying the Rouché Theorem one finds that the functions $f(x) = a_0(z)x$ and $p_z(x)$ have the same number of zeros in the open unit disc. To prove the inequality $|a_0(z)| > |a_{-1}(z) + a_1(z)|$ we observe that

$$\begin{aligned}
|a_0(z)| &\geq |a_{0,0} - 1| - |a_{0,-1}z^{-1}| - |a_{0,1}z| \\
&= -(a_{0,0} - 1) - a_{0,-1} - a_{0,1} \\
&= a_{-1,-1} + a_{-1,0} + a_{-1,1} + a_{1,-1} + a_{1,0} + a_{1,1} \\
&\geq |a_{-1,-1}z^{-1} + a_{-1,0} + a_{-1,1}z + a_{1,-1}z^{-1} + a_{1,0} + a_{1,1}z|,
\end{aligned}$$

where at least one of the two above inequalities is strict because of condition 2. □

Under the hypotheses of Theorem 3.4, the operators G and R also belong to $\mathfrak{QT}_1$, that is $G = T(g) + E_g$ and $R = T(r) + E_r$, with $g(z), r(z) \in \mathcal{W}_1$ and $E_g, E_r \in \mathcal{E}$; this doesn't always happen, as we can see in the following

Example 3.1. Let Z be the down-shift operator having ones in the lower diagonal and zeros elsewhere, let $e_1 = [1, 0, 0, \dots]^T$ and define $A_{-1} = e_1 e_1^T$, $A_0 = \frac{1}{2}Z$, $A_1 = \frac{1}{2}\left(I - e_1 e_1^T\right)$. Observe that the matrix $A_{-1} + A_0 + A_1$ is stochastic, $A_i \in \mathfrak{QT}_1$, for $i = -1, 0, 1$ and that the quadratic equations have minimal nonnegative solutions

$$R = \frac{1}{2}\begin{bmatrix} 0 & 0 \\ \mathbf{1} & \left(I - \frac{1}{2}Z\right) \end{bmatrix}, \quad G = \mathbf{1}e_1^T,$$

on the other hand G and R do not belong to $\mathfrak{QT}_1$ since their corrections to the Toeplitz part are neither in $\mathcal{F}$ nor have a bounded 2-norm.

Now we present a necessary condition that has to be satisfied in order to guarantee that the operator G belongs to $\mathfrak{QT}_1$.

Proposition 3.1. *Under the assumptions of Theorem 3.4, let* $\varphi(z) = z^{-1}A_{-1} + (A_0 - I) + zA_1$ *with* $A_i \in \mathfrak{QT}_1$. *Let* $a_i(z)$ *be the symbols associated with the blocks* A_i, *let* $g(z)$ *be the minimal nonnegative (coefficient-wise) Laurent series such that*

$$a_{-1}(z) + a_0(z)g(z) + a_1(z)g(z)^2 = g(z), \quad g(z) = \sum_{i \in \mathbb{Z}} g_i z^i.$$

If the minimal nonnegative solution G *of* (2.8) *belongs to* $\mathfrak{QT}_1$, *then* $g(1) = 1$.

Proof. Assume that $G = T(g) + E_g \in \mathfrak{QT}_1$, where $T(g)$ is the Toeplitz part and $E_g \in \mathcal{F}$. Since G verifies (2.8), then the symbol of $T(g)$ needs to be $g(z)$.

Since we assumed that the process is positive recurrent, then $G\mathbf{1} = \mathbf{1}$ and we have $E_g\mathbf{1} \geq \mathbf{1} - T(g)\mathbf{1} \geq \epsilon\mathbf{1}$, with $\epsilon = 1 - g(1)$. If $\epsilon > 0$ then every row of E_g has sum of moduli at least ϵ, and therefore $E_g \notin \mathcal{F}$, which leads to a contradiction. □

In the previous example, the operator G it is such that its symbol satisfies $g(1) = 0$, therefore, for the conditions we have presented, it can't belong to $\mathfrak{QT}_1$.

3.4.1 Computation of π

We now discuss the practical computation of the invariant vector π, that represents the steady state vector of the process. As we have seen, we have

$$\pi_0 = \pi_0(B_0 + B_1 G), \quad \pi_1 = \pi_0 B_1 A_1^{-1} R, \quad \pi_n = \pi_{n-1}R = \pi_1 R^{n-1}.$$

This reduces the problem of computing π to the computation of R and π_0. Notice that both π and π_0 are infinite vectors. In particular, $\pi_0 \in \ell^1(\mathbb{N})$, so that, for any

$\epsilon > 0$, there exists an index after which all its entries are smaller than ϵ in magnitude, we assume that ϵ has been fixed once and for all, and we are only interested in computing the components of π of magnitude larger than ϵ.

If the solutions G and R are in $\mathcal{QT}_1$, the infinite vectors π_n can be easily computed. The idea is to apply a power method on the operator $M := B_0 + B_1 G \in \mathcal{QT}_1$ that is we consider the following iteration of infinite raw vectors

$$\begin{cases} x^{(0)} = \mathbf{e}_1, \\ x^{(k+1)} = \frac{x^{(k)} M}{\|x^{(k)} M\|} \; k = 0, 1, 2, \ldots. \end{cases}$$

The iterations stop when the norm of the difference of two consecutive vectors become smaller than a fixed tolerance. The starting vector $x^{(0)}$ can also be set equal to any other vector with norm equal to 1.

An alternative method. Using a suitable reblocking, one can interpret the matrix M as numerically block tridiagonal and Toeplitz, with the only exception of the first block row. More precisely, by choosing sufficiently large blocks $\widehat{M}_i$, M_i, we can rephrase the problem as follows:

$$\pi_0 M = \begin{bmatrix} \pi_0^{(0)} \pi_0^{(1)} \ldots \end{bmatrix} \begin{bmatrix} \widehat{M}_0 & \widehat{M}_1 & & & \\ M_{-1} & M_0 & M_1 & & \\ & M_{-1} & M_0 & M_1 & \\ & & \ddots & \ddots & \ddots \end{bmatrix} = \begin{bmatrix} \pi_0^{(0)} \pi_0^{(1)} \ldots \end{bmatrix} = \pi.$$

The size m of the blocks M_i is chosen as $m = \max\{b_l, b_u\}$ where b_l, b_u are the lower and upper bandwidth of M after the truncation with the relative threshold ϵ. In particular, the matrix M represents the transition matrix of a QBD with finite dimensional blocks. Since the original process is positive recurrent, then the process associated with the matrix M can be seen as a restricted version of the original one, that is still positive recurrent. Therefore, solving the previous system consists in computing the steady state vector of a QBD, for which $\pi_0^{(0)}(\widehat{M}_0 + \widehat{M}_1 G_M) = \pi_0^{(0)}$ and $\pi_0^{(k)} = \pi_0^{(k-1)} R_M$, where R_M and G_M are the minimal nonnegative solutions of the matrix equations:

$$R_M^2 M_{-1} + R_M M_0 + M_1 = R_M, \quad M_{-1} + M_0 G_M + M_1 G_M^2 = G_M.$$

Since the spectral radius of R_M is smaller than 1, one can give explicit estimates of the number of non negligible components $\pi_0^{(k)}$ with respect to ϵ. In our numerical experiments we stop when $\|\pi_0^{(k)}\|_\infty < \epsilon \|\pi_0^{(0)}\|_\infty$. Matrices R_M and G_M can be computed by applying Cyclic Reduction and the finite dimensional vector $\pi_0^{(0)}$ can be computed by applying a standard method for approximating the Perron vector of a nonnegative matrix. Once π_0 is computed, the other entries can be recovered by right multiplication by R.

Chapter 4

Compensation approach

In this chapter we consider our problem modeled as a continuous time Markov process (double QBD process) on the pairs $\mathbf{x} = (m, n) \in S = \mathbb{Z}_+^2$ with the transition rates shown in Figure 1.1. We make the assumption that the Markov process is irreducible and we impose the following equations about the transition rates in order to avoid some pathological cases:

- Non-zero rate to the South: $q_{-1-1} + q_{0-1} + q_{1-1} > 0$.
- Non-zero rate to the West: $q_{-1-1} + q_{-10} + q_{-11} > 0$.
- Non-zero reflecting rate for the horizontal axis: $h_{-11} + h_{01} + h_{11} > 0$.
- Non-zero reflecting rate for the vertical axis: $v_{11} + v_{10} + v_{1-1} > 0$.
- Non-zero reflecting rate out of the origin: $r_{01} + r_{11} + r_{10} > 0$.

The equilibrium equations for $\pi_{m,n}$ can be found by equating for each state the rate into and the rate out of that state. These equations are formulated below, here for the special cases $0 \leq m, n \leq 1$,

$$
\begin{aligned}
q\pi_{1,1} &= \sum_{s=-1}^{0} \sum_{t=-1}^{0} q_{st}\pi_{1-s,1-t} + \sum_{s=-1}^{0} h_{s1}\pi_{s+2,0} + \sum_{t=-1}^{0} v_{1t}\pi_{0,t+2} + r_{11}\pi_{0,0}, \\
h\pi_{1,0} &= q_{0-1}\pi_{1,1} + q_{-1-1}\pi_{2,1} + h_{-10}\pi_{2,0} + v_{1-1}\pi_{0,1} + r_{10}\pi_{0,0}, \\
v\pi_{0,1} &= q_{-10}\pi_{1,1} + q_{-1-1}\pi_{1,2} + v_{0-1}\pi_{0,2} + h_{-11}\pi_{1,0} + r_{01}\pi_{0,0}, \\
r\pi_{0,0} &= q_{-1-1}\pi_{1,1} + h_{-10}\pi_{1,0} + v_{0-1}\pi_{0,1},
\end{aligned}
$$

and here for $m, n \geq 2$

$$q\pi_{m,n} = \sum_{s=-1}^{1} \sum_{t=-1}^{1} q_{st}\pi_{m-s,n-t}, \tag{4.1a}$$

$$q\pi_{1,n} = \sum_{s=-1}^{0} \sum_{t=-1}^{1} q_{st}\pi_{1-s,n-t} + \sum_{t=-1}^{1} v_{1t}\pi_{0,n-t}, \tag{4.1b}$$

$$v\pi_{0,n} = \sum_{t=-1}^{1} q_{-1t}\pi_{1,n-t} + \sum_{t=-1}^{1} v_{0t}\pi_{0,n-t}, \tag{4.1c}$$

$$q\pi_{m,1} = \sum_{s=-1}^{1} \sum_{t=-1}^{0} q_{st}\pi_{m-s,1-t} + \sum_{s=-1}^{1} h_{s1}\pi_{m-s,0}, \tag{4.1d}$$

$$h\pi_{m,0} = \sum_{s=-1}^{1} q_{s-1}\pi_{m-s,1} + \sum_{s=-1}^{1} h_{s0}\pi_{m-s,0}. \tag{4.1e}$$

The compensation approach, introduced for the first time in [2] and [1], constructs a formal solution of the equilibrium equations (4.1) by using linear combinations of products $\alpha^m\beta^n$ satisfying equation (4.1a) in the interior of the state space. Inserting $\alpha^m\beta^n$ into (4.1a) and then dividing both sides of that equation by the common factor $\alpha^{m-1}\beta^{n-1}$ leads to the following characterization.

Lemma 4.1. *The product $\alpha^m\beta^n$ is a solution of* (4.1a) *if and only if α and β satisfy*

$$\alpha\beta q = \alpha^2 q_{-11} + \alpha q_{01} + q_{11} + \beta q_{10} + \beta^2 q_{1-1} + \alpha\beta^2 q_{0-1} + \alpha^2\beta^2 q_{-1-1} + \alpha^2\beta q_{-10} \tag{4.2}$$

Any linear combination of products $\alpha^m\beta^n$ with α and β satisfying (4.2), is a solution of (4.1a). Our purpose is to find a linear combination of such kind of products also satisfying the other equations in (4.1).

In order to construct this linear combination we start from arbitrary α_0 and β_0 satisfying (4.2) (we will see later how to find α_0 and β_0) and we suppose that $\alpha_0^m\beta_0^n$ violates the vertical boundary conditions (4.1b) and (4.1c).

Now we try to find α, β and c_1 with α and β satisfying (4.2) and such that $\alpha_0^m\beta_0^n + c_1\alpha^m\beta^n$ satisfies the vertical boundary conditions. Inserting this linear combination into (4.1b) and (4.1c), in order to satisfy the equations we obtain, we are forced to take

$$\beta = \beta_0, \quad \alpha = \alpha_1,$$

where α_1 is the other root of the quadratic equation (4.2) with $\beta = \beta_0$ fixed. Moreover dividing these equations by the common factor β_0^{n-1} leads to two linear equations for c_1, which have, in general, no solution. Therefore, we introduce an extra coefficient

by considering

$$\begin{cases} \alpha_0^m \beta_0^n + c_1 \alpha_1^m \beta_0^n & \text{for } m > 0, n > 0, \\ e_0 \beta_0^n & \text{for } m = 0, n > 0. \end{cases}$$

Inserting this form into the vertical boundary conditions and then dividing by the common factor β_0^{n-1} leads to two linear equations for c_1 and e_0, which can readily be solved using Cramer's rule. The resulting expressions can be simplified by using (4.1). This procedure is generalized in the following lemma. The second part formulates the analogue for the horizontal boundary.

