TU/e Technische Universiteit
Eindhoven
University of Technology

Where innovation starts

Analysing shortest expected delay routing for Erlang servers

Adan, I.J.B.F.; Wessels, J.

Please check the document version of this publication:

• A submitted manuscript is the author's version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

EINDHOVEN UNIVERSITY OF TECHNOLOGY
Department of Mathematics and Computing Science

Memorandum COSOR 93-38

Analysing shortest expected delay routing for Erlang servers

I.J.B.F. Adan
J. Wessels

Eindhoven, November 1993
The Netherlands

Analysing shortest expected delay routing for Erlang servers

I.J.B.F. Adan[1]   J. Wessels[1,2]

November 26, 1993

Abstract

The queueing problem with a Poisson arrival stream and two identical Erlang
servers is analysed for the queueing discipline based on shortest expected delay. This
queueing problem may be represented as a random walk on the integer grid in the
first quadrant of the plane. In the paper it is shown that the equilibrium distribution
of this random walk can be written as a countable linear combination of product
forms. This linear combination is constructed in a compensation procedure. In this
case the compensation procedure is essentially more complicated than in other cases
where the same idea was exploited. The reason for the complications is that in this
case the boundary consists of several layers which in turn is caused by the fact that
transitions starting in inner states are not restricted to end in neighbouring states.

Good starting solutions for the compensation procedure are found by solving
the shortest expected delay problem with the same service distributions but with
instantaneous jockeying.

It is also shown that the results can be used for an efficient computation of
relevant performance criteria.

1   Introduction

The *shortest queue problem* is one of the most intensively studied queueing problems, even
in its simplest formulation with a Poisson stream of arriving customers and two identical
exponential servers. In this simple formulation the shortest queue problem is an example
of a two-dimensional random walk on an unbounded part of the plane. Since it appeared
to be hard to solve these simply formulated problems, it has been tried again and again
to obtain more insight in the equilibrium behaviour of two-dimensional random walks in
general and the shortest queue problem in particular.

The shortest queue problem was formulated originally by Haight in [12], however, the
first major step towards its analysis was made by Kingman in [14] and Flatto and McKean
in [10]. Using a uniformization approach they proved that the generating function of the
equilibrium distribution is meromorphic and they established explicit relations for the
poles and residues. As a consequence, partial fraction decomposition of the generating
function would in principle lead to a representation of the equilibrium distribution as

---

[1]Eindhoven University of Technology, Department of Mathematics and Computing Science, Eind-
hoven, The Netherlands.

[2]International Institute for Applied Systems Analysis, Laxenburg, Austria.

an infinite linear combination of product forms. Regrettably, however, it appeared to be practically impossible to obtain more than two terms of the representation in this way. Also other analytic approaches took the generating function as a starting point which lead to interesting analytic results, but not to explicit solutions for the equilibrium distribution (compare, for instance, Cohen and Boxma [8]). Another approach would have been to exploit more directly the representation of the equilibrium distribution as a linear combination of product forms in order to obtain a more explicit expression. This may have been tried in the course of the years, but the idea was only successful when it was combined with an idea of how to construct subsequent terms of the linear combination (cf. [1]). The essence of the latter idea is that there is an uncountable number of product forms satisfying the equilibrium equations for states in the inner area; of these product forms a selection should be made such that a particular linear combination of this selection not only satisfies the equilibrium equations for inner states, but also for all boundary states. In [1] it has been worked out in detail how this selection can be made via a compensation argument: alternately, product forms are added which make the current solution a solution for the equilibrium equations on the horizontal and on the vertical boundary respectively. In this way the current solution always needs compensation to turn it into a solution on the other boundary, however, after compensation it stops being a solution on the first boundary, which requires again compensation, etc. In [1] it has been proved that the constructed solution converges absolutely.

From a practical point of view the shortest queue problem with Poisson arrivals and identical exponential servers is not the most interesting one. Therefore, the aim of the present paper is to extend the compensation method to the case of two identical Erlang-r servers. We will do so for the situation of *shortest expected delay routing* which means that the jobs of all customers are considered as consisting of r exponentially distributed subjobs and a new job is joining the queue with the lowest number of subjobs to be executed (in case the numbers of subjobs are equal, either queue is joined with probability $\frac{1}{2}$). Hordijk and Koole [13] and Weber [16] have shown that, within certain constraints, this way of routing is optimal. For us, however, the important aspect is that the model can be represented by a two-dimensional random walk in the first quadrant of the plane: Define i and j as the numbers of subjobs in both queues and define the state (m, n) by $m = \min(i, j), n = |i - j|$. Now the process on (m, n) is a random walk on the grid points with $m \geq 0, n \geq 0$. This random walk is homogeneous for $m > 0, n \geq r$. On the line $m = 0$ its mechanism is just the truncation of the main mechanism. On each of the lines $n = 0, 1, \ldots, r - 1$ its behaviour is homogeneous, but different on different lines (compare fig.1, where the transition diagram is depicted for an instream rate λ and three subjobs with mean 1 for each job).

The analysis of the described model is of practical value, but also theoretically significant, since the model essentially differs from the other models for which the compensation method has been developed so far. In [4] it has been investigated for which random walks on the first quadrant of the plane a compensation approach could be developed. In order to keep the analysis tractable, a restriction was made to random walks with transitions to neighbouring states only. For this restricted class of random walks it was proven, that the essential condition for the compensation approach is that no transitions are possible from inner states to the North, the North-East and the East. The model to be treated in the present paper satisfies the latter condition for $n \geq r$ (compare fig.1), but it does not belong to the restricted class considered in [4], since for inner states the transitions are not restricted to neighbouring states. At first sight the difference is not essential, although it will make the analysis more cumbersome. The main difference seems to be

2

that now not only the states on the line $n = 0$ form the horizontal boundary, but also states with $n = 1, \ldots, n = r - 1$ are part of the horizontal boundary. However, it will appear that the compensation procedure will get an essentially more complex form. The reason for this phenomenon is that the equilibrium equations for the states on the lines $m = 0, \ldots, m = r - 1$ all differ from the equations in the real inner states, which implies that compensation on the vertical boundary sets r requirements and therefore appears to require r compensatory terms. Each of these terms then requires a compensation on the horizontal boundary.

The compensation procedure is developed in Section 3 after a formulation of the equilibrium equations in Section 2. Section 4 treats the question of convergence and Section 5 is devoted to finding starting solutions for the compensation process, which lead to useful linear combinations. The more complex compensation procedure essentially complicates the quest for feasible starting solutions. In fact, it becomes more and more apparent that finding feasible starting solutions is the essential difficulty of the compensation approach. In the first exploration of the compensation approach for the symmetric shortest queue problems (cf.[1]), this aspect of finding a starting solution still looked like a simple question, since there was a natural candidate, but later explorations placed this aspect more and more in the center (cf. [3, 4, 6, 7]). So, Section 5 may well be called the key part of this paper. In fact the problem is solved by the solution of a variant of our problem in which instantaneous jockeying is allowed. This solution is a finite linear combination of products. Section 6 gives the finishing touch by first investigating the risk of degeneration of the compensation procedure and then using the results of the previous sections to prove that, indeed, the equilibrium probabilities can be represented by an absolutely convergent linear combination of countably many product forms for which the factors and the coefficients can easily be computed. Section 7 shows, as an example, that the moments of the waiting time and the waiting time distribution can be expressed as the sum of infinite series, which can easily be approximated. Numerical aspects of the computation of the equilibrium probabilities are treated in Section 8. Section 9 gives some numerical results and, finally, Section 10 is devoted to some conclusions and comments on the further development of the compensation approach in general as well as on the use of this approach for models related to the present one.

2 The model and the equilibrium equations

We consider a system with two identical parallel servers. The service times are Erlang-r distributed. Jobs arrive in a Poisson stream with intensity λ. Without loss of generality, we may assume that the Erlang-distributions have scale parameter 1. To guarantee stability, we require:

$$\lambda r < 2.$$

An arriving job can be thought of as consisting of r identical subjobs, where each subjob requires an exponentially distributed service time with unit mean. Arriving jobs join the queue with the *smallest number of subjobs*, and in case the number of subjobs in the two queues is equal, they join either queue with probability $\frac{1}{2}$.

This queueing system can be represented by a continuous-time Markov process, whose natural state space consists of the pairs (i, j) where i and j are the numbers of subjobs in each queue. In order to obtain a homogeneous random walk in the first quadrant, we will use the variables m and n instead of i and j, where $m = \min(i, j)$ and $n = |j - i|$.

Let $\{p_{m,n}\}$ be the equilibrium distribution. For the model with Erlang-3 servers the transition-rate diagram is depicted in Figure 1.
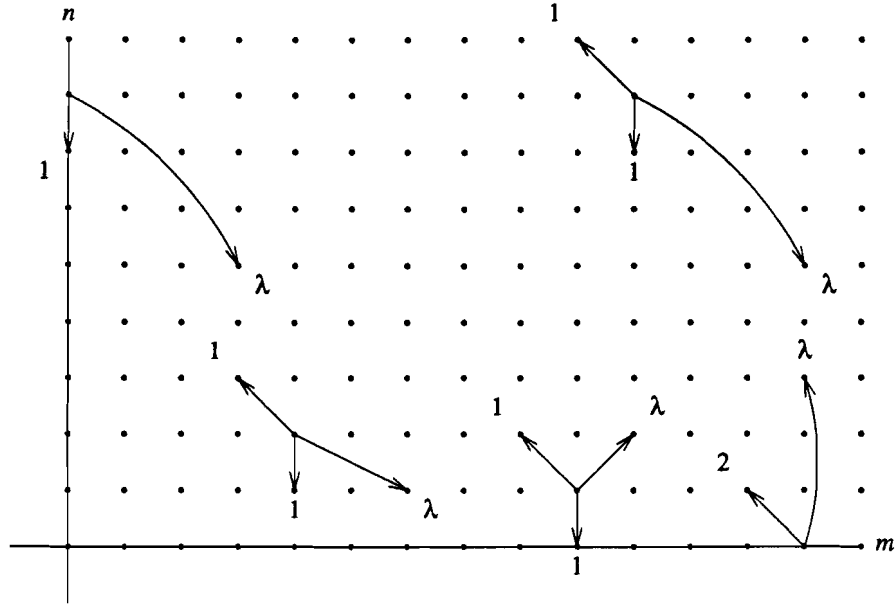
Figure 1: *The transition-rate diagram for the queueing model with Erlang-3 servers. The transition rates on the vertical boundary are the truncation of the rates in the inner region. The horizontal boundary consists of 3 layers.*

The equilibrium probabilities satisfy the following relations, which are obtained by equating for each state the average number of times per time unit the state is reached and the average number of times per time unit the state is left. Since the states with $m < r$ cannot result from a transition triggered by the arrival of a new job (compare Figure 1), the 'most regular' equation is only obtained for states (m,n) with $m \geq r, n > r$.

$$p_{m,n}(2 + \lambda) = p_{m-r,n+r}\lambda + p_{m,n+1} + p_{m+1,n-1}, \qquad (2.1)$$

$$m \geq r, n > r;$$

$$p_{m,n}(2 + \lambda) = p_{m,n+1} + p_{m+1,n-1}, \qquad (2.2)$$

$$0 < m < r, n > r;$$

$$p_{0,n}(1 + \lambda) = p_{0,n+1} + p_{1,n-1}, \qquad (2.3)$$

$$n > r;$$

$$p_{m,n}(2 + \lambda) = p_{m-r,n+r}\lambda + p_{m,n+1} + p_{m+1,n-1} + p_{m-r+n,r-n}\lambda, \qquad (2.4)$$

$$m \geq r, 1 < n \leq r;$$

$$p_{m,1}(2 + \lambda) = p_{m-r,1+r}\lambda + p_{m,2} + p_{m+1,0}2 + p_{m-r+1,r-1}\lambda, \qquad (2.5)$$

$$m \geq r;$$

$$p_{m,0}(2 + \lambda) = p_{m-r,r}\lambda + p_{m,1}, \qquad (2.6)$$

$$m \geq r.$$

4

The equations in states (m, n) with $0 \leq m < r$ and $0 \leq n \leq r$ are left out because of their minor importance to the analysis. The equations (2.1) form the inner conditions, the equations (2.2) and (2.3) form the vertical boundary conditions and, finally, the equations (2.4), (2.5) and (2.6) form the horizontal boundary conditions. In the following sections we shall try to prove that the equilibrium probabilities $p_{m,n}$ can be expressed as an infinite sum of products of powers, i.e., there are α_i, β_i and c_i such that for all m and n,

$$p_{m,n} = \sum_{i=0}^{\infty} c_i \alpha_i^m \beta_i^n \,.$$

3 The compensation approach

In this section we will try to construct a formal solution to the set of equations (2.1)–(2.6) by combining products $\alpha^m \beta^n$ satisfying equation (2.1) for the interior points of the grid.

Inserting the product $\alpha^m \beta^n$ into equation (2.1) and then dividing both sides of that equation by the common factor $\alpha^{m-r} \beta^{n-1}$ leads to the following characterization:

Lemma 3.1 *The product $\alpha^m \beta^n$ is a solution of equation (2.1) if α and β satisfy*

$$\alpha^r \beta (2 + \lambda) = \beta^{r+1} \lambda + \alpha^r \beta^2 + \alpha^{r+1} \,. \tag{3.1}$$

Any linear combination of products $\alpha^m \beta^n$ with α and β satisfying equation (3.1) is a solution of equation (2.1). The next step is to construct such linear combinations which also satisfy (2.2)–(2.6).

Let us start by considering an arbitrary product $\alpha^m \beta^n$ with α, β satisfying equation (3.1). Most likely, this form will not satisfy the vertical boundary conditions (2.2)–(2.3). The straightforward compensation idea implies the addition of a compensating term $c \tilde{\alpha}^m \tilde{\beta}^n$ such that

$$\alpha^m \beta^n + c \tilde{\alpha}^m \tilde{\beta}^n$$

satisfies (2.1)–(2.3). Insertion of this linear combination into (2.2)–(2.3) yields r equations of the form

$$\beta^{n-1} A + c \tilde{\beta}^{n-1} B = 0 \qquad \text{for } n > r \,.$$

This condition requires that

$$\tilde{\beta} = \beta \,.$$

Since $\tilde{\alpha}$ and $\tilde{\beta}$ have to satisfy (3.1), it follows that $\tilde{\alpha}$ and α must be two of the $r + 1$ roots of (3.1) for the given β, only leaving c for fulfilling r requirements. Hence, this choice does not provide sufficient freedom to adopt the compensating term to the requirements. However, we may also use the other $r - 1$ roots of (3.1) for the given β, resulting in the following compensation procedure which is slightly more complicated:

Try to find $c_1, \ldots, c_r$ such that the linear combination

$$\alpha^m \beta^n + c_1 \alpha_1^m \beta^n + \ldots + c_r \alpha_r^m \beta^n$$

satisfies (2.2)–(2.3), where $\alpha, \alpha_1, \ldots, \alpha_r$ are the $r + 1$ roots of equation (3.1) for given β.

5

So, each term satisfies (2.1). If we note that (2.2) and (2.3) are just versions of (2.1) with one and two terms missing respectively, we can insert the linear combination into (2.2) and (2.3) and simplify by exploiting (2.1), which leads to:

$$\alpha^{m-r}\beta^{n+r}\lambda + \sum_{i=1}^{r} c_i \alpha_i^{m-r}\beta^{n+r}\lambda = 0, \qquad 0 < m < r, n > r; \qquad (3.2)$$

$$\alpha^{-r}\beta^{n+r}\lambda + \sum_{i=1}^{r} c_i \alpha_i^{-r}\beta^{n+r}\lambda = \beta^n + \sum_{i=1}^{r} c_i \beta^n, \qquad n > r. \qquad (3.3)$$

By dividing these equations by $\beta^{n-m}\lambda$ we get

$$\left(\frac{\beta}{\alpha}\right)^{r-m} + \sum_{i=1}^{r} c_i \left(\frac{\beta}{\alpha_i}\right)^{r-m} = 0, \qquad 0 < m < r;$$

$$\left(\frac{\beta}{\alpha}\right)^{r} + \sum_{i=1}^{r} c_i \left(\frac{\beta}{\alpha_i}\right)^{r} = \frac{1}{\lambda} + \sum_{i=1}^{r} c_i \frac{1}{\lambda}.$$

These equations for the coefficients $c_1, \ldots, c_r$ are of a *Vandermonde-type*, and therefore, may be solved explicitly. The solution can be simplified by using that the product of the roots $\alpha, \alpha_1, \ldots, \alpha_r$ is equal to $(-1)^r \beta^r \lambda$. This compensation procedure is summarized in the following lemma.

Lemma 3.2 *Let $\alpha_0, \alpha_1, \ldots, \alpha_r$ be the roots of equation (3.1) for given β. Then the sum*

$$\alpha_0^m \beta^n + c_1 \alpha_1^m \beta^n + \ldots + c_r \alpha_r^m \beta^n$$

satisfies the conditions (2.1)–(2.3) if the coefficients $c_1, \ldots, c_r$ are given by

$$c_i = -\frac{\left(\frac{\alpha_i}{\beta} - 1\right)\prod_{j\neq i}\left(\frac{\beta}{\alpha_0} - \frac{\beta}{\alpha_j}\right)}{\left(\frac{\alpha_0}{\beta} - 1\right)\prod_{j\neq i}\left(\frac{\beta}{\alpha_i} - \frac{\beta}{\alpha_j}\right)} \qquad for\ i = 1, \ldots, r,$$

where the index j in the two products runs through $1, \ldots, r$.

To satisfy the horizontal boundary conditions (2.4)–(2.6) we try to follow a similar approach starting with the addition of $d\hat{\alpha}^m \hat{\beta}^n$ instead of $c\tilde{\alpha}^m \tilde{\beta}^n$ to the original term $\alpha^m \beta^n$. This requires that $\hat{\alpha}$ has to be equal to α and that $\hat{\beta}$ and β have to be roots of equation (3.1) for the given α, leaving d to satisfy $r + 1$ requirements, which, clearly, is not sufficient. To create sufficient freedom we introduce extra coefficients for the solution in states (m, n) with $n = 0, 1, \ldots, r - 1$ instead of using the other $r - 1$ roots of equation (3.1) for the given α, by considering

$$\begin{array}{ll} \alpha^m \beta^n + d\alpha^m \hat{\beta}^n & \text{for } m \geq 0, n \geq r \text{ and} \\ e_n \alpha^m & \text{for } m \geq 0, n = 0, \ldots, r - 1. \end{array} \qquad (3.4)$$

The analogous procedure would not have worked for the vertical boundary, since we would have been forced by the inner conditions (2.1) for the states on the lines $m = r, r + 1, \ldots, m + r - 1$ to set the coefficients of β^n equal to $\alpha^m + c\tilde{\alpha}^m$, yielding insufficient freedom to satisfy the vertical boundary conditions (2.2)–(2.3). In the case of the horizontal boundary, however, the lines $n = 0, 1, \ldots, r - 1$ only allow transitions to the line $n = r$

and not to any state with $n > r$. Because of this feature, the equations (2.1) for inner states do not restrict the freedom of choice for the e_n. Note that the way of compensating of (3.4) was also used in [4], the vertical compensation, however, is of a new type. Later it will become clear why we don't use the approach of Lemma 3.2 also for the horizontal boundary. It has to do with convergence of the solutions.

Insertion of the terms (3.4) into the conditions (2.4)–(2.6) and then dividing by the common factor α^{m-r} yields $r+1$ linear equations for the coefficients $e_0, \ldots, e_{r-1}, d$, which may readily be solved. This procedure is summarized in the following lemma.

Lemma 3.3 *Let β_0 and β_1 be roots of equation (3.1) for given α. Then the terms*

$$\alpha^m \beta_0^n + d\alpha^m \beta_1^n \qquad for\ m \geq 0,\ n \geq r\ and$$
$$e_n \alpha^m \qquad for\ m \geq 0,\ n = 0, \ldots, r-1.$$

satisfy the conditions (2.1) and (2.4)–(2.6) if the coefficients $e_0, \ldots, e_{r-1}, d$ are the solution of the following $r+1$ linear equations:

$$A(\alpha) \begin{pmatrix} e_0 \\ \vdots \\ e_{r-1} \end{pmatrix} + C(\alpha, \beta_1)d\beta_1^r + C(\alpha, \beta_0)\beta_0^r = 0,$$

where the $(r+1) \times r$-matrix $A(\alpha)$ and the $(r+1)$-column vector $C(\alpha, \beta)$ are given by

$$A(\alpha) = \begin{pmatrix} \kappa\alpha^r & \alpha^r & 0 & \ldots & 0 & 0 & 0 \\ 2\alpha^r & \kappa\alpha^{r-1} & \alpha^{r-1} & & 0 & 0 & 0 \\ 0 & \alpha^{r-1} & \kappa\alpha^{r-2} & & 0 & 0 & 0 \\ \vdots & & & & & & \vdots \\ 0 & 0 & 0 & & \alpha^3 & \kappa\alpha^2 & \alpha^2 \\ 0 & 0 & 0 & & 0 & \alpha^2 & \kappa\alpha \\ 0 & 0 & 0 & \ldots & 0 & 0 & \alpha \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 & \ldots & 0 & 0 & 0 \\ 0 & 0 & 0 & & 0 & 0 & \lambda \\ 0 & 0 & 0 & & 0 & \lambda & 0 \\ \vdots & & & & & & \vdots \\ 0 & 0 & \lambda & & 0 & 0 & 0 \\ 0 & \lambda & 0 & & 0 & 0 & 0 \\ \lambda & 0 & 0 & \ldots & 0 & 0 & 0 \end{pmatrix},$$

$$C(\alpha, \beta) = \begin{pmatrix} \lambda \\ \lambda\frac{\beta}{\alpha} \\ \lambda(\frac{\beta}{\alpha})^2 \\ \vdots \\ \lambda(\frac{\beta}{\alpha})^{r-2} \\ \lambda(\frac{\beta}{\alpha})^{r-1} + \alpha \\ -\frac{\alpha}{\beta} \end{pmatrix},$$

with $\kappa = -(2+\lambda)$.

