Quasi-birth-death processes with an explicit rate matrix

TU/e technische universiteit eindhoven

SPOR-Report 2004-08

Quasi-birth-death processes with an explicit rate matrix

J.S.H. van Leeuwaarden
E.M.M. Winands

SPOR-Report
Reports in Statistics, Probability and Operations Research

Eindhoven, May 2004
The Netherlands

/department of mathematics and computing science

SPOR-Report
Reports in Statistics, Probability and Operations Research

Eindhoven University of Technology
Department of Mathematics and Computing Science
Probability theory, Statistics and Operations research
P.O. Box 513
5600 MB Eindhoven - The Netherlands

Secretariat: Main Building 9.10
Telephone: + 31 40 247 3130
E-mail: wscosor@win.tue.nl
Internet: http://www.win.tue.nl/math/bs/cosor.html

ISSN 1567-5211

Quasi-birth-death processes with an explicit rate matrix

J.S.H. van Leeuwaarden[1] & E.M.M. Winands[2]

[1] EURANDOM
P.O. Box 513, 5600 MB Eindhoven, The Netherlands

[2] Department of Mathematics and Computer Science and
Department of Technology Management
Eindhoven University of Technology
P.O. Box 513, 5600 MB Eindhoven, The Netherlands

May 2004

Abstract

We determine the equilibrium distribution for a class of quasi-birth-death (QBD) processes using the matrix-geometric method, which requires the determination of the rate matrix R. In contrast to most QBD processes, the class under consideration allows for an explicit description of R, yielding an explicit description of the equilibrium distribution. We obtain R by exploiting its probabilistic interpretation and show that the problem of finding each element of R reduces to counting lattice paths in the transition diagram. The counting problem is resolved using an extension of the classical Ballot theorem. Throughout the paper, we give examples of queueing models that fit into the class.

Key words: Quasi-birth-death processes, matrix-geometric method, equilibrium distribution, rate matrix, lattice path counting.

1 Introduction

Consider a Markov process on the two-dimensional state space $\{(i,j)|i \geq 0, 0 \leq j \leq H\}$, and refer by level n to the set of states $\{(n,0),(n,1),\ldots,(n,H)\}$. Such a Markov process is called a homogeneous quasi-birth-death (QBD) process when one-step transitions are restricted to states in the same level or in two adjacent levels, and the transition rates are assumed to be level independent.

A well-known method for finding the stationary distribution of QBD processes is the matrix-geometric method. With $\pi(i,j)$ the stationary probability of the process being in state (i,j), and using the vector notation $\pi_n = (\pi(n,0),\ldots,\pi(n,H))$, the probability vectors can be expressed as

$$\pi_{n+1} = \pi_n R, \quad n \geq 1, \tag{1}$$

where the so-called rate matrix R is the minimal nonnegative solution of a nonlinear matrix equation. In most applications, R needs to be computed by using an iterative algorithm. We present, however, a class of QBD processes for which the rate matrix R can be determined explicitly, based on probabilistic arguments. Each element of R can be found separately by monitoring the QBD process from the time it leaves a certain level until it returns to that same level for the first time. For the class to be presented, this reduces to counting lattice paths in the transition diagram, which can be done using an extension of the classic Ballot theorem, see e.g. [5] or [8]. We give various examples of QBD processes that fall within the presented class.

2 Matrix-geometric method

2.1 QBD processes and matrix-geometric solution

For a homogeneous QBD process as described in Section 1, we order the states lexicographically, i.e.

$$\{(0,0),\ldots,(0,H),(1,0),\ldots,(1,H),\ldots,(n,0),\ldots,(n,H),\ldots\},$$

and assume that the infinitesimal generator Q has the following block tridiagonal structure:

$$Q=\begin{pmatrix} B_1 & B_0 & 0 & 0 & 0 & \ldots \\ B_2 & A_1 & A_0 & 0 & 0 & \ldots \\ 0 & A_2 & A_1 & A_0 & 0 & \ldots \\ 0 & 0 & A_2 & A_1 & A_0 & \\ \vdots & \vdots & & \ddots & \ddots & \ddots \end{pmatrix}, \tag{2}$$

where A_0, A_1 and A_2 are square matrices of order $H+1$. The matrices A_0, A_2, B_0 and B_2 are nonnegative and the matrices B_1 and A_1 have nonnegative off-diagonal elements and strictly negative diagonals. We denote the diagonal elements of B_1 and A_1 by Δ, which are such that the row sums of Q equal zero.

The QBD process driven by Q is ergodic if and only if it satisfies the mean drift condition (see [6])

$$\omega A_0 e < \omega A_2 e, \tag{3}$$

where $\omega = (\omega_0, \ldots, \omega_H)$ is the equilibrium distribution of the generator $A_0 + A_1 + A_2$ and e the unity vector. When (3) is satisfied, the stationary distribution of the QBD process exists. Denoting by $\pi(i,j)$ the stationary probability of the process being in state (i,j), and using the vector notation $\pi_n = (\pi(n,0),\ldots,\pi(n,H))$, the balance equations of the QBD process are given by

$$\pi_{n-1}A_0 + \pi_n A_1 + \pi_{n+1}A_2 = 0, \quad n \geq 2, \tag{4}$$

and

$$\begin{aligned} \pi_0 B_1 + \pi_1 B_2 &= 0, & (5)\\ \pi_0 B_0 + \pi_1 A_1 + \pi_2 A_2 &= 0. & (6) \end{aligned}$$

Introducing the rate matrix R as the minimal nonnegative solution of the nonlinear matrix equation

$$A_0 + RA_1 + R^2 A_2 = 0, \tag{7}$$

it can be proved that the equilibrium probabilities satisfy (see e.g. [6])

$$\pi_{n+1} = \pi_n R, \quad n \geq 1. \tag{8}$$

The vectors π_0 and π_1 follow from the boundary conditions (5-6) and the normalization condition

$$\sum_{i=0}^{\infty}\sum_{j=0}^{H}\pi(i,j) = \pi_0 e + \pi_1 (I-R)^{-1} e = 1, \tag{9}$$

where I represents the identity matrix.

In order to obtain the stationary distribution, one should thus determine the rate matrix R. Several iterative procedures exist for solving (7). For example, the modified SS method (see e.g. [2]) uses the following scheme

$$R^{(k+1)} = -(A_0 + R^{(k)^2} A_2) A_1^{-1}, \quad k = 0, 1, \ldots, \tag{10}$$

starting with $R^{(0)}$ a matrix of zero-entries only. An overview of other algorithms is given in [3]. However, for the class of QBD process to be described in Section 3, no approximate method like (10) is needed to determine R.