Lemma 4.2. *Let x_1 and x_2 be the roots of the quadratic equation (4.1) for fixed β and let y_1 and y_2 be the roots of (4.1) for fixed α. Then*

(i) the quantity

$$z_{mn} = \begin{cases} x_1^m \beta^n + c x_2^m \beta^n & \text{for } m > 0, n > 0, \\ e \beta^n & \text{for } m = 0, n > 0, \end{cases}$$

satisfies (4.1a), (4.1b) *and* (4.1c) *with c and e given by*

$$c = -\frac{x_2^{-1}\left(\beta^2 v_{1-1} + \beta v_{10} + v_{11}\right) + v_{01} + \beta^2 v_{0-1} - \beta v}{x_1^{-1}\left(\beta^2 v_{1-1} + \beta v_{10} + v_{11}\right) + v_{01} + \beta^2 v_{0-1} - \beta v}, \tag{4.3}$$

$$e = -\frac{\left(\beta^2 q_{1-1} + \beta q_{10} + q_{11}\right)\left(x_2^{-1} - x_1^{-1}\right)}{x_1^{-1}\left(\beta^2 v_{1-1} + \beta v_{10} + v_{11}\right) + v_{01} + \beta^2 v_{0-1} - \beta v}, \tag{4.4}$$

(ii) the quantity

$$w_{mn} = \begin{cases} \alpha^m y_1^n + d \alpha^m y_2^n & \text{for } m > 0, n > 0, \\ f \alpha^n & \text{for } m > 0, n = 0, \end{cases}$$

satisfies (4.1a), (4.1d) *and* (4.1e) *with d and f given by*

$$d = -\frac{y_2^{-1}\left(\alpha^2 h_{1-1} + \alpha h_{10} + h_{11}\right) + h_{01} + \alpha^2 h_{0-1} - \alpha h}{y_1^{-1}\left(\alpha^2 h_{1-1} + \alpha h_{10} + h_{11}\right) + h_{01} + \alpha^2 h_{0-1} - \alpha h}, \tag{4.5}$$

$$f = -\frac{\left(\alpha^2 q_{1-1} + \alpha q_{10} + q_{11}\right)\left(y_2^{-1} - y_1^{-1}\right)}{y_1^{-1}\left(\alpha^2 h_{1-1} + \alpha h_{10} + h_{11}\right) + h_{01} + \alpha^2 h_{0-1} - \alpha h}. \tag{4.6}$$

Proof. Inserting this form into the vertical boundary conditions we obtain

$$\begin{cases} q x_1^m \beta^n &= \sum_{s=-1}^{0} \sum_{t=-1}^{1} q_{st} (x_1^{1-s} \beta^{n-t} + c x_2^{1-s} \beta^{n-t}) + \sum_{t=-1}^{1} v_{1t} e \beta^{n-t}, \\ q e \beta^n &= \sum_{t=-1}^{1} q_{-1t} (x_1 \beta^{n-t} + c x_2 \beta^{n-t}) + \sum_{t=-1}^{1} v_{0t} e \beta^{n-t}. \end{cases}$$

This two equations can both be divided by the common factor β^{n-1}, this leads to the following linear system in the unknowns e and c

$$\begin{cases} qx_1^m\beta & = \sum_{s=-1}^{0}\sum_{t=-1}^{1} q_{st}(x_1^{1-s}\beta^{1-t} + cx_2^{1-s}\beta^{1-t}) + \sum_{t=-1}^{1} v_{1t}e\beta^{1-t}, \\ qe\beta & = \sum_{t=-1}^{1} q_{-1t}(x_1\beta^{1-t} + cx_2\beta^{1-t}) + \sum_{t=-1}^{1} v_{0t}e\beta^{1-t}. \end{cases}$$

Now, by applying Cramer's rule on both the unknowns, with some simple manipulation, we obtain the formulas in (i). An analogue approach to the horizontal boundary leads to the formulas in (ii). □

We added $c_1\alpha_1^m\beta_0^n$ to compensate for the error of $\alpha_0^m\beta_0^n$ on the vertical boundary and by doing so introduced a new error on the horizontal boundary, since $c_1\alpha_1^m\beta_0^n$ violates these boundary conditions. To compensate for this error we add $c_1d_1\alpha_1^m\beta_1^n$ where β_1 is the other root of (4.2) with $\alpha = \alpha_1$ and d_1 descends from the previous lemma. However, this term violates the vertical boundary conditions, so we have to add again a term, and so on. Thus the compensation of $\alpha_0^m\beta_0^n$ on the vertical boundary generates an infinite sequence of compensation terms. An analogous sequence is generated by starting the compensation of $\alpha_0^m\beta_0^n$ on the horizontal boundary. This idea leads to the following bi-infinite sum of product terms

$$\cdots + \underbrace{c_{-1}d_{-1}\alpha_{-1}^m\beta_{-1}^n + \overbrace{d_{-1}c_0\alpha_0^m\beta_{-1}^n + c_0d_0\alpha_0^m\beta_0^n}^{H} + \overbrace{c_1d_0\alpha_1^m\beta_0^n + c_1d_1\alpha_1^m\beta_1^n}^{H} + \ldots}_{V}$$

Each term in the sum satisfies (4.1a), each sum of two terms with the same β factor satisfies the vertical boundary conditions (4.1b) and (4.1c), each sum of two terms with the same α factor satisfies the horizontal boundary conditions (4.1d) and (4.1e). Since the equilibrium equations are linear, we can conclude that the sum below formally satisfies the equations (4.1).

For all $m > 0$ and $n > 0$, let us define

$$x_{mn}(\alpha_0, \beta_0) = \sum_{i=-\infty}^{\infty} d_i(c_i\alpha_i^m + c_{i+1}\alpha_{i+1}^m)\beta_i^n \qquad \text{(pairs with same } \beta \text{ factor)} \quad (4.7)$$

$$= \sum_{i=-\infty}^{\infty} c_{i+1}(d_i\beta_i^m + d_{i+1}\beta_{i+1}^m)\alpha_{i+1}^m \quad \text{(pairs with same } \alpha \text{ factor)}. \quad (4.8)$$

Moreover for the horizontal and vertical boundaries let us define

$$x_{0n}(\alpha_0,\beta_0)=\sum_{i=-\infty}^{\infty} d_i e_i \beta_i^n \quad for \quad n>0, \tag{4.9}$$

$$x_{m0}(\alpha_0,\beta_0)=\sum_{i=-\infty}^{\infty} c_{i+1} f_{i+1} \alpha_{i+1}^m \quad for \quad m>0. \tag{4.10}$$

The coefficients sequence is generated such that for α_i and α_{i+1} are the roots of (4.2) with fixed $\beta=\beta_i$, β_i and β_{i+1} are the roots of (4.2) with fixed $\alpha=\alpha_{i+1}$ and the terms c_i, d_i, e_i, f_i satisfy the following recursive formulas, obtained thanks to Lemma 4.2, where we initially set $c_0=1$ and $d_0=1$

$$c_{i+1}=-\frac{\alpha_{i+1}^{-1}\left(\beta_i^2 v_{1-1}+\beta_i v_{10}+v_{11}\right)+v_{01}+\beta_i^2 v_{0-1}-\beta_i v}{\alpha_i^{-1}\left(\beta_i^2 v_{1-1}+\beta_i v_{10}+v_{11}\right)+v_{01}+\beta_i^2 v_{0-1}-\beta_i v} c_i \quad i\in\mathbb{Z}, \tag{4.11}$$

$$e_i=-\frac{\left(\beta_i^2 q_{1-1}+\beta_i q_{10}+q_{11}\right)\left(\alpha_{i+1}^{-1}-\alpha_i^{-1}\right)}{\alpha_i^{-1}\left(\beta_i^2 v_{1-1}+\beta_i v_{10}+v_{11}\right)+v_{01}+\beta_i^2 v_{0-1}-\beta_i v} c_i \quad i\geq 0, \tag{4.12}$$

$$e_i=-\frac{\left(\beta_i^2 q_{1-1}+\beta_i q_{10}+q_{11}\right)\left(\alpha_i^{-1}-\alpha_{i+1}^{-1}\right)}{\alpha_{i+1}^{-1}\left(\beta_i^2 v_{1-1}+\beta_i v_{10}+v_{11}\right)+v_{01}+\beta_i^2 v_{0-1}-\beta_i v} c_{i+1} \quad i<0, \tag{4.13}$$

$$d_{i+1}=-\frac{\beta_{i+1}^{-1}\left(\alpha_{i+1}^2 h_{-11}+\alpha_{i+1} h_{01}+h_{11}\right)+h_{10}+\alpha_{i+1}^2 h_{-10}-\alpha_{i+1} h}{\beta_i^{-1}\left(\alpha_{i+1}^2 h_{-11}+\alpha_{i+1} h_{01}+h_{11}\right)+h_{10}+\alpha_{i+1}^2 h_{-10}-\alpha_{i+1} h} d_i \quad i\in\mathbb{Z}, \tag{4.14}$$

$$f_{i+1}=-\frac{\left(\alpha_{i+1}^2 q_{-11}+\alpha_{i+1} q_{01}+q_{11}\right)\left(\beta_{i+1}^{-1}-\beta_i^{-1}\right)}{\beta_i^{-1}\left(\alpha_{i+1}^2 h_{-11}+\alpha_{i+1} h_{01}+h_{11}\right)+h_{10}+\alpha_{i+1}^2 h_{-10}-\alpha_{i+1} h} d_i \quad i\geq 0, \tag{4.15}$$

$$f_{i+1}=-\frac{\left(\alpha_{i+1}^2 q_{-11}+\alpha_{i+1} q_{01}+q_{11}\right)\left(\beta_i^{-1}-\beta_{i+1}^{-1}\right)}{\beta_{i+1}^{-1}\left(\alpha_{i+1}^2 h_{-11}+\alpha_{i+1} h_{01}+h_{11}\right)+h_{10}+\alpha_{i+1}^2 h_{-10}-\alpha_{i+1} v} d_{i+1} \quad i<0. \tag{4.16}$$

Each solution $x_{mn}(\alpha_0,\beta_0)$ has its own sequence $\{\alpha_i,\beta_i\}$ depending on the initial values α_0 and β_0, and its associated sequence of coefficients $\{c_i,d_i,e_i,f_i\}$; For any pair α_0,β_0 satisfying equation (4.2) the series $x_{mn}(\alpha_0,\beta_0)$ formally satisfies the equations (4.1).

4.1 Necessary Conditions for Convergence

Under certain conditions compensation fails. This happens if for some value of i the equation (4.2) with fixed $\beta=\beta_i$ or fixed $\alpha=\alpha_{i+1}$ reduces to a linear equation,

so the necessary second root does not exist. If the second root is equal to the first one, then it can be verified (see [2]) that the compensation procedure constructs the null solution. Furthermore, compensation fails if for some value of i the denominator in the definition of the coefficients vanishes.

Let us suppose in this section that for the initial α_0 and β_0, in at least one direction compensation is always possible. We want to know under what conditions the infinite sum $x_{mn}(\alpha_0, \beta_0)$ converges. To aid convergence of $x_{mn}(\alpha_0, \beta_0)$ for fixed m and n, we require that α_i and β_i tend to zero as $|i|$ tends to infinity. To aid convergence of $x_{mn}(\alpha_0, \beta_0)$ over all values m and n (necessary for normalization), we require that $|\alpha_i|, |\beta_i| < 1$ for all i.

Since α_i and α_{i+1} are the roots of quadratic equation (4.2) with $\beta = \beta_i$, we have

$$\alpha_i\alpha_{i+1} = \frac{\beta_i^2 q_{1-1} + \beta_i q_{10} + q_{11}}{\beta_i^2 q_{-1-1} + \beta_i q_{-10} + q_{-11}}, \quad \alpha_i + \alpha_{i+1} = \frac{-\beta_i^2 q_{0-1} + \beta_i q - q_{01}}{\beta_i^2 q_{-1-1} + \beta_i q_{-10} + q_{-11}},$$

in the same way, β_i and β_{i+1} are the roots of quadratic equation (4.2) with $\alpha = \alpha_i$, we have

$$\beta_i\beta_{i+1} = \frac{\alpha_{i+1}^2 q_{-11} + \alpha_{i+1} q_{01} + q_{11}}{\alpha_{i+1}^2 q_{-1-1} + \alpha_{i+1} q_{0-1} + q_{1-1}}, \quad \beta_i + \beta_{i+1} = \frac{-\alpha_{i+1}^2 q_{-10} + \alpha_{i+1} q - q_{10}}{\alpha_{i+1}^2 q_{-1-1} + \alpha_{i+1} q_{0-1} + q_{1-1}}.$$

From the equations above, we deduce that

$$q_{11} = q_{01} = q_{10} = 0 \tag{4.17}$$

is a necessary condition for convergence to zero of α_i and β_i.