Lemma 3.2 and 3.3 show that an arbitrary solution of equation (2.1) can be compensated in such a way that it either becomes a solution to (2.1), (2.2), (2.3) (Lemma 3.2) or a solution to (2.1), (2.4), (2.5), (2.6) (Lemma 3.3). One might hope that alternate use of these two compensation procedures would in the long run turn some solution of equation (2.1) into a solution of (2.1)–(2.6), at least if the compensating terms converge to zero sufficiently fast. For the time being, we do not attend to the convergence problem, but only define the formal solution which can be constructed with this compensation approach.

Starting with an arbitrary product form solution $\alpha_0^m \beta_0^n$ of equation (2.1), we add $c_1 \alpha_1^m \beta_0^n + \ldots + c_r \alpha_r^m \beta_0^n$ to compensate for the error of $\alpha_0^m \beta_0^n$ on the vertical boundary and, by doing so, we introduce r new errors on the horizontal boundary, since each of the terms $c_1 \alpha_1^m \beta_0^n, \ldots, c_r \alpha_r^m \beta_0^n$ violates these boundary conditions. An essential feature is now that we compensate each of these errors individually on the horizontal boundary. To compensate for the error of $c_1 \alpha_1^m \beta_0^n$ we consider

$$c_1 \alpha_1^m \beta_0^n + c_1 d_1 \alpha_1^m \beta_1^n \qquad \text{for } m \geq 0, \, n \geq r \text{ and}$$
$$c_1 e_{1,n} \alpha_1^m \qquad \text{for } m \geq 0, \, n = 0, \ldots, r-1$$

where β_1 is one of the r other roots of (3.1) with $\alpha = \alpha_1$, and choose $e_{1,0}, \ldots, e_{1,r-1}, d_1$ such that these terms satisfy the horizontal boundary conditions (2.4), (2.5) and (2.6). The same procedure is used to compensate for the terms $c_2 \alpha_2^m \beta_0^n, \ldots, c_r \alpha_r^m \beta_0^n$. However, the new terms $c_1 d_1 \alpha_1^m \beta_1^n, \ldots, c_r d_r \alpha_r^m \beta_r^n$ violate the vertical boundary conditions, so we have to add again terms, 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. The resulting sum has, due to the compensation on the vertical boundary, the structure of an r-fold tree, which is illustrated in Figure 2.

By definition we have $c_0 = 1$ and the coefficient d_0 will be defined later on. Each term in the sum satisfies (2.1), each sum of $r + 1$ terms with the same β-factor satisfies the vertical boundary conditions (2.2)–(2.3) and each sum of two terms with the same α-factor satisfies the horizontal boundary conditions (2.4)–(2.6). Since the equilibrium equations are linear, we can conclude that the infinite sum formally satisfies the equations (2.1)–(2.6). Let us define $x_{m,n}(\alpha_0, \beta_0)$ as the infinite sum of compensation terms. For all $m \geq 0$, $n \geq r$ set

$$
\begin{aligned}
x_{m,n}(\alpha_0, \beta_0) &= \sum_{i=0}^{\infty} d_i \left(c_i \alpha_i^m + \sum_{j=1}^{r} c_{ri+j} \alpha_{ri+j}^m \right) \beta_i^n \\
&\quad + \sum_{i=0}^{-\infty} d_{i-1} \left(c_i \alpha_i^m + \sum_{j=1}^{r} c_{ri-j} \alpha_{ri-j}^m \right) \beta_{i-1}^n \\
&\qquad \text{(terms with same } \beta\text{-factor)}, \qquad\qquad (3.5)
\end{aligned}
$$

$$
\begin{aligned}
&= \sum_{i=0}^{\infty} \sum_{j=1}^{r} c_{ri+j} \left(d_i \beta_i^n + d_{ri+j} \beta_{ri+j}^n \right) \alpha_{ri+j}^m \\
&\quad + c_0 \left(d_0 \beta_0^n + d_{-1} \beta_{-1}^n \right) \alpha_0^m \\
&\quad + \sum_{i=0}^{-\infty} \sum_{j=1}^{r} c_{ri-j} \left(d_{i-1} \beta_{i-1}^n + d_{ri-j-1} \beta_{ri-j-1}^n \right) \alpha_{ri-j}^m \\
&\qquad \text{(terms with same } \alpha\text{-factor)}. \qquad\qquad (3.6)
\end{aligned}
$$

The representations (3.5) and (3.6) reflect the compensation on the vertical and horizontal boundaries, respectively. The compensation on the horizontal boundary requires the introduction of new coefficients for the terms in $x_{m,n}(\alpha_0, \beta_0)$ with $n < r$. For all $m \geq 0$, $n = 0, 1, \ldots, r-1$ set

$$x_{m,n}(\alpha_0, \beta_0) = \sum_{i=-\infty}^{\infty} c_i e_{i,n} \alpha_i^m . \qquad\qquad (3.7)$$

Below we formulate recurrence relations for α_i, β_i, c_i, d_i, $e_{i,0}, \ldots, e_{i,r-1}$.
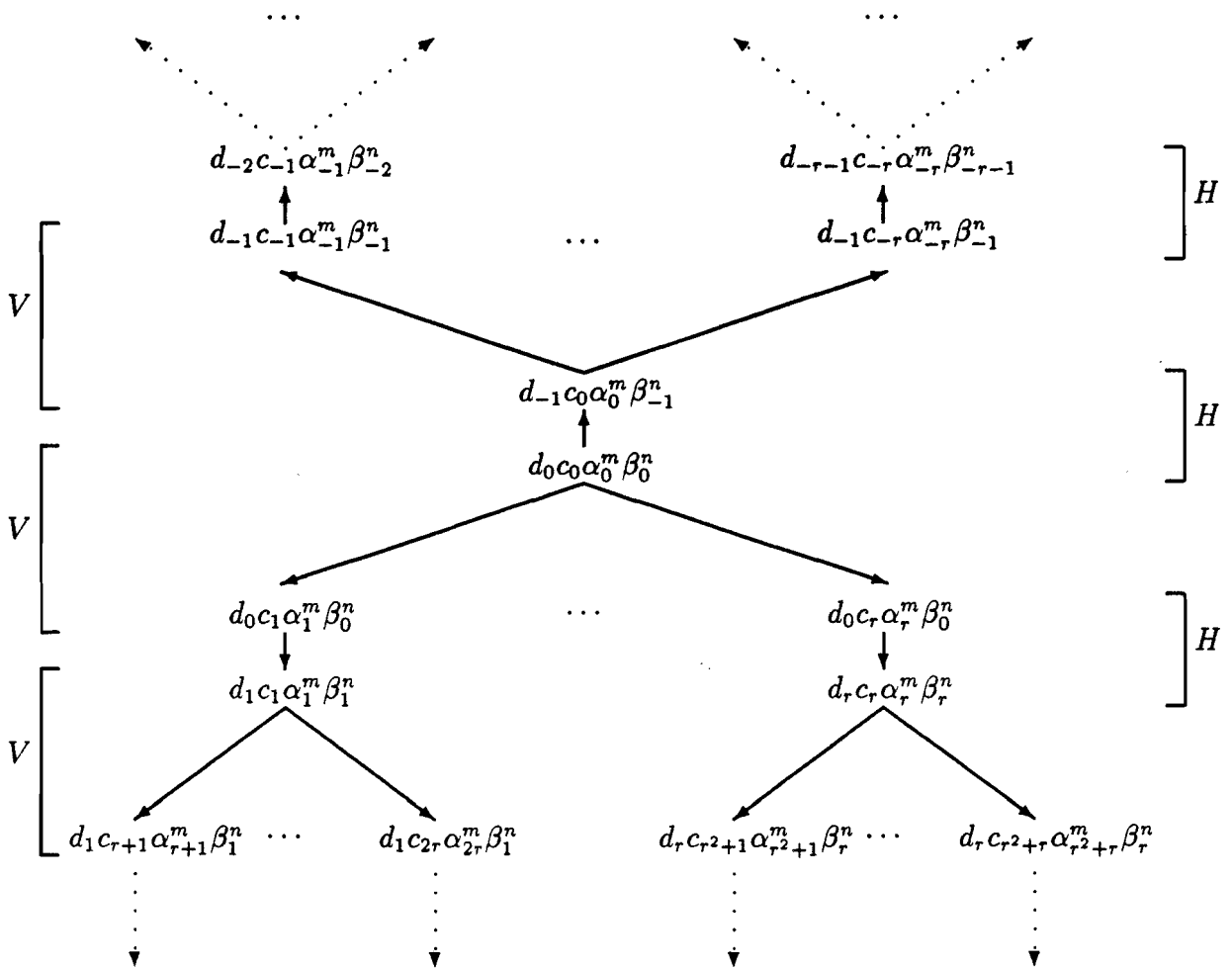
8

Figure 2: *The resulting infinite sum of compensation terms. Sums of $r+1$ terms with the same β-factor satisfy the vertical boundary conditions (V) and sums of two terms with the same α-factor satisfy the horizontal boundary conditions (H).*

The numbers α_i and β_i can be represented in an r-fold tree which is depicted in Figure 3 (compare Figure 2).

For given initial roots α_0 and β_0 of equation (3.1), the numbers α_i and β_i with $i \geq 0$ are generated such that for all $i \geq 0$ the r descendants $\alpha_{ri+1}, \ldots, \alpha_{ri+r}$ of β_i and its predecessor α_i are the $r+1$ roots of (3.1) with fixed $\beta = \beta_i$, and the descendant β_{ri+j} of α_{ri+j}, $j = 1, \ldots, r$ and its predecessor β_i are two roots of (3.1) with fixed $\alpha = \alpha_{ri+j}$. Notice that β_{ri+j} is not uniquely determined, since we can choose among r candidates. In the next section we will show that the convergence question determines the selection of the appropriate root for β. There it will also become clear why we do not use the more elegant way of compensating at the vertical boundary also for the horizontal boundary: that would force us to use all the roots for β, which raises convergence problems.

The numbers α_i and β_i with $i \leq 0$ are generated analogously, starting with β_{-1} being one of the r roots other than β_0 of equation (3.1) with $\alpha = \alpha_0$.

Initially set

$$c_0 = 1.$$

9
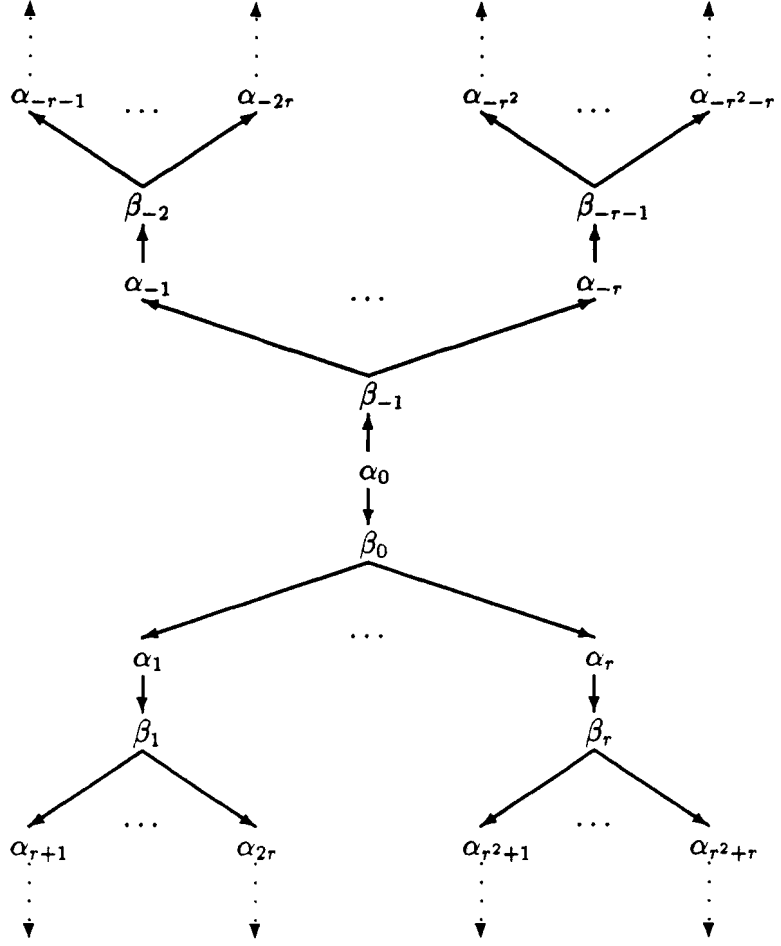
Figure 3: *The r-fold tree structure of the sequence of α_i and β_i.*

The coefficients c_i with $i > 0$ are generated such that for all $i \geq 0$ the sum

$$(c_i \alpha_i^m + \sum_{j=1}^{r} c_{ri+j} \alpha_{ri+j}^m) \beta_i^n$$

satisfies the vertical boundary conditions (2.2)–(2.3). Hence, applying Lemma 3.2, the coefficients $c_{ri+1}, \ldots, c_{ri+r}$ for $i \geq 0$ can be obtained from c_i by

$$c_{ri+j} = -\frac{\left(\frac{\alpha_{ri+j}}{\beta_i} - 1\right) \prod_{k \neq j} \left(\frac{\beta_i}{\alpha_i} - \frac{\beta_i}{\alpha_{ri+k}}\right)}{\left(\frac{\alpha_i}{\beta_i} - 1\right) \prod_{k \neq j} \left(\frac{\beta_i}{\alpha_{ri+j}} - \frac{\beta_i}{\alpha_{ri+k}}\right)} c_i \qquad \text{for } j = 1, \ldots, r, \qquad (3.8)$$

where the index k in the two products runs through $1, \ldots, r$. Similar recurrence relations can be formulated for the coefficients c_i with $i < 0$.

Initially the coefficients d_0, d_{-1}, $e_{0,0}, \ldots, e_{0,r-1}$ are determined such that the sequence $p_{m,n}$ given by

$$(d_0 \beta_0^n + d_{-1} \beta_{-1}^n) \alpha_0^m \qquad \text{for } m \geq 0,\ n \geq r \text{ and}$$
$$e_{0,n} \alpha_0^m \qquad \text{for } m \geq 0,\ n = 0, \ldots, r - 1$$

10

satisfies the horizontal boundary conditions (2.4)–(2.6). This means that $d_0, d_{-1}, e_{0,0}, \ldots, e_{0,r-1}$ have to be a *nonnull* solution of

$$A(\alpha_0)\begin{pmatrix} e_{0,0} \\ \vdots \\ e_{0,r-1} \end{pmatrix} + C(\alpha_0, \beta_{-1})d_{-1}\beta_{-1}^r + C(\alpha_0, \beta_0)d_0\beta_0^r = 0. \qquad (3.9)$$

The coefficients $d_i, e_{i,0}, \ldots, e_{i,r-1}$ with $i > 0$ are generated such that for all $i \geq 0$ and $j = 1, \ldots, r$ the sequence $p_{m,n}$ given by

$$\begin{array}{ll} (d_i\beta_i^n + d_{ri+j}\beta_{ri+j}^n)\alpha_{ri+j}^m & \text{for } m \geq 0, \, n \geq r \text{ and} \\ e_{ri+j,n}\alpha_{ri+j}^m & \text{for } m \geq 0, \, n = 0, \ldots, r - 1 \end{array}$$

satisfies the horizontal boundary conditions (2.4)–(2.6). Hence, applying Lemma 3.3, the coefficients $d_{ri+j}, e_{ri+j,0}, \ldots, e_{ri+j,r-1}$ for $i \geq 0$ and $j = 1, \ldots, r$ are the solution of

$$A(\alpha_{ri+j})\begin{pmatrix} e_{ri+j,0} \\ \vdots \\ e_{ri+j,r-1} \end{pmatrix} + C(\alpha_{ri+j}, \beta_{ri+j})d_{ri+j}\beta_{ri+j}^r + C(\alpha_{ri+j}, \beta_i)d_i\beta_i^r = 0. \qquad (3.10)$$

Similar recurrence relations can be formulated for $d_{i-1}, e_{i,0}, \ldots, e_{i,r-1}$ with $i < 0$.

This concludes the definition of $x_{m,n}(\alpha_0, \beta_0)$. For any pair of roots α_0, β_0 of equation (3.1) the series $x_{m,n}(\alpha_0, \beta_0)$ formally satisfies the equations (2.1)–(2.6). Below, it will be shown that $x_{m,n}(\alpha_0, \beta_0)$ also satisfies the equilibrium equations in the points (m, n) with $0 \leq m < r, 0 < n \leq r$ and $m + n \geq r$. The equations in these points are given by

$$p_{0,r}(1 + \lambda) = p_{0,r+1} + p_{1,r-1} + p_{0,0}\lambda; \qquad (3.11)$$

$$p_{m,n}(2 + \lambda) = p_{m,n+1} + p_{m+1,n-1} + p_{m-r+n,r-n}\lambda, \qquad (3.12)$$

$$0 < m < r - 1, 1 < n < r, m + n \geq r;$$

$$p_{r-1,1}(2 + \lambda) = p_{r-1,2} + p_{r,0}2 + p_{0,r-1}\lambda. \qquad (3.13)$$

Compared to the horizontal boundary conditions (2.4)–(2.6), only the term $p_{m-r,n+r}\lambda$ at the right-hand side is missing and in (3.11) an extra term $p_{0,r}$ at the left-hand side is missing. Using (3.2)–(3.3), it is readily verified that for $p_{m,n} = x_{m,n}(\alpha_0, \beta_0)$ the missing terms vanish. So it is harmless for $p_{m,n} = x_{m,n}(\alpha_0, \beta_0)$ to add these terms to (3.11)–(3.13), and by doing so, we get equations of the same form as the equations (2.4)–(2.6), implying that $x_{m,n}(\alpha_0, \beta_0)$ indeed satisfies the conditions (3.11)–(3.13).

Hence, we can conclude that $x_{m,n}(\alpha_0, \beta_0)$ formally satisfies all equilibrium equations, except for the ones in the points (m, n) with $m + n < r$. So there are $r(r + 1)/2$ boundary equations to be satisfied yet. If we can find $r(r + 1)/2$ pairs α_0, β_0 for which $x_{m,n}(\alpha_0, \beta_0)$ converges, then, by linearly combining these solutions, we can construct a solution that also satisfies the remaining boundary conditions. In the next section it will be investigated for what α_0, β_0 the series $x_{m,n}(\alpha_0, \beta_0)$ converges.

4   Convergence results

For convergence of $x_{m,n}(\alpha_0, \beta_0)$ for *fixed* m and n we need that the compensation terms converge sufficiently fast to zero as i tends to plus or minus infinity. So it would help if

we are able to show that α_i and β_i converge to zero as i tends to plus or minus infinity. For convergence of the sum of $x_{m,n}(\alpha_0, \beta_0)$ over *all values* m and n (necessary for normalization) we need that $|\alpha_i| < 1$, $|\beta_i| < 1$ for all i. In this section we investigate whether α_i and β_i indeed converge to zero and remain inside the open unit disk.

The numbers α_i and β_i are defined as roots of equation (3.1). The next lemma formulates some useful properties of the roots of equation (3.1).

Lemma 4.1 *For each fixed α satisfying $0 < |\alpha| < 1$, equation (3.1) has exactly one root β with $0 < |\beta| < R|\alpha|$ and r roots β with $|\beta| > |\alpha|$, where R is the positive root, less than one, of the equation*

$$\lambda x^{r+1} + x^2 - (2 + \lambda)x + 1 = 0.$$

For each fixed β satisfying $0 < |\beta| < 1$, equation (3.1) has exactly r simple roots α with $0 < |\alpha| < |\beta|$ and one root α with $|\alpha| > |\beta|$.

Proof We first prove the first part of the lemma. Let α be fixed and satisfy $0 < |\alpha| < 1$. By dividing equation (3.1) by α^{r+1} and introducing $z = \beta/\alpha$ we may rewrite (3.1) as $f(z) + g(z) = 0$, where $f(z) = \lambda z^{r+1} + \alpha z^2$ and $g(z) = -(2 + \lambda)z + 1$. It follows for all z with $z \neq 0$ that

$$|f(z)| - |g(z)| < \lambda|z|^{r+1} + |z|^2 - (2 + \lambda)|z| + 1.$$

It is readily verified that the right-hand side $h(x) = \lambda x^{r+1} + x^2 - (2 + \lambda)x + 1$ is convex for $x \geq 0$, $h(0) = 1 > 0$, $h(1) = 0$ and $h'(1) = r\lambda > 0$. Hence, $h(x) = 0$ has two positive roots, namely $x = 1$ and one in $(0, 1)$, say $x = R$. Subsequently applying Rouché's Theorem to the circles $|z| = R$ and $|z| = 1$ proves the first part of the lemma.