2.2 QBD processes with an explicit rate matrix

To the best of our knowledge, only one class of QBD processes with an explicit description of the rate matrix R has appeared in the literature. This class consists of QBD processes for which A_0 or A_2 is of rank 1, see [7]. Then, the following result holds:

Theorem 1. *Assume (3) is satisfied and $A_2 = \nu \cdot \alpha$, where ν is a column vector and α a row vector normalized by $\alpha e = 1$. Then*

$$R = -A_0 (A_1 + A_0 e \alpha)^{-1}. \tag{11}$$

Proof Substituting $A_2 = \nu \cdot \alpha$ into (4) yields

$$\pi_{n-1} A_0 + \pi_n A_1 + \pi_{n+1} \nu \cdot \alpha = 0, \quad n \geq 2. \tag{12}$$

To eliminate π_{n+1} from (12) we employ the following relation that results from equating the flow between level n and $n+1$, i.e.

$$\pi_n A_0 e = \pi_{n+1} A_2 e = \pi_{n+1} \nu \cdot \alpha e = \pi_{n+1} \nu, \quad n \geq 2. \tag{13}$$

Hence, substituting (13) into (12) gives

$$\pi_n = \pi_{n-1}[-A_0 (A_1 + A_0 e \alpha)^{-1}], \quad n \geq 2, \tag{14}$$

which completes the proof. □

In case A_0 is the product of a column vector and a row vector a similar property holds, where we refer the reader to [7] for more details. In the next section we expand the class of QBD processes with an explicit rate matrix by exploiting an elementwise probabilistic interpretation of R.

3 Main result

We represent R as

$$R = \begin{pmatrix} R_{00} & R_{01} & \dots & R_{0H} \\ R_{10} & R_{11} & \dots & R_{1H} \\ \vdots & \vdots & & \vdots \\ R_{H0} & R_{H1} & \dots & R_{HH} \end{pmatrix}, \tag{15}$$

and define an *excursion* as

Definition 1. *For an arbitrary level n, an excursion is defined as the time elapsing from the moment the QBD process leaves an initial state in level n until the time of the first return of the process to level n. The excursion should always leave level n.*

Then, the following property can be shown to hold (see [4] p. 142 and [6] p. 8):

Property 1. *Element R_{jk} represents the expected time spent in state $(n+1,k)$ during an excursion with initial state (n,j), expressed in the expected time spent in state (n,j).*

An elementwise interpretation of R as given in Property 1, however, is hardly applicable since describing all possible ways in which an excursion starting from (n,j) might visit $(n+1,k)$ is a complex or even impossible task. We now, however, describe a class of QBD processes for which this can be done. We consider QBD processes for which in state (i,j), $i \geq 2, 0 \leq j \leq H-1$, the following transitions are possible (see Figure 1):

- from (i,j) to $(i-1,j+1)$ with rate $f(-1,1)$;
- from (i,j) to $(i,j+1)$ with rate $f(0,1)$;
- from (i,j) to $(i+1,j+1)$ with rate $f(1,1)$;
- from (i,j) to $(i+1,j)$ with rate $f(1,0)$.

We next refer to the set of states (i,H), $i \geq 2$, as the boundary states. For these boundary states we allow the following transitions:

- from (i,H) to $(i+1,H)$ with rate $f_H(1,0)$;
- from (i,H) to $(i-1,H)$ with rate $f_H(-1,0)$.

The main reason that these QBD processes are suitable for applying Property 1 is that an excursion visits non-boundary states at most once. Therefore, an excursion will either end or reach one of the boundary states in a finite number of steps. Once an excursion visits one of the boundary states, it will visit boundary states only until it returns to level n (and the excursion ends).

We note that although the rates from state (i,j) do not depend on j, the approach as presented in this section can be easily extended to cases for which the rates do depend on j. For ease of presentation, we choose to formulate Property 1 as follows:

Property 2. *Element R_{jk} can be described as*

$$R_{jk} = q_{jk} \cdot \mathbb{E}X_k \cdot \frac{[A_1]_{jj}}{[A_1]_{kk}}, \tag{16}$$

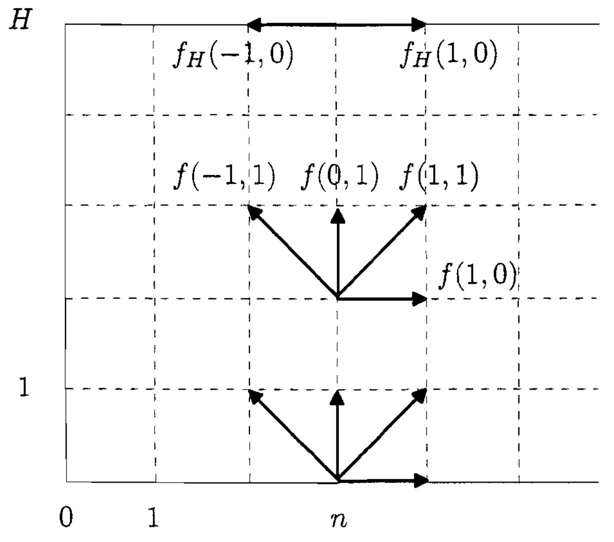


Figure 1: The transition rates out of state (i, j), $i \geq 2, 0 \leq j \leq H-1$ for the class considered.

where, for an excursion starting from (n, j), q_{jk} denotes the probability that this excursion reaches state $(n+1, k)$, $\mathbb{E}X_k$ represents the expected number of visits to state $(n+1, k)$ during the whole excursion given state $(n+1, k)$ is reached at least once, and $[A]_{jj}$ denotes the j-th diagonal element of a matrix A. When visited, the expected time spent in state (n, j) and (n, k) (for $n \geq 2$) equals $1/[A_1]_{jj}$ and $1/[A_1]_{kk}$, respectively.