We suppose from now on that this condition is satisfied and moreover, to exclude pathological cases, that there is a rate component to the south west, that is

$$q_{-10} + q_{-1-1} + q_{0-1} > 0, \tag{4.18}$$

with the assumption (4.17), the process can be expressed in a visual way in figure 4.1.

By assumption (4.17), equation (4.2) simplifies to

$$\alpha\beta q = \alpha^2 q_{-11} + \beta^2 q_{1-1} + \alpha\beta^2 q_{0-1} + \alpha^2\beta^2 q_{-1-1} + \alpha^2\beta q_{-10}, \tag{4.19}$$

about which we state the the following

Lemma 4.3. *For each fixed α such that $0 < |\alpha| < 1$, equation (4.19) has exactly one root with modulus smaller than $|\alpha|$ and one root of modulus larger than $|\alpha|$. The same holds with α and β interchanged.*

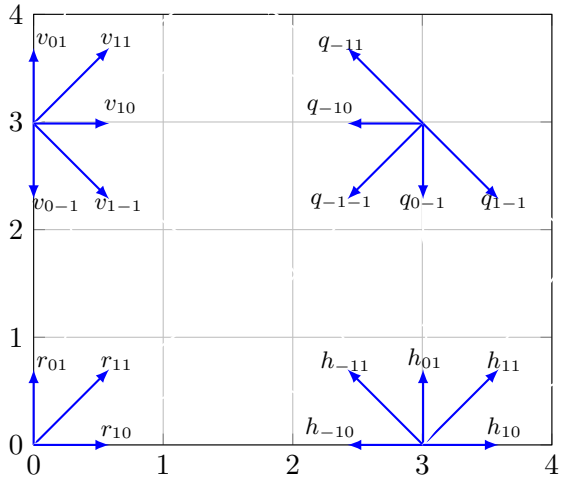


Figure 4.1: Transition rates of the random walk with condition (4.17).

Proof. Dividing (4.19) by α^2 and applying the change of variable $z = \frac{\beta}{\alpha}$, we obtain $f(z) + g(z) = 0$, where

$$f(z) = z^2\left(\alpha^2 q_{-1-1} + \alpha q_{0-1} + q_{1-1}\right), \quad g(z) = -z\left(q - \alpha q_{-10}\right) + q_{-11}.$$

Then for all z such that $|z| = 1$, we have

$$|f(z)| \leq q_{-1-1} + q_{0-1} + q_{1-1}, \quad |g(z)| \geq q_{-1-1} + q_{0-1} + q_{1-1},$$

where, by all previous assumptions, at least one of the inequalities is strict. We conclude by applying Rouché's theorem to $f(z)$ and $g(z)$ above $\partial\mathbb{B}$. □

Corollary 4.1. *Let α_0 and β_0 be roots of equation (4.19) satisfying $1 > |\alpha_0| > |\beta_0| > 0$. Then there exists a negative value of i for which $|\alpha_i| \geq 1$ or $|\beta_i| \geq 1$ and*

$$1 > |\alpha_{i+1}| > |\beta_{i+1}| > \cdots > |\alpha_0| > |\beta_0| > |\alpha_1| > |\beta_1| \ldots$$

Moreover the sequences $\{\alpha_i\}$ and $\{\beta_i\}$ tend to zero as i tend to $+\infty$. A similar result holds if $1 > |\beta_0| > |\alpha_0| > 0$.

Proof. The monotonicity follows directly from Lemma 4.3. To prove that there exists a negative value of i for which $|\alpha_i| \geq 1$ or $|\beta_i| \geq 1$, we also need information about the β-roots of equation (4.19) for fixed α with $|\alpha_i| = 1$.

For fixed α with $|\alpha_i| = 1$ and $\alpha \neq 1, -1$ it follows by applying Rouché's theorem, with the same notation as in the proof of Lemma 4.3, that equation $f(z) + g(z) = 0$ has one root z with $|z| < 1$ and one root z with $|z| > 1$. The same result holds for

$\alpha = -1$ or $\alpha = 1$ if at least one of the rates q_{-10} and q_{0-1} is positive. For $\alpha = 1$ or $\alpha = -1$ and, if $q_{-10} = q_{0-1} = 0$, the equation $f(z) + g(z) = 0$ is solved by $z = 1$ and $z = q_{-11}\left(q_{-1-1} + q_{0-1} + q_{1-1}\right)^{-1}$, respectively.

Hence, if $q_{-11} < q_{-1-1}+q_{0-1}+q_{1-1}$ we can define $z(\alpha)$ as the root of $f(z)+g(z) = 0$ for fixed α with $|\alpha| < 1$, which satisfies $z(\alpha) < 1$. Since $z(\alpha)$ is continuous, the maximum of $|z(\alpha)|$ for $|\alpha| < 1$ exists and is less than one.

So $|\frac{\beta_i}{\alpha_i}| = |z(\alpha_i)| \leq \max_{|\alpha_i|\leq 1}|z(\alpha)| < 1$ as long as $\alpha_i < 1$. This proves that $|\alpha_i|$ and $|\beta_i|$ decrease exponentially fast to zero as i tends to infinity, and that $|\alpha_i| \geq 1$ or $|\beta_i| \geq 1$ for some negative value of i.

If $q_{-11} \geq q-1-1+q_{0-1}+q_{1-1}$, then from previous assumptions we obtain the inequality $q_{1-1} < q_{-1-1} + q_{-10} + q_{-11}\cdot$ Hence, we can repeat the arguments above by considering the roots of equation (4.19) for fixed β instead of fixed α. □

When the sequence of α_i and β_i is started with roots α_0 and β_0 of (4.19) such that $1 > |\alpha_0| > 0$ and $1 > |\beta_0| > 0$, then by previous lemma, $|\alpha_0| > |\beta_0|$ or $|\alpha_0| < |\beta_0|$. Hence, by the corollary, $|\alpha_i|$ and $|\beta_i|$ decrease to zero in at least one direction. In the other direction $|\alpha_i|$ and $|\beta_i|$ increase and eventually $|\alpha_i| \geq 1$ and $|\beta_i| \geq 1$ for some i. Therefore we can't meet the convergence requirement in that direction, unless c_i or d_i vanishes for some i before $|\alpha_i| \geq 1$ or $|\beta_i| \geq 1$.

After renumbering the terms, this amounts to the requirement that the initial product $\alpha_0^m \beta_0^n$ fits the horizontal boundary conditions with $\alpha_0 > |\beta_0|$ or the vertical boundary conditions with $\alpha_0 < \beta_0$; in the first case we have $d_{-1} = 0$, in the second case we have $c_1 = 0$. Pairs α_0, β_0 satisfying these requirements will be called *feasible pairs*.

Definition 4.1. A pair α_0, β_0 is called feasible if:

- α_0 and β_0 are roots of (4.19) with $1 > |\alpha_0| > 0$ and $1 > |\beta_0| > 0$,
- if $|\alpha_0| > |\beta_0|$, then $d_{-1} = 0$,
- if $|\alpha_0| < |\beta_0|$, then $c_1 = 0$.

4.1.1 On the existence of feasible pairs

Now we ask under what conditions the existence of feasible pairs is ensured, how many they are and whether they are real or complex. In this section we give the answer to all these questions, but we don't prove the results we give, all the proofs can be found in [2].

Moreover all the results concern the feasible pairs with respect to the horizontal boundary, that is we only consider roots α_0, β_0 of (4.19) with $1 > |\alpha_0| > |\beta_0| > 0$ and $d_{-1} = 0$. Feasibility with respect to the vertical boundary can be treated similarly.

The first result we give is about the maximum number of feasible pairs related to the boundary rates.

Theorem 4.1. *There exists at most two feasible pairs with respect to horizontal boundary, these pairs are always real. In particular we distinguish the following cases:*

1. *if $h_{11} > 0$, then there are at most two pairs. If there are exactly two pairs, then at least one of the α_0 and one of the β_0 must be positive;*
2. *if $h_{11} = 0$ and $h_{01} + h_{10} > 0$, then there is at most one pair. These roots are positive;*
3. *if $h_{11} = h_{01} = h_{10} = 0$, then there are no pairs.*

Let us observe that, for fixed α, equation (4.19) is solved by

$$y_{\pm}(\alpha) = \alpha \frac{q - \alpha q_{-10} \pm \sqrt{(q - \alpha q_{-10})^2 - 4\left(\alpha^2 q_{-1-1} + \alpha q_{0-1} + q_{1-1}\right) q_{-11}}}{2\left(\alpha^2 q_{-1-1} + \alpha q_{0-1} + q_{1-1}\right)}, \quad (4.20)$$

in a very similar way we can consider $x_{\pm}(\beta)$, the root of (4.19) for fixed β. Consider the following function in the variable α

$$f(\alpha) := \frac{\alpha^2 h_{-11} + \alpha h_{01} + h_{11}}{y_{+}(\alpha)} + \alpha^2 h_{-10} + h_{10},$$

the following theorem gives a condition for the existence of feasible pairs which involves the function $f(\alpha)$.

Theorem 4.2. *If $h_{11}+h_{01}+h_{10} > 0$, then the maximum number of feasible pairs with respect to the horizontal boundary is obtained if and only if the following condition is satisfied:*

$$q_{-1-1} + q_{0-1} + q_{1-1} > q_{-11} \Rightarrow h < f'(1). \quad (4.21)$$

In particular we distinguish the following cases:

1. *if $h_{11} > 0$, then there are two pairs. One α_0 is the solution of the equation*

$$h\alpha = f(\alpha) \quad (4.22)$$

   in the interval $(0, 1)$, the other α_0 is its solution in $(-1, 0)$;

2. *if $h_{11} = 0$ and $h_{01} + h_{10} > 0$, then there is one pair and its α_0 is the solution of (4.22) in $(0, 1)$.*

4.2 Convergence Theorem

As we have seen in the previous section, there are at most two feasible pairs with respect to the horizontal boundary, we denote them by

$$(\alpha_+, y_-(\alpha_+)), \quad (\alpha_-, y_-(\alpha_-)),$$

where α_+ is the solution of (4.22) in $(0,1)$ and α_- is its solution in $(-1,0)$. In the same way the feasible pairs on the vertical boundary are denoted by

$$(\beta_+, x_-(\beta_+)), \quad (\beta_-, x_-(\beta_-)),$$

where β_+ is the solution of the β-equivalent of (4.22) in $(0,1)$ and α_- is its solution in $(-1,0)$. For $\alpha_0 = \alpha_+$ and $\beta_0 = y_-(\alpha_+)$ we abbreviate the notation $x_{mn}(\alpha_0, \beta_0)$ to $x_{mn}(\alpha_+)$, similar abbreviations are used for the other feasible pairs.

The formal solutions $x_{mn}(\alpha_0, \beta_0)$ with feasible initial pairs simplify with respect to the forms in (4.7) and (4.8). If we take $\alpha_0 = \alpha_+$ and $\beta_0 = y_-(\alpha_+)$, then $d_{-1} = 0$ and so $d_i = f_i = 0$ for all $i < 0$. Then for $m, n > 0$ the series $x_{mn}(\alpha_+)$ simplifies to

$$\begin{aligned} x_{mn}(\alpha_+) &= \sum_{i=0}^{\infty} d_i (c_i \alpha_i^m + c_{i+1} \alpha_{i+1}^m) \beta_i^n \\ &= d_0 c_0 \beta_0^n \alpha_0^m + \sum_{i=0}^{\infty} c_{i+1} (d_i \beta_i^n + d_{i+1} \beta_{i+1}^n) \alpha_{i+1}^m, \end{aligned}$$

while the boundary series become

$$x_{0n}(\alpha_+) = \sum_{i=0}^{\infty} d_i e_i \beta_i^n, \quad x_{m0}(\alpha_+) = c_0 f_0^n \alpha_0^m + \sum_{i=0}^{\infty} c_{i+1} f_{i+1} \alpha_{i+1}^m;$$

when the sequences $\{\alpha_i\}$ and $\{\beta_i\}$ are initialized with $\alpha_0 = \alpha_-$ and $\beta_0 = y_-(\alpha_-)$, the solution $x_{mn}(\alpha_-)$ simplifies accordingly.