Now let β be fixed and satisfy $0 < |\beta| < 1$. By dividing equation (3.1) by β^{r+1} and introducing $u = \alpha/\beta$, we may rewrite (3.1) as $\hat{f}(u) + \hat{g}(u) = 0$, where $\hat{f}(u) = u^{r+1} + \beta u^r$ and $\hat{g}(u) = -(2 + \lambda)u^r + \lambda$. Analogously to the first part of the proof, application of Rouché's Theorem yields that $\hat{f}(u) + \hat{g}(u)$ has r zeros inside the circle $|u| = 1$. It is easily seen that the derivative of $\hat{f}(u) + \hat{g}(u)$ only vanishes for

$$\tilde{u} = \frac{r}{r+1}(2 + \lambda - \beta).$$

Since

$$|\hat{f}(\tilde{u}) + \hat{g}(\tilde{u})| = |\frac{1}{r}\tilde{u}^{r+1} - \lambda| \geq \frac{1}{r}|\tilde{u}|^{r+1} - \lambda > \frac{1}{r}\left(\frac{r}{r+1}(1 + \lambda)\right)^{r+1} - \lambda \geq 0,$$

we can conclude that all zeros of $\hat{f}(u) + \hat{g}(u)$ are simple, which completes the proof of the second part of the lemma. $\square$

For the convergence of $x_{m,m}(\alpha_0, \beta_0)$ it is helpful to choose α_i, β_i as small as possible. Therefore, starting with α_0 satisfying $0 < |\alpha_0| < 1$, we choose β_0 as the root of (3.1) with $0 < |\beta_0| < |\alpha_0|$. Then Lemma 4.1 directly yields that for all $i \geq 0$ it is possible to choose β_i as the smallest root in absolute value of (3.1) with fixed $\alpha = \alpha_i$. This implies that α_i and β_i decrease exponentially fast to zero (at least with rate R) as i tends to plus infinity and that $|\alpha_i| < 1$, $|\beta_i| < 1$ for all $i \geq 0$. However, in the upper part of the tree of numbers α_i and β_i, it appears that α_i and β_i do not remain inside the open unit disk. Since $|\beta_0| < |\alpha_0|$, it holds that $|\beta_{-1}| > |\alpha_0|$, so there is exactly one α_j with $j = -1, \ldots, -r$ such that $|\alpha_j| > |\beta_{-1}|$, say α_{-1}. Now we have $|\beta_{-2}| > |\alpha_{-1}|$ (for the

12

other α_{-j} with $j = 2, \ldots, r$ it is possible to choose $|\beta_{-j-1}| < |\alpha_{-j}|)$. Hence, repeating the argument above, we find a path starting in α_0 along which α_k and β_k are monotonously increasing in absolute value, the modulus of the ratio of β_k and its α-predecessor being at least R^{-1}. So this path eventually ends in some β_{i-1} with $i \leq 0$ for which $|\beta_{i-1}| \geq 1$ or one of its descendants α_{ri-j}, $j = 1, \ldots, r$ satisfies $|\alpha_{r-j}| \geq 1$. In the first case, the addition of the compensating term $c_i d_{i-1} \beta_{i-1}^n \alpha_i^m$ would lead to a solution which cannot be normalized. Since c_i is nonzero (cf. (3.8)) this can only be repaired by requiring that $d_{i-1} = 0$. In the second case, we have to prevent the addition of $d_{i-1} c_{ri-j} \alpha_{ri-j}^m \beta_{i-1}^n$. Since c_{ri-j} is nonzero, this again implies $d_{i-1} = 0$. After renumbering the terms this amounts to the requirement $d_{-1} = 0$. This means that the initial product $d_0 c_0 \alpha_0^m \beta_0^n$ has to satisfy the horizontal boundary conditions.

Hence, summarizing our findings, to guarantee for the initial roots α_0, β_0 satisfying $1 > |\alpha_0| > |\beta_0| > 0$ that the numbers α_i and β_i converge to zero and remain inside the open unit disk, we have to require that $d_0 \alpha_0^m \beta_0^n$ satisfies the horizontal boundary conditions, i.e. there are nonnull coefficients $d_0, e_{0,0}, \ldots, e_{0,r-1}$ such that $d_0 \beta_0^n \alpha_0^m$ for $m \geq 0$, $n \geq r$ and $e_{0,n} \alpha_0^m$ for $m \geq 0$, $n = 0, \ldots, r-1$ satisfy the horizontal boundary conditions (2.4)-(2.6). In that case, namely, we only have to generate the lower part of the tree of compensating terms. To prevent that in the lower part of the tree α_i or β_i runs out of the open unit disk, we choose β_i as the *smallest root in absolute value* of equation (3.1) for fixed $\alpha = \alpha_i$ for all $i \geq 0$. Then the numbers α_i and β_i converge (exponentially fast) to zero as i tends to plus infinity and remain inside the open unit disk for all $i \geq 0$. Pairs α_0, β_0 which satisfy these requirements will be called *feasible pairs*. Since β_0 is uniquely determined once α_0 is given, we will simply speak of feasible α_0's without mentioning the corresponding β_0's.

The question arises whether there exist feasible α_0's and if so, how they may be found. In the next section we will try to answer this question.

5 The quest for feasible α_0's

Insertion of the sequence $p_{m,n}$ given by $d_0 \beta_0^n \alpha_0^m$ for $m \geq 0$, $n \geq r$ and $e_{0,n} \alpha_0^m$ for $m \geq 0$, $n = 0, \ldots, r-1$ into the horizontal boundary conditions (2.4)-(2.6), where β_0 is the smallest root of equation (3.1) for $\alpha = \alpha_0$ with $|\alpha_0| < 1$, and then dividing the resulting equations by common powers of α_0 yields the following set of equations for α_0 and the coefficients $d_0, e_{0,0}, \ldots, e_{0,r-1}$ (see (3.9) with $d_{-1} = 0$):

$$A(\alpha_0) \begin{pmatrix} e_{0,0} \\ \vdots \\ e_{0,r-1} \end{pmatrix} + C(\alpha_0, \beta_0) d_0 \beta_0^r = 0. \tag{5.1}$$

The feasible α_0's are the ones for which this set of equations has a nonnull solution. The analysis of this set of equations, however, seems to constitute a difficult analytical problem. Therefore, we will follow another approach to find the feasible α_0's, which is more elegant from an analytical as well as computational point of view. This approach is inspired by the following result for the exponential case $r = 1$. For $r = 1$ it appeared that there is a relation between the equilibrium distributions of the standard shortest queue problem and the shortest queue problem with instantaneous jockeying (i.e. a job jumps to the other queue as soon as this would improve its perspectives). In [2] it has been shown that the equilibrium probabilities $\tilde{p}_{m,n}$ of the *instantaneous jockeying problem* can be expressed as

$$\tilde{p}_{m,n} = f_n \alpha^m, \qquad m \geq 0, 0 \leq n \leq 1, (m,n) \neq (0,0),$$

for $\alpha = \rho^2$ and some f_0 and f_1. Remarkably, α appears to be the feasible α_0 for the case $r = 1$ (see [1]). In this section it will be investigated whether also for the general case the solution of the instantaneous jockeying problem produces the feasible α_0's.

For general r, the jockeying problem is characterized by the property that as soon as the difference between the number of *subjobs* in the two queues exceeds r, then one job, or equivalently a batch of r subjobs jumps from the longer to the shorter queue. For the model with Erlang-3 servers the transition-rate diagram is depicted in Figure 4 (compare Figure 1).
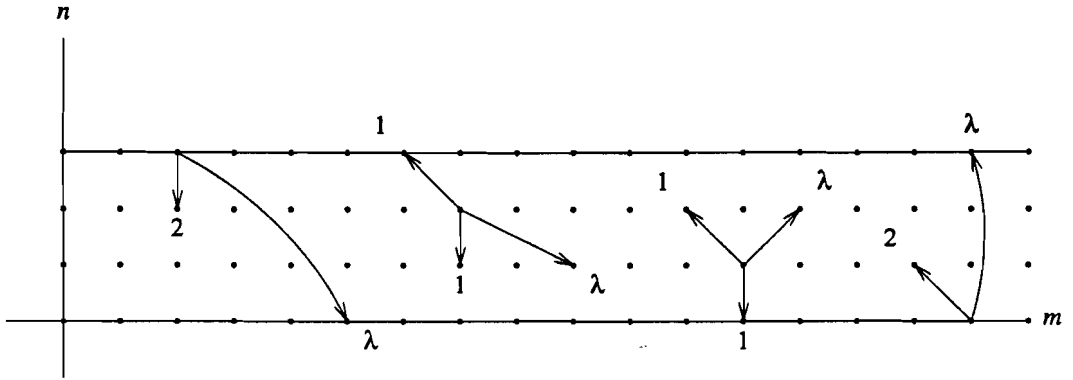

Figure 4: *The transition-rate diagram for the jockeying model with Erlang-3 servers.*

We will first show that the probabilities $\tilde{p}_{m,n}$ for the jockeying problem can be expressed as a finite sum of geometric terms,

$$\tilde{p}_{m,n} = \sum_\alpha f_n(\alpha)\alpha^m, \qquad m \geq 0,\ 0 \leq n \leq r,\ (m,n) \neq (0,0),$$

where α runs through a set of $r(r+1)/2$ possible values and then we show that these α's indeed produce the feasible α_0's.

5.1   Analysing the jockeying problem

For the points (m,n) with $m > 0$, $0 \leq n \leq r$ and $m + n \geq r$ the probabilities $\tilde{p}_{m,n}$ for the jockeying problem satisfy the following relations:

$$\tilde{p}_{m,0}(2 + \lambda) = \tilde{p}_{m,1} + \tilde{p}_{m-r,r}\lambda\,; \tag{5.2}$$

$$\tilde{p}_{m,1}(2 + \lambda) = \tilde{p}_{m,2} + \tilde{p}_{m+1,0}2 + \tilde{p}_{m-r+1,r-1}\lambda\,; \tag{5.3}$$

$$\tilde{p}_{m,n}(2 + \lambda) = \tilde{p}_{m,n+1} + \tilde{p}_{m+1,n-1} + \tilde{p}_{m-r+n,r-n}\lambda, \qquad 1 < n < r - 1\,; \tag{5.4}$$

$$\tilde{p}_{m,r-1}(2 + \lambda) = \tilde{p}_{m,r}2 + \tilde{p}_{m+1,r-2} + \tilde{p}_{m-1,1}\lambda\,; \tag{5.5}$$

$$\tilde{p}_{m,r}(2 + \lambda) = \tilde{p}_{m+1,r-1} + \tilde{p}_{m,0}\lambda\,. \tag{5.6}$$

These relations are valid for $r \geq 3$. For the special cases $r = 1$ and $r = 2$ it is easily seen how these relations should be adapted. In this subsection we will attempt to construct a solution of the jockeying problem by combining geometric terms of the form

$$\tilde{p}_{m,n} = f_n\alpha^m, \tag{5.7}$$

which all fit the equations (5.2)–(5.6) for the interior points. Inserting the form (5.7) into (5.2)–(5.6) and then dividing the resulting equations by common powers of α leads to the following set of $r+1$ equations for $f_0, \ldots, f_r$ and α.

$$\tilde{A}(\alpha) \begin{pmatrix} f_0 \\ \vdots \\ f_r \end{pmatrix} = 0, \tag{5.8}$$

where the $(r+1) \times (r+1)$-matrix $\tilde{A}(\alpha)$ is composed of the matrix $A(\alpha)$ and an extra column:

$$\tilde{A}(\alpha) = \begin{pmatrix} \kappa\alpha^r & \alpha^r & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \lambda \\ 2\alpha^r & \kappa\alpha^{r-1} & \alpha^{r-1} & 0 & 0 & & 0 & 0 & \lambda & 0 \\ 0 & \alpha^{r-1} & \kappa\alpha^{r-2} & \alpha^{r-2} & 0 & & 0 & \lambda & 0 & 0 \\ 0 & 0 & \alpha^{r-2} & \kappa\alpha^{r-3} & \alpha^{r-3} & & \lambda & 0 & 0 & 0 \\ \vdots & & & & & & & & & \vdots \\ 0 & 0 & \lambda & 0 & 0 & & \alpha^3 & \kappa\alpha^2 & \alpha^2 & 0 \\ 0 & \lambda & 0 & 0 & 0 & & 0 & \alpha^2 & \kappa\alpha & 2\alpha \\ \lambda & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & \alpha & \kappa \end{pmatrix},$$

with $\kappa = -(2 + \lambda)$. Now we would like to find the α's for which the set of equations (5.8) has a nonnull solution $f_0, \ldots, f_r$. However, it is more convenient to consider the symmetric equations obtained by insertion of the form

$$\tilde{p}_{m,n} = g_n \gamma^{2m+n} \tag{5.9}$$

into the equations (5.1)–(5.5). This leads to the following set of equations.

$$G(\gamma) \begin{pmatrix} g_0 \\ \vdots \\ g_r \end{pmatrix} = 0, \tag{5.10}$$

where the $(r+1) \times (r+1)$-matrix $G(\gamma)$ is given by

$$G(\gamma) = \begin{pmatrix} \kappa\gamma^r & \gamma^{r+1} & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \lambda \\ 2\gamma^{r+1} & \kappa\gamma^r & \gamma^{r+1} & 0 & 0 & & 0 & 0 & \lambda & 0 \\ 0 & \gamma^{r+1} & \kappa\gamma^r & \gamma^{r+1} & 0 & & 0 & \lambda & 0 & 0 \\ 0 & 0 & \gamma^{r+1} & \kappa\gamma^r & \gamma^{r+1} & & \lambda & 0 & 0 & 0 \\ \vdots & & & & & & & & & \vdots \\ 0 & 0 & \lambda & 0 & 0 & & \gamma^{r+1} & \kappa\gamma^r & \gamma^{r+1} & 0 \\ 0 & \lambda & 0 & 0 & 0 & & 0 & \gamma^{r+1} & \kappa\gamma^r & 2\gamma^{r+1} \\ \lambda & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & \gamma^{r+1} & \kappa\gamma^r \end{pmatrix},$$

Of course, the geometric forms (5.7) and (5.9) are equivalent; the substitution

$$\alpha = \gamma^2, \qquad f_n = g_n \gamma^n, \tag{5.11}$$

transforms (5.9) to (5.7). We now try to find the γ's with $|\gamma| < 1$ for which equation (5.10) has a nonnull solution $g_0, \ldots, g_r$, or in other words, for which the columns of $G(\gamma)$

are dependent. Instead, we consider the equivalent problem of finding the γ's for which the *rows* of $G(\gamma)$ are dependent, i.e., there is a nonnull solution $a_0, \ldots, a_r$ of the equations

$$(a_0, \ldots, a_r)G(\gamma) = 0 . \tag{5.12}$$

Such γ's will be called *feasible*. In the following two subsections we shall consider the cases r is *odd* and r is *even* separately. In both cases $r(r+1)$ feasible γ's will be found, with the property that for each feasible γ also $-\gamma$ is feasible. By (5.11), the square of each feasible γ yields an α for which (5.7) has a nonnull solution. Let $\mathcal{F}$ be the set of squared feasible γ's. Then for each choice of the coefficients $k(\alpha)$ the linear combination

$$\tilde{p}_{m,n} = \sum_\alpha k(\alpha)f_n(\alpha)\alpha^m , \qquad m \geq 0, \, 0 \leq n \leq r, \, (m,n) \neq (0,0) ,$$

satisfies the equations (5.2)–(5.6), where α runs through the set $\mathcal{F}$ and $f_0(\alpha), \ldots, f_r(\alpha)$ is a nonnull solution of (5.8). The remaining equilibrium equations are the ones in the states (m,n) with $m + n < r$ and the one in $(0,r)$. These equations form a linear, homogeneous system for the unknowns $k(\alpha)$ and the unknown quantity $\tilde{p}_{0,0}$. The number of equations is equal to the number of unknowns. Hence, since the system of equations is dependent, there is a nonnull solution $\tilde{p}_{m,n}$ and normalization of the $\tilde{p}_{m,n}$ produces the equilibrium distribution. These findings are summarized in the following theorem, for which the proof will be completed in the subsections 5.1.1 and 5.1.2 by finding sufficient numbers of feasible γ's.

Theorem 5.1 (Jockeying problem) *For all states (m,n) with $m \geq 0$, $0 \leq n \leq r$ and $(m,n) \neq (0,0)$ it holds that*

$$\tilde{p}_{m,n} = \sum_\alpha k(\alpha)f_n(\alpha)\alpha^m ,$$

where α runs through the set $\mathcal{F}$, $f_0(\alpha), \ldots, f_r(\alpha)$ is a nonnull solution of (5.8) and $k(\alpha)$ is some appropriately chosen coefficient.

5.1.1 Finding the feasible γ's in case $r = 2k + 1$

For each sequence $a_0, \ldots, a_r$ satisfying the equations (5.12) it follows from the symmetry of these equations that the reversed sequence $a_r, \ldots, a_0$ also satisfies (5.12). Hence,

$$(\tilde{a}_0, \ldots, \tilde{a}_r) = (a_0, \ldots, a_r) + (a_r, \ldots, a_0)$$

and

$$(\hat{a}_0, \ldots, \hat{a}_r) = (a_0, \ldots, a_r) - (a_r, \ldots, a_0)$$

are also solutions of (5.12). The first solution satisfies $\tilde{a}_j = \tilde{a}_{r-j}$, the second one satisfies $\hat{a}_j = -\hat{a}_{r-j}$ and at least one of the two solutions is nonnull if the original solution $a_0, \ldots, a_r$ is nonnull. Hence, we may conclude that for each feasible γ there is a nonnull solution $a_0, \ldots, a_r$ of (5.12) with $a_j = a_{r-j}$ or $a_j = -a_{r-j}$. In the first case (5.12) simplifies to

$$(a_0, \ldots, a_k)G^+(\gamma) = 0 \tag{5.13}$$

and in the second case it simplifies to

$$(a_0, \ldots, a_k)G^-(\gamma) = 0 , \tag{5.14}$$

where the $(k+1) \times (k+1)$-matrices $G^+(\gamma)$ and $G^-(\gamma)$ are given by

$$G^+(\gamma) = \begin{pmatrix} \kappa\gamma^r + \lambda & \gamma^{r+1} & 0 & 0 & \cdots & 0 & 0 & 0 \\ 2\gamma^{r+1} & \kappa\gamma^r + \lambda & \gamma^{r+1} & 0 & & 0 & 0 & 0 \\ 0 & \gamma^{r+1} & \kappa\gamma^r + \lambda & \gamma^{r+1} & & 0 & 0 & 0 \\ \vdots & & & & & & & \vdots \\ 0 & 0 & 0 & 0 & & \gamma^{r+1} & \kappa\gamma^r + \lambda & \gamma^{r+1} \\ 0 & 0 & 0 & 0 & \cdots & 0 & \gamma^{r+1} & \kappa\gamma^r + \gamma^{r+1} + \lambda \end{pmatrix} ;$$

$$G^-(\gamma) = \begin{pmatrix} \kappa\gamma^r - \lambda & \gamma^{r+1} & 0 & 0 & \cdots & 0 & 0 & 0 \\ 2\gamma^{r+1} & \kappa\gamma^r - \lambda & \gamma^{r+1} & 0 & & 0 & 0 & 0 \\ 0 & \gamma^{r+1} & \kappa\gamma^r - \lambda & \gamma^{r+1} & & 0 & 0 & 0 \\ \vdots & & & & & & & \vdots \\ 0 & 0 & 0 & 0 & & \gamma^{r+1} & \kappa\gamma^r - \lambda & \gamma^{r+1} \\ 0 & 0 & 0 & 0 & \cdots & 0 & \gamma^{r+1} & \kappa\gamma^r - \gamma^{r+1} - \lambda \end{pmatrix} .$$

Since r is odd, it directly follows that if $a_0, \ldots, a_k$ is a solution of (5.13) with $\gamma = \tilde{\gamma}$, then $a_0, -a_1, a_2, \ldots, (-1)^k a_k$ satisfies (5.14) with $\gamma = -\tilde{\gamma}$ (the same holds with (5.13) and (5.14) interchanged). Hence, it suffices to find the γ's with $|\gamma| < 1$ for which (5.13) has a nonnull solution. The problem of finding these γ's can be translated to an eigenvalue problem as follows. By dividing $G^+(\gamma)$ by γ^{r+1} and introducing

$$-z = \frac{\kappa\gamma^r + \lambda}{\gamma^{r+1}} , \tag{5.15}$$

the matrix $G^+(\gamma)$ is transformed to the $(k+1) \times (k+1)$-matrix $H^+ - zI$, where I is the $(k+1) \times (k+1)$ identity matrix and H^+ is given by

$$H^+ = \begin{pmatrix} 0 & 1 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 2 & 0 & 1 & 0 & & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & & 0 & 0 & 0 & 0 \\ \vdots & & & & & & & & \vdots \\ 0 & 0 & 0 & 0 & & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 1 & 1 \end{pmatrix} \tag{5.16}$$

for all $k > 0$ and by $H^+ = (2)$ for $k = 0$. We now have to find the z's for which $H^+ - zI$ is singular, i.e. we have to find the eigenvalues of H^+. The feasible γ's may then be found from relation (5.15). The following lemma formulates properties of the eigenvalues of H^+.

Lemma 5.1 *All eigenvalues of the $(k{+}1){\times}(k{+}1)$-matrix H^+ defined by (5.16) are real-valued and simple. The largest eigenvalue is 2 and the other ones are in absolute value strictly less than 2.*

Proof The case $k = 0$ is trivial. Now suppose that $k > 0$. Dividing the first column of the determinant $|H^+ - zI|$ by $\sqrt{2}$ and multiplying the first row by $\sqrt{2}$ yields $|H^+ - zI| = |T - zI|$ where T is the $(k{+}1){\times}(k{+}1)$ *symmetric* matrix given by

$$
T = \begin{pmatrix}
0 & \sqrt{2} & 0 & 0 & \ldots & 0 & 0 & 0 & 0 \\
\sqrt{2} & 0 & 1 & 0 & & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 1 & & 0 & 0 & 0 & 0 \\
\vdots & & & & & & & \vdots & \\
0 & 0 & 0 & 0 & & 1 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & & 0 & 1 & 0 & 1 \\
0 & 0 & 0 & 0 & \ldots & 0 & 0 & 1 & 1
\end{pmatrix}.
$$

So the spectrum of H^+ is identical to the one of T. Since T is symmetric, all eigenvalues of T are real and T can be reduced to diagonal form. Furthermore, since all elements on the lower subdiagonal of T are nonzero and all elements below that subdiagonal are zero, we can conclude that the dimension of each eigenspace is exactly one. Hence, all eigenvalues of T, and thus also the ones of H^+, are real and simple, which proves the first part of the lemma.

To prove the second part of the lemma, notice that the spectral radius of H^+ is bounded by $\| H^+ \|_1 = 2$. Hence, all eigenvalues of H^+ are in absolute value bounded by 2. The sum of the rows of $H^+ - 2I$ vanishes, so 2 is an eigenvalue of H^+. It remains to show that -2 is not an eigenvalue of H^+. Let us consider the equations

$$
(a_0, \ldots, a_k)(H^+ + 2I) = 0,
$$

or

$$
\begin{aligned}
2a_0 + 2a_1 &= 0; \\
a_{n-1} + 2a_n + a_{n+1} &= 0, \qquad n = 1, \ldots, k-1; \\
a_{k-1} + 3a_k &= 0.
\end{aligned}
$$

The first k equations yield $a_n = (-1)^n a_0$ and then the final one that $a_0 = 0$. Hence, $H^+ + 2I$ is nonsingular, and thus -2 is not an eigenvalue, which completes the proof of the lemma. $\square$

For each eigenvalue σ of H^+ the roots inside the unit circle of equation (5.15) with $z = \sigma$ produce feasible γ's. The following lemma states that this equation has exactly r simple roots inside the unit circle. The proof of the lemma is similar to that of Lemma 4.1, and therefore it is omitted.