First realize that for the introduced class of QBD processes, an excursion starting from state (n, j) cannot visit states $(n+1, k)$, $k < j$. Hence, we can restrict ourselves to describing q_{jk}, $j \geq 0$, $j \leq k < H$. We do this by summing over all possible paths an excursion might follow, each path multiplied by its probability of occurrence. By conditioning on the number of steps in $(0,1)$ and $(1,1)$ direction, denoted by r and s, respectively, it follows that

$$q_{jk} = \sum_{r=0}^{k-j} \sum_{s=0}^{\lfloor \frac{k-j-r+1}{2} \rfloor} L(k-j|r,s)P(k-j|r,s), \qquad j \geq 0, \quad j \leq k \leq H-1, \tag{17}$$

where $L(k-j|r,s)$ denotes the number of paths from (n, j) to $(n+1, k)$ without returning to level n containing r steps in $(0,1)$ direction, and s steps in $(1,1)$ direction, and $P(k-j|r,s)$ the probability of each such path. Note that given the current state, the probability of going to the next state only depends on the current state. Due to the exponential residence time in each state, $P(k-j|r,s)$ is given by

$$P(k-j|r,s) = \varphi(0,1)^r \varphi(1,1)^s \varphi(-1,1)^{k-j-r-s} \varphi(1,0)^{k-j-r-2s+1}, \tag{18}$$

with

$$\varphi(x,y) := \frac{f(x,y)}{\sum_x \sum_y f(x,y)}, \quad \varphi_H(x,y) := \frac{f_H(x,y)}{\sum_x \sum_y f_H(x,y)}. \tag{19}$$

This leaves us to determine $L(k-j|r,s)$, for which we invoke the following result on lattice path counting:

Theorem 2. *For $j \leq k,\ k \leq H-1$, the number $L(k-j|r,s)$ of lattice paths from (n,j) to $(n+1,k)$ without returning to level n having $r \in \{0,1,\dots,k-j\}$ and $s \in \{0,1,\dots,\lfloor\frac{k-j-r+1}{2}\rfloor\}$ steps in $(0,1)$ and $(1,1)$ direction, respectively, equals*

$$L(k-j|r,s) = \frac{1}{k-j-r-s+1}\binom{2k-2j-2r-2s}{k-j-r-s}\binom{2k-2j-r-2s}{r}\binom{k-j-r-s+1}{s}. \tag{20}$$

Proof See Appendix A. □

The number of paths that cause an excursion starting from state (n,j) to hit one of the boundary states (and eventually hit state $(n+1,H)$) is infinite. Therefore, we derive the probabilities q_{jH}, $j = 1,\dots,H-1$ by subtracting from 1 the probability of an excursion *not* reaching state $(n+1,H)$. Hence, we have

$$q_{jH} = \varphi(1,0) + \varphi(1,1) - \varphi(-1,1)\sum_{k=j}^{H-1} q_{jk}, \quad j = 1,\dots,H-1, \tag{21}$$

and trivially $q_{HH} = \varphi_H(1,0)$.

Observe that $\mathbb{E}X_k = 1$, $k = 0,\dots,H-1$, because once visited, these states cannot be visited again without returning to level n. For state $(n+1,H)$ this is different. Once an excursion visits state $(n+1,H)$, the excursion moves to state $(n+2,H)$ w.p. $\varphi_H(1,0)$, and this would imply an additional visit to state $(n+1,H)$, while the excursion returns to level n w.p. $\varphi_H(-1,0) = 1 - \varphi_H(1,0)$, which would end the excursion. We thus have

$$\mathbb{E}X_H = 1 + \varphi_H(1,0)\mathbb{E}X_H \quad \Leftrightarrow \quad \mathbb{E}X_H = \frac{1}{\varphi_H(-1,0)}. \tag{22}$$

All elements of the rate matrix R are now fully specified, and so the stationary distribution for this class of QBD processes can be obtained.

3.1 Examples

We now present some examples of QBD processes that fit within the class as described in Section 3. For all these processes, the R matrix can thus be determined explicitly.

Example 1

Consider a two-station tandem queue. Customers arrive at station 1 according to a Poisson process with rate λ_1, and upon service completion at station 1 they join the queue at station 2. After service completion at station 2 they leave the system. Customers can also arrive directly at station 2 with Poisson rate λ_2. These customers only require service at station 2 and then leave the system. A third optional Poisson arrival stream with rate λ_3 yields a new customer at each of the stations, see Figures 2 and 3. Customers require an exponentially distributed service time at station j with mean $1/\mu_j$. A special feature of the model is that a single server, working at rate 1, moves between the two queues and gives preemptive priority to customers at station 1. Also, the second station possesses a finite buffer of capacity H. Customers that arrive at station 2 when the buffer is fully occupied are removed from the system. The system can then be described as a QBD process with states (i,j), where i is the number of customers at station 1 and j the number of customers at station 2.

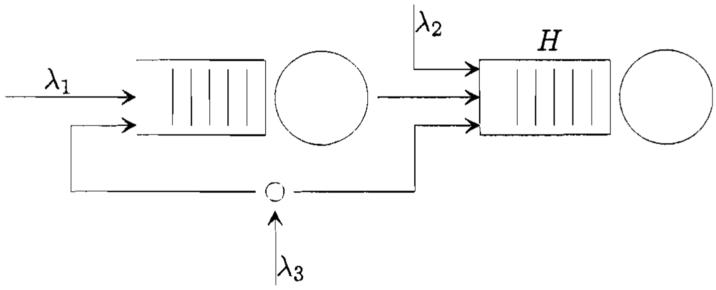


Figure 2: Example 1, sketch of two-station tandem queue with finite buffer at station 2 of size H.

Hence, the stationary distribution should satisfy the balance equations (4) and (5-6), where

$$A_0 = B_0 = \begin{pmatrix} \lambda_1 & \lambda_3 & & \\ & \ddots & \ddots & \\ & & \lambda_1 & \lambda_3 \\ & & & \lambda_1+\lambda_3 \end{pmatrix}, \quad A_1 = \begin{pmatrix} \Delta & \lambda_2 & & \\ & \ddots & \ddots & \\ & & \Delta & \lambda_2 \\ & & & \Delta \end{pmatrix},$$

and

$$A_2 = B_2 = \begin{pmatrix} 0 & \mu_1 & & \\ & \ddots & \ddots & \\ & & 0 & \mu_1 \\ & & & \mu_1 \end{pmatrix}, \quad B_1 = \begin{pmatrix} \Delta & \lambda_2 & & & \\ \mu_2 & \Delta & \lambda_2 & & \\ & \ddots & \ddots & \ddots & \\ & & \mu_2 & \Delta & \lambda_2 \\ & & & \mu_2 & \Delta \end{pmatrix}.$$

From (3) the following condition should hold for the stationary distribution to exist:

$$\lambda_1 + \lambda_3 < \mu_1. \tag{23}$$

It is easily verified that this model fits into the classification of the present section. We have $f(-1,1) = \mu_1$, $f(0,1) = \lambda_2$, $f(1,1) = \lambda_3$, $f(1,0) = \lambda_1$, $f_H(-1,0) = \mu_1$ and $f_H(1,0) = \lambda_1+\lambda_3$.