If we take $\alpha_0 = x_-(\beta_+)$ and $\beta_0 = \beta_+$, then $c_1 = 0$ and so $c_i = e_i = 0$ for all $i > 0$. Then for $m, n > 0$ the series $x_{mn}(\beta_+)$ simplifies to

$$\begin{aligned} x_{mn}(\beta_+) &= d_0 c_0 \beta_0^n \alpha_0^m + \sum_{i=-\infty}^{-1} d_i (c_i \alpha_i^m + c_{i+1} \alpha_{i+1}^m) \beta_i^n \\ &= \sum_{i=-\infty}^{-1} c_{i+1} (d_i \beta_i^n + d_{i+1} \beta_{i+1}^n) \alpha_{i+1}^m, \end{aligned}$$

while the boundary series become

$$x_{0n}(\beta_+) = d_0 e_0 \beta_0^n + \sum_{i=-\infty}^{-1} d_i e_i \beta_i^n, \quad x_{m0}(\beta_+) = \sum_{i=-\infty}^{-1} c_{i+1} f_{i+1} \alpha_{i+1}^m;$$

when the sequences $\{\alpha_i\}$ and $\{\beta_i\}$ are initialized with $\alpha_0 = x_-(\beta_-)$ and $\beta_0 = \beta_-$, the solution $x_{mn}(\beta_-)$ simplifies accordingly.

Theorem 4.3. *Let us assume that* (4.17), (4.18), (4.21) *hold, as well as the conditions we presented at the beginning of the chapter in order to avoid pathological cases. Then there exists an integer $N \in \mathbb{Z}_+$ such that for $n + m > N$ the solutions for different feasible pairs are linearly independent and the invariant measure can be written as*

$$\pi_{m,n} = \sum_{(\alpha_0,\beta_0)} c(\alpha_0, \beta_0) x_{mn}(\alpha_0, \beta_0), \tag{4.23}$$

where (α_0, β_0) runs through the set of at most four feasible pairs (two with respect to the horizontal boundary and two with respect to vertical boundary), $c(\alpha_0, \beta_0)$ is an appropriately chosen coefficient and $x_{mn}(\alpha_0, \beta_0)$ is obtained as explained above.

The complete proof of the previous theorem can be found in [2] as well as the following results concerning the explicit calculation of the integer $N \in \mathbb{Z}_+$ that appears in Theorem 4.3.

Lemma 4.4. *Consider the feasible initial pair given by $\alpha_0 = \alpha_+$ and $\beta_0 = y_-(\alpha_+)$, then we have the following convergence results:*

$$\lim_{i\to\infty} \frac{\beta_i}{\alpha_i} = \frac{1}{A_+}, \quad \lim_{i\to\infty} \frac{\alpha_{i+1}}{\beta_i} = A_-, \quad \lim_{i\to\infty} \frac{c_{i+1}}{c_i} = -\gamma, \quad \lim_{i\to\infty} \frac{d_{i+1}}{d_i} = -\eta,$$

where

$$A_\pm := \frac{q \pm \sqrt{q^2 - 4q_{1-1}q_{-11}}}{2q_{-11}}$$

$$\gamma := \begin{cases} A_+A_-^{-1} & \text{if} \quad v_{11} > 0, \\ \left(A_-^{-1}v_{10} + v_{01}\right)\left(A_+^{-1}v_{10} + v_{01}\right)^{-1} & \text{if} \quad v_{11} = 0, v_{10} + v_{01} > 0, \\ \left(A_-^{-1}v_{1-1} - v\right)\left(A_+^{-1}v_{1-1} - v\right)^{-1} & \text{if} \quad v_{11} = v_{10} = v_{01} = 0, \end{cases}$$

$$\eta := \begin{cases} A_+A_-^{-1} & \text{if} \quad h_{11} > 0, \\ \left(A_+h_{01} + h_{10}\right)\left(A_-h_{01} + v_{10}\right)^{-1} & \text{if} \quad h_{11} = 0, h_{01} + h_{10} > 0, \\ \left(A_+h_{-11} - h\right)\left(A_-h_{-11} - h\right)^{-1} & \text{if} \quad h_{11} = h_{01} = h_{10} = 0. \end{cases}$$

Theorem 4.4. *The integer $N \in \mathbb{Z}_+$ that appears in Theorem 4.3 is the smallest such that*

$$|\gamma\eta| \left(\frac{A_-}{A_+}\right)^{N+1} < 1.$$

Let us observe that if $h_{01} + h_{11} + h_{10} > 0$ and $v_{01} + v_{11} + v_{10} > 0$, then we have $1 \le \gamma, \eta \le \frac{A_+}{A_-}$, so $N \le 2$. In particular if $h_{11}, v_{11} > 0$, then $N = 2$. However in the general case N can be arbitrarily large as it is we can see in the following

Example 4.1. Consider the process illustrated in Figure 4.2, with $0 < \delta < 1$. For this example it can be readily verified that

$$A_{\pm} = \frac{3}{4} \pm \frac{1}{4}\sqrt{1+8\delta}, \quad \eta = 1, \quad \gamma = \frac{1+\sqrt{1+8\delta}}{1-\sqrt{1+8\delta}}.$$

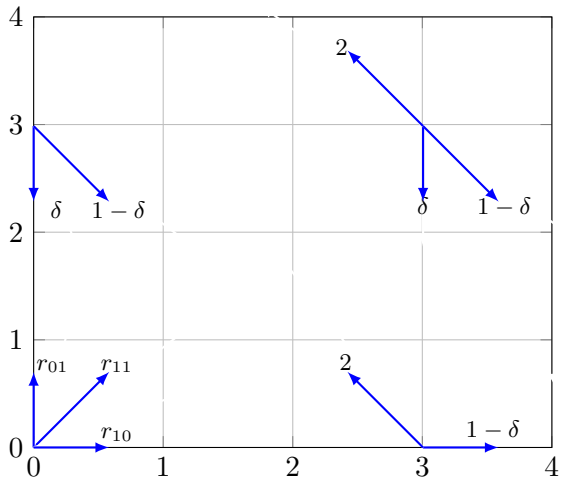


Figure 4.2: An example for which N can be arbitrarily large.

Hence, the integer N is the smallest for which

$$\frac{\sqrt{1+8\delta}+1}{\sqrt{1+8\delta}-1}\left[\frac{3-\sqrt{1+8\delta}}{3+\sqrt{1+8\delta}}\right]^{N+1} < 1,$$

from this inequality it follows that $\lim_{\delta\to 0} N = \infty$.

In order to simplify the notation, we will make the following assumption:

- the sum in equation (4.23) runs through a single pair (α_0, β_0) obtained from the horizontal boundary;
- the relative coefficient $c(\alpha_0, \beta_0)$ can be set equal to 1;
- equation (4.23) is valid for $n + m > 0$, that is $N = 0$.

With these assumptions, we can rewrite the expressions of $\pi_{m,n}$ as

$$\pi_{m,n} = \sum_{i=0}^{\infty} d_i(c_i\alpha_i^m + c_{i+1}\alpha_{i+1}^m)\beta_i^n = d_0c_0\beta_0^n\alpha_0^m + \sum_{i=1}^{\infty} c_i(d_{i-1}\beta_{i-1}^n + d_i\beta_i^n)\alpha_i^m,$$

$$\pi_{0,n} = \sum_{i=0}^{\infty} d_ie_i\beta_i^n, \quad \pi_{m,0} = \sum_{i=0}^{\infty} c_if_i\alpha_i^m; \tag{4.24}$$

4.3 Boundary value problem

The boundary value problem is an analytic method which is used in literature to approach some two-dimensional random walks restricted to the first quadrant. The first step to describe is to define the bivariate *probability generating function*

$$\Pi(x,y) = \sum_{m=0}^{\infty}\sum_{n=0}^{\infty} \pi_{m,n} x^m y^n,$$

with $|x| < 1$ and $|y| < 1$.

The probability generating function of the position of a homogeneous nearest neighbour random walk satisfies a functional equation of the form

$$Q(x,y)\Pi(x,y) + H(x,y)\Pi(x,0) + V(x,y)\Pi(0,y) + R(x,y)\Pi(0,0) = 0,$$

with $Q(x,y)$, $H(x,y)$, $V(x,y)$ and $R(x,y)$ known bivariate polynomials in x and y, depending only on the parameters of the random walk. In particular,

$$\begin{aligned}
Q(x,y) &= xy\left[\sum_{s=-1}^{1}\sum_{t=-1}^{1} x^s y^t q_{st} - q\right],\\
H(x,y) &= xy\left[\sum_{s=-1}^{1}\sum_{t=-1}^{1} x^s y^t (q_{st} - h_{st}) - (q-h)\right],\\
V(x,y) &= xy\left[\sum_{s=-1}^{1}\sum_{t=-1}^{1} x^s y^t (q_{st} - v_{st}) - (q-v)\right],\\
R(x,y) &= xy\left[\sum_{s=-1}^{1}\sum_{t=-1}^{1} x^s y^t (q_{st} + h_{st} + v_{st} - r_{st}) - (q+h+v-r)\right].
\end{aligned}$$

Let us observe that these polynomials are closely related to the equilibrium equations for $\pi_{m,n}$, indeed setting $Q\left(\alpha^{-1},\beta^{-1}\right) = 0$ reduces to exactly equation (4.19), furthermore, $H\left(\alpha^{-1},\beta^{-1}\right) = 0$ reduces to exactly the balance equations for the horizontal boundary, and similarly $V\left(\alpha^{-1},\beta^{-1}\right) = 0$ reduces to exactly the balance equations for the vertical boundary.

In the following we relate the ideas and results from the compensation approach to the matrix geometric approach by utilizing the tools of the boundary value method. Furthermore we achieve to connect the three approaches and gain valuable insight on the analytic and probabilistic interpretation of the terms appearing in the invariant measure. First and foremost, we show that the sequences $\{\alpha_i\}$ and $\{\beta_i\}$ are connected with the eigenvalues and left eigenvectors of matrix R.

To do this, let us consider the operator $U_\alpha = (u_{in})_{i,n=0,1,2,\dots}$ whose rows are defined by $u_i = [u_{i0}, u_{i1}, u_{i2}, \dots]$, with $u_{i0} = c_i f_i$, $i \geq 0$, $u_{0n} = d_0 c_0 \beta_0^n$, with $n \geq 1$ and $u_{in} = c_i(d_{i-1}\beta_{i-1}^n + d_i \beta_i^n)$, with $i, n \geq 1$.

We can write the elements of the invariant vector $\pi_{m,n}$ for $m \geq 1$ and $n \geq 0$ in function of the elements of U_α in the following way

$$\pi_{m,n} = \sum_{i=0}^{\infty} \alpha_i^m u_{in}. \tag{4.25}$$

Let us consider

$$V_\alpha = \begin{bmatrix} 1 & 1 & 1 & 1 & \dots \\ \alpha_0 & \alpha_1 & \alpha_2 & \alpha_3 & \dots \\ \alpha_0^2 & \alpha_1^2 & \alpha_2^2 & \alpha_3^2 & \dots \\ \alpha_0^3 & \alpha_1^3 & \alpha_2^3 & \alpha_3^3 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}, \; D_\alpha = \begin{bmatrix} \alpha_0 & & & & \\ & \alpha_1 & & & \\ & & \alpha_2 & & \\ & & & \alpha_3 & \\ & & & & \ddots \end{bmatrix},$$

the operator V_α is the Vandermonde operator associated to the sequence $\{\alpha_i\}$, D_α is the diagonal operator with the sequence $\{\alpha_i\}$ on the diagonal. The relation (4.25) can be written in a operator form as

$$V_\alpha D_\alpha U_\alpha = \begin{bmatrix} \pi_1^{(L)} \\ \pi_2^{(L)} \\ \pi_3^{(L)} \\ \vdots \end{bmatrix},$$

where for each m fixed we have $\pi_m^{(L)} = [\pi_{m,0}\ \pi_{m,1}\ \pi_{m,2}\ \dots]$.

Now we multiply this equation on the right by the operator $R^{(L)}$, which is the minimal solution of

$$A_1^{(L)} + R^{(L)} A_0^{(L)} + \left(R^{(L)}\right)^2 A_{-1}^{(L)} = 0,$$

and that satisfy $\pi_m^{(L)} R^{(L)} = \pi_{m+1}^{(L)}$, we get

$$V_\alpha D_\alpha U_\alpha R^{(L)} = \begin{bmatrix} \pi_1^{(L)} \\ \pi_2^{(L)} \\ \pi_3^{(L)} \\ \vdots \end{bmatrix} R^{(L)} = \begin{bmatrix} \pi_2^{(L)} \\ \pi_3^{(L)} \\ \pi_4^{(L)} \\ \vdots \end{bmatrix} = V_\alpha D_\alpha^2 U_\alpha. \tag{4.26}$$

From this equation we would like to conclude that $U_\alpha R^{(L)} = D_\alpha U_\alpha$, but $V_\alpha D_\alpha$ is not invertible (as operator from ℓ^1 to itself or from ℓ^∞ to itself). However in the

following proposition, in which we rework a result of [12], we prove this property, which states that the $\{\alpha_i\}$ are the eigenvalues of $R^{(L)}$ and the rows of U_α are the eigenvectors, with some additional hypothesis.