Lemma 5.2 *For given $r \geq 1$ and $|\sigma| \leq 2$ the equation*

$$
\sigma\gamma^{r+1} + \kappa\gamma^r + \lambda = 0 \tag{5.17}
$$

has exactly r simple roots inside the unit circle.

18

This concludes the determination of the set of γ's with $|\gamma| < 1$ for which (5.13) has a nonnull solution. As mentioned before, the set of γ's for which (5.14) has a nonnull solution may be found from the first set by multiplying each γ in that set by -1. The set $\mathcal{F}$ is obtained by squaring the γ's in these two sets. So, to produce $\mathcal{F}$, it suffices to square the γ's in the first set only. For each γ in the first set, i.e. γ is a root inside the unit circle of (5.17) for some eigenvalue σ, it holds that $-\gamma$ is not in this set, since for each eigenvalue $\tilde{\sigma}$ of H^+ we get (recall that r is odd)

$$|\tilde{\sigma}(-\gamma)^{r+1} + \kappa(-\gamma)^r + \lambda| = |-2\kappa\gamma^r + (\tilde{\sigma} - \sigma)\gamma^{r+1}| \geq (2|\kappa| - 4)|\gamma|^{r+1} \geq 2\lambda|\gamma|^{r+1} > 0\,,$$

where the equality follows by substituting $\lambda = -\kappa\gamma^r - \sigma\gamma^{r+1}$. Hence, we can conclude that $\mathcal{F}$ may be obtained by squaring the γ's in the first set only, yielding $r(k+1) = r(r+1)/2$ possible values for α. This is formulated in the following lemma.

Lemma 5.3 (Characterization of $\mathcal{F}$ in case $r = 2k + 1$) *The set $\mathcal{F}$ may be obtained by squaring the roots γ with $|\gamma| < 1$ of the equations (5.17) where σ runs through the set of eigenvalues of the $(k+1) \times (k+1)$-matrix H^+ defined by (5.16).*

5.1.2 Finding the feasible γ's in case $r = 2k$

Analogously to the previous subsection, we may conclude, by the symmetry of the set of equations (5.12), that, if there is a nonnull solution $a_0, \ldots, a_r$ to (5.12), then there is also a solution with $a_j = a_{r-j}$. So (5.12) simplifies to

$$(a_0, \ldots, a_k)G^+(\gamma) = 0\,, \tag{5.18}$$

or with $a_j = -a_{r-j}$, in which case $a_k = 0$ and (5.12) simplifies to

$$(a_0, \ldots, a_{k-1})G^-(\gamma) = 0\,, \tag{5.19}$$

where the $(k+1) \times (k+1)$-matrix $G^+(\gamma)$ is given by

$$G^+(\gamma) = \begin{pmatrix} \kappa\gamma^r + \lambda & \gamma^{r+1} & 0 & 0 & \ldots & 0 & 0 & 0 \\ 2\gamma^{r+1} & \kappa\gamma^r + \lambda & \gamma^{r+1} & 0 & & 0 & 0 & 0 \\ 0 & \gamma^{r+1} & \kappa\gamma^r + \lambda & \gamma^{r+1} & & 0 & 0 & 0 \\ \vdots & & & & & & & \vdots \\ 0 & 0 & 0 & 0 & & \gamma^{r+1} & \kappa\gamma^r + \lambda & 2\gamma^{r+1} \\ 0 & 0 & 0 & 0 & \ldots & 0 & \gamma^{r+1} & \kappa\gamma^r + \lambda \end{pmatrix}\,,$$

and the $k \times k$-matrix $G^-(\gamma)$ is given by

$$G^-(\gamma) = \begin{pmatrix} \kappa\gamma^r - \lambda & \gamma^{r+1} & 0 & 0 & \ldots & 0 & 0 & 0 \\ 2\gamma^{r+1} & \kappa\gamma^r - \lambda & \gamma^{r+1} & 0 & & 0 & 0 & 0 \\ 0 & \gamma^{r+1} & \kappa\gamma^r - \lambda & \gamma^{r+1} & & 0 & 0 & 0 \\ \vdots & & & & & & & \vdots \\ 0 & 0 & 0 & 0 & & \gamma^{r+1} & \kappa\gamma^r - \lambda & \gamma^{r+1} \\ 0 & 0 & 0 & 0 & \ldots & 0 & \gamma^{r+1} & \kappa\gamma^r - \lambda \end{pmatrix}\,.$$

The problem of finding the γ's with $|\gamma| < 1$ for which (5.18) or (5.19) has a nonnull solution may again be transformed to an eigenvalue problem. By dividing $G^+(\gamma)$ by γ^{r+1} and introducing

$$-v = \frac{\kappa\gamma^r + \lambda}{\gamma^{r+1}}\,, \tag{5.20}$$

the matrix $G^+(\gamma)$ is transformed to the $(k+1) \times (k+1)$-matrix $H^+ - vI$, where I is the $(k+1) \times (k+1)$ identity matrix and H^+ is given by

$$H^+ = \begin{pmatrix} 0 & 1 & 0 & 0 & \ldots & 0 & 0 & 0 & 0 \\ 2 & 0 & 1 & 0 & & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & & 0 & 0 & 0 & 0 \\ \vdots & & & & & & & & \vdots \\ 0 & 0 & 0 & 0 & & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & & 0 & 1 & 0 & 2 \\ 0 & 0 & 0 & 0 & \ldots & 0 & 0 & 1 & 0 \end{pmatrix} . \tag{5.21}$$

Similarly, by dividing $G^-(\gamma)$ by γ^{r+1} and introducing

$$-w = \frac{\kappa \gamma^r - \lambda}{\gamma^{r+1}} , \tag{5.22}$$

the matrix $G^-(\gamma)$ is transformed to the $k \times k$-matrix $H^- - wI$, where I is the $k \times k$ identity matrix and H^- is given by

$$H^- = \begin{pmatrix} 0 & 1 & 0 & 0 & \ldots & 0 & 0 & 0 & 0 \\ 2 & 0 & 1 & 0 & & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & & 0 & 0 & 0 & 0 \\ \vdots & & & & & & & & \vdots \\ 0 & 0 & 0 & 0 & & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & \ldots & 0 & 0 & 1 & 0 \end{pmatrix} . \tag{5.23}$$

We now have to find the v's and w's for which $H^+ - vI$ and $H^- - wI$ respectively are singular, i.e. we have to find the eigenvalues of H^+ and H^-. The feasible γ's may then be found from the relations (5.20) and (5.22). The following lemma formulates properties of the eigenvalues of H^+ and H^-.

Lemma 5.4 *All eigenvalues of the $(k+1) \times (k+1)$-matrix H^+ defined by (5.21) and the $k \times k$-matrix H^- defined by (5.23) are real-valued, simple and bounded by 2. Both 2 and -2 are eigenvalues of H^+. All eigenvalues of H^- are in absolute value strictly less than 2. For each eigenvalue σ of H^+ or H^- it holds that $-\sigma$ is also an eigenvalue.*

Proof Multiplying the even rows and the odd columns of the matrices $H^+ - vI$ and $H^- - wI$ by -1 yields $H^+ + vI$ and $H^- + wI$ respectively, hence we get for the determinants $|H^+ - vI| = (-1)^{k+1}|H^+ + vI|$ and $|H^- - wI| = (-1)^k|H^- + wI|$, implying that for each eigenvalue σ of H^+ or H^- also $-\sigma$ is an eigenvalue. The rest of the lemma can be proved similarly to Lemma 5.1. □

For each eigenvalue σ of H^+ we get by substituting $v = \sigma$ in (5.20) the equation

$$\sigma \gamma^{r+1} + \kappa \gamma^r + \lambda = 0 . \tag{5.24}$$

By Lemma 5.2 this equation has exactly r simple roots inside the unit circle, each root producing a feasible γ. The roots of (5.24) for the eigenvalue $-\sigma$ follow by multiplying the ones of (5.24) for σ by -1. Hence, since $\mathcal{F}$ consists of squared feasible γ's, it suffices

to consider the nonnegative eigenvalues of H^+ only. It is readily verified for γ satisfying (5.24) for some $\sigma > 0$ that $-\gamma$ cannot satisfy (5.24) for some $\tilde{\sigma} \geq 0$. Of course, if γ satisfies (5.24) for $\sigma = 0$, then also $-\gamma$ satisfies that equation. Notice that $\sigma = 0$ is only an eigenvalue of H^+ if k if even, otherwise it is an eigenvalue of H^-.

Similar remarks can be made for the roots of the equations

$$\tau\gamma^{r+1} + \kappa\gamma^r - \lambda = 0, \tag{5.25}$$

where τ runs through the set of eigenvalues of H^-. Furthermore, for each γ with $|\gamma| < 1$ satisfying (5.25) for some eigenvalue τ, it holds that $\pm\gamma$ cannot satisfy (5.24) for some eigenvalue σ, since (recall that r is even)

$$|\sigma(\pm\gamma)^{r+1} + \kappa(\pm\gamma)^r + \lambda| = |2\kappa\gamma^r + (\sigma(\pm 1)^{r+1} + \tau)\gamma^{r+1}| \geq (2|\kappa| - 4)|\gamma|^{r+1} \geq 2\lambda|\gamma|^{r+1} > 0,$$

where the equality follows by substituting $\lambda = \kappa\gamma^r + \tau\gamma^{r+1}$. Hence, we can conclude that $\mathcal{F}$ may be obtained by squaring the roots inside the unit circle of the equations (5.24) and (5.25), where σ and τ run through the set of nonnegative eigenvalues of H^+ and H^- respectively, yielding $(r + 1)k = r(r + 1)/2$ possible values for α. This is formulated in the following lemma.

Lemma 5.5 (Characterization of $\mathcal{F}$ in case $r = 2k$) *The set $\mathcal{F}$ may be obtained by squaring the roots γ with $|\gamma| < 1$ of the equations (5.24) and (5.25) where σ runs through the set of nonnegative eigenvalues of the $(k+1)\times(k+1)$-matrix H^+ defined by (5.21) and τ runs through the set of nonnegative eigenvalues of the $k \times k$-matrix H^- defined by (5.23).*

5.2   Characterizing feasible α_0's

We now return to the problem of finding the feasible α_0's for the original problem with no jockeying (see the beginning of Section 5). The following theorem formulates the remarkable result that the set $\mathcal{F}$ yields the feasible α_0's.

Theorem 5.2 *The set $\mathcal{F}$ is the set of feasible α_0's for the original problem without jockeying.*

Proof   We will first show that each α_0 in the set $\mathcal{F}$ is feasible. Then we will show that $\mathcal{F}$ yields *all* feasible values for α_0.

Part 1. *Each $\alpha_0 \in \mathcal{F}$ is feasible.*

To establish the first part we have to show for each α_0 in the set $\mathcal{F}$ that there are nonnull coefficients $d_0, e_{0,0}, \ldots, e_{0,r-1}$ such that the sequence $p_{m,n}$ given by

$$\begin{array}{ll} d_0\alpha_0^m\beta_0^n & \text{for } m \geq 0,\, n \geq r \text{ and} \\ e_{0,n}\alpha_0^m & \text{for } m \geq 0,\, 0 \leq n < r \end{array} \tag{5.26}$$

satisfies the horizontal boundary conditions (2.4)–(2.6), where β_0 is the smallest root of equation (3.1) with $\alpha = \alpha_0$. Since α_0 is in the set $\mathcal{F}$ it holds that

$$\alpha_0 = \gamma^2$$

for some feasible γ. Hence, we can rewrite the sequence $p_{m,n}$ in the form (cf. (5.9))

$$p_{m,n} = g_n\gamma^{2m+n}, \qquad m \geq 0,\, n \geq 0, \tag{5.27}$$

where

$$g_n = g_r \left(\frac{\beta_0}{\gamma}\right)^{n-r}, \qquad n \geq r \tag{5.28}$$

and

$$g_n = \begin{cases} d_0 \left(\frac{\beta_0}{\gamma}\right)^r, & n = r; \\ e_{0,n} \left(\frac{1}{\gamma}\right)^n, & 0 \leq n < r. \end{cases}$$

The original problem of finding nonnull coefficients $d_0, e_{0,0}, \ldots, e_{0,r-1}$ such that the sequence $p_{m,n}$ given by (5.26) satisfies the conditions (2.4)–(2.6) can now be translated to finding a nonnull sequence g_n satisfying (5.28) such that the sequence $p_{m,n}$ given by (5.27) satisfies the conditions (2.4)–(2.6).

Insertion of the form (5.27) into the inner conditions (2.1) and the horizontal boundary conditions (2.4)–(2.6) and then dividing by common powers of γ leads to the following set of equations for the sequence g_n.

$$\tilde{G}(\gamma) \begin{pmatrix} g_0 \\ g_1 \\ \vdots \end{pmatrix} = 0, \tag{5.29}$$

where the countable matrix $\tilde{G}(\gamma)$ is given by (cf. the matrix $G(\gamma)$)

$$\tilde{G}(\gamma) = \left(\begin{array}{ccccccccc|ccc} \kappa\gamma^r & \gamma^{r+1} & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \lambda & 0 & 0 & \cdots \\ 2\gamma^{r+1} & \kappa\gamma^r & \gamma^{r+1} & 0 & 0 & & 0 & 0 & \lambda & 0 & \lambda & 0 \\ 0 & \gamma^{r+1} & \kappa\gamma^r & \gamma^{r+1} & 0 & & 0 & \lambda & 0 & 0 & 0 & \lambda \\ 0 & 0 & \gamma^{r+1} & \kappa\gamma^r & \gamma^{r+1} & & \lambda & 0 & 0 & 0 & 0 & 0 \\ \vdots & & & & & & & & & & & \vdots \\ 0 & 0 & \lambda & 0 & 0 & & \gamma^{r+1} & \kappa\gamma^r & \gamma^{r+1} & 0 & 0 & 0 \\ 0 & \lambda & 0 & 0 & 0 & & 0 & \gamma^{r+1} & \kappa\gamma^r & \gamma^{r+1} & 0 & 0 \\ \lambda & 0 & 0 & 0 & 0 & & 0 & 0 & \gamma^{r+1} & \kappa\gamma^r & \gamma^{r+1} & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & & 0 & 0 & 0 & \gamma^{r+1} & \kappa\gamma^r & \gamma^{r+1} \\ 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \gamma^{r+1} & \kappa\gamma^r & \cdots \\ \vdots & & & & & & & & & & & \vdots \end{array} \right),$$

where the upper left block is of dimensions $r+1 \times r+1$. Since the sequence g_n for $n \geq r$ is given by (5.28), all equations in (5.29), except for the first $r + 1$ equations, are satisfied for each choice of g_r. Substitution of (5.28) into the first $r + 1$ equations of (5.29) yields a system of $r + 1$ homogeneous linear equations for $g_0, \ldots, g_r$. It now has to be shown that this system of equations has a nonnull solution.

First we shall show that the rows of $\tilde{G}(\gamma)$ are *dependent*, i.e. there is a nonnull *bounded* solution $a_0, a_1, \ldots$ of

$$(a_0, a_1, \ldots)\tilde{G}(\gamma) = 0. \tag{5.30}$$

Since γ is feasible, it follows that the equations (5.12) have a nonnull solution $a_0, \ldots, a_r$. It is possible to extend this sequence periodically (with period $2r$) to all nonnegative integers, such that it satisfies (5.30). Namely, by defining for all $k = 1, 2, \ldots$

$$a_{kr+j} = a_{kr-j}, \qquad j = 0, 1, \ldots, r,$$

it is easily verified that the resulting sequence $a_0, a_1, \ldots$ is a nonnull bounded solution of the system of equations (5.30).

We can now prove that there indeed exist nonnull numbers $g_0, \ldots, g_r$ such that the first $r + 1$ equations in (5.29) are satisfied. At least one of $a_0, \ldots, a_r$ is nonzero, say a_k. By substituting (5.28) into the first $r + 1$ equations of (5.29) and omitting the k-th equation we get a system of r homogeneous linear equations for $g_0, \ldots, g_r$. This system has a nonnull solution. Now it remains to prove that for this solution the k-th equation in (5.29) is also satisfied. Since $a_0, a_1, \ldots$ satisfies (5.30) we have for all $j = 0, 1, \ldots$ that

$$\sum_{i=0}^{\infty} a_i \tilde{g}_{ij} = 0 \,,$$

where the $\tilde{g}_{ij}$'s are the elements of the matrix $\tilde{G}(\gamma)$. Multiplying the j-th equation by g_j and then summing over all j yields

$$\sum_{j=0}^{\infty} \sum_{i=0}^{\infty} a_i \tilde{g}_{ij} g_j = 0 \,. \tag{5.31}$$

Since $|\beta_0| < |\alpha_0| = |\gamma|^2 < |\gamma|$, so $|\beta_0/\gamma| < 1$, it follows from (5.28) that

$$\sum_{j=0}^{\infty} |g_j| < \infty \,.$$

Hence, since the sequence $a_0, a_1, \ldots$ is bounded, we get

$$\sum_{j=0}^{\infty} \sum_{i=0}^{\infty} |a_i \tilde{g}_{ij} g_j| \leq \sup_i |a_i| \sum_{j=0}^{\infty} \sum_{i=0}^{\infty} |\tilde{g}_{ij}| |g_j| \leq 2|\kappa| \sup_i |a_i| \sum_{j=0}^{\infty} |g_j| < \infty \,.$$

So we may change summations in the double sum (5.31) yielding

$$\sum_{i=0}^{\infty} a_i \sum_{j=0}^{\infty} \tilde{g}_{ij} g_j = 0 \,. \tag{5.32}$$

The sequence g_j satisfies all equations in (5.29), except maybe for the k-th equation, so the equality (5.32) reduces to

$$a_k \sum_{j=0}^{\infty} \tilde{g}_{kj} g_j = 0 \,.$$

Hence, since a_k is nonzero, we can conclude that the k-th equation is also satisfied. This completes the first part of the proof.

Part 2. $\mathcal{F}$ yields all feasible values.

To finally prove that $\mathcal{F}$ produces *all* feasible values for α_0 we introduce an irreducible Markov process on a slightly different grid, namely

$$\{(m, n) | m \geq 0, n \geq 0\} \cup \{(m, n) | m < 0, m + n \geq r\} \,.$$

The transition rates from states $(m; n)$ with $n < r$ are identical to the ones for the original process. From the other states the transition rates are identical to the ones from the interior points of the original process, where it is assumed that transitions from
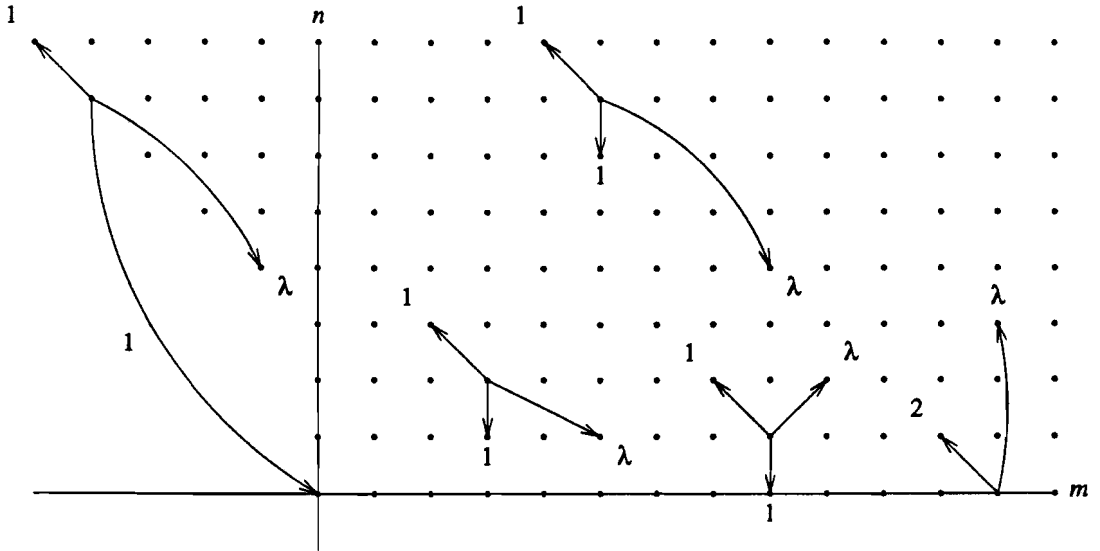
23

Figure 5: *The transition-rate diagram for the Markov process in the proof of Theorem 5.2 on the grid $\{(m,n)|m \geq 0, n \geq 0\} \cup \{(m,n)|m < 0, m+n \geq r\}$ for the case $r = 3$.*

$(m, r - m)$ with $m < 0$ to $(m, r - m - 1)$, which is not a grid point, are redirected to the origin. The transition rate diagram for the case $r = 3$ is depicted in Figure 5 (compare Figure 1).

The equilibrium equation in state (m,n) with $m + n \geq r$ is given by (2.1) for $n > r$, (2.4) for $1 < n \leq r$, (2.5) for $n = 1$ and (2.6) for $n = 0$. Hence, for each feasible α_0 the sequence $y_{m,n}(\alpha_0)$ given by

$$d_0 \beta_0^n \alpha_0^m \quad \text{for } m + n \geq r, n \geq r \text{ and}$$
$$e_{0,n} \alpha_0^m \quad \text{for } m \geq 0, 0 \leq n < r,$$

where β_0 is the smallest root of (3.1) with $\alpha = \alpha_0$ and the coefficients $d_0, e_{0,0}, \ldots, e_{0,r-1}$ are a nonnull solution of (5.1), satisfies the equilibrium equations in all states with $m + n \geq r$. The sequences $y_{m,n}(\alpha_0)$ with α_0 in the set $\mathcal{F}$ may be linearly combined such that the remaining equations in the states with $m + n < r$ are also satisfied. Substitution of the linear combination

$$p_{m,n} = \sum_{\alpha_0 \in \mathcal{F}} k(\alpha_0) y_{m,n}(\alpha_0)$$

into the equilibrium equations in the states (m,n) with $0 < m + n < r$ yields a system of $r(r + 1)/2 - 1$ homogeneous linear equations for the $r(r + 1)/2$ unknown coefficients $k(\alpha_0)$. So this system has a nonnull solution. The equation in $(0,0)$ is also satisfied, since insertion of $p_{m,n}$ into the equations in states $(m,n) \neq (0,0)$ and then summing over these equations and changing summations exactly yields the desired equation. Since $|\beta_0| < |\alpha_0| < 1$ the sum of $y_{m,n}(\alpha_0)$ over all states absolutely converges, which implies that changing summations is indeed allowed. Hence, $p_{m,n}$ is an absolutely convergent solution of all equilibrium equations. It is a nonnull solution, because the sequences $y_{m,n}(\alpha_0)$ for different feasible α_0's are linearly independent (see Lemma 6.1 in the next section).