The strategy to serve station 1 even when the buffer at station 2 is full seems wasteful, since the customer served at station 1 is immediately rejected at station 2. A better strategy would be to give priority to station 1 except when the buffer is full, in case one customer is served at station 2. After its service completion, service is continued at station 1. This strategy leads to an explicit R matrix as well. However, it requires some additional reasoning similar to the approach presented in Section 4 for some advanced models.

Example 2

Consider a machine processing jobs in the *first come first served* order. Jobs arrive according to a Poisson process with rate λ and have exponentially distributed service requirements with rate μ. The machine is turned off when the queue is empty and it is turned on again upon arrival of a new job. The set-up time consists of H exponential phases with parameter θ. Jobs are only processed, when the machine is in phase H. The system can be described as a QBD process with states (i, j), where i represents the number of jobs in the queue and j denotes the set-up phase of the machine. The stationary distribution should satisfy the balance equations

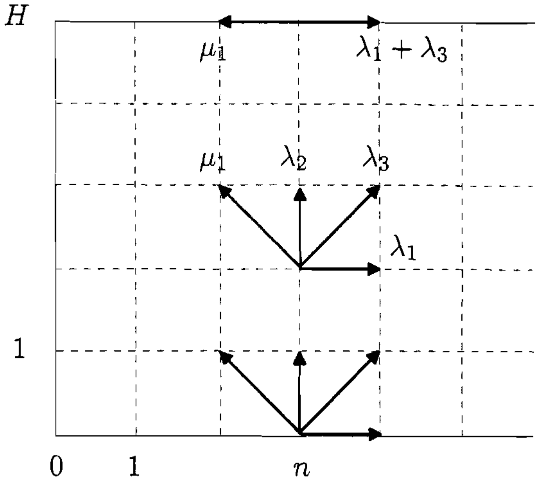


Figure 3: Transition diagram of Example 1: Tandem queue with priority for jobs at station 1, finite buffer at station 2 of size H.

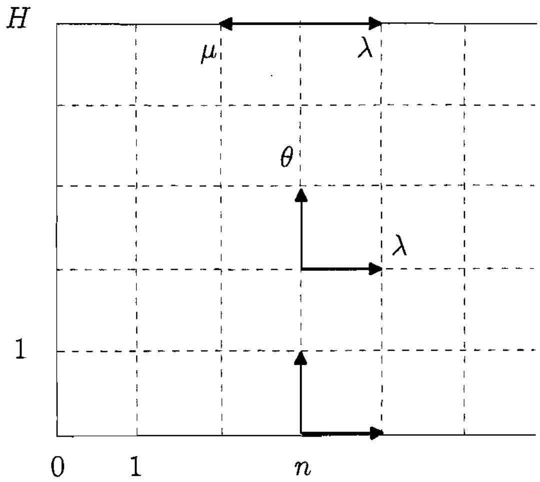


Figure 4: Transition diagram of Example 2: Machine with set-up time consisting of H exponential phases.

(4) and (5-6), where $B_0 = \begin{pmatrix} \lambda & 0 & \cdots & 0 \end{pmatrix}$, $B_1 = \Delta$, $B_2 = \begin{pmatrix} 0 & \cdots & 0 & \mu \end{pmatrix}'$, and

$$A_0 = \begin{pmatrix} \lambda & & & \\ & \lambda & & \\ & & \ddots & \\ & & & \lambda \end{pmatrix}, A_1 = \begin{pmatrix} \Delta & \theta & & \\ & \ddots & \ddots & \\ & & \Delta & \theta \\ & & & \Delta \end{pmatrix}, A_2 = \begin{pmatrix} 0 & & & \\ & \ddots & & \\ & & 0 & \\ & & & \mu \end{pmatrix}.$$

From (3) it follows that the mean drift condition is given by

$$\lambda < \mu. \tag{24}$$

This model clearly fits into the classification of the present section. We have $f(0,1) = \theta$, $f(1,0) = \lambda$, $f_H(-1,0) = \mu$, $f_H(1,0) = \lambda$ and all other transition rates are equal to zero.

Example 3

Consider a machine that warms up during the production of the first H jobs. While warming up, the machine produces at rate θ. After the completion of the H-th job, the production rate increases from θ to μ. The system can again be described as a QBD process with states (i,j), where i represents the number of jobs in the queue and j denotes the set-up phase of the machine. The stationary distribution should satisfy the balance equations (4) and (5-6), where $B_0 = \begin{pmatrix} \lambda & 0 & \cdots & 0 \end{pmatrix}$, $B_1 = \Delta$, $B_2 = \begin{pmatrix} \theta & \cdots & \theta & \mu \end{pmatrix}'$, and

$$A_0 = \begin{pmatrix} \lambda & & & \\ & \lambda & & \\ & & \ddots & \\ & & & \lambda \end{pmatrix}, A_1 = \begin{pmatrix} \Delta & & & \\ & \ddots & & \\ & & \Delta & \\ & & & \Delta \end{pmatrix}, A_2 = \begin{pmatrix} 0 & \theta & & \\ & \ddots & \ddots & \\ & & 0 & \theta \\ & & & \mu \end{pmatrix}.$$

The mean drift condition (3) is given by

$$\lambda < \mu. \tag{25}$$

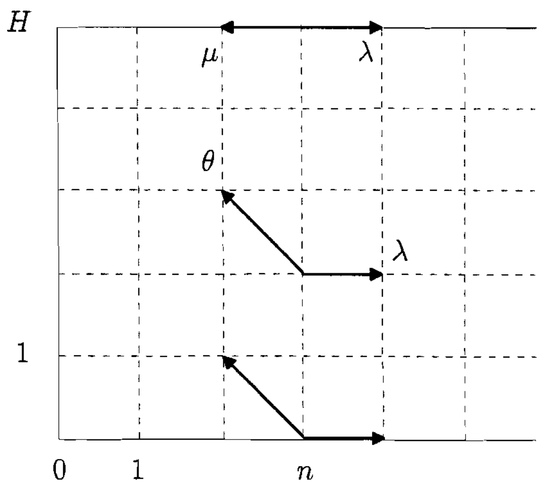


Figure 5: Transition diagram of Example 3: Machine that warms up while producing the first H products. After finishing the H-th product, the rate increases from θ to μ.