Proposition 4.1. *Let us suppose that the following limitation hold*

$$\|u_i\left(\alpha_i I - R^{(L)}\right)\|_\infty < \frac{1}{|\alpha_i|}, \tag{4.27}$$

then the terms $\{\alpha_i\}$ with $i \geq 0$ constitute the different eigenvalues of the operator $R^{(L)}$. For an eigenvalue α_i, the corresponding left eigenvector of the operator $R^{(L)}$ is u_i, with $i \geq 0$.

Proof. If we consider in equation 4.26 the m-th rows, we obtain

$$\sum_{i=0}^{\infty} \alpha_i^m u_i \left(\alpha_i I - R^{(L)}\right) = 0, \tag{4.28}$$

for $m > 0$. Dividing each term of this series by α_0^m, we get

$$-u_0\left(\alpha_0 I - R^{(L)}\right) = \sum_{i=1}^{\infty} \left(\frac{\alpha_i}{\alpha_0}\right)^m u_i \left(\alpha_i I - R^{(L)}\right).$$

Thus, by taking the norms, we obtain

$$\begin{aligned} 0 \leq \|u_0\left(\alpha_0 I - R^{(L)}\right)\|_\infty &\leq \inf_{m>0} \sum_{i=1}^{\infty} \left|\frac{\alpha_i}{\alpha_0}\right|^m \|u_i\left(\alpha_i I - R^{(L)}\right)\|_\infty \\ &= \lim_{m\to\infty} \sum_{i=1}^{\infty} \left|\frac{\alpha_i}{\alpha_0}\right|^m \|u_i\left(\alpha_i I - R^{(L)}\right)\|_\infty \\ &= \sum_{i=1}^{\infty} \lim_{m\to\infty} \left|\frac{\alpha_i}{\alpha_0}\right|^m \|u_i\left(\alpha_i I - R^{(L)}\right)\|_\infty = 0, \end{aligned}$$

where the fact that the infimum is equal to the limit is allowed by the monotone convergence of the sequence $\{\alpha_i\}$. Moreover the exchange of the series and the limit is justified from hypotheses 4.27, indeed from this it follows that

$$\sum_{i=1}^{\infty} \left|\frac{\alpha_i}{\alpha_0}\right|^m \|u_i\left(\alpha_i I - R^{(L)}\right)\|_\infty \leq \frac{1}{\alpha_0} \sum_{i=1}^{\infty} \left|\frac{\alpha_i}{\alpha_0}\right|^{m-1} \leq \frac{1}{\alpha_0} \sum_{i=1}^{\infty} \left|\frac{\alpha_i}{\alpha_0}\right|,$$

and we can apply the Lebesgue Theorem.

Therefore we conclude that $u_0\left(\alpha_0 I - R^{(L)}\right) = 0$ and, recursively, that

$$u_i\left(\alpha_i I - R^{(L)}\right) = 0,$$

for $i \geq 0$, which implies the statement of the proposition. □

In a very similar way as before, let us consider the operator $U_\beta = (\tilde{u}_{im})_{i,m=0,1,2,\ldots}$ whose rows are defined by $\tilde{u}_i = [\tilde{u}_{i0}\ \tilde{u}_{i1}\ \tilde{u}_{i2}\ \ldots]$, with $\tilde{u}_{i0} = d_i e_i$, $i \geq 0$ and $\tilde{u}_{im} = d_i(c_{i-1}\alpha_{i-1}^m + c_i\alpha_i^m)$, with $m \geq 1$ and $i \geq 0$.

We can write the elements of the invariant vector $\pi_{m,n}$ for $m \geq 0$ and $n \geq 1$ in function of the elements of U_β in the following way

$$\pi_{m,n} = \sum_{i=0}^{\infty} \beta_i^n \tilde{u}_{im}. \tag{4.29}$$

As we did before, we consider

$$V_\beta = \begin{bmatrix} 1 & 1 & 1 & 1 & \ldots \\ \beta_0 & \beta_1 & \beta_2 & \beta_3 & \ldots \\ \beta_0^2 & \beta_1^2 & \beta_2^2 & \beta_3^2 & \ldots \\ \beta_0^3 & \beta_1^3 & \beta_2^3 & \beta_3^3 & \ldots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix},\ D_\beta = \begin{bmatrix} \beta_0 & & & & \\ & \beta_1 & & & \\ & & \beta_2 & & \\ & & & \beta_3 & \\ & & & & \ddots \end{bmatrix},$$

The relation (4.29) can be written in a operator form as

$$V_\beta D_\beta U_\beta = \begin{bmatrix} \pi_1^{(A)} \\ \pi_2^{(A)} \\ \pi_3^{(A)} \\ \vdots \end{bmatrix},$$

where for each n fixed we have $\pi_n^{(A)} = [\pi_{0,n}\ \pi_{1,n}\ \pi_{2,n}\ \ldots]$.

Now we multiply this equation on the right by the operator $R^{(A)}$, which is the minimal solution of

$$A_1^{(A)} + R^{(A)} A_0^{(A)} + \left(R^{(A)}\right)^2 A_{-1}^{(A)} = 0,$$

and that satisfy $\pi_n^{(A)} R^{(A)} = \pi_{n+1}^{(A)}$, we get

$$V_\beta D_\beta U_\beta R^{(A)} = \begin{bmatrix} \pi_1^{(A)} \\ \pi_2^{(A)} \\ \pi_3^{(A)} \\ \vdots \end{bmatrix} R^{(A)} = \begin{bmatrix} \pi_2^{(A)} \\ \pi_3^{(A)} \\ \pi_4^{(A)} \\ \vdots \end{bmatrix} = V_\beta D_\beta^2 U_\beta. \tag{4.30}$$

Again we can't directly conclude that $U_\beta R^{(A)} = D_\beta S$ because $V_\beta D_\beta$ has not an inverse, but we can obtain the following Proposition that is analogous to 4.1.

Proposition 4.2. *Let us suppose that the following limitation hold*

$$\|\tilde{u}_i \left(\beta_i I - R^{(A)}\right)\|_\infty < \frac{1}{|\beta_i|}, \tag{4.31}$$

then the terms $\{\beta_i\}$ *with* $i \geq 0$ *constitute the different eigenvalues of the operator* $R^{(A)}$*. For an eigenvalue* β_i*, the corresponding left eigenvector of the operator* $R^{(A)}$ *is* $\tilde{u}_i$*, with* $i \geq 0$*.*

We formulate below a proposition that describes the form of the resolvent operator, which will appear naturally when calculating the bivariate probability generating function.

Corollary 4.2. *The resolvent operator of the matrix* $R^{(L)}$ *can be computed in terms of the eigenvalues* $\{\alpha_i\}$ *and the corresponding eigenvectors* $\{u_i\}$ *as follows:*

$$\pi_1^{(A)}(\alpha I - R^{(L)})^{-1} = \sum_{i=0}^{\infty} \frac{\alpha_i}{\alpha - \alpha_i} u_i. \tag{4.32}$$

Proof. First, note that

$$\Pi(x,y) = \sum_{m=0}^{\infty} x^m \pi_m^{(L)} \left[1 \; y \; y^2 \dots\right]^T = \left[\Pi_{:,0}(x) \; \Pi_{:,1}(x) \; \Pi_{:,2}(x) \dots\right] \left[1 \; y \; y^2 \dots\right]^T,$$

where $\Pi_{:,n}(x) = \sum_{m=0}^{\infty} x^m \pi_{m,n}$ for $m \geq 0$. On the one hand, from the definition of $\Pi_{:,n}(x)$ and from equation (4.24), we have, for $m > 0$,

$$\begin{aligned}
\left[\Pi_{:,0}(x) \; \Pi_{:,1}(x) \; \Pi_{:,2}(x) \dots\right] &= \left[\sum_{m=0}^{\infty} x^m \pi_{m,0} \; \sum_{m=0}^{\infty} x^m \pi_{m,1} \; \sum_{m=0}^{\infty} x^m \pi_{m,2} \dots\right] \\
&= \pi_0^{(L)} + \sum_{m=1}^{\infty} x^m \sum_{i=0}^{\infty} \alpha_i^m u_i \\
&= \pi_0^{(L)} \sum_{i=0}^{\infty} \frac{\alpha_i}{x^{-1} - \alpha_i} u_i.
\end{aligned}$$

On the other hand, from equation (2.2), we obtain, for $m > 0$,

$$\begin{aligned}
\left[\Pi_{:,0}(x) \; \Pi_{:,1}(x) \; \Pi_{:,2}(x) \dots\right] &= \sum_{m=0}^{\infty} x^m \pi_m^{(L)} \\
&= \pi_0^{(L)} + \sum_{m=1}^{\infty} x^m \pi_1^{(L)} \left(R^{(L)}\right)^{m-1} \\
&= \pi_0^{(L)} \left(x^{-1} I - R^{(L)}\right)^{-1}.
\end{aligned}$$

Setting $x = \alpha^{-1}$ in the last two equations and equating them we obtain exactly equation (4.32). □

Again an analogous Corollary can be obtain for the operator $R^{(A)}$ and the sequence $\{\beta_i\}$.

Corollary 4.3. *The resolvent operator of the matrix $R^{(A)}$ can be computed in terms of the eigenvalues $\{\beta_i\}$ and the corresponding eigenvectors $\{\tilde{u}_i\}$ as follows:*

$$\pi_1^{(A)}(\beta I - R^{(A)})^{-1} = \sum_{i=0}^{\infty} \frac{\beta_i}{\beta - \beta_i} \tilde{u}_i. \tag{4.33}$$

4.4 Calculation of the invariant measure

In this section, we turn our focus to the calculation of all the coefficients that are involved into the representation of π, beginning from the sequences $\{\alpha_i\}$ and $\{\beta_i\}$. To this purpose, we first compute the bivariate probability generating function in terms of the resolvent of the operator $R^{(L)}$.

$$\begin{aligned}
\Pi(x,y) &= \sum_{m=0}^{\infty} x^m \pi_m^{(L)} \left[1\ y\ y^2 \dots\right]^T \\
&= \left(\pi_0 + \sum_{m=1}^{\infty} x^m \pi_1^{(L)} \left(R^{(L)}\right)^{m-1}\right) \left[1\ y\ y^2 \dots\right]^T \\
&= \left(\pi_0^{(L)} + \pi_1^{(L)} \left(x^{-1} I - R^{(L)}\right)^{-1}\right) \left[1\ y\ y^2 \dots\right]^T.
\end{aligned}$$

Substituting the above result in the functional equation of the boundary value method, yields after some straightforward manipulations

$$\begin{aligned}
&\pi_1^{(L)} \left(x^{-1} I - R^{(L)}\right)^{-1} \left(Q(x,y) \left[1\ y\ y^2 \dots\right]^T + H(x,y) \left[1\ 0\ 0 \dots\right]^T\right) \\
&= -\pi_0^{(L)} \left(\left(Q(x,y) + V(x,y)\right) \left[1\ y\ y^2 \dots\right]^T + \left(H(x,y) + R(x,y)\right) \left[1\ 0\ 0 \dots\right]^T\right).
\end{aligned}$$

Using the result of Corollary (4.2) in the above expression immediately yields

$$\begin{aligned}
&\sum_{i=0}^{\infty} \frac{\alpha_i}{x^{-1} - \alpha_i} u_i \left(Q(x,y) \left[1\ y\ y^2 \dots\right]^T + H(x,y) \left[1\ 0\ 0 \dots\right]^T\right) \\
&= -\pi_0^{(L)} \left(\left(Q(x,y) + V(x,y)\right) \left[1\ y\ y^2 \dots\right]^T + \left(H(x,y) + R(x,y)\right) \left[1\ 0\ 0 \dots\right]^T\right).
\end{aligned}$$

Equivalently, by defining $\mathbf{e}_j$ to be an infinite-dimension column vector with a 1 in the j-th position and 0 elsewhere, the above equation reduces to

$$\sum_{i=0}^{\infty} \frac{\alpha_i}{x^{-1}-\alpha_i} u_i \left(Q(x,y) \sum_{j=1}^{\infty} y^{j-1} \mathbf{e}_j + H(x,y)\mathbf{e}_1 \right) = -\pi_0^{(L)} \left((Q(x,y)+V(x,y)) \sum_{j=1}^{\infty} y^{j-1}\mathbf{e}_j + (H(x,y)+R(x,y))\,\mathbf{e}_1 \right).$$

After straightforward calculations, the above can be equivalently written as

$$\sum_{i=0}^{\infty} \frac{\alpha_i}{x^{-1}-\alpha_i} \sum_{j=1}^{\infty} \left(Q(x,y) y^{j-1} u_{ij-1} + H(x,y) u_{i0} \mathbb{1}(j=1) \right) = -\sum_{j=1}^{\infty} \left((Q(x,y)+V(x,y))\, y^{j-1} \pi_{0,j-1} + (H(x,y)+R(x,y))\, \pi_{0,0} \mathbb{1}(j=1) \right). \tag{4.34}$$

For the recursive calculation of the sequences $\{\alpha_i\}$ and $\{\beta_i\}$, the main idea lies on defining the zero couples (x,y), such that $|x|,|y|<1$, such that $Q(x,y)=0$. To this purpose, we multiply (4.34) by $x^{-1}-\alpha_0$ and then we take the limit as x goes to α_0^{-1}, this leads to

$$\sum_{j=1}^{\infty} \left(Q(\alpha_0^{-1},y) y^{j-1} u_{ij-1} + H(\alpha_0^{-1},y) u_{i0} \mathbb{1}(j=1) \right) = 0.$$

Restricting the investigation on the set of y-roots that satisfy $Q(\alpha_0^{-1},y)=0$, the above equation yields that $H(\alpha_0^{-1},y)=0$, since $p_{00}\neq 0$. Thus, choosing the α_0 such that $Q(\alpha_0^{-1},y)=0$ and $H(\alpha_0^{-1},y)=0$ reveals the starting solution for the iterative calculation of the sequences $\{\alpha_i\}$ and $\{\beta_i\}$.