24

Hence, from a result of Foster ([11], Theorem 1) it follows that the Markov process in Figure 5 is ergodic and normalization of $p_{m,n}$ produces the equilibrium distribution. Since the equilibrium distribution of an ergodic Markov process is unique and for different α_0's the sequences $y_{m,n}(\alpha_0)$ are linearly independent, there cannot be more than $r(r+1)/2$ feasible α_0's. Hence, we may conclude that $\mathcal{F}$ is the set of all feasible values of α_0. This completes the second part of the proof. □

6   The finishing touch

In this section we first investigate under what conditions the construction of formal solutions with feasible initial pairs does not fail, because of degeneration of the coefficients of the compensating terms involved. Then we will prove that the formal solutions constitute absolutely convergent series, except perhaps in a neighborhood around the origin. Finally, the results obtained are summarized in the main result of this paper.

Each formal solution $x_{m,n}(\alpha_0, \beta_0)$ has its *own sequence* $\{\alpha_i, \beta_i\}$ depending on the initial value α_0, and its own associated sequence of coefficients $\{c_i, d_i, e_{i,0}, \ldots, e_{i,r-1}\}$ (see Section 3). In the previous section we have found the feasible values for α_0. For each feasible α_0 it holds that $d_{-1} = 0$ and thus $d_{i-1} = e_{i,0} = \ldots = e_{i,r-1} = 0$ for all $i < 0$. Hence the upper part of the tree of compensating terms vanishes (see Figure 2). Since the corresponding β_0 is uniquely determined as the smallest root of equation (3.1) with $\alpha = \alpha_0$, we may abbreviate the notation $x_{m,n}(\alpha_0, \beta_0)$ to $x_{m,n}(\alpha_0)$. So for each feasible α_0 the series $x_{m,n}(\alpha_0)$ simplifies for $m \geq 0$ and $n \geq r$ to (see (3.5) and (3.6))

$$x_{m,n}(\alpha_0) = \sum_{i=0}^{\infty} d_i(c_i\alpha_i^m + \sum_{j=1}^{r} c_{ri+j}\alpha_{ri+j}^m)\beta_i^n \tag{6.1}$$

$$= c_0 d_0 \beta_0^n \alpha_0^m + \sum_{i=0}^{\infty}\sum_{j=1}^{r} c_{ri+j}(d_i\beta_i^n + d_{ri+j}\beta_{ri+j}^n)\alpha_{ri+j}^m \tag{6.2}$$

and for $m \geq 0$ and $n = 0, \ldots, r-1$ to (see (3.7))

$$x_{m,n}(\alpha_0) = \sum_{i=0}^{\infty} c_i e_{i,n} \alpha_i^m . \tag{6.3}$$

We will now investigate whether the construction of $x_{m,n}(\alpha_0)$ with feasible α_0 can possibly fail. From Lemma 4.1 we can conclude that the generation of the coefficients c_i with $i > 0$ never breaks down because of a vanishing denominator in (3.8). So the compensation on the horizontal boundary always works. It follows from (3.10) that the generation of the coefficients $d_{ri+j}, e_{ri+j,0}, \ldots, e_{ri+j,r-1}$ with $i \geq 0$ and $j = 1, \ldots, r$ possibly fails when the *homogeneous* equations

$$A(\alpha_{ri+j}) \begin{pmatrix} e_{ri+j,0} \\ \vdots \\ e_{ri+j,r-1} \end{pmatrix} + C(\alpha_{ri+j}, \beta_{ri+j})d_{ri+j}\beta_{ri+j}^r = 0$$

are linearly dependent, i.e. they have a nonnull solution. This means that the compensating terms $d_{ri+j}\beta_{ri+j}^n\alpha_{ri+j}^m$ for $m \geq 0$, $n \geq r$ and $e_{ri+j,n}\alpha_{ri+j}^m$ for $m \geq 0$, $n = 0, \ldots, r-1$ can be fitted to the horizontal boundary conditions, so that α_{ri+j} is feasible. This really may occur, as simple examples show. Of course, this complication completely vanishes if

the term $d_i\beta_i^n\alpha_{ri+j}^m$ itself can be fitted to the horizontal boundary conditions, i.e. if there exist coefficients $e_{ri+j,0}, \ldots, e_{ri+j,r-1}$ such that

$$A(\alpha_{ri+j})\begin{pmatrix} e_{ri+j,0} \\ \vdots \\ e_{ri+j,r-1} \end{pmatrix} + C(\alpha_{ri+j},\beta_i)d_i\beta_i^r = 0$$

The question arises whether this is possible. Below we will show that the answer is confirmative and in fact, it appears to be *always* possible when α_{ri+j} is feasible.

So, let us suppose that α_{ri+j} is feasible, say

$$\alpha_{ri+j} = \gamma^2$$

for some feasible γ. We now wonder whether there are coefficients $e_{ri+j,0}, \ldots, e_{ri+j,r-1}$ such that the sequence $p_{m,n}$ given by

$$\begin{array}{ll} d_i\alpha_{ri+j}^m\beta_i^n & \text{for } m \geq 0, n \geq r \text{ and} \\ e_{ri+j,n}\alpha_0^m & \text{for } m \geq 0, 0 \leq n < r \end{array} \qquad (6.4)$$

satisfies the horizontal boundary conditions (2.4)–(2.6). If $|\beta_i| < |\gamma|$, then this can be established by following exactly the same lines as in the first part of the proof of Theorem 5.2, where, since the coefficient d_i is nonzero, it also has to be shown for the solution $g_0, g_1, \ldots$ of (5.29) that g_r is nonzero.

To prove that g_r is nonzero, we show that the assumption $g_r = 0$ leads to a contradiction. By (5.28), this assumption implies that $g_n = 0$ for all $n \geq r$. It then follows that the sequence $g_0, \ldots, g_{r-1}, g_r$ (with $g_r = 0$) satisfies the equations (5.10). Consequently,

$$(\tilde{g}_0, \ldots, \tilde{g}_r) = (g_0, g_1, \ldots, g_{r-1}, 0) + (0, g_{r-1}, \ldots, g_1, g_0)$$

and

$$(\hat{g}_0, \ldots, \hat{g}_r) = (g_0, g_1, \ldots, g_{r-1}, 0) - (0, g_{r-1}, \ldots, g_1, g_0)$$

also satisfy (5.10). Now $\tilde{g}_j \equiv 0$ or $\hat{g}_j \equiv 0$, since otherwise, both $G^+(\gamma)$ and $G^-(\gamma)$ would be singular, which is not possible, as we have seen in the subsections 5.1.1 and 5.1.2. Hence, we obtain that $g_0 = 0$. This implies, together with $g_r = 0$, that $g_j \equiv 0$, which obviously is a contradiction. So, indeed, we may conclude that g_r is nonzero. Remark that by using the same arguments it follows that for each feasible starting solution the coefficient d_0 is nonzero. So we may set $d_0 = 1$ and then solve the equations (5.1) for the coefficients $e_{0,0}, \ldots, e_{0,r-1}$.

Now it remains to show that $|\beta_i| < |\gamma|$. To do so, it will be helpful to first derive bounds for β_i and γ.

Definition 6.1 *For* $0 \leq \sigma \leq 2$ *define* $U(\sigma)$ *as the root on* $(0,1)$ *of (cf. (5.17))*

$$\sigma x^{r+1} + \kappa x^r + \lambda = 0$$

and $L(\sigma)$ *as the root on* $(0,1)$ *of*

$$-\sigma x^{r+1} + \kappa x^r + \lambda = 0.$$

26

Lemma 6.1

(i) *For each γ with $|\gamma| < 1$ satisfying (cf. (5.17) and (5.24))*

$$\sigma\gamma^{r+1} + \kappa\gamma^r + \lambda = 0 \tag{6.5}$$

or (cf. (5.25))

$$\sigma\gamma^{r+1} + \kappa\gamma^r - \lambda = 0, \tag{6.6}$$

where $|\sigma| \leq 2$, it holds that

$$L(|\sigma|) \leq |\gamma| \leq U(|\sigma|);$$

(ii) *If $0 \leq \sigma_1 \leq \sigma_2 \leq 2$, then $L(\sigma_2) \leq L(\sigma_1) \leq U(\sigma_1) \leq U(\sigma_2)$.*

Proof First observe that for $|\sigma| \leq 2$ it holds that

$$|\sigma|x^{r+1} + \kappa x^r + \lambda \quad > \quad 0, \qquad 0 \leq x < U(|\sigma|);$$
$$< \quad 0, \qquad U(|\sigma|) < x \leq 1.$$

Now, for each γ with $|\gamma| < 1$ satisfying (6.5) or (6.6) we obtain

$$-\kappa|\gamma|^r = |\kappa\gamma^r| = |\pm\lambda + \sigma\gamma^{r+1}| \leq \lambda + |\sigma||\gamma|^{r+1}$$

so

$$|\sigma||\gamma|^{r+1} + \kappa|\gamma|^r + \lambda \geq 0,$$

from which it directly follows that $|\gamma| \leq U(|\sigma|)$. The other inequality can be proved similarly and the proof of part (ii) is straightforward. $\qquad\square$

Corollary *For all feasible γ's it holds that $L(2) \leq |\gamma| \leq U(2)$.*

Let us from here on abbreviate $L(2)$ and $U(2)$ by L and U respectively. In Figure 6 the feasible γ's together with the bounds L and U are depicted for $r = 10$ and $\lambda = 0.1$. In this case the feasible γ's are the roots inside the unit circle of the equations (5.24) and (5.25) where σ and τ run through the set of 3 nonnegative eigenvalues of H^+ and H^- respectively (cf. Lemma 5.5). The roots are rather insensitive to changes in σ and τ respectively. This explains why the feasible γ's in Figure 6 are clustered in groups of 3.

The Corollary gives a lower bound for $|\gamma|$, namely

$$|\gamma| \geq L.$$

To get an upper bound for $|\beta_i|$, remark that it follows from Lemma 4.1 that $|\beta_i| < R|\alpha_i|$. Combining this inequality with $|\alpha_i| \leq |\alpha_0| \leq U^2$ yields

$$|\beta_i| < RU^2.$$

Hence, we are done once the following lemma has been established.
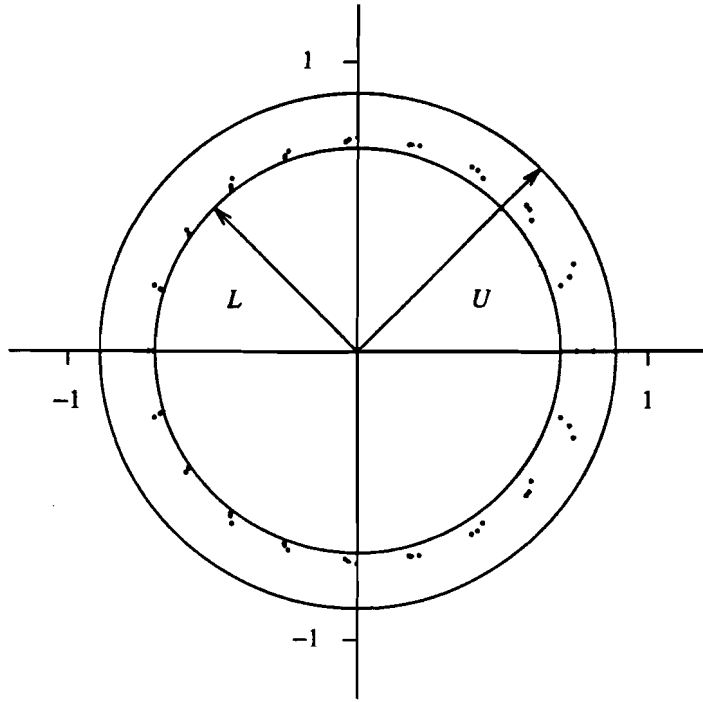
Figure 6: *The feasible γ's together with the bounds L and U for the case that $r = 10$ and $\lambda = 0.1$.*

Lemma 6.2 $L > RU^2$.

Proof The numbers U, L and R are subsequently defined as the roots on $(0,1)$ of

$$rx^{r+1} - (r + \rho)x^r + \rho = 0;$$ (6.7)

$$-rx^{r+1} - (r + \rho)x^r + \rho = 0;$$ (6.8)

$$\rho x^{r+1} - (r + \rho)x + r(x^2 + 1)/2 = 0,$$ (6.9)

where κ and λ have been eliminated by substituting $\kappa = -(2 + \lambda)$ and $\lambda = 2\rho/r$ respectively. One of the problems to establish the inequality in the lemma is that L, R and U are implicitly defined. Therefore we first derive an explicit upper bound for U. Since $x = 1$ satisfies (6.7), this equation can be rewritten as

$$(x - 1)(rx^r - \rho(x^{r-1} + \ldots + x + 1)) = 0.$$

Hence, U satisfies

$$rU^r - \rho(U^{r-1} + \ldots + U + 1) = 0,$$

from which it follows that

$$U = \sqrt[r]{\rho(U^{r-1} + \ldots + U + 1)/r} \leq \sqrt[r]{\rho}.$$

We will now prove that

$$L > R\sqrt[r]{\rho^2},$$ (6.10)

28

which, of course, implies the desired inequality. Now (6.10) holds iff for $x = R \sqrt[r]{\rho^2}$ the left-hand side of (6.8) is positive, i.e.

$$-rR^r\rho^2 R \sqrt[r]{\rho^2} - (r + \rho)R^r\rho^2 + \rho > 0,$$

which definitely holds if

$$-rR^r\rho^2 - (r + 1)R^r\rho^2 + \rho > 0,$$

or equivalently,

$$R < \frac{1}{\sqrt[r]{(2r + 1)\rho}}. \tag{6.11}$$

This inequality is trivial for $\rho \leq 1/(2r + 1)$. For $r = 1$ we have $R = 1/(2\rho + 1)$, so in this case it is clear that (6.11) is valid for all $0 < \rho < 1$. Hence, it remains to prove that (6.11) is valid for $1/(2r + 1) < \rho < 1$ and all $r > 1$. This proceeds straightforwardly as will be shown below.

Inequality (6.11) holds iff for $x = 1/\sqrt[r]{(2r + 1)\rho}$ the left-hand side of (6.9) is negative, which, by dividing by x, amounts to

$$\frac{1}{2r + 1} - (r + \rho) + \frac{r}{2}\left(\frac{1}{\sqrt[r]{(2r + 1)\rho}} + \sqrt[r]{(2r + 1)\rho}\right) < 0. \tag{6.12}$$

By substituting $y = \sqrt[r]{(2r + 1)\rho}$ in (6.12) and defining

$$f(y) = \frac{1}{2r + 1} - r - \frac{y^r}{2r + 1} + \frac{r}{2}(y^{-1} + y),$$

we thus have to show for $r > 1$ that $f(y) < 0$ for $y > 1$. Since $f(1) = 0$, we are done once it has been established that $f'(y) < 0$ for $y > 1$. Now $f'(y)$ is strictly concave for $y > 1$ having its maximum at

$$\tilde{y} = \left(\frac{2r + 1}{r - 1}\right)^{\frac{1}{r+1}}$$

where

$$f'(\tilde{y}) = \frac{r}{2}\left(1 - \frac{r + 1}{r - 1}\frac{1}{\tilde{y}^2}\right). \tag{6.13}$$

Hence, if $f'(\tilde{y}) < 0$, then we may conclude that indeed $f'(y)$ is negative for all $y > 1$. Equality (6.13) implies that $f'(\tilde{y}) < 0$ if

$$\tilde{y}^2 = \left(\frac{2r + 1}{r - 1}\right)^{\frac{2}{r+1}} < \frac{r + 1}{r - 1},$$

or, equivalently, by raising both sides to the power $(r + 1)/2$ and multiplying both sides with $(r - 1)/(r + 1)$,

$$\frac{2r + 1}{r + 1} < \left(\frac{r + 1}{r - 1}\right)^{\frac{r-1}{2}} = \left(1 + \frac{2}{r - 1}\right)^{\frac{r-1}{2}}. \tag{6.14}$$

It is readily verified that inequality (6.14) is valid for $r = 2$. Since the left-hand side of (6.14) is less than 2 and the right-hand side is increasing and equal to 2 for $r = 3$, it follows that (6.14) is valid for all $r > 1$. This completes the proof of the lemma. $\square$

Summarizing, we have found that the construction of $x_{m,n}(\alpha_0)$ with feasible α_0 never degenerates. We now try to prove that the series (6.1) converges absolutely. We need absolute convergence to guarantee that rearranging of terms is feasible (see (6.2)). The series (6.1) however, might possibly *diverge* in states near the origin of the state space, but we will prove:

Theorem 6.1 *There exists an integer $N \geq r$ such that for all feasible α_0's the following holds:*

(i) *The series $\sum_{i=0}^{\infty} d_i c_i \alpha_i^m \beta_i^n$ and $\sum_{i=0}^{\infty} d_i c_{ri+j} \alpha_{ri+j}^m \beta_i^n$ with $j = 1, \ldots, r$, which add up to $x_{m,n}(\alpha_0, \beta_0)$ by definition for $m \geq 0$, $n \geq r$, converge absolutely for all $m \geq 0$, $n \geq r$ with $m + n \geq N$;*

(ii) *The series $\sum_{i=0}^{\infty} c_i e_{i,n} \alpha_i^m$, which defines $x_{m,n}(\alpha_0, \beta_0)$ for $m \geq 0$, $0 \leq n < r$, converge absolutely for all $m \geq N - r$, $0 \leq n < r$;*

(iii) $\sum_{(m,n) \in \mathcal{A}} |x_{m,n}(\alpha_0, \beta_0)| < \infty$, *where $\mathcal{A}$ is the set of states on which the series in (i) and (ii) converge absolutely, i.e.*

$$\mathcal{A} = \{(m,n) | m \geq 0, n \geq r, m + n \geq N\} \cup \{(m,n) | m \geq N - r, 0 \leq n < r\}.$$

In the following lemma we will first formulate the limiting behaviour of the sequence $\{\alpha_i, \beta_i\}_{i=0}^{\infty}$ and the associated sequence of coefficients $\{c_i, d_i, e_{i,0}, \ldots, e_{i,r-1}\}_{i=0}^{\infty}$. With these results the proof of the theorem itself appears to be simple. If $d_i = 0$ for some $i \geq 0$, then the complete subtree of compensating terms starting with $d_i c_i \alpha_i^m \beta_i^n$ vanishes. Hence, to prove Theorem 6.1 it suffices to consider the case that d_i never vanishes.

Lemma 6.3 *Let α_0 satisfy $0 < |\alpha_0| < 1$ and let β_i be the smallest root in absolute value of equation (3.1) with $\alpha = \alpha_i$ for all $i \geq 0$. Suppose that the equations (3.10) for the coefficients $d_{ri+j}, e_{ri+j,0}, \ldots, e_{ri+j,r-1}$ have a solution for all $i \geq 0$ and $j = 1, \ldots, r$, so that the generation of these coefficients never breaks down, and suppose that d_i never vanishes. Then we have:*

(i) *If $i \to \infty$, then*

$$\frac{\beta_i}{\alpha_i} \to \frac{1}{A_0}, \tag{6.15}$$

$$\frac{\alpha_{ri+j}}{\beta_i} \to A_j, \qquad j = 1, \ldots, r \tag{6.16}$$

where $A_1, \ldots, A_r$ are the r simple roots inside the unit circle and A_0 is the root outside the unit circle of the equation $z^r(2 + \lambda) = \lambda + z^{r+1}$;

(ii) *If $i \to \infty$, then*

$$\frac{c_{ri+j}}{c_i} \to C_j \overset{\text{def}}{=} -\frac{(A_j - 1) \prod_{k \neq j} \left(\frac{1}{A_0} - \frac{1}{A_k}\right)}{(A_0 - 1) \prod_{k \neq j} \left(\frac{1}{A_j} - \frac{1}{A_k}\right)}, \qquad j = 1, \ldots, r$$

where the index k in the two products runs through $1, \ldots, r$;

30

(iii) *If $i \to \infty$, then*

$$\frac{d_{ri+j}}{d_i} \to -\left(\frac{A_0}{A_j}\right)^r,$$

$$\frac{e_{ri+j,r-k}}{d_i\beta_i^r} \to -\left(\frac{1}{A_j^k} - \frac{1}{A_0^k}\right), \qquad k = 1,\ldots,r-1,$$

$$\frac{e_{ri+j,0}}{d_i\beta_i^r} \to -\frac{A_0 - A_j}{\lambda}.$$

Proof By dividing equation (3.1) by β^{r+1} and introducing $z = \alpha/\beta$ we get

$$z^r(2+\lambda) = \lambda + \beta z + z^{r+1}. \tag{6.17}$$

For $\beta = \beta_i$ this equation has roots $z_0 = \alpha_i/\beta_i$ and $z_j = \alpha_{ri+j}/\beta_i$, $j = 1,\ldots,r$. From Lemma 4.1 it follows that as $i \to \infty$, then α_i and $\beta_i \to 0$, and since the roots of (6.17) continuously depend on the parameter β, this implies that the roots z_i converge to the roots of equation (6.17) with $\beta = 0$. Similar to the proof of Lemma 4.1 it can be shown that equation (6.17) with $\beta = 0$ has r simple roots inside the unit circle and one outside the unit circle. This completes the proof of part (i). The proof of part (ii) directly follows from expression (3.8) and the limits (6.15)–(6.16). From the limiting behaviour of α_i and β_i we obtain for the matrix $A(\alpha_{ri+j})$ and the vectors $C(\alpha_{ri+j},\beta_{ri+j})$ and $C(\alpha_{ri+j},\beta_i)$ in the equations (3.10) that as $i \to \infty$,

$$A(\alpha_{ri+j}) \to \begin{pmatrix} 0 & 0 & \ldots & 0 & 0 \\ 0 & 0 & & 0 & \lambda \\ 0 & 0 & & \lambda & 0 \\ \vdots & & & & \vdots \\ 0 & \lambda & & 0 & 0 \\ \lambda & 0 & \ldots & 0 & 0 \end{pmatrix},$$

$$C(\alpha_{ri+j},\beta_{ri+j}) \to \begin{pmatrix} \lambda \\ \lambda\frac{1}{A_0} \\ \lambda\frac{1}{A_0^2} \\ \vdots \\ \lambda\frac{1}{A_0^{r-1}} \\ -A_0 \end{pmatrix}, \quad C(\alpha_{ri+j},\beta_i) \to \begin{pmatrix} \lambda \\ \lambda\frac{1}{A_j} \\ \lambda\frac{1}{A_j^2} \\ \vdots \\ \lambda\frac{1}{A_j^{r-1}} \\ -A_j \end{pmatrix},$$

from which part (iii) readily follows. $\square$

Proof of Theorem 6.1 If $d_i = 0$ for some $i > 0$, then the complete subtree of compensating terms starting with $d_i c_i \alpha_i^m \beta_i^n$ vanishes. Hence, it suffices to consider the case that d_i never vanishes. To prove Theorem 6.1 in this case, consider a *fixed* $m \geq 0$ and $n \geq r$. Then by Lemma 6.1, as $i \to \infty$ and $j = 1, \ldots, r$,

$$\frac{|d_i c_{ri+j} \alpha_{ri+j}^m \beta_i^n|}{|d_i c_i \alpha_i^m \beta_i^n|} = \frac{|c_{ri+j} \alpha_{ri+j}^m|}{|c_i \alpha_i^m|} \to |C_j| \left(\frac{|A_j|}{|A_0|}\right)^m \tag{6.18}$$

and

$$\frac{|d_{ri+j} c_{ri+j} \alpha_{ri+j}^m \beta_{ri+j}^n|}{|d_i c_{ri+j} \alpha_{ri+j}^m \beta_i^n|} = \frac{|d_{ri+j} \beta_{ri+j}^n|}{|d_i \beta_i^n|} \to \left(\frac{|A_j|}{|A_0|}\right)^{n-r} . \tag{6.19}$$

To formulate the condition for absolute convergence of $x_{m,n}(\alpha_0)$ we introduce the notion of a *positive geometric r-fold tree*.