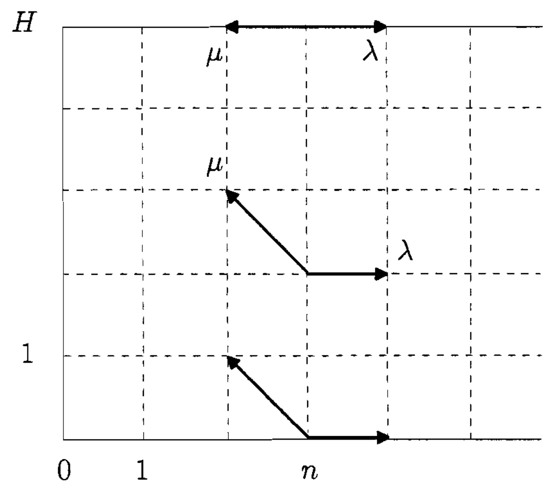


Figure 6: Transition diagram of Example 4: Joint process of queue length and number of customers served during a busy period in an $M/M/1$ queue.

It is easily verified that this model fits into the classification of the present section. More specifically, we have $f(-1,1) = \theta$, $f(1,0) = \lambda$, $f_H(-1,0) = \mu$, $f_H(1,0) = \lambda$ and all other transition rates are equal to zero.

Example 4

Consider an ordinary $M/M/1$ queue with arrival rate λ and service rate μ. Suppose we are interested in the joint equilibrium distribution of the number of customers waiting in the queue (or in the service position) and the number of customers served so far in the ongoing busy period. By counting only the first H customers served in a busy period, this example can be modelled as a QBD process with states (i, j), where i is the number of customers waiting in the queue and j the number of customers served so far in the ongoing busy period. The stationary distribution should once more satisfy the equations (4) and (5), with $B_0 = \begin{pmatrix} \lambda & 0 & \cdots & 0 \end{pmatrix}$, $B_1 = \Delta$, $B_2 = \begin{pmatrix} \mu & \mu & \cdots & \mu \end{pmatrix}'$, and

$$A_0 = \begin{pmatrix} \lambda & & & \\ & \lambda & & \\ & & \ddots & \\ & & & \lambda \end{pmatrix}, A_1 = \begin{pmatrix} \Delta & & & \\ & \ddots & & \\ & & \Delta & \\ & & & \Delta \end{pmatrix}, A_2 = \begin{pmatrix} 0 & \mu & & \\ & \ddots & \ddots & \\ & & 0 & \mu \\ & & & \mu \end{pmatrix}.$$

The stability condition of this model (cf. (3)) is obviously given by

$$\lambda < \mu. \tag{26}$$

This example fits into the classification of the present section. This means that we have $f(-1,1) = \mu$, $f(1,0) = \lambda$, $f_H(-1,0) = \mu$, $f_H(1,0) = \lambda$ and all other transition rates are equal to zero.

4 Extensions

We now extend the class of processes as presented in Section 3. Firstly, observe that for determining the element R_{jk} from Property 1, the following two numbers should be tractable:

- The total number of possible paths that an excursion starting from (n, j) can take before it reaches state $(n+1, k)$ for the first time.
- The expected number of times an excursion visits state $(n+1, k)$ given the excursion is in state $(n+1, k)$ for the first time.

Note that the latter does not depend on the initial state (n, j). We impose conditions on the transition rates in the interior of the state space that ensure that both quantities can be calculated. In particular, we can distinguish the following two classes:

Class 1: The transition rates going out of the non-boundary states that belong to one of the four subsets displayed in Figure 7.

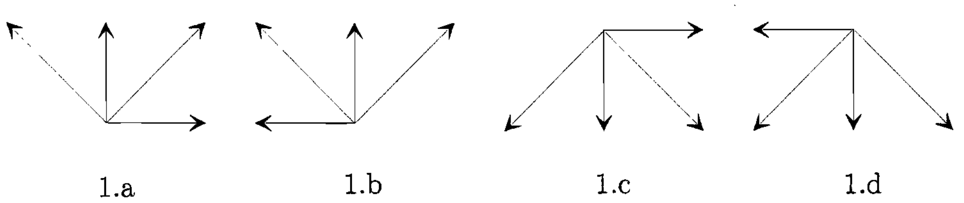


Figure 7: The transition rates going out of the non-boundary states for Class 1.

Class 2: The transition rates going out of the non-boundary states that belong to one of the four subsets displayed in Figure 8.

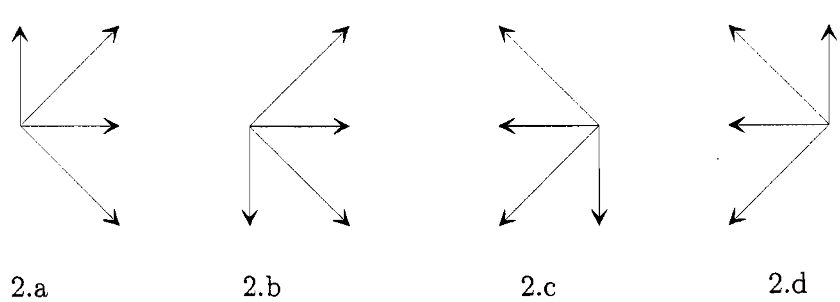


Figure 8: The transition rates going out of the non-boundary states for Class 2.

Note that the set of transitions denoted by 1.a is precisely the set analyzed in Section 3. The other feasible sets are merely *rotations* of this set. The main reason for these classes to be suitable for applying Property 1 is that an excursion cannot visit the same state again without visiting one of the boundary states in the meantime.

Class 2 is closely related to the class for which either A_0 or A_2 is of rank 1. That is, for a process in Class 2 to be interesting or recurrent, at the boundary states the process should have a rate of opposite direction as the non-boundary states. If, for the states in a certain level, the process has this opposite rate in just one of the two sets of boundary states, (i,j) with $j=0$ or $j=H$, the requirement that either A_0 or A_2 is of rank 1 is satisfied. Of course, the class with either A_0 or A_2 of rank 1 is much larger than Class 2. However, for the models that fall in Class 2, one could determine each element of R separately, purely based on lattice path counting, an example of which is given in the next section.

4.1 Examples

We now present some more examples of QBD processes that fit within the classes as described in Section 4, and for which small adjustments to the analysis as presented in Section 3 again lead to an explicit description of the R matrix.