The existence of the solution α_0 inside the unit disk is equivalent to the existence of feasible solution for the compensation approach and it is ensured by theorem 4.2.

As we have already noticed equation (4.19) can be written as $Q(\alpha^{-1},\beta^{-1})=0$, because of this the bivariate polynomial $Q(x,y)$ can be factorized by means of (4.20) as

$$Q(x,y) = f(x)(y-y_+(x))(y-y_-(x)) = g(y)(x-x_+(y))(x-x_-(y)),$$

for some polynomial f and g of one variable. This yields

$$y_+(\alpha_i^{-1})=\beta_i^{-1}, \quad y_-(\alpha_i^{-1})=\beta_{i-1}^{-1}, \quad x_+(\beta_i^{-1})=\alpha_{i-1}^{-1}, \quad x_-(\beta_i^{-1})=\alpha_i^{-1}.$$

The above equations produce explicitly the sequence $\{\alpha_i\}$ and $\{\beta_i\}$ with "forward" operators y_+ and x_+ and "backward" operators y_- and x_- constructed as the solutions of the equation $Q(x, y) = 0$ with respect to y and x as shown in Figure 4.3.

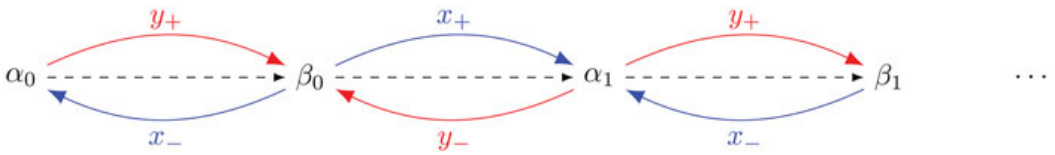


Figure 4.3: The recursive structure of the product-form terms.

The building of the vector π is completed by calculating the other coefficients involved with equations (4.11)-(4.15).

Chapter 5

Numerical Results

In this chapter we compare the algorithms obtained from the two approaches presented in the thesis. The examples that we consider are taken from [12] and [16], all the codes are written in Matlab as well as the scripts of all the experiments, only the quadratic equations system is solved symbolically with the software Sage.

5.1 Join the Shortest Queue

Consider a system with two identical servers. Jobs arrive according to a Poisson stream with rate 2ρ where $0 < \rho < 1$. On arrival a job joins the shortest queue, in the case of a tie it chooses either queue with probability $\frac{1}{2}$. The jobs require exponentially distributed service times with unit mean, the service times are supposed to be independent. This model is known as the symmetric shortest queue model.

Such a queueing system is modelled as a Markov process with states $(w_1, w_2) \in \mathbb{Z}_+^2$, where w_i is the number of customers at queue i, including a customer possibly in service. By defining $m = \min(w_1, w_2)$ and $n = w_2 - w_1$, one transforms the state space from a homogeneous random walk in the quadrant to a homogeneous random walk in the half-plane, where the two quadrants are symmetrical. The transition rate diagram of the Markov process for $n, m \geq 0$ is shown in Figure 5.1.

The boundary value problem polynomials in this specific case become

$$\begin{aligned} Q(x,y) &= y^2 + 2\rho x^2 + x - xy(2 + 2\rho), \\ H(x,y) &= \rho x y^2 - 2\rho x^2 - x + xy(1 + \rho), \\ V(x,y) &= -y^2 + xy, \end{aligned}$$

we can find that the solution of the systems

$$\begin{cases} Q(x,y) = 0, \\ H(x,y) = 0, \end{cases} \qquad \begin{cases} Q(x,y) = 0, \\ V(x,y) = 0, \end{cases}$$

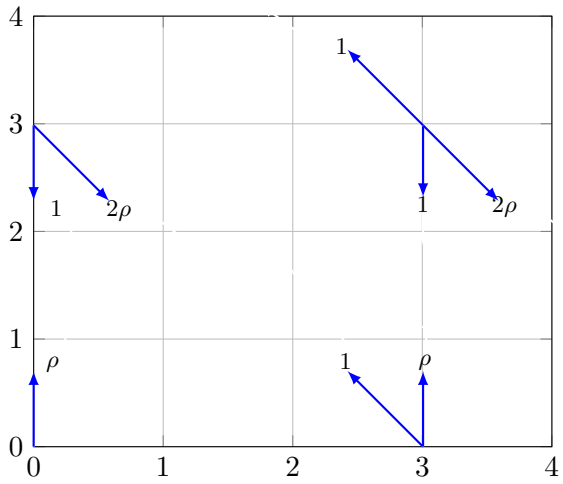


Figure 5.1: Transition rates for the JSQ model.

are

$$\left(-\frac{1}{2\rho},0\right),\ (0,0),\ \left(\frac{1}{\rho^2},\frac{1}{\rho}\right),\ (1,1),\ \left(-\frac{1}{2\rho},-\frac{\rho+1}{\rho}\right),$$
$$(0,0),\ \left(-\frac{1}{2\rho},0\right),\ (1,1),$$

so we get only the following feasible pair

$$(\alpha_0,\beta_0)=\left(\rho^2,\frac{\rho^2}{\rho+2}\right),$$

and it is related to the horizontal boundary.

Moreover, it can be easily verified that

$$A_\pm=\rho+1\pm\sqrt{\rho^2+1},\quad \eta=\frac{A_+}{A_-},\quad \gamma=\frac{2\rho A_-^{-1}-(2\rho+1)}{2\rho A_+^{-1}-(2\rho+1)}=\frac{1-A_-}{1-A_+},$$

so the integer N, which appears in Theorem 4.4, is the smallest integer such that

$$|\gamma\eta|\left(\frac{A_-}{A_+}\right)^{N+1}=\left|\frac{1-A_-}{1-A_+}\right|\left(\frac{A_-}{A_+}\right)^{N}=\left|\frac{\rho-\sqrt{\rho^2+1}}{\rho+\sqrt{\rho^2+1}}\right|\left(\frac{A_-}{A_+}\right)^{N}<1,$$

it is evident that in this case $N=0$.

From these considerations, we obtain that the compensation approach can be applied, in particular equation (4.23) reduces to just one series and produces the

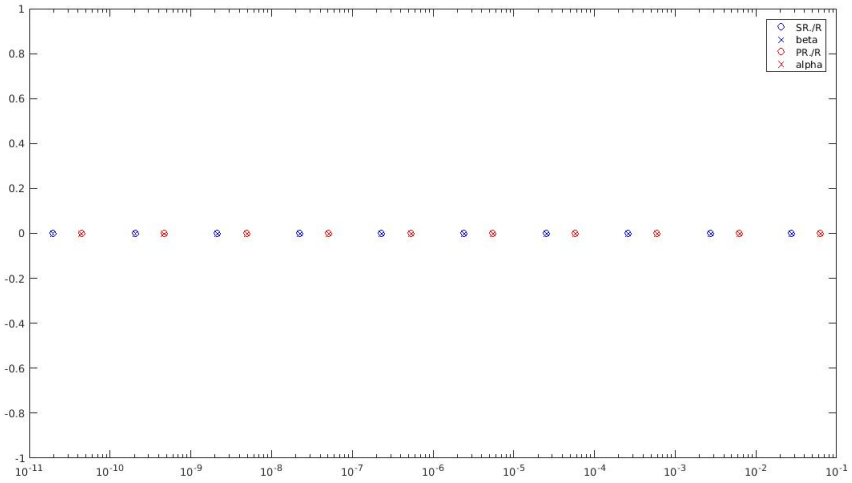


Figure 5.2: The sequences $\{\alpha_i\}$ and $\{\beta_i\}$ in the complex plane. The real axis is in logarithmic scale.

invariant vector component $\pi_{m,n}$ for $m, n > 0$, the remaining component $\pi_{0,0}$ is obtained using (4.1).

In Figure 5.2 we can see, in red, the first 10 elements of the sequence $\{\alpha_i\}$ and, in blue, the first 10 elements of the sequence $\{\beta_i\}$, let us observe as they converge exponentially fast to 0. The circles are obtained by multiplying the first 10 rows of the eigenvectors operators U_β and U_α built with the compensation approach by the operators $R^{(L)}$ and $R^{(A)}$ obtained with the Cyclic Reduction in the $\mathcal{QT}_1$ arithmetic, and then dividing the first components of these vectors by the first components of the selected rows of U_β and U_α.

5.1.1 Relative error

The absolute error is defined by means of the quantities

$$\|\pi Q\|_\infty \quad \text{or} \quad \|\pi P - \pi\|_\infty,$$

depending on which model we are using (Markov process or Markov chain, respectively). It is not a good measure of how much precisely we are calculating π, because, since the invariant vector is stochastic, then it has a decay to 0, so its components become very small in magnitude; the absolute error may report an error which appears small in general but actually is as small as the value we are calculating.

Because of this, it is more convenient to introduce the relative error which is defined as

$$\|(\pi Q) \oslash \pi\|_\infty \quad \text{or} \quad \|(\pi P - \pi) \oslash \pi\|_\infty,$$

where $\oslash$ is the operation of componentwise division.