Definition 6.2 *The sequence* $n_0, n_1, \ldots$ *is called a positive geometric r-fold tree if:*

(i) *The sequence* $n_0, n_1, \ldots$ *has an r-fold tree structure as depicted in Figure 7;*

(ii) $n_0 > 0$;

(iii) *There are positive numbers* $R_1, \ldots, R_r$ *such that* $n_{ri+j} = R_j n_i$ *for all* $i \geq 0$ *and* $j = 1, \ldots, r$.
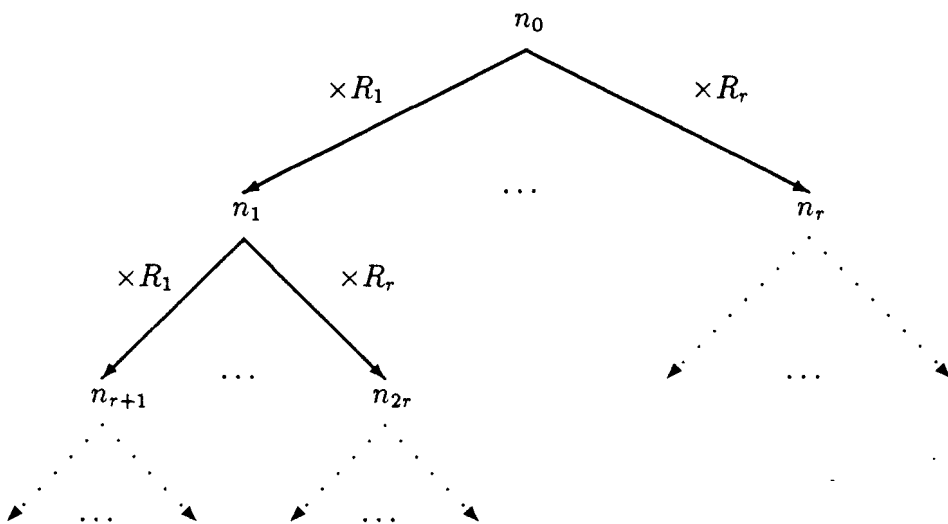
Figure 7: *The r-fold tree structure of the sequence* n_i.

For an r-fold tree $n_0, n_1, \ldots$ it is easy to prove the following necessary and sufficient condition for convergence of $\sum_{i=0}^{\infty} n_i$.

Lemma 6.4 *For a positive geometric r-fold tree* $n_0, n_1, \ldots$ *with rates* $R_1, \ldots, R_r$ *it holds that* $\sum_{i=0}^{\infty} n_i < \infty$ *iff* $R_1 + \ldots + R_r < 1$.

The compensating terms in the lower part of the tree in Figure 2 add up to $x_{m,n}(\alpha_0)$ (the upper part vanishes for feasible α_0's). This tree may be split up into two subtrees, one consisting of the terms $d_i c_i \alpha_i^m \beta_i^n$ with $i \geq 0$ and the other of the terms $d_i c_{ri+j} \alpha_{ri+j}^m \beta_i^n$ with $i \geq 0$ and $j = 1, \ldots, r$. From the limits (6.18)–(6.19) it readily follows that the tree of terms $|d_i c_i \alpha_i^m \beta_i^n|$ asymptotically behaves as a positive geometrical r-fold tree with rates

$$R_j(m + n) = |C_j| \left(\frac{|A_j|}{|A_0|}\right)^{m+n-r} , \qquad j = 1, \ldots, r .$$

32

The same holds for the tree of terms $|d_i c_{ri+j} \alpha_{ri+j}^m \beta_i^n|$. With Lemma 6.2 it is now straight-forward to show that the sum of the terms in the two trees converges if $R_1(m+n) + \ldots + R_r(m+n) < 1$.

Similarly, by Lemma 6.1, we obtain for fixed values of $m \geq 0$ and $n = 0, \ldots, r-1$ that as $i \to \infty$ and $j = 1, \ldots, r$,

$$\frac{|c_{ri+j} e_{ri+j,n} \alpha_{ri+j}^m|}{|c_i e_{i,n} \alpha_i^m|} \to |C_j| \left(\frac{|A_j|}{|A_0|}\right)^m .$$

Hence, the tree of compensating terms $|c_i e_i \alpha_i^m|$ with $i \geq 0$ asymptotically behaves as a positive geometrical r-fold tree with rates $R_j(m)$, $j = 1, \ldots, r$. So the sum of the terms converges if $R_1(m) + \ldots + R_r(m) < 1$.

We can now define the integer N mentioned in Theorem 6.1 as follows.

Definition 6.3 N *is the smallest integer* $\geq r$ *for which* $R_1(N) + \ldots + R_r(N) < 1$.

We finally prove Theorem 6.1(iii). Insertion of the series (6.1) and (6.3) yields

$$\sum_{(m,n)\in\mathcal{A}} |x_{m,n}(\alpha_0, \beta_0)| = \sum_{m=0}^{N-r-1} \sum_{n=N-m}^{\infty} |x_{m,n}(\alpha_0, \beta_0)| + \sum_{m=N-r}^{\infty} \sum_{n=r}^{\infty} |x_{m,n}(\alpha_0, \beta_0)|$$

$$+ \sum_{n=0}^{r-1} \sum_{m=N-r}^{\infty} |x_{m,n}(\alpha_0, \beta_0)|$$

$$\leq \sum_{m=0}^{N-r-1} \sum_{n=N-m}^{\infty} \sum_{i=0}^{\infty} \left(|d_i c_i \alpha_i^m \beta_i^n| + \sum_{j=1}^{r} |d_i c_{ri+j} \alpha_{ri+j}^m \beta_i^n|\right)$$

$$+ \sum_{m=N-r}^{\infty} \sum_{n=r}^{\infty} \sum_{i=0}^{\infty} \left(|d_i c_i \alpha_i^m \beta_i^n| + \sum_{j=1}^{r} |d_i c_{ri+j} \alpha_{ri+j}^m \beta_i^n|\right)$$

$$+ \sum_{n=0}^{r-1} \sum_{m=N-r}^{\infty} \sum_{i=0}^{\infty} |c_i e_{i,n} \alpha_i^m|$$

$$= \sum_{m=0}^{N-r-1} \sum_{i=0}^{\infty} \left(\frac{|d_i c_i \alpha_i^m \beta_i^{N-m}|}{1 - |\beta_i|} + \sum_{j=1}^{r} \frac{|d_i c_{ri+j} \alpha_{ri+j}^m \beta_i^{N-m}|}{1 - |\beta_i|}\right)$$

$$+ \sum_{i=0}^{\infty} \left(\frac{|d_i c_i \alpha_i^{N-r} \beta_i^r|}{(1 - |\alpha_i|)(1 - |\beta_i|)} + \sum_{j=1}^{r} \frac{|d_i c_{ri+j} \alpha_{ri+j}^{N-r} \beta_i^r|}{(1 - |\alpha_i|)(1 - |\beta_i|)}\right)$$

$$+ \sum_{n=0}^{r-1} \sum_{i=0}^{\infty} \frac{|c_i e_{i,n} \alpha_i^{N-r}|}{1 - |\alpha_i|}$$

$$\leq \frac{1}{1 - |\beta_0|} \sum_{m=0}^{N-r-1} \sum_{i=0}^{\infty} \left(|d_i c_i \alpha_i^m \beta_i^{N-m}| + \sum_{j=1}^{r} |d_i c_{ri+j} \alpha_{ri+j}^m \beta_i^{N-m}|\right)$$

$$+ \frac{1}{(1 - |\beta_0|)(1 - |\alpha_0|)} \sum_{i=0}^{\infty} \left(|d_i c_i \alpha_i^{N-r} \beta_i^r| + \sum_{j=1}^{r} |d_i c_{ri+j} \alpha_{ri+j}^{N-r} \beta_i^r|\right)$$

$$+ \frac{1}{1 - |\alpha_0|} \sum_{n=0}^{r-1} \sum_{i=0}^{\infty} |c_i e_{i,n} \alpha_i^{N-r}|$$

$$< \infty ,$$

since each of the series converges by virtue of the parts (i) and (ii) of Theorem 6.1. This completes the proof of Theorem 6.1. □

We can now prove our main result stating that for states (m, n) in the set $\mathcal{A}$ the equilibrium probabilities $p_{m,n}$ can be expressed as a linear combination of the series $x_{m,n}(\alpha_0)$ with α_0 running through the set $\mathcal{F}$ of feasible values. In Figure 8 the set $\mathcal{A}$ is depicted for the special case that $r = 3$ and $N = 6$.
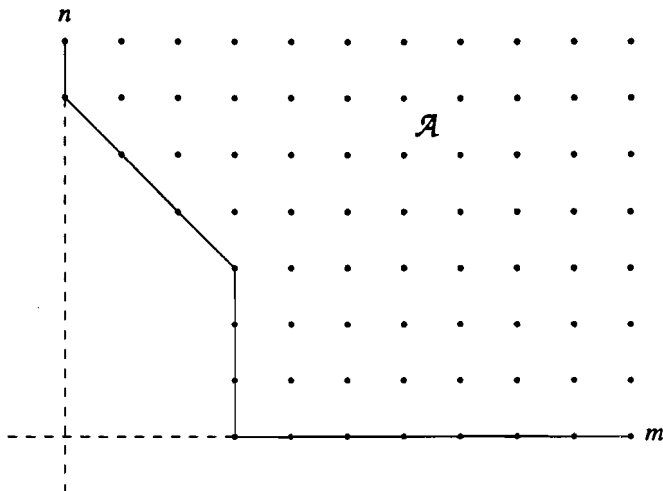

Figure 8: *The set $\mathcal{A}$ on which the series $x_{m,n}(\alpha_0, \beta_0)$ converge absolutely, for the special case $r = 3$ and $N = 6$.*

Theorem 6.2 (Main result) *For all states (m, n) in the set $\mathcal{A}$ it holds that*

$$p_{m,n} = \sum_{\alpha_0 \in \mathcal{F}} k(\alpha_0) x_{m,n}(\alpha_0)$$

for appropriately chosen coefficients $k(\alpha_0)$.

Proof Define for all $(m, n) \in \mathcal{A}$

$$p_{m,n} = \sum_{\alpha_0 \in \mathcal{F}} k(\alpha_0) x_{m,n}(\alpha_0) \, ,$$

where the coefficients $k(\alpha_0)$ still have to be selected appropriately. These $p_{m,n}$ satisfy the equilibrium equations at least for all states $m + n \geq N$, since these equations only contain $p_{m,n}$ with $(m, n) \in \mathcal{A}$. The remaining equations to be satisfied are the ones in states (m, n) with $m + n < N$. These equations form a linear, homogeneous system for the unknowns $k(\alpha_0)$ and the unknown quantities $p_{m,n}$ with (m, n) not in $\mathcal{A}$. For $N = r$ all states are contained in $\mathcal{A}$, so there are no undefined $p_{m,n}$. For $N > r$, there are $\frac{1}{2}N(N+1) - \frac{1}{2}r(r+1)$ states not in $\mathcal{A}$. Therefore, the number of equations is always equal to the number of unknowns, viz. $\frac{1}{2}N(N+1)$. Hence, by first omitting the equation in $(0, 0)$ say, the reduced system has a nonnull solution. The equation in $(0, 0)$ is automatically satisfied, since inserting the solution $p_{m,n}$ into the equations in states $(m, n) \neq (0, 0)$ and then summing over these equations and changing summations exactly yields the desired equation. Changing summations is allowed by the absolute convergence stated in Theorem

34

6.1(iii). Hence, $p_{m,n}$ is an absolutely convergent solution of all equilibrium equations. It remains to show that $p_{m,n}$ is a nonnull solution. This readily follows if at least one of the quantities $p_{m,n}$ with (m,n) not in $\mathcal{A}$ is nonnull. If all these quantities are null, then at least one of the coefficients $k(\alpha_0)$ must be nonnull. In this case the property that the sequences $x_{m,n}(\alpha_0)$ for different feasible α_0's are *linearly independent* on the set $\mathcal{A}$ implies that $p_{m,n}$ is a nonnull solution. The independence will be proved as a corollary to another result, which will be proved in the next lemma. Let's take the independence for granted at the moment. Then we can finish the proof of the theorem as follows. From a result of Foster ([11], Theorem 1) we may now conclude that the Markov process is ergodic and normalization of the $p_{m,n}$ produces the equilibrium probabilities. $\square$

Lemma 6.5 *Let the sequence* $a_0, a_1, \ldots$ *satisfy* $0 < |a_i| < 1$ *for* $i \geq 0$, $a_i \neq a_j$ *for* $i \neq j$ *and* $a_i \to 0$ *as* $i \to \infty$. *Define for all* $m \geq 0$

$$x_m = \sum_{i=0}^{\infty} k_i a_i^m \,,$$

with $\sum_{i=0}^{\infty} |k_i| < \infty$. *Then* $x_m \equiv 0 \iff k_i \equiv 0$.

Proof Define

$$f(z) = \sum_{m=0}^{\infty} x_m z^m \,, \qquad |z| \leq 1 \,.$$

By inserting the series for x_m and then changing summations, it readily follows that $f(z)$ may be continued to the meromorphic function

$$f(z) = \sum_{i=0}^{\infty} \frac{k_i}{1 - a_i z} \,, \qquad z \in \mathcal{C} \setminus \left\{ \frac{1}{a_0}, \frac{1}{a_1}, \ldots \right\} \,.$$

Let

$$\delta_j = \inf_{i \neq j} \left| \frac{1}{a_i} - \frac{1}{a_j} \right| \,, \qquad j \geq 0 \,,$$

so δ_j is the distance from the pole $1/a_j$ of $f(z)$ to its other poles. Since all a_i are distinct and $a_i \to 0$ as $i \to \infty$, it follows that $\delta_j > 0$. Use of Cauchy's Theorem of Residues yields

$$x_m \equiv 0 \;\Rightarrow\; f(z) \equiv 0 \;\Rightarrow\; 0 = \frac{1}{2\pi i} \int_{|z - \frac{1}{a_j}| = \frac{\delta_j}{2}} f(z)dz = \frac{k_j}{a_j} \;\Rightarrow\; k_j = 0 \;\; (j \geq 0) \,.$$

The implication $k_i \equiv 0 \Rightarrow x_m \equiv 0$ is trivial. $\square$

Corollary *The sequences* $x_{m,n}(\alpha_0)$ *for different feasible* α_0's *are linearly independent on the set* $\mathcal{A}$.

Proof We will show that the assumption that the sum

$$\sum_{\alpha_0 \in \mathcal{F}} l(\alpha_0) x_{m,n}(\alpha_0)$$

vanishes for all (m,n) with $m + n \geq N$ implies that $l(\alpha_0) = 0$ for all α_0 in $\mathcal{F}$. The reversed implication is trivial.

For fixed n the sum above is an infinite sum of powers α_i^m. Let us first consider an α_0 in $\mathcal{F}$, say $\tilde{\alpha}_0$, which is maximal in absolute value. The power $\tilde{\alpha}_0^m$ occurs exactly once in that sum, with coefficient $l(\tilde{\alpha}_0)d_0\beta_0^n$ if $n \geq r$ and $l(\tilde{\alpha}_0)e_{0,n}$ if $n < r$. Hence, from Lemma 6.3 we obtain for $\tilde{\alpha}_0$ that

$$l(\tilde{\alpha}_0)d_0 = l(\tilde{\alpha}_0)e_{0,0} = \ldots = l(\tilde{\alpha}_0)e_{0,r-1} = 0 \,,$$

which implies, since not all coefficients $d_0, e_{0,0}, \ldots, e_{0,r-1}$ are zero, that $l(\tilde{\alpha}_0) = 0$. By subsequently applying the same arguments to the (or a) second largest α_0 in $\mathcal{F}$ and so on, it follows that $l(\alpha_0) = 0$ for all α_0 in $\mathcal{F}$. $\qquad\square$

We end this section with a remark on the normalization equation. By inserting into this equation the series for $p_{m,n}$ with (m,n) in $\mathcal{A}$ we get

$$
\begin{aligned}
1 &= \sum_{(m,n)\notin\mathcal{A}} p_{m,n} + \sum_{(m,n)\in\mathcal{A}} \sum_{\alpha_0\in\mathcal{F}} k(\alpha_0)x_{m,n}(\alpha_0) \\
&= \sum_{(m,n)\notin\mathcal{A}} p_{m,n} + \sum_{\alpha_0\in\mathcal{F}} k(\alpha_0) \sum_{(m,n)\in\mathcal{A}} x_{m,n}(\alpha_0) \,.
\end{aligned}
\tag{6.20}
$$

By substituting the series (6.2)–(6.3), the infinite sum with $x_{m,n}(\alpha_0)$ expands to

$$
\begin{aligned}
\sum_{(m,n)\in\mathcal{A}} x_{m,n}(\alpha_0) &= \sum_{m=0}^{N-r-1} c_0 \frac{d_0\beta_0^{N-m}}{1-\beta_0}\alpha_0^m + c_0 \frac{d_0\beta_0^r}{1-\beta_0}\frac{\alpha_0^{N-r}}{1-\alpha_0} \\
&+ \sum_{m=0}^{N-r-1}\sum_{i=0}^{\infty}\sum_{j=1}^{r} c_{ri+j} \left(\frac{d_i\beta_i^{N-m}}{1-\beta_i} + \frac{d_{ri+j}\beta_{ri+j}^{N-m}}{1-\beta_{ri+j}} \right) \alpha_{ri+j}^m \\
&+ \sum_{i=0}^{\infty}\sum_{j=1}^{r} c_{ri+j} \left(\frac{d_i\beta_i^r}{1-\beta_i} + \frac{d_{ri+j}\beta_{ri+j}^r}{1-\beta_{ri+j}} \right) \frac{\alpha_{ri+j}^{N-r}}{1-\alpha_{ri+j}} + \sum_{i=0}^{\infty} c_i \sum_{n=0}^{r-1} e_{i,n}\frac{\alpha_i^{N-r}}{1-\alpha_i} \,.
\end{aligned}
$$

In the next section we will demonstrate that in the same way expressions for global performance measures (like the mean waiting time) can be derived.

7 Expressions for global performance measures

In this section we will show that the series for the probabilities $p_{m,n}$, as developed in the previous sections, lead to similar series for the mean waiting time and the complementary waiting time distribution.