Example 5

Consider products that are produced in two phases. The first phase is standard and identical for all products. The second phase is customer specific. At most H half-finished products can be stored. The production of the two phases is done by a common tight resource. The first phase takes an exponential time with mean $1/\mu_1$, the second phase is exponential with mean $1/\mu_2$. Orders arrive for one item at a time according to a Poisson process with rate λ. This model has also been studied in Adan & van der Wal [1] by using the spectral decomposition method.

The system can be described as a QBD process with states (i,j), where i is the number of orders in the system and j the number of half-finished products on stock or in use. So state $(1,2)$ denotes the situation with 1 order in the system for which the production resource is processing phase 2, and 1 half-finished product on the shelf. If phase 2 is completed, the state changes to $(0,1)$ and the resource continues with producing phase 1 products until the limit H is reached or a new order arrives. So, a new order preempts the stock production. In state $(i,0)$ the resource is working on phase 1. If phase 1 is completed, the state changes to $(i,1)$ and the resource continues with phase 2.

The stationary distribution should satisfy the balance equations (4) and (5-6), where

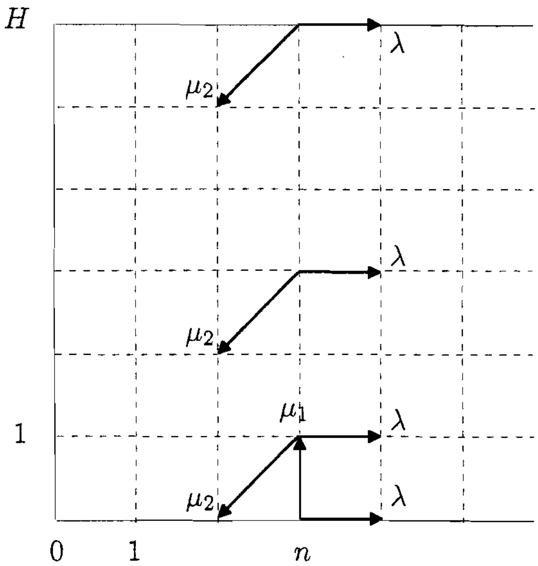
$$A_0 = B_0 = \begin{pmatrix} \lambda & & & \\ & \lambda & & \\ & & \ddots & \\ & & & \lambda \end{pmatrix}, \quad A_1 = \begin{pmatrix} \Delta & \mu_1 & & \\ & \ddots & & \\ & & \Delta & \\ & & & \Delta \end{pmatrix},$$

and

$$A_2 = B_2 = \begin{pmatrix} 0 & & & \\ \mu_2 & 0 & & \\ & \ddots & \ddots & \\ & & \mu_2 & 0 \end{pmatrix}, \quad B_1 = \begin{pmatrix} \Delta & \mu_1 & & \\ & \ddots & \ddots & \\ & & \Delta & \mu_1 \\ & & & \Delta \end{pmatrix}.$$

The mean drift condition (see (3)) now reads

$$\frac{\lambda}{\mu_1} + \frac{\lambda}{\mu_2} < 1, \tag{27}$$



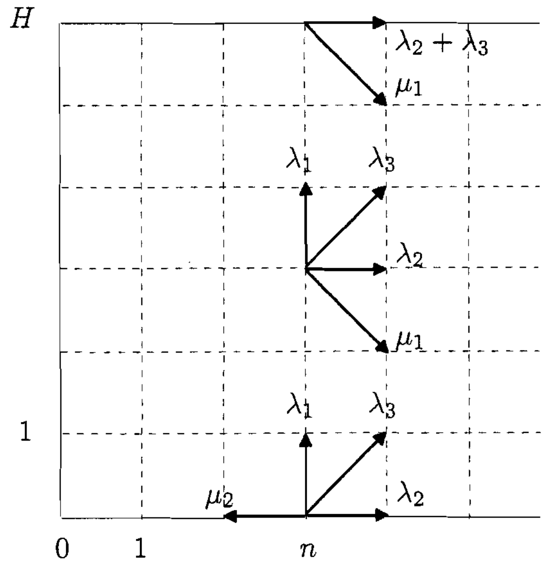


Figure 9: Transition diagram of Example 5: Products produced in two phases, phase 1 produced on a stock with a maximum size of H.

Figure 10: Transition diagram of Example 6: Tandem queue with priority for jobs at station 1, finite buffer at station 1 of size H.

which should be satisfied for the stationary distribution to exist.

The present queueing system is part of Class 1 as defined in Section 4. We determine its equilibrium distribution via a slight modification of the analysis introduced in Section 3. From (17), we evidently have

$$q_{jk} = \varphi(1,0)^{j-k+1}\varphi(-1,-1)^{j-k}L(j-k|0,0), \quad j \geq k, \ k \geq 2. \tag{28}$$

We can calculate q_{j1}, the probability of reaching state $(n+1,1)$ from state (n,j) without returning to level n, by determining the probability of not reaching state $(n+1,1)$ and subtract this from 1, yielding

$$q_{j1} = \varphi(1,0) - \varphi(-1,-1)\sum_{k=2}^{j} q_{jk}, \quad j = 1,\dots,H. \tag{29}$$

Similar reasoning yields

$$q_{j0} = \varphi(1,0) - \varphi(-1,-1)\sum_{k=1}^{j} q_{jk}, \quad j = 1,\dots,H. \tag{30}$$

Furthermore, it is readily found that

$$q_{00} = \frac{\lambda}{\lambda + \mu_1}. \tag{31}$$

Observe that $\mathbb{E}X_k = 1$ for $k = 2,\dots,H$, because once visited, these states cannot be visited again without returning to level n. For $k = 0$ we obtain

$$\mathbb{P}(X_0 = n) = \Big(1 - \frac{\mu_1}{\lambda+\mu_1}\varphi(-1,-1)\Big)^{n-1}\frac{\mu_1}{\lambda+\mu_1}\varphi(-1,-1), \quad n \geq 1. \tag{32}$$

This can be interpreted as follows. Once state $(n+1,0)$ has been reached, the only way in which the system does not return to this state without crossing level n is finishing phase 1 and 2 of the order in service before a new order arrives. So we have

$$\mathbb{E}X_0 = \sum_{n=1}^{\infty} n\mathbb{P}(X_0 = n) = \frac{\lambda + \mu_1}{\mu_1 \varphi(-1,-1)}. \tag{33}$$

For $k = 1$ we find by similar reasoning (using $\varphi(1,0) + \varphi(-1,-1) = 1$)

$$\mathbb{E}X_1 = \sum_{n=1}^{\infty} n\varphi(1,0)^{n-1}\varphi(-1,-1) = \frac{1}{\varphi(-1,-1)}. \tag{34}$$

Now all elements of (16) are known, and thus R is known, with which the complete stationary distribution of the QBD process can be determined.