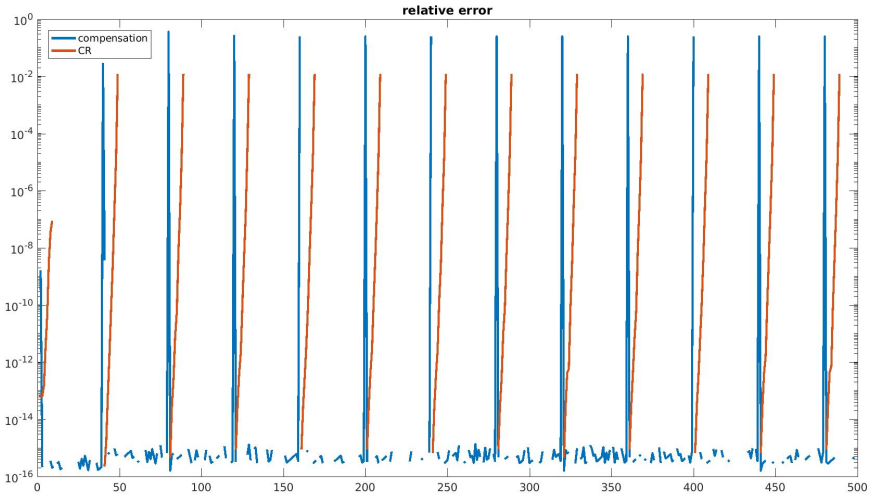
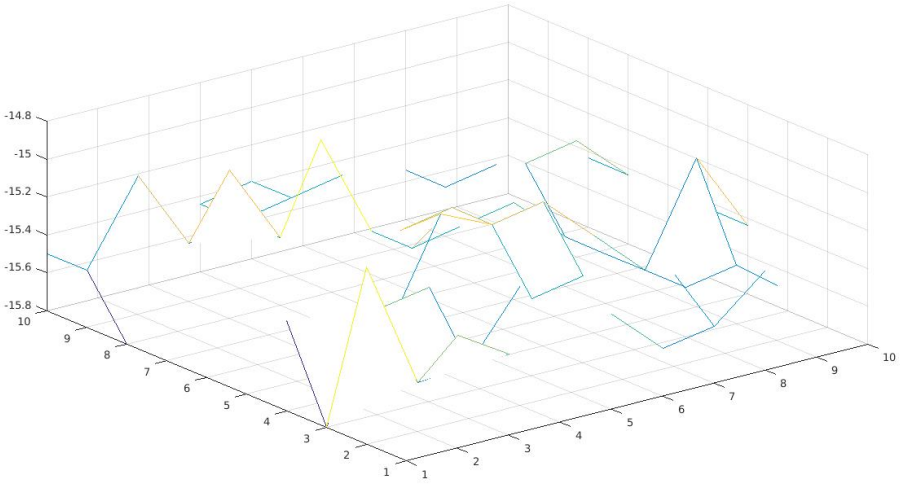


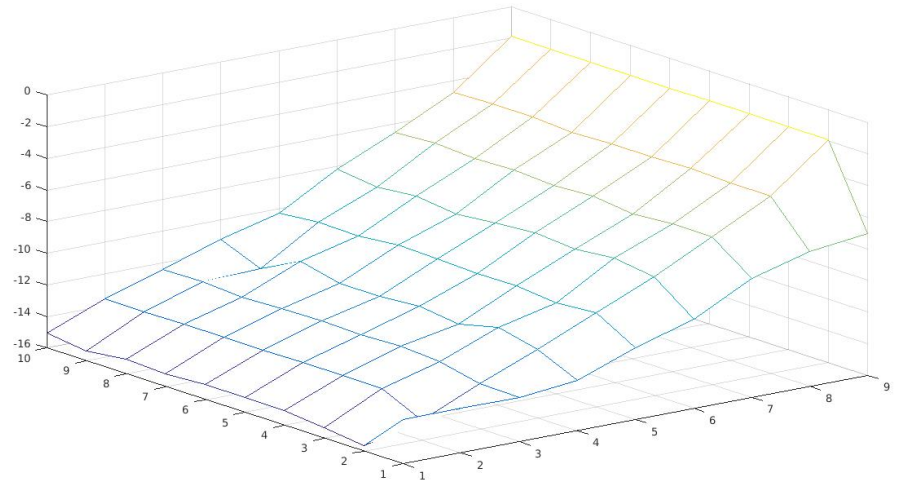
Figure 5.3: Relative errors of the two methods for $\rho = \frac{1}{4}$. The vertical axis is in logarithmic scale.

Our goal is to compare the relative errors produced by Cyclic Reduction and compensation approach. In the first experiment we fix $\rho = \frac{1}{4}$ in the JSQ, we calculate a truncation of the invariant vector with both the methods, and finally we plot the vectors $|(\pi Q) \oslash \pi|$ and $|(\pi P - \pi) \oslash \pi|$, in this context the absolute value is meant component-wise. As we can see in Figure 5.3, the error of the compensation approach is in general smaller than the error of the Cyclic Reduction, in particular in this case, some components of the error vector are not calculated because of the truncations in the $\mathfrak{QT}_1$ arithmetic. Moreover for both the methods we can observe an increase of the relative error where the components of π become smaller.

In Figure 5.4 we can observe the three-dimensional plots of the relative error of both methods for $\pi_{m,n}$ with $m, n \leq 10$, we limit to this square because it turns out to be the smallest square with all components smaller than the machine precision outside of it. As we can see the relative error for these components of π is almost equal to the machine precision in the case of the compensation approach, while in the case of the Cyclic Reduction in certain components it is not calculated because of truncation problem and when it is calculated it reaches the maximum value of 10^{-2} which is still acceptable.



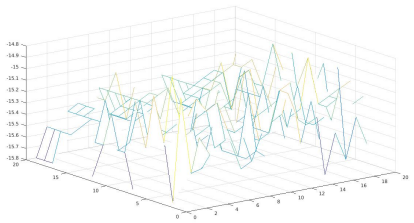
(a) Relative error of the compensation approach.



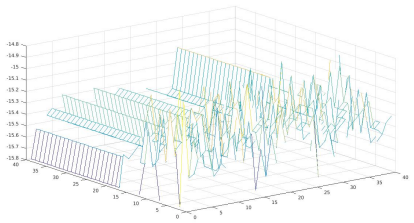
(b) Relative error of the Cyclic Reduction algorithm.

Figure 5.4: The relative errors for the JSQ example with $\rho = \frac{1}{4}$ for $\pi_{m,n}$ with $m, n \leq 10$. Both the graphics are in logarithmic scale on the vertical axes.

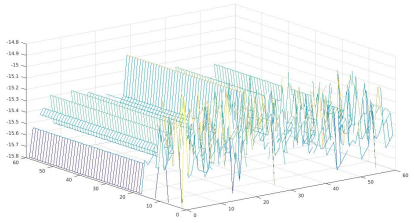
From these remarks it is evident that the compensation approach achieves better results than the Cyclic Reduction algorithm from the point of view of the relative error. This is due to the fact that in the compensation approach the vector π is calculated by truncating a series for each component. The larger is the number of terms we use in the truncation, the smaller the errors (relative and absolute). In Figure 5.5 we can see the three dimensional plots of the relative error produced with the compensation approach in which each series is truncated after 80 terms.



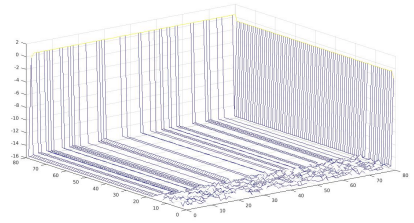
(a) $N_t = 20$.



(b) $N_t = 40$.



(c) $N_t = 60$.



(d) $N_t = 80$.

Figure 5.5: The relative errors of the compensation approach for the JSQ example with $\rho = \frac{1}{4}$ for $\pi_{m,n}$ with $m, n \leq N_t$ for successive values of N_t. All the graphics are in logarithmic scale on the vertical axes.

5.1.2 Behaviour on null recurrent state

In this section we study what are the effects of changing the value of the parameter ρ on the behaviour of both the approaches, in particular we are interested in the case $\rho = 1$ which produces a null recurrent process.

Before doing this, we first observe that in the cases $\rho \geq \frac{1}{2}$, the Cyclic Reduction algorithm has a problem related to the size of the correction part of the operator $A_0^{(h)}$. Indeed, as the iterations go on, this size becomes so large that exceeds the available memory.

We can overcome this problem by modeling the JSQ with the anti-lexicographic order instead of the lexicographic that is the one considered until now.

Remark 5.1. Changing the model according to the way we order the states (lexicographic or anti-lexicographic) is equivalent to consider $\Sigma P \Sigma^T$ instead of the transition operator P, where Σ is the permutation operator that swaps the two factors of a Kronecker product, that is

$$\Sigma (A \otimes B) \Sigma^T = B \otimes A,$$

for each couple of operators A and B. It is also equivalent to just simply consider the random walk in the quarter plane with the axes swapped.

It's interesting to note that, with this trick, the Cyclic Reduction manages to compute $R^{(L)}$ even when $\rho \geq 1$ that represent the null recurrent and transient states of the JSQ. The problem we have in these cases appears in the computation of π, indeed the operator on which we apply the power methods to compute π_0 has the first two larger eigenvalues very close to each other, so the iterations are very slow and we are forced to arrest them with lower precision.

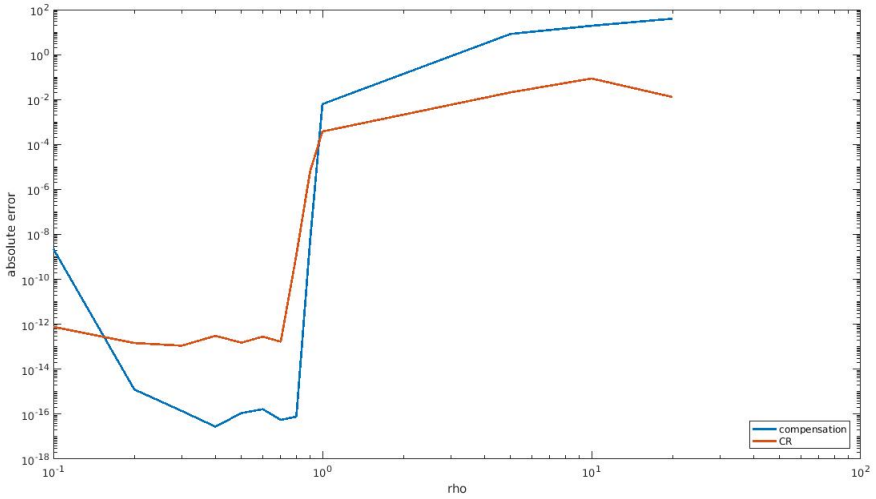


Figure 5.6: Absolute errors of the two methods in function of the parameter ρ. Both the axis are in logarithmic scale.

As we can see in Figure 5.6, both the algorithms increase their absolute errors in correspondence of the case $\rho = 1$ but this is due to different reasons.

As we said, the Cyclic Reduction computes without any problem the solution of the quadratic operator equations, but when $\rho \geq 1$ the power method we use to compute π_0 is very slow. Therefore, in order to make it finish, we have to increase its stopping threshold and this induces an increasing of the absolute error.

On the other hand the compensation approach always computes the same truncated series which simply become meaningless in the cases $\rho \geq 1$.

5.2 Two-demand Model

Double queue arises when customers arriving at the system simultaneously place two demands on two different servers working independently. The customer arrivals form a Poisson process with rate 1, and the service times of the two servers are independent and exponential with rates a and b, respectively. Let $X_1(t)$ and $X_2(t)$ represent the number of customers waiting or in service at time t in queues 1 and 2, respectively. We consider the two-dimensional Markov chain $(X_1(t), X_2(t))$ with state space $\mathbb{Z}_+^2$, viewed as a QBD process, whose transition are shown in Figure 5.7.

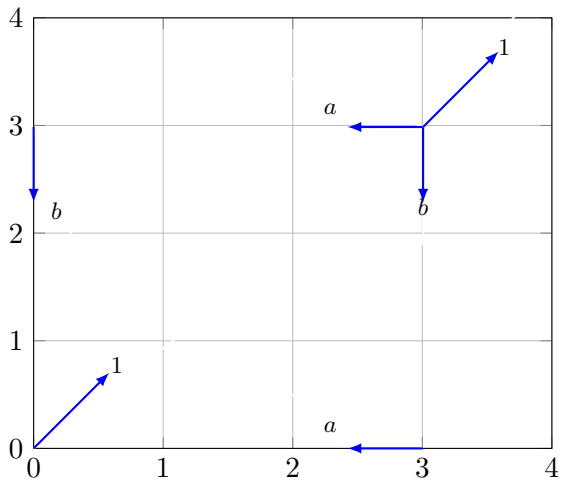


Figure 5.7: Transition rates for the two-demand model.

As we can see from Figure 5.7, this problem has a strongly symmetric structure, it is easy to note that changing the order adopted for the model is equivalent to swapping the parameters a and b, moreover it is evident that in the case $a = b$ the model is invariant for reflection on the bisector.

This symmetry has consequences also from the experimental point of view. Indeed it turns out that when $b > a$ the Cyclic Reduction applied to the model with the lexicographic order has problems related to the available memory that are analogous to the ones of the JSQ for $\rho = \frac{1}{2}$. The same kind of problems are found in the case $a > b$ with the anti-lexicographic order.

Because of these observations we fix $b \leq a$ and we work only with the model with the lexicographic order.

As pointed out in [16], the Markov chain related to this example is positive recurrent when $a > 1$ and $b > 1$. In the following experiment we consider $B_0 + B_1 G$ that is the operator on which we apply the power method, and we calculate the eigenvalues λ_i of its truncation. In Table 5.1 we can see how the ratio $\omega = \frac{\lambda_1}{\lambda_2}$ between the two largest eigenvalues, which measures the power method's convergence speed, changes in relation with the value of the parameters a and b. The variable κ represent the lowest integer such that $|\omega|^\kappa$ is lower than the machine precision. As we expected, it seems that the closer we get to the null recurrent and transient states $a, b \leq 1$, the more iterations of the power method are needed.

a	b	ω	κ
10	10	0.5105	55
10	5	0.6718	93
5	5	0.6057	74
5	2	0.9017	357
5	1	0.9703	1222
2	1	0.9581	861
2	0.5	0.9597	896
0.5	0.5	0.9514	740

Table 5.1: The module of the ratio between the two largest eigenvalues of the power method operator related to the value of the parameters a and b.