The waiting time W of an arbitrary job is given by

$$W = S_1 + \ldots + S_M \,,$$

where M is the number of subjobs in the queue with the smallest number of subjobs on arrival and $S_1, S_2, \ldots$ are exponentially distributed random variables with unit mean and independent of M. By conditioning on M and using the property that Poisson arrivals see time averages (see e.g. Wolff [17]) we find

$$
EW = \sum_{(m,n)} m p_{m,n} \,,
\tag{7.1}
$$

$$
P(W > t) = \sum_{(m,n)} F_m^c(t) p_{m,n} \qquad (t > 0) \,,
\tag{7.2}
$$

where $F_m^c(t)$ is the complementary Erlang-m distribution with scale parameter 1, i.e.

$$F_m^c(t) = \sum_{j=0}^{m-1} e^{-t} \frac{t^j}{j!}.$$

Insertion of the series for $p_{m,n}$ with $(m,n) \in \mathcal{A}$ into (7.1) and (7.2) yields

$$
\begin{aligned}
EW &= \sum_{(m,n)\notin\mathcal{A}} m p_{m,n} + \sum_{(m,n)\in\mathcal{A}} \sum_{\alpha_0\in\mathcal{F}} k(\alpha_0) m x_{m,n}(\alpha_0) \\
&= \sum_{(m,n)\notin\mathcal{A}} m p_{m,n} + \sum_{\alpha_0\in\mathcal{F}} k(\alpha_0) \sum_{(m,n)\in\mathcal{A}} m x_{m,n}(\alpha_0)
\end{aligned}
\tag{7.3}
$$

and

$$P(W > t) = \sum_{(m,n)\notin\mathcal{A}} F_m^c(t) p_{m,n} + \sum_{\alpha_0\in\mathcal{F}} k(\alpha_0) \sum_{(m,n)\in\mathcal{A}} F_m^c(t) x_{m,n}(\alpha_0) \tag{7.4}$$

By substituting the series (6.2)–(6.3), the sums with $x_{m,n}(\alpha_0)$ in (7.3) and (7.4) expand to

$$
\begin{aligned}
\sum_{(m,n)\in\mathcal{A}} m x_{m,n}(\alpha_0) &= \sum_{m=0}^{N-r-1} m c_0 \frac{d_0 \beta_0^{N-m}}{1-\beta_0} \alpha_0^m + c_0 \frac{d_0 \beta_0^r}{1-\beta_0} A(\alpha_0) \\
&\quad + \sum_{m=0}^{N-r-1} m \sum_{i=0}^{\infty} \sum_{j=1}^{r} c_{ri+j} \left(\frac{d_i \beta_i^{N-m}}{1-\beta_i} + \frac{d_{ri+j}\beta_{ri+j}^{N-m}}{1-\beta_{ri+j}} \right) \alpha_{ri+j}^m \\
&\quad + \sum_{i=0}^{\infty} \sum_{j=1}^{r} c_{ri+j} \left(\frac{d_i \beta_i^r}{1-\beta_i} + \frac{d_{ri+j}\beta_{ri+j}^r}{1-\beta_{ri+j}} \right) A(\alpha_{ri+j}) \\
&\quad + \sum_{i=0}^{\infty} c_i \sum_{n=0}^{r-1} e_{i,n} A(\alpha_i)
\end{aligned}
\tag{7.5}
$$

and

$$
\begin{aligned}
\sum_{(m,n)\in\mathcal{A}} F_m^c(t) x_{m,n}(\alpha_0) &= \sum_{m=0}^{N-r-1} F_m^c(t) c_0 \frac{d_0 \beta_0^{N-m}}{1-\beta_0} \alpha_0^m + c_0 \frac{d_0 \beta_0^r}{1-\beta_0} B(\alpha_0) \\
&\quad + \sum_{m=0}^{N-r-1} F_m^c(t) \sum_{i=0}^{\infty} \sum_{j=1}^{r} c_{ri+j} \left(\frac{d_i \beta_i^{N-m}}{1-\beta_i} + \frac{d_{ri+j}\beta_{ri+j}^{N-m}}{1-\beta_{ri+j}} \right) \alpha_{ri+j}^m \\
&\quad + \sum_{i=0}^{\infty} \sum_{j=1}^{r} c_{ri+j} \left(\frac{d_i \beta_i^r}{1-\beta_i} + \frac{d_{ri+j}\beta_{ri+j}^r}{1-\beta_{ri+j}} \right) B(\alpha_{ri+j}) \\
&\quad + \sum_{i=0}^{\infty} c_i \sum_{n=0}^{r-1} e_{i,n} B(\alpha_i),
\end{aligned}
\tag{7.6}
$$

where $A(\alpha)$ and $B(\alpha)$ are defined as

$$A(\alpha) = \frac{\alpha + (N-r)(1-\alpha)}{(1-\alpha)^2} \alpha^{N-r},$$

$$B(\alpha) = \frac{F_{N-r}^c(t)\alpha^{N-r} + e^{(\alpha-1)t}(1 - F_{N-r}^c(\alpha t))\alpha}{1-\alpha}.$$

Of course, similar expressions may be obtained for the second or higher moments of the waiting time.

37

8   Some numerical aspects

This section is devoted to numerical aspects of the series for the probabilities $p_{m,n}$.

One of the numerical problems of using the series $x_{m,n}(\alpha_0)$ for the computation of the probabilities $p_{m,n}$ is the generation of the numbers α_i and β_i. Starting with α_0 the numbers α_i and β_i with $i \geq 0$ are generated such that for all $i \geq 0$ the β_i is the root inside the circle $|\beta| = |\alpha_i|$ of

$$\alpha_i^r \beta (2 + \lambda) = \beta^{r+1} \lambda + \alpha_i^r \beta^2 + \alpha_i^{r+1} \tag{8.1}$$

and the numbers $\alpha_{ri+1}, \ldots, \alpha_{ri+r}$ are the roots inside the circle $|\alpha| = |\beta_i|$ of

$$\alpha^r \beta_i (2 + \lambda) = \beta_i^{r+1} \lambda + \alpha^r \beta_i^2 + \alpha^{r+1} \, . \tag{8.2}$$

Hence, we have to calculate zeros of the polynomials in (8.1)–(8.2). In general, this constitutes a difficult numerical problem. Luckily, in our case we can transform the equations (8.1)–(8.2) to contraction equations, which can be solved efficiently. By dividing equation (8.1) by α_i^{r+1} and introducing $y = \beta/\alpha_i$ we can rewrite equation (8.1) as

$$y = F(y, \alpha_i) \, , \tag{8.3}$$

where

$$F(y, \alpha) = \frac{1}{2 + \lambda} (\lambda y^{r+1} + \alpha y^2 + 1) \, .$$

Analogously, by dividing equation (8.2) by β_i^{r+1} and introducing $z = \alpha/\beta_i$, equation (8.2) can be rewritten as

$$z = \phi G(z, \beta_i) \, , \tag{8.4}$$

where ϕ satisfies $\phi^r = 1$ and

$$G(z, \beta) = \sqrt[r]{\frac{\lambda}{2 + \lambda - z - \beta}} \, .$$

The following lemma states that both F and G are contractions.

Lemma 8.1

(i) *For all α with $|\alpha| < 1$ the function $F(y, \alpha)$ is a contraction on $|y| \leq Y$, where Y is the positive root, less than one, of the equation*

$$y = F(y, |\alpha|) \, . \tag{8.5}$$

(ii) *For all β with $|\beta| < 1$ the function $\phi G(z, \alpha)$ is a contraction on $|z| \leq Z$, where ϕ satisfies $\phi^r = 1$ and Z is the positive root, less than one, of the equation*

$$z = G(z, |\beta|) \, . \tag{8.6}$$

Proof It is readily verified that $F(y, |\alpha|)$ is convex for $0 \leq y \leq 1$, $F(0, |\alpha|) > 0$ and $F(1, |\alpha|) < 1$. Hence, there is, for each α with $|\alpha| < 1$, a unique Y with $0 < Y < 1$ satisfying

$$Y = F(Y, |\alpha|) \, .$$

38

For $|y| \le Y$s, it follows that

$$|F(y,\alpha)| \le F(|y|,|\alpha|) \le F(Y,|\alpha|) = Y ,$$

so, F maps $|y| \le Y$ into itself. Further, for $|y| \le Y$ we have

$$\left| \frac{\partial}{\partial y} F(y,\alpha) \right| \le \frac{\partial}{\partial y} F(|y|,|\alpha|) \le \frac{\partial}{\partial y} F(Y,|\alpha|) < 1 .$$

Hence, for $|y_1|, |y_2| \le Y$, we get

$$
\begin{aligned}
|F(y_1,\alpha) - F(y_2,\alpha)| &= \left| \int_0^1 \frac{\partial}{\partial y} F(y_1 + t(y_2 - y_1),\alpha)(y_2 - y_1) dt \right| \\
&\le |y_1 - y_2| \max_{0 \le t \le 1} \left| \frac{\partial}{\partial y} F(y_1 + t(y_2 - y_1),\alpha) \right| \\
&\le |y_1 - y_2| \frac{\partial}{\partial y} F(Y,|\alpha|) .
\end{aligned}
$$

which proves part (i) of the lemma. Part (ii) can be proved along the same lines. $\square$

By Lemma 8.1(i), equation (8.3) has exactly one root $\tilde{y}$ with $|\tilde{y}| \le Y$ and this root may be found by using the iteration scheme

$$y_{n+1} = F(y_n, \alpha_i), \qquad n = 0, 1, \dots$$

starting with $y_0 = 0$. The sequence $y_0, y_1, \dots$ converges exponentially fast to $\tilde{y}$, at least with rate $\frac{\partial}{\partial y} F(Y, |\alpha_i|)$. The desired β_i is then given by $\beta_i = \tilde{y}\alpha_i$. Analogously, by denoting the roots of $\phi^r = 1$ by $\phi_1, \dots, \phi_r$, it follows from Lemma 8.1(ii) that for each $j = 1, \dots, r$ equation (8.4) with $\phi = \phi_j$ has exactly one root $\tilde{z}_j$ satisfying $|\tilde{z}_j| \le Z$ and this root may be found by using the iteration scheme

$$z_{n+1} = \phi_j G(z_n, \beta_i), \qquad n = 0, 1, \dots$$

starting with $z_0 = 0$. The desired α_{ri+j} is then given by $\alpha_{ri+j} = \tilde{z}_j \beta_i$.

Another numerical problem is the calculation of the feasible values of α_0. According to Lemmas 5.3 and 5.5, for the calculation of the feasible values we first have to find the eigenvalues of the matrices H^+ and H^-. The eigenvalues are the zeros of the associated characteristic polynomials. For the characteristic polynomials of H^+ and H^- the following relations may be derived. Define for $k > 0$ the $k \times k$-matrix H_k by

$$H_k = \begin{pmatrix}
0 & 1 & 0 & 0 & \dots & 0 & 0 & 0 & 0 \\
2 & 0 & 1 & 0 & & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 1 & & 0 & 0 & 0 & 0 \\
\vdots & & & & & & & & \vdots \\
0 & 0 & 0 & 0 & & 1 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & & 0 & 1 & 0 & 1 \\
0 & 0 & 0 & 0 & \dots & 0 & 0 & 1 & 0
\end{pmatrix}$$

and let $h_k(x)$ be the characteristic polynomial of H_k. Then it is readily verified that $h_k(x)$ satisfies the following recurrence relations:

$$
\begin{aligned}
h_1(x) &= -x\,; \\
h_2(x) &= x^2 - 2\,; \\
h_k(x) &= -x h_{k-1}(x) - h_{k-2}(x)\,, \qquad k > 2\,.
\end{aligned}
$$

In the following lemma we formulate relations between the characteristic polynomials $h_k(x)$ and the ones of H^+ and H^-.

Lemma 8.2

(i) *For $k \geq 0$ the characteristic polynomial $h_{k+1}^+(x)$ of the $(k+1) \times (k+1)$-matrix H^+ defined by (5.16) is given by*

$$
\begin{aligned}
h_1^+(x) &= 2 - x\,; \\
h_2^+(x) &= x^2 - x - 2\,; \\
h_{k+1}^+(x) &= (1-x) h_k(x) - h_{k-1}(x)\,, \qquad k > 1\,.
\end{aligned}
$$

(ii) *For $k > 0$ the characteristic polynomial $h_{k+1}^+(x)$ of the $(k+1) \times (k+1)$-matrix H^+ defined by (5.21) is given by*

$$
\begin{aligned}
h_2^+(x) &= x^2 - 4\,; \\
h_{k+1}^+(x) &= -x h_k(x) - 2 h_{k-1}(x)\,, \qquad k > 1\,,
\end{aligned}
$$

and since the $k \times k$-matrix H^- defined by (5.23) is identical to H_k, its characteristic polynomial is given by $h_k(x)$.

For the case that r is odd, we obtain from Lemma 5.1 that all zeros of the characteristic polynomial of the matrix H^+ defined by (5.16) are simple, real-valued and in absolute value bounded by 2. Hence, these zeros can easily be computed by using a bisection method. Then, by Lemma 5.3, we can produce the feasible values of α_0 by squaring the roots γ with $|\gamma| < 1$ of the equations (5.17) where σ runs through the set of zeros of the characteristic polynomial of H^+. Equation (5.17) can be rewritten as

$$
\gamma = \phi \tilde{G}(\gamma, \sigma)
$$

where ϕ satisfies $\phi^r = 1$ and $\tilde{G}(\gamma, \sigma)$ is given by

$$
\tilde{G}(\gamma, \sigma) = \sqrt[r]{\frac{-\lambda}{\kappa + \sigma \gamma}}\,.
$$

It is readily verified that $\tilde{G}(\gamma, 2)$ is convex for $0 \leq \gamma \leq 1$, $\tilde{G}(0, 2) > 0$, $\tilde{G}(1, 2) = 1$ and $\frac{\partial}{\partial \gamma} \tilde{G}(1, 2) = 2/r\lambda > 1$. So there is a unique Γ with $0 < \Gamma < 1$ satisfying

$$
\Gamma = \tilde{G}(\Gamma, 2)\,.
$$

Similar to the proof of Lemma 8.1 it can be shown that $\phi \tilde{G}(\gamma, \sigma)$ is a contraction on $|\gamma| \leq \Gamma$. Hence, the zeros of equation (5.17) can be produced efficiently by using the scheme

$$
\gamma_{n+1} = \phi \tilde{G}(\gamma_n, \sigma)\,, \qquad n = 0, 1, \ldots
$$

starting with $\gamma_0 = 0$, where σ runs through the set of eigenvalues of H^+ and ϕ through the set of roots of $\phi^r = 1$.

In case r is even, the feasible values of α_0 can be produced similarly.

The terms $c_{ri+j}(d_i\beta_i^n + d_{ri+j}\beta_{ri+j}^n)\alpha_{ri+j}^m$, which add up to $x_{m,n}(\alpha_0)$ for $m \geq 0$ and $n \geq r$ (see (6.2)), may be represented in the r-fold tree depicted in Figure 9.
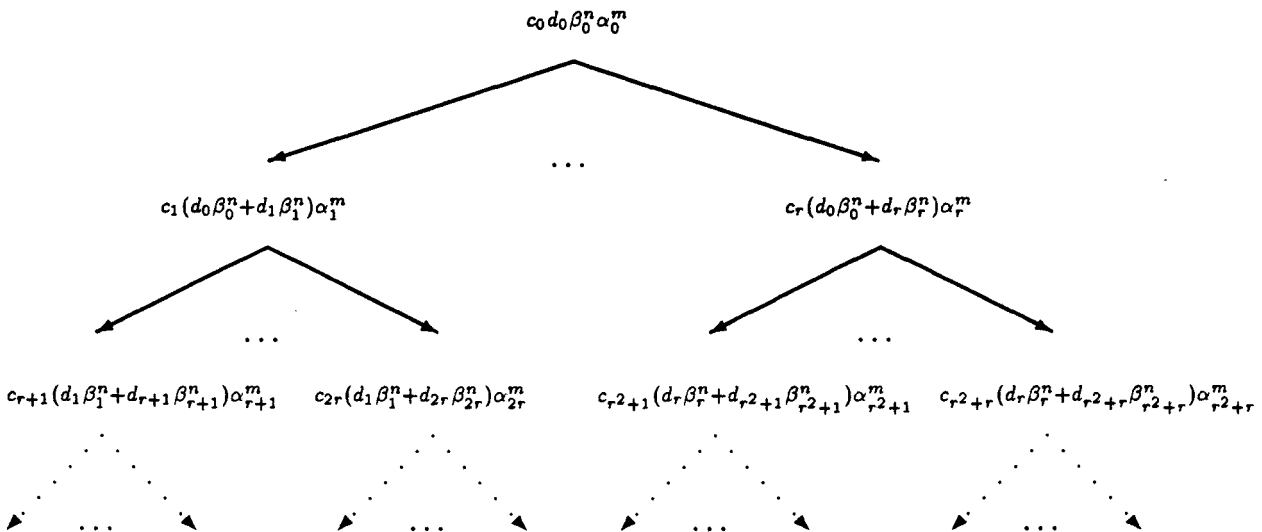

Figure 9: *The r-fold tree structure of the terms in the infinite sum $x_{m,n}(\alpha_0)$ given by (6.2).*

The infinite sum of the terms in this tree can be approximated by the finite sum of the terms at the first l levels. The number of terms at the first l levels equals $1 + r + \ldots + r^{l-1}$. This number increases very fast as l increases. Luckely, in Section 6 it has been shown for $m \geq 0$ and $n \geq r$ that the terms in $x_{m,n}(\alpha_0)$ behave asymptotically as a geometrical r-fold tree with rates

$$R_j(m+n) = |C_j| \left(\frac{|A_j|}{|A_0|}\right)^{m+n-r} , \qquad j = 1, \ldots, r. \qquad (8.7)$$

Since $|A_j|/|A_0| < 1$ for $j = 1, \ldots, r$ it follows that the terms in $x_{m,n}(\alpha_0)$ converge exponentially fast to zero. So, usually, already a few levels in the tree in Figure 9 provide an accurate approximation for $x_{m,n}(\alpha_0)$. Moreover, we may conclude from (8.7) that the convergence to zero of the terms in $x_{m,n}(\alpha_0, \beta_0)$ is faster for states further away from the origin. This feature can be exploited for numerical purposes; the trees of compensating terms are only used for the calculation of $p_{m,n}$ if $m + n$ is sufficiently large, that is, if the sum $R_1(m+n) + \ldots + R_r(m+n)$ is sufficiently small. The remaining quantities $p_{m,n}$ can then be solved from the equilibrium equations. The rate at which the sum $R_1(m+n) + \ldots + R_r(m+n)$ becomes small depends on the ratios $|A_1|/|A_0|, \ldots, |A_r|/|A_0|$. The equation

$$z^r(2 + \lambda) = \lambda + z^{r+1} ,$$

the roots of which define $A_0, \ldots, A_r$ (see Lemma 6.3), has exactly two positive roots, one larger and the other less than one. The larger root is A_0 and one of the roots $A_1, \ldots, A_r$ is the smaller one, A_1 say. It may readily be verified that $A_0 > 2$ and for $j = 1, \ldots, r$ that $|A_j| \leq A_1$, so

$$\frac{|A_j|}{|A_0|} \leq \frac{A_1}{A_0} < \frac{1}{2} .$$

In Table 1 the ratio A_1/A_0 is listed for increasing values of r and the workload $\rho = r\lambda/2$.

r	ρ	A_1	A_0	A_1/A_0
2	0.1	0.231	2.077	0.111
	0.5	0.500	2.414	0.207
	0.9	0.630	2.784	0.226
	0.95	0.642	2.832	0.227
5	0.1	0.481	2.039	0.236
	0.5	0.665	2.196	0.303
	0.9	0.740	2.355	0.314
	0.95	0.747	2.375	0.315
10	0.1	0.656	2.020	0.325
	0.5	0.772	2.100	0.368
	0.9	0.817	2.180	0.375
	0.95	0.821	2.190	0.375

Table 1: *The ratios A_1/A_0 for increasing values of r and the workload $\rho = r\lambda/2$. Here A_1 and A_0 are the positive roots with $A_1 < A_0$ of the equation $z^r(2 + \lambda) = \lambda + z^{r+1}$.*

Table 1 illustrates that A_1/A_0 is small, which implies that the sum $R_1(m + n) + \ldots + R_r(m + n)$ will already be small for $m + n$ being a little bit larger than r.

Finally, we remark that for high values of r (say, $r \geq 15$) the system of linear equations for the coefficients $k(\alpha_0)$ is 'close' to singular. In that case, the accurate computation of the $k(\alpha_0)$ is delicate. The reason is that for high values of r the feasible α_0's are closely clustered (see e.g. Figure 6; recall that the feasible α_0's are produced by squaring the feasible γ's). Hence, the set of solutions $x_{m,n}(\alpha_0)$ is nearly dependent, and thus also the system of linear equations for the coefficients $k(\alpha_0)$. Maybe, it helps to transform the solution base to a better conditioned base.

9   Numerical results

This section is devoted to numerical results. The evaluation of performance measures requires the computation of infinite sums (cf. (7.5) and (7.6)). The coefficients $k(\alpha_0)$ and the probabilities $p_{m,n}$ with (m,n) not in $\mathcal{A}$ (cf. (7.3) and (7.4)) are also required. They have to be solved from the equilibrium equations in states (m,n) with $m + n < N$ and the normalization equation. To set up these equations one has to compute the infinite sums $x_{m,n}(\alpha_0)$ for all (m,n) with $m \geq N - r$ and $m + n < N$, but also for all (m,n) with $m + n = N$, because these states have an incoming rate into one of the states (m,n) with $m + n < N$. All infinite sums will be approximated sufficiently close by partial sums. So, to evaluate performance measures, the following steps may be distinguished:

(i) Select a number l with $l \geq 1$ and approximate, for all states (m,n) in the set
$$\{(m,n)|m \geq N - r, n \geq 0, m + n < N\} \cup \{(m,n)|0 \leq m \leq N, m + n = N\},$$
the infinite sum $x_{m,n}(\alpha_0)$ by the terms of the first l levels (see Figure 9), i.e.

$$\hat{x}_{m,n}(\alpha_0) = c_0 d_0 \beta_0^n \alpha_0^m + \sum_{i=0}^{\frac{r^{l-1}-r}{r-1}} \sum_{j=1}^{r} c_{ri+j}(d_i \beta_i^n + d_{ri+j}\beta_{ri+j}^n)\alpha_{ri+j}^m, \qquad m \geq 0, n \geq r;$$

$$\hat{x}_{m,n}(\alpha_0) = \sum_{i=0}^{\frac{r^{l-1}-r}{r-1}} c_i e_{i,n} \alpha_i^m, \qquad m \geq 0, 0 \leq n < r.$$

Do the same for the infinite sums in the normalization equation (6.20).

(ii) Solve the coefficients $k(\alpha_0)$ and the probabilities $p_{m,n}$, with (m,n) in the set $\{(m,n)|0 \leq m < N-r, n \geq 0, m+n < N\}$, from the boundary equations and the normalization equation.

(iii) Compute approximations for performance measures by approximating the infinite sums occurring in the expressions for the performance measures (cf. (7.5) and (7.6)) by the terms at the first l levels.

(iv) Set $l = l + 1$ and perform the same computations as in (i)–(iii).

(v) Stop if only negligible improvement is seen for the performance measures. Otherwise repeat step (iv).

As it has already been mentioned in the previous section it is numerically sensible to replace N by a higher value $\tilde{N}$ because convergence of the infinite sums is faster for states further away from the origin (cf. Table 1). This is demonstrated in Table 2 for the case of Erlang-2 servers. We list the probabilities $p_{0,0}$ and $\sum_{n=0}^{\infty} p_{0,n}$ with an accuracy of 0.1% for increasing values of the occupation rate $\rho = \lambda r/2$ and $\tilde{N}$. The value of l indicates the number of levels of the trees required to reach the desired accuracy. The minimum value of $\tilde{N}$ is N which equals 2 in each example.