Example 6

Like Example 1 (see Figure 2), but we now assume that station 1 instead of station 2 has a finite buffer of size H, so that the system can be described as a QBD process with states (i,j), where i is the number of customers at station 2, and j the number of customers at station 1. Hence, the stationary distribution should satisfy the balance equations (4) and (5-6), where

$$A_0 = B_0 = \begin{pmatrix} \lambda_2 & \lambda_3 & & & \\ \mu_1 & \lambda_2 & \lambda_3 & & \\ & \ddots & \ddots & \ddots & \\ & & \mu_1 & \lambda_2 & \lambda_3 \\ & & & \mu_1 & \lambda_2 + \lambda_3 \end{pmatrix}, \quad A_1 = \begin{pmatrix} \Delta & \lambda_1 & & \\ & \ddots & \ddots & \\ & & \Delta & \lambda_1 \\ & & & \Delta \end{pmatrix},$$

and

$$A_2 = B_2 = \begin{pmatrix} \mu_2 & & & \\ & 0 & & \\ & & \ddots & \\ & & & 0 \end{pmatrix}, \quad B_1 = \begin{pmatrix} \Delta & \lambda_1 & & \\ & \ddots & \ddots & \\ & & \Delta & \lambda_1 \\ & & & \Delta \end{pmatrix}.$$

From (3) the following condition should hold for the process to be stable, and the stationary distribution to exist:

$$\lambda_2 + \lambda_3 + (1 - \xi_0)\mu_1 < \xi_0 \mu_2, \tag{35}$$

where ξ_0 is the probability of an empty system in an $M/M/1/H$ queue, with arrival rate $\lambda_1 + \lambda_3$ and service rate μ_1. Obviously, the present model is part of Class 2 as introduced in Section 4. However, it also has a matrix A_2 of rank 1, and an explicit expression for R, as given by (11), thus exists. We now show how one can determine each element of R separately, purely based on tracking down all possible paths of an excursion.

For this system, we know that every excursion starting from level n that enters level $n+1$ will hit state $(n+1,0)$ at some point, since it can only return from level $n+1$ to n through state $(n+1,0)$. We therefore again slightly adapt the formulation of Property 1.

Property 3. *Element R_{jk} is given by*

$$R_{jk} = (m_j \mathbb{E}Y_k + n_{jk}) \frac{[A_1]_{jj}}{[A_1]_{kk}}, \tag{36}$$

where m_j is the probability that an excursion that starts from state (n,j) enters level $n+1$, Y_k the number of times an excursion visits state $(n+1,k)$ assuming it is in state $(n+1,0)$, and n_{jk} the probability that an excursion reaches state $(n+1,k)$ before it reaches state $(n+1,0)$.

From the transitions state diagram it readily follows that

$$m_H = 1, \quad m_0 = \frac{\lambda_2+\lambda_3}{\lambda_1+\lambda_2+\lambda_3+\mu_2}, \quad m_j = 1-\varphi(0,1),\ j=1,\dots,H-1, \tag{37}$$

which leaves us to determine $\mathbb{E}Y_k$ and n_{jk}.

We first consider $\mathbb{E}Y_0$. Assume that the excursion is in state $(n+1,0)$ for the first time. Then, with probability

$$\beta_0 := \frac{\lambda_1+\lambda_2+\lambda_3}{\lambda_1+\lambda_2+\lambda_3+\mu_2} \tag{38}$$

the excursion goes to a state other than $(n,0)$, and so the excursion continues. Evidently, the excursion will then hit $(n+1,0)$ for a second time and the same reasoning will hold. Thus,

$$\mathbb{E}Y_0 = 1+\beta_0\mathbb{E}Y_0 \quad \Leftrightarrow \quad \mathbb{E}Y_0 = \frac{1}{1-\beta_0}. \tag{39}$$

We now turn to $\mathbb{E}Y_k$ for $k \geq 1$. Again assume that the excursion is in state $(n+1,0)$ for the first time. Then, with probability

$$\beta_k := \frac{\lambda_1}{\lambda_1+\lambda_2+\lambda_3+\mu_2}\varphi(0,1)^{k-1} \tag{40}$$

the excursion will hit state $(n+1,k)$. With probability β_0 the excursion ends up in state $(n+1,0)$ for the second time, irrespective of whether or not it visited state $(n+1,k)$ meanwhile. We thus have

$$\mathbb{E}Y_k = \beta_k+\beta_0\mathbb{E}Y_k \quad \Leftrightarrow \quad \mathbb{E}Y_k = \frac{\beta_k}{1-\beta_0}, \quad k=1,\dots,H. \tag{41}$$

For the description of the n_{jk} we again use the notation as introduced in (3). Carefully studying the transition diagram then yields

$$\begin{aligned}
n_{j,j-1} &= \varphi(1,-1), && (42)\\
n_{jj} &= \varphi(1,0)+\varphi(1,-1)\varphi(0,1), && (43)\\
n_{jk} &= [\varphi(1,1)+\varphi(1,0)\varphi(0,1)+\varphi(1,-1)\varphi(0,1)^2]\varphi(0,1)^{k-j-1}, && (44)
\end{aligned}$$

for $j=2,\dots,H-1,\ k=j+1,\dots,H$. We further have, for $k=2,\dots,H$,

$$\begin{aligned}
n_{01} &= \lambda_3/(\lambda_1+\lambda_2+\lambda_3+\mu_2), && (45)\\
n_{0k} &= n_{01}\varphi(0,1)^{k-1}, && (46)\\
n_{1k} &= \varphi(0,1)^{k-2}[\varphi(1,1)+\varphi(1,0)\varphi(0,1)], && (47)
\end{aligned}$$

and

$$n_{11} = \varphi(1,0), \quad n_{H,H-1} = \frac{\mu_1}{\lambda_2+\lambda_3+\mu_1}, \quad n_{HH} = \frac{\lambda_2+\lambda_3}{\lambda_2+\lambda_3+\mu_1}+\frac{\mu_1\varphi(0,1)}{\lambda_2+\lambda_3+\mu_1}. \tag{48}$$

In cases other than (42-48), n_{jk} equals zero. Again, all elements of the rate matrix R are fully specified, and so the stationary distribution of the QBD process can be obtained.