5.3 Failures of the compensation approach

As we have seen in previous sections, the compensation approach produces more accurate results than the $\mathcal{QT}_1$ arithmetic Cyclic Reduction algorithm. Its weakness is that we can't apply it to every random walk. Indeed, as in the case of the two demand model, the conditions (4.17) are not satisfied and the convergence of the compensation approach series is not ensured.

Furthermore this is not the only case in which the compensation fails, indeed let us consider the process whose transitions are shown in Figure 5.8.

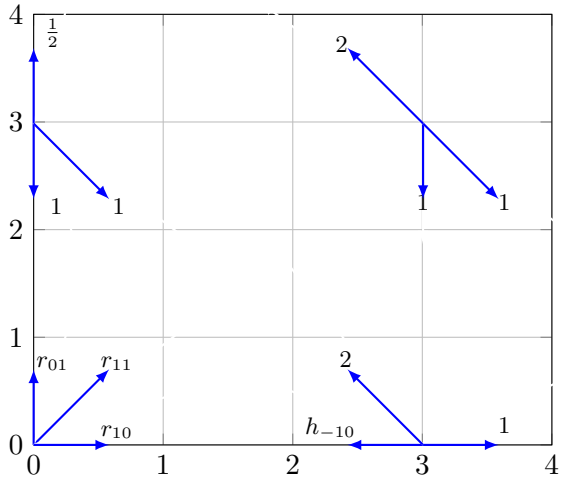


Figure 5.8: Transition rates of a random walk for which the compensation fails.

The boundary value problem polynomials in this specific case become

$$\begin{aligned} Q(x,y) &= 2y^2 + x^2 + x - 4xy, \\ H(x,y) &= x^2 y - x^2 - x + xy, \\ V(x,y) &= \frac{1}{2}y^2 x - 2y^2 + \frac{3}{2}xy, \end{aligned}$$

it's easy to verify that

$$(\alpha_0, \beta_0) = \left(\frac{1}{2}, \frac{1}{3}\right)$$

is a feasible pair related to the horizontal boundary.

Although this process satisfies conditions (4.17), the compensation approach stops at the first iteration since the denominator in the computation of c_1 in formula (4.11) becomes

$$\frac{\beta_0^2}{\alpha_0} + \frac{1}{2} + \beta_0^2 - \frac{5}{2}\beta_0 = 0.$$

On the other side the Cyclic Reduction algorithm has not this kind of limitation in its application and it succeeds in calculating the invariant vector of probability of the process in Figure 5.8. The relative error is similar to the one of the JSQ example as we can see in Figure 5.9.

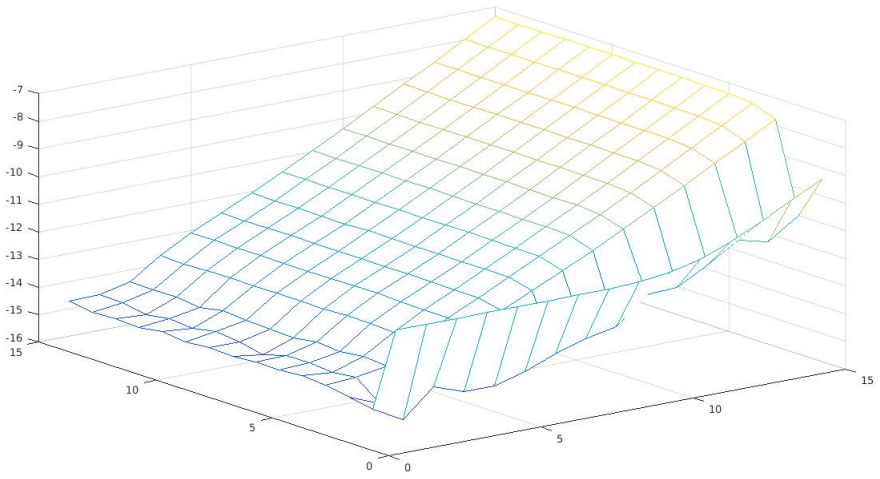


Figure 5.9: Three dimensional plot of the relative error of the Cyclic Reduction. The vertical axes is in logarithmic scale.

5.4 Conclusions

From the comparison between the two approaches, it turns out that the applicability of the compensation approach is more restricted than the operator approach. Indeed, in addition to the restriction on the transitions expressed in (4.17), we have seen other examples for which the compensation approach, for different reasons, fails while the operator approach has not any problem.

On the other hand the compensation approach shows a better performance in terms of accuracy. In fact it provides approximation with very small relative error, whereas the operator approach in the current implementation does not maintain a uniform bound to the relative error, even though the absolute error in the approximation is quite small.

Another difference concerns the behaviour on null recurrent states. In these cases the compensation approach computes the series in the same way as in the positive recurrent cases, but these series simply become meaningless, indeed the absolute error is significant. In these cases the absolute error is significant also for the operator approach but this happens because the two largest eigenvalues of the operator $B_0 + B_1 G$ are very close to each other and so the power method becomes very slow. Despite of this, the operators R and G are computed with large precision even in the null recurrent cases with the Cyclic Reduction algorithm.

Appendix

Now we report the code of the Matlab function for the compensation approach. For what concernes the operator approach, in [7] we can find a complete Matlab toolbox for the $\mathcal{QT}_1$ arithmetic as well as the implementation of the Cyclic Reduction algorithm.

```
function [alpha,beta,c,d,e,f,Pi]=compensation(q,h,v,r,N,a0,bm1)
sq = sum(sum(q));
sh = sum(sum(h));
sr = sum(sum(r));
sv = sum(sum(v));
beta = zeros(N,1);
c = zeros(N,1);
d = zeros(N,1);
e = zeros(N,1);
f = zeros(N,1);
alpha = zeros(N,1);
alpha(1) = a0;
c(1) = 1;
d(1) = 1;


for k=1:N
    den = q(3,1)/(alpha(k)^2)+q(2,1)/alpha(k)+q(1,1);
    deltax = (q(1,2)-sq/alpha(k))^2-4*q(1,3)*den;
    beta(k) = (2*q(1,3))/(sq/alpha(k)-q(1,2)+sqrt(deltax));
    deltay = (q(2,1)-sq/beta(k))^2-4*q(3,1)*(q(1,3)/(beta(k)^2)
                                            + q(1,2)/beta(k)+q(1,1));
    alpha(k+1) = 2*q(3,1)/(sq/beta(k)-q(2,1)+sqrt(deltay));

    num = (beta(k)^2*v(3,1)+beta(k)*v(3,2)+v(3,3))/alpha(k+1)
                                    + v(2,3)+beta(k)^2*v(2,1)-beta(k)*sv;
    den = (beta(k)^2*v(3,1)+beta(k)*v(3,2)+v(3,3))/alpha(k)
                                    + v(2,3)+beta(k)^2*v(2,1)-beta(k)*sv;
    c(k+1) = -num*c(k)/den;

    num = (beta(k)^2*q(3,1) + beta(k)*q(3,2)
                                     + q(3,3))*(1/alpha(k+1)-1/alpha(k));
    e(k) = -num*c(k)/den;
```

```
34
35      if k>1
36
37          num = (alpha(k)^2*h(1,3)+alpha(k)*h(2,3)+h(3,3))/beta(k)
38                                      + h(3,2)+alpha(k)^2*h(1,2)-alpha(k)*sh;
39          den = (alpha(k)^2*h(1,3)+alpha(k)*h(2,3)+h(3,3))/beta(k-1)
40                                      + h(3,2)+alpha(k)^2*h(1,2)-alpha(k)*sh;
41          d(k) = -num*d(k-1)/den;
42
43          num = (alpha(k)^2*q(1,3) + alpha(k)*q(2,3)
44                                          + q(3,3))*(1/beta(k)-1/beta(k-1));
45          f(k) = -num*d(k-1)/den;
46      end
47  end
48
49  num = (alpha(1)^2*q(1,3)+alpha(1)*q(2,3)+q(3,3))*(1/bm1-1/beta(1));
50  den = (alpha(1)^2*h(1,3) + alpha(1)*h(2,3)+h(3,3))/beta(1)
51                                      + h(3,2)+alpha(1)^2*h(1,2)-alpha(1)*sh;
52  f(1) = -num/den;
53
54
55  Pi = zeros(N);
56
57  for mm=2:N
58      for nn=2:N
59          for i=1:N
60              Pi(mm,nn) = Pi(mm,nn) + d(i)*(c(i)*alpha(i)^(mm-1)
61                                  + c(i+1)*alpha(i+1)^(mm-1))*beta(i)^(nn-1);
62          end
63      end
64  end
65
66  for nn=2:N
67      for i=1:N
68          Pi(nn,1) = Pi(nn,1) + c(i)*f(i)*alpha(i)^(nn-1);
69          Pi(1,nn) = Pi(1,nn) + d(i)*e(i)*beta(i)^(nn-1);
70      end
71  end
72
73
74  Pi(1,1) = (q(1,1)*Pi(2,2)+h(1,2)*Pi(2,1)+v(2,1)*Pi(1,2))/sr;
75  Pi = Pi/sum(sum(Pi));
```

Bibliography

[1] I. J.-B. F. Adan, J. Wessels, and W. H. M. Zijm. A compensation approach for two-dimensional Markov processes. *Adv. in Appl. Probab.*, 25(4):783–817, 1993.

[2] Ivo Jean-Baptiste François Adan. *A compensation approach for queueing problems.* Technische Universiteit Eindhoven, Eindhoven, 1991. Dissertation, Technische Universiteit Eindhoven, Eindhoven, 1991, With a Dutch summary.

[3] D. A. Bini, L. Gemignani, and B. Meini. Computations with infinite Toeplitz matrices and polynomials. *Linear Algebra Appl.*, 343/344:21–61, 2002. Special issue on structured and infinite systems of linear equations.

[4] D. A. Bini, G. Latouche, and B. Meini. *Numerical methods for structured Markov chains.* Numerical Mathematics and Scientific Computation. Oxford University Press, New York, 2005. Oxford Science Publications.

[5] DA Bini, S Massei, B Meini, and L Robol. On quadratic matrix equations with infinite size coefficients encountered in qbd stochastic processes. *Numerical Linear Algebra with Applications*, 2017.

[6] Dario Bini, Stefano Massei, and Beatrice Meini. Semi-infinite quasi-toeplitz matrices with applications to qbd stochastic processes. *Mathematics of Computation*, 2018.

[7] Dario A Bini, Stefano Massei, and Leonardo Robol. Quasi-toeplitz matrix arithmetic: a matlab toolbox. *arXiv preprint arXiv:1801.08158*, 2018.

[8] Dario A. Bini and Beatrice Meini. The cyclic reduction algorithm: from Poisson equation to stochastic processes and beyond. In memoriam of Gene H. Golub. *Numer. Algorithms*, 51(1):23–60, 2009.

[9] Albrecht Böttcher and Sergei M. Grudsky. *Spectral properties of banded Toeplitz matrices.* Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2005.

[10] Haim Brezis. *Functional analysis, Sobolev spaces and partial differential equations*. Universitext. Springer, New York, 2011.

[11] Guy Fayolle, Roudolf Iasnogorodski, and Vadim Malyshev. *Random walks in the quarter plane*, volume 40 of *Probability Theory and Stochastic Modelling*. Springer, Cham, second edition, 2017. Algebraic methods, boundary value problems, applications to queueing systems and analytic combinatorics.

[12] Stella Kapodistria and Zbigniew Palmowski. Matrix geometric approach for random walks: Stability condition and equilibrium distribution. *Stoch. Models*, 33(4):572–597, 2017.

[13] G. Latouche and V. Ramaswami. *Introduction to matrix analytic methods in stochastic modeling*. ASA-SIAM Series on Statistics and Applied Probability. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA; American Statistical Association, Alexandria, VA, 1999.

[14] A. S. Markus. *Introduction to the spectral theory of polynomial operator pencils*, volume 71 of *Translations of Mathematical Monographs*. American Mathematical Society, Providence, RI, 1988. Translated from the Russian by H. H. McFaden, Translation edited by Ben Silver, With an appendix by M. V. Keldysh.

[15] Masakiyo Miyazawa. Light tail asymptotics in multidimensional reflecting processes for queueing networks. *TOP*, 19(2):233–299, 2011.

[16] Allan J. Motyer and Peter G. Taylor. Decay rates for quasi-birth-and-death processes with countably many phases and tridiagonal block generators. *Adv. in Appl. Probab.*, 38(2):522–544, 2006.

[17] J. R. Norris. *Markov chains*, volume 2 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 1998. Reprint of 1997 original.

[18] Alexandre Ostrowski. Recherches sur la méthode de Graeffe et les zéros des polynomes et des séries de Laurent. *Acta Math.*, 72:99–155, 1940.