ρ	$\tilde{N}$	l	$p_{0,0}$	$\sum_{n=0}^{\infty} p_{0,n}$
0.8	2	12	0.09708	0.3029
	3	4	0.09709	0.3029
0.9	2	11	0.04399	0.1560
	3	4	0.04398	0.1560
0.99	2	11	0.004005	0.01510
	3	4	0.004005	0.01510

Table 2: *The probabilities $p_{0,0}$ and $\sum_{n=0}^{\infty} p_{0,n}$ for the Erlang-2 queueing system with an accuracy of 0.1% for increasing values of the occupation rate $\rho = \lambda r/2$ and $\tilde{N}$. The value of l indicates the number of levels of the trees required to reach the desired accuracy. The minimum value of $\tilde{N}$ is N which equals 2 in each example.*

Table 2 illustrates that the number of levels of the trees, that needs to be evaluated to reach the desired accuracy, decreases very fast as $\tilde{N}$ increases. When realizing that the first l levels involve $(r^l - 1)/(r - 1)$ terms (due to the r-fold tree structure) we may conclude that the numerical effort to compute the partial sums considerably decreases as $\tilde{N}$ is a little larger than N. Of course, the price is that a larger set of linear equations for the coefficients $k(\alpha_0)$ has to be solved, but this extra effort is easily compensated by the advantage of efficiently computing the infinite sums.

In Table 3 we list the probability $p_{0,0}$ of the system being empty, the delay probability π_W ($= 1 - \sum_{n=0}^{\infty} p_{0,n}$ by PASTA) and the *normalized* first two moments and the squared coefficient of variation $c^2_{W|W>0}$ of the conditional waiting time $W|W > 0$ for increasing values of ρ and r. The normalized k-th moment of the conditional waiting time is defined as the quotient of $E(W^k|W > 0)$ and the k-th power of the mean service time ($= r^k$). All quantities are computed with a relative accuracy of 0.1%. The value of l indicates the number of levels of the trees required to reach the desired accuracy. In the computations N has been replaced by the somewhat higher value $\tilde{N}$.

r	ρ	$p_{0,0}$	π_W	$\frac{E(W|W>0)}{r}$	$\frac{E(W^2|W>0)}{r^2}$	$c^2_{W|W>0}$	l	$\tilde{N} - r$	$N - r$
1	0.1	0.8175	0.01753	1.010	2.042	1.0000	3	1	0
	0.5	0.3160	0.3160	1.349	3.631	0.9947	4	1	0
	0.9	0.04225	0.8422	5.313	56.05	0.9855	4	1	0
	0.95	0.01996	0.9200	10.31	211.7	0.9905	4	1	0
	0.99	0.003808	0.9838	50.31	5057	0.9977	4	1	0
2	0.1	0.8177	0.01768	0.6475	0.7990	0.9061	3	1	0
	0.5	0.3193	0.3193	0.9375	1.665	0.8946	3	1	0
	0.9	0.04398	0.8440	3.923	30.17	0.9608	4	1	0
	0.95	0.02089	0.9209	7.672	116.5	0.9785	4	1	0
	0.99	0.004005	0.9840	37.67	2832	0.9953	4	1	0
3	0.1	0.8177	0.01774	0.5409	0.5331	0.8222	3	1	0
	0.5	0.3208	0.3208	0.8051	1.188	0.8329	3	1	0
	0.9	0.04486	0.8449	3.460	23.33	0.9479	4	1	0
	0.95	0.02137	0.9214	6.793	91.02	0.9723	4	1	0
	0.99	0.004108	0.9841	33.46	2233	0.9942	4	1	0
4	0.1	0.8178	0.01776	0.4903	0.4241	0.7639	3	1	0
	0.5	0.3216	0.3216	0.7399	0.9816	0.7929	3	1	0
	0.9	0.04539	0.8454	3.229	20.23	0.9401	4	1	0
	0.95	0.02166	0.9217	6.354	79.47	0.9685	4	1	0
	0.99	0.004171	0.9842	31.35	1960	0.9934	4	1	0
5	0.1	0.8178	0.01778	0.4609	0.3659	0.7224	3	2	0
	0.5	0.3221	0.3221	0.7011	0.8678	0.7653	3	2	0
	0.9	0.04574	0.8457	3.091	18.49	0.9349	4	2	0
	0.95	0.02186	0.9219	6.090	72.92	0.9660	4	2	0
	0.99	0.004214	0.9842	30.09	1804	0.9930	4	2	0
10	0.1	0.8178	0.01779	0.4040	0.2646	0.6215	3	3	1
	0.5	0.3231	0.3231	0.6242	0.6624	0.7003	3	3	0
	0.9	0.04653	0.8465	2.814	15.23	0.9230	3	3	0
	0.95	0.02230	0.9223	5.563	60.67	0.9660	3	3	0
	0.99	0.004311	0.9843	27.56	1513	0.9919	3	3	0

Table 3: *The probability $p_{0,0}$, the delay probability π_W and the normalized first two moments and the squared coefficient of variation $c^2_{W|W>0}$ of the conditional waiting time $W|W > 0$ for increasing values of ρ and r. All quantities are computed with a relative accuracy of 0.1%. The value of l indicates the number of levels of the trees required to reach the desired accuracy. In the computations N has been replaced by $\tilde{N}$.*

The examples in Table 3 show that the probabilities and performance measures can be computed efficiently; already a small number of terms, namely $(r^l - 1)/(r - 1)$ terms, is sufficient to reach high accuracy. Note from the results in Table 3 that the probabilities

$p_{0,0}$ and π_W are rather insensitive to the service time distribution and that the normalized first two moments of the conditional waiting time $W|W > 0$ are decreasing in $1/r$, i.e. the squared coefficient of variation of the service time. Furthermore, the normalized mean conditional waiting time appears to be nearly linear in $1/r$. This is of course well-known for the $M|G|1$ queue, but apparently it is also approximately valid for the shortest delay problem.

In Table 4 the waiting time probability $P(W > t)$ is listed for increasing values of t, ρ and r. The probabilities are computed with a relative accuracy of 0.1%. The values of t are restricted to multiples of the mean service time r. The exact values are compared with a two-moment approximation of the conditional waiting time probabilities. For all examples displayed in Table 4 it holds that $c^2_{W|W>0}$ exceeds $\frac{1}{2}$ (and is less than 1; see Table 3), so that we can fit a so-called K_2 distribution with *gamma normalization* to the first two moments of $W|W > 0$ (alternatively, we might have used an $E_{1,2}$ distribution; see Tijms [15] for details).

		$P(W > t)$					
		$t = r$		$t = 5r$		$t = 10r$	
r	ρ	exa	app	exa	app	exa	app
1	0.9	0.7005	0.6999	0.3296	0.3298	0.1275	0.1277
	0.95	0.8371	0.8359	0.5684	0.5683	0.3491	0.3496
2	0.9	0.6645	0.6608	0.2351	0.2366	0.06385	0.06385
	0.95	0.8157	0.8112	0.4822	0.4834	0.2494	0.2505
3	0.9	0.6468	0.6421	0.1971	0.1989	0.04453	0.04435
	0.95	0.8052	0.7992	0.4434	0.4452	0.2101	0.2112
4	0.9	0.6361	0.6312	0.1769	0.1787	0.03569	0.03542
	0.95	0.7988	0.7921	0.4213	0.4236	0.1893	0.1903
5	0.9	0.6288	0.6240	0.1644	0.1661	0.03074	0.03041
	0.95	0.7946	0.7875	0.4072	0.4097	0.1765	0.1765
10	0.9	0.6118	0.6078	0.1388	0.1403	0.02174	0.02132
	0.95	0.7845	0.7770	0.3764	0.3794	0.1504	0.1512

Table 4: *The waiting time probability $P(W > t)$ for increasing values of t, ρ and r. The values of t are restricted to multiples of the mean service time r. The exact values are computed with a relative accuracy of 0.1% and are listed in the columns with heading 'exa'. The two-moment approximations are listed in the columns with heading 'app'.*

The results in Table 4 indicate that the two-moment approximation is quite accurate. This is also illustrated in Figure 10 where for the Erlang-4 system with load 0.95 the distribution $P(W > t)$ together with the two-moment approximation are depicted for a wide range of t values. Of course, from a practical point of view, we also like to have good approximations for the first two moments of the conditional waiting time, since then we are able to use the two-moment approximation to produce fast and accurate approximations for the conditional waiting time distribution. A good approximation for the first moment is available (it is nearly linear in $1/r$; see the remark below Table 3); more research is required to find a good approximation for the second moment.

We end this section by demonstrating that also for extremely high traffic intensities the exact solution behaves well numerically. For the Erlang-2 and Erlang-4 system with workload $\rho = 0.999$, we list in Table 5 the probability $p_{0,0}$, the delay probability π_W, the normalized first two moments and the squared coefficient of variation $c^2_{W|W>0}$ of the conditional waiting time $W|W > 0$ and the waiting time probability $P(W > 100r)$.
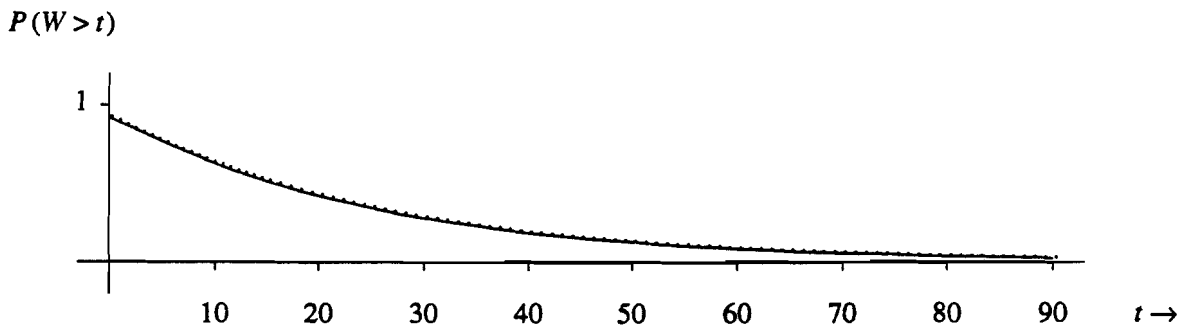
$P(W > t)$

Figure 10: *The exact waiting time probabilities $P(W > t)$ (solid line) together with the two-moment approximation (dotted line) for the Erlang-4 system with load 0.95.*

As before, all quantities are computed with a relative accuracy of 0.1%; the value of l indicates the number of levels of the trees required to reach the desired accuracy and in the computations N has been replaced by $\tilde{N}$.

r	$p_{0,0}$	π_W	$\frac{E(W|W>0)}{r}$	$c^2_{W|W>0}$	$P(W > 100r)$	l	$\tilde{N} - r$	$N - r$
2	0.000397	0.9984	375.2	0.9995	0.7649	4	1	0
5	0.000418	0.9984	300.1	0.9993	0.7156	4	2	0

Table 5: *The probability $p_{0,0}$, the delay probability π_W, the normalized mean and the squared coefficient of variation $c^2_{W|W>0}$ of the conditional waiting time $W|W > 0$ and the waiting time probability $P(W > 100r)$ for $\rho = 0.999$ and $r = 2$ and 5, respectively. All quantities are computed with a relative accuracy of 0.1%. The value of l indicates the number of levels of the trees required to reach the desired accuracy. In the computations N has been replaced by $\tilde{N}$.*

10 Conclusions and comments

The primary conclusion of the present paper may be that the compensation method, which was originally developed for the shortest queue problem with exponentially distributed service times and Poisson arrivals (cf. [1]), can indeed be extended to the case of Erlang distributed service times with expected delay routing. The compensation steps themselves become more complicated and also the finding of starting values α_0 requires more skill. Nevertheless, the final result has the same pleasant analytic and algorithmic features as the result for the simpler case. Numerically, the nice features are the following: a simple algorithm for the computation of the equilibrium probabilities; the series of products of powers converge exponentially fast and the convergence is faster for states further away from the origin; similar features for performance criteria like expected residence time. The algorithms in [1, 3] for exponential servers not only produce approximations but also upper and lower bounds which approach each other fastly; this also seems to be possible for Erlang servers although it has not been worked out in the present paper. In the exponential case it could be proven that the series $x_{m,n}(\alpha_0)$ converge absolutely for every $m, n \geq 0$. In the present case we could only prove absolute convergence on the set $\mathcal{A}$, which was determined by N. If $N = r$, then $\mathcal{A}$ represents the whole first quadrant of the

46

grid, but for $N > r$, there are also states for which $x_{m,n}(\alpha_0)$ was formally defined, but may not converge absolutely. A similar feature was already encountered in the case of the asymmetic shortest queue problem (cf. [2]) and also in the case of general random walks on the first quadrant with transitions to neighbours only (cf. [4]). The remaining question is of course: how large can N be? Or more precisely: how much larger than r can N be? Numerical evidence indicates that N is usually equal to r, but sometimes it is equal to $r + 1$. So, we might conjecture that there is indeed an upper limit for N, for instance $r + 1$ (cf. Table 3).

From a numerical point of view it may even be sensible to replace N by an even higher value than the prescribed one, since this speeds up convergence (at the price of having to solve a larger set of linear equations). One may add here that N was determined in a search for necessary and sufficient conditions for absolute convergence. It is quite well possible that for states not in $\mathcal{A}$ there is common convergence if the series are defined appropriately, namely by keeping terms together which are each other's complex conjugates. Our conjecture is that with this amendment all $x_{m,n}(\alpha_0)$ are convergent.

Another numerical complication is provided by the r-foldedness of the compensation trees. Luckily, it is usually not necessary to investigate many levels of these trees. Moreover, the amount of work can be kept within bounds by only climbing higher in the tree where it is really necessary (cf. [3]).

The problem in the present paper is a particular case of a whole class of related models. Its advantage is that it can be described in a useful way by a two-dimensional random walk with nice features. With *useful* we mean here that residence time characteristics can be obtained from the equilibrium probabilities. With *nice features* we mean here that the random walk is still sufficiently homogeneous. If we would allow the subjobs of an arriving job to split up over the available queues, then we would destroy the usefulness, although the analogous analysis can still be made. If we would take the shortest-queue-routing, then we can still represent the process by a random walk, however, we would loose homogeneity in m for fixed $n < r$. It is most likely that in this case the problem can still be solved with the compensation approach, however, one would need extra freedom because of the lack of homogeneity on the horizontal boundary levels. A good way of constructing the extra freedom might be to introduce separate state variables indicating how many subjobs of a job in service are still waiting for execution. In this way the state space is becoming more complicated, which leads to basic solutions of the form

$$\alpha^m \beta^n \gamma^k \delta^l \text{ with } k, l = 0, \ldots, r - 1.$$

Such an approach has appeared to be effective for the $E_s/E_r/c$-problem (cf. [5]). The same holds for the replacement of the Poisson arrival process, which might be replaced by Erlang distributed interarrival times: an extra factor ε^i might help, like in [5]. It seems natural to replace the pure Erlang service times by mixtures of Erlang distributions with the same scale parameter. With the class of distributions we have the ability to approximate any distribution sufficiently close. This model can be described by a two-dimensional random walk with the same nice features as the model with pure Erlang servers. However, finding feasible starting solutions is then essentially more complicated and this problem cannot be solved anymore by the solution of the jockeying model.

For obtaining the feasible starting solutions, the instantaneous jockeying model was used. In fact, the interpretation as instantaneous jockeying was not the essential feature, although it was good to see how the jockeying model could be solved relatively straight-forwardly. The essential feature, however, was that the jockeying model represented an ideally truncated version of the original problem and that, in this way, we obtained all

feasible starting solutions for the original problem. Since the more heuristic procedures of other cases, can all be interpreted similarly, one may conjecture that this truncation procedure provides a general approach to procuring the feasible starting solutions.

Of course, it is also possible to consider other, possibly more realistic, variants of the jockeying model. Indeed, similar approaches as the one in Section 5 seem workable in such cases.

The compensation approach clearly has its limitations. The most important one being that transitions to North, North-East and East are already forbidden in the case with transitions to neighbours only (cf. [4]). The strong feature, however, of the compensation approach is that it helps in finding such conditions for getting elegant solutions and that it provides constructive methods for obtaining the elegant solutions in case the conditions are satisfied. In fact, the two-dimensional random walks with transitions to neighbours only have been studied very intensively over the years with the help of generating functions, but the results mentioned above could only be produced in that way with the results of the compensation approach as guidelines (cf. Cohen [9]). Possibly, also the results of the present paper can be recognized in properties of the generating functions.

References

[1] I.J.B.F. Adan, J. Wessels, W.H.M. Zijm. Analysis of the symmetric shortest queue problem. *Stochastic Models* **6** (1990) 691–713.

[2] I.J.B.F. Adan, J. Wessels, W.H.M. Zijm. Analysis of the asymmetric shortest queue problem with threshold jockeying. *Stochastic Models* **7** (1991) 615–627.

[3] I.J.B.F. Adan, J. Wessels, W.H.M. Zijm. Analysis of the asymmetric shortest queue problem. *Queueing Systems* **8** (1991) 1–58.

[4] I.J.B.F. Adan, J. Wessels, W.H.M. Zijm. A compensation approach for two-dimensional Markov processes. *Advances of Applied Probability* (1993) (to appear).

[5] I.J.B.F. Adan, W.A. van de Waarsenburg, J. Wessels. Analysing $E_k/E_r/c$ queues. Submitted for publication.

[6] I.J.B.F. Adan, G.J. van Houtum, J. Wessels, W.H.M. Zijm. A compensation procedure for multiprogramming queues. *O.R.-Spektrum* **15** (1993) 95–106.

[7] O.J. Boxma, G.J. van Houtum. The compensation approach applied to a 2×2 switch. *Probab. Engineer. Inform. Sci.* (to appear).

[8] J.W. Cohen, O.J. Boxma. *Boundary value problems in queueing system analysis.* North-Holland, Amsterdam 1983.

[9] J.W. Cohen. On a class of two-dimensional nearest neighbouring random walks. CWI report BS-R9220 (1992). To appear in *J. Appl. Prob.*, Festschrift for Lajos Takacs.

[10] L.Flatto, H.P. McKean. Two queues in parallel. *Comm. Pure Appl. Math.* **30** (1977) 255–263.

[11] F.G. Foster. On the stochastic matrices associated with certain queueing processes. *Ann. Math. Statist.* **24** (1953) 355–360.

[12] F.A. Haight. Two queues in parallel. *Biometrika* **45** (1958) 401–410.

[13] A. Hordijk, G. Koole. On the optimality of the generalized shortest queue policy. *Prob. Engineer. Inform. Sci.* **4** (1990) 477–487.

[14] J.F.C. Kingman. Two similar queues in parallel. *Ann. Math. Statist.* **32** (1961) 1314–1323.

[15] H.C. Tijms. *Stochastic modelling and analysis: a computational approach.* John Wiley, Chichester 1986.

[16] R.R. Weber. On the optimal assignment of customers to parallel queues. *J. Appl. Prob.* **15** (1978) 406–413.

[17] R.W. Wolff. Poisson arrivals see time averages. *Opns. Res.* **30** (1982) 223–231.

List of COSOR-memoranda - 1993

Number	Month	Author	Title
93-01	January	P. v.d. Laan C. v. Eeden	Subset selection for the best of two populations: Tables of the expected subset size
93-02	January	R.J.G. Wilms J.G.F. Thiemann	Characterizations of shift-invariant distributions based on summation modulo one.
93-03	February	Jan Beirlant John H.J. Einmahl	Asymptotic confidence intervals for the length of the shortt under random censoring.
93-04	February	E. Balas J. K. Lenstra A. Vazacopoulos	One machine scheduling with delayed precedence constraints
93-05	March	A.A. Stoorvogel J.H.A. Ludlage	The discrete time minimum entropy H_∞ control problem
93-06	March	H.J.C. Huijberts C.H. Moog	Controlled invariance of nonlinear systems: nonexact forms speak louder than exact forms
93-07	March	Marinus Veldhorst	A linear time algorithm to schedule trees with communication delays optimally on two machines
93-08	March	Stan van Hoesel Antoon Kolen	A class of strong valid inequalities for the discrete lot-sizing and scheduling problem
93-09	March	F.P.A. Coolen	Bayesian decision theory with imprecise prior probabilities applied to replacement problems
93-10	March	A.W.J. Kolen A.H.G. Rinnooy Kan C.P.M. van Hoesel A.P.M. Wagelmans	Sensitivity analysis of list scheduling heuristics
93-11	March	A.A. Stoorvogel J.H.A. Ludlage	Squaring-down and the problems of almost-zeros for continuous-time systems
93-12	April	Paul van der Laan	The efficiency of subset selection of an ε-best uniform population relative to selection of the best one
93-13	April	R.J.G. Wilms	On the limiting distribution of fractional parts of extreme order statistics

Number	Month	Author	Title
93-14	May	L.C.G.J.M. Habets	On the Genericity of Stabilizability for Time-Day Systems
93-15	June	P. van der Laan C. van Eeden	Subset selection with a generalized selection goal based on a loss function
93-16	June	A.A. Stoorvogel A. Saberi B.M. Chen	The Discrete-time H_∞ Control Problem with Strictly Proper Measurement Feedback
93-17	June	J. Beirlant J.H.J. Einmahl	Maximal type test statistics based on conditional processes
93-18	July	F.P.A. Coolen	Decision making with imprecise probabilities
93-19	July	J.A. Hoogeveen J.K. Lenstra B. Veltman	Three, four, five, six or the Complexity of Scheduling with Communication Delays
93-20	July	J.A. Hoogeveen J.K. Lenstra B. Veltman	Preemptive scheduling in a two-stage multiprocessor flow shop is NP-hard
93-21	July	P. van der Laan C. van Eeden	Some generalized subset selection procedures
93-22	July	R.J.G. Wilms	Infinite divisible and stable distributions modulo 1
93-23	July	J.H.J. Einmahl F.H. Ruymgaart	Tail processes under heavy random censorship with applications
93-24	August	F.W. Steutel	Probabilistic methods in problems of applied analysis