5 Conclusions

QBD processes can be analyzed with the matrix-geometric method, for which one should derive the solution to a non-linear matrix equation. This solution, referred to as the rate matrix R, is usually determined by some numerical iterative procedure. For a few cases though, R is known to have an explicit description, due to the special structure of the transition diagram.

In this paper we have presented a class of QBD processes, other than the aforementioned cases, for which also an explicit description of the rate matrix R can be given. We do this by exploiting a probabilistic interpretation of each element of R. For the present class, the problem of finding each element of R then reduces to counting lattice paths in the transition diagram. The counting problem has been resolved using an extension of the classic Ballot theorem. We have also given a variety of examples of queueing models that fit into the class of QBD processes as presented.

Acknowledgment

The authors like to thank Ivo Adan for valuable comments.

A Proof of Theorem 2

We start from a classic result known as the Ballot Theorem:

Lemma 1. *In an election candidates A and B receive a and b votes ($a > b$), respectively. If it is assumed that all orderings are equally likely, the number of ways in which A's votes always exceed B's votes equals*

$$\frac{a-b}{a+b}\binom{a+b}{b}. \tag{49}$$

Proof See Chapter 5 of [8]. □

We now turn to the transition diagram given in Figure 1 and prove Theorem 2 in three steps. First, we present a lemma that gives the number of paths leading from (n,j) to $(n+1,k)$ without returning to level n for the case only $(1,0)$ and $(-1,1)$ steps are allowed. In the second lemma, we permit $(0,1)$ steps as well. Finally, we prove Theorem 2, which covers $(1,0)$, $(-1,1)$, $(0,1)$ and $(1,1)$ steps.

Lemma 2. *For $j \le k$, $k \le H-1$, the number $L_1(k-j)$ of lattice paths from (n,j) to $(n+1,k)$ with $(1,0)$ and $(-1,1)$ steps without returning to level n is given by*

$$L_1(k-j) = \frac{1}{k-j+1}\binom{2k-2j}{k-j}. \tag{50}$$

Proof Consider an arbitrary path from (n,j) to $(n+1,k)$ with $(1,0)$ and $(-1,1)$ steps not returning to the level n. Each such path consists of exactly $k-j+1$ $(1,0)$ steps and $k-j$ $(-1,1)$ steps. The number of $(1,0)$ steps should always exceed the number of $(-1,1)$ steps, otherwise the excursion would have returned to level n. So we can apply the Ballot theorem with $a = k-j+1$ and $b = k-j$. Now, (50) follows after some elementary computations. □

Lemma 3. *For $j \leq k,\ k \leq H-1$, the number $L_2(k-j|r)$ of lattice paths from (n,j) to $(n+1,k)$ with $(1,0)$, $(-1,1)$ and $(0,1)$ steps without returning to level n having $r \in \{0,1,\ldots,k-j\}$ steps in $(0,1)$ direction equals*

$$L_2(k-j|r) = \frac{1}{k-j-r+1}\binom{2k-2j-2r}{k-j-r}\binom{2k-2j-r}{r}. \tag{51}$$

Proof Consider an arbitrary path from (n,j) to $(n+1,k)$ with r $(0,1)$ steps, that does not return to level n. By removing these $(0,1)$ steps we deduce a partial path from (n,j) to $(n+1,k-r)$ only using $(1,0)$ and $(-1,1)$ steps. By Lemma 2 the number of such paths is equal to $L_1(k-j-r)$. Each partial path passes through $2k-2j-2r+1$ lattice points (excluding the point (n,j)). The r $(0,1)$ steps can take place in either one of these points, which is equivalent to putting r balls into $2k-2j-2r+1$ bins. The number of ways to do so equals $\binom{2k-2j-r}{r}$ and this completes the proof. □

Theorem 2. *For $j \leq k,\ k \leq H-1$, the number $L(k-j|r,s)$ of lattice paths from (n,j) to $(n+1,k)$ without returning to level n having $r \in \{0,1,\ldots,k-j\}$ and $s \in \{0,1,\ldots,\lfloor\frac{k-j-r+1}{2}\rfloor\}$ steps in $(0,1)$ and $(1,1)$ direction, respectively, equals*

$$L(k-j|r,s) = \frac{1}{k-j-r-s+1}\binom{2k-2j-2r-2s}{k-j-r-s}\binom{2k-2j-r-2s}{r}\binom{k-j-r-s+1}{s}. \tag{52}$$

Proof Consider an arbitrary path from (n,j) to $(n+1,k)$ with r $(0,1)$ steps and s $(1,1)$ steps not returning to the level n. We decompose each $(1,1)$ step into a horizontal component $(1,0)$ and a vertical component $(0,1)$. By leaving out the vertical steps (and components), we consider the partial path from (n,j) to $(n+1,k-r-s)$ with steps $(1,0)$ and $(-1,1)$. The number of such paths is given by $L_1(k-j-r-s)$. We extend each partial path by

1. placing r $(0,1)$ steps in the $2k-2j-2r-2s+1$ lattice points excluding the point (n,j) that are crossed. The number of ways to do so equals $\binom{2k-2j-2s-r}{r}$.
2. extending s $(1,0)$ steps from the $k-j-r-s+1$ available by a vertical component. This extension can be done in $\binom{k-j-r-s+1}{s}$ ways.

This concludes the proof. □

References

[1] I.J.B.F. Adan, J. van der Wal (1998). Combining make to order and make to stock, *OR Spektrum* **20**: 73-81.

[2] L. Gun (1989). Experimental results on matrix-analytical solutions techniques - extensions and comparisons, *Stochastic Models* **5** (4): 669-682.

[3] G. Latouche, V. Ramaswami (1993). A logarithmic reduction algorithm for quasi-birth-and-death processes, *Journal of Applied Probability* **30**: 650-674.

[4] G. Latouche, V. Ramaswami (1999). *Introduction to Matrix Analytic Methods in Stochastic Modeling*, SIAM, Philadelphia.

[5] S.G. Mohanty (1979). *Lattice Path Counting and Applications*, Acadamic Press, London.

[6] M.F. Neuts (1981). *Matrix-geometric Solutions in Stochastic Models, An Algorithmic Approach*, The Johns Hopkins Press, Baltimore.

[7] V. Ramaswami, G. Latouche (1986). A general class of Markov processes with explicit matrix geometric solutions, *OR Spektrum* **8**: 209-218.

[8] W.A. Whitworth (1959). *Choice and Chance; With One Thousand Exercises*, fifth edition, Hafner Publishing Co., New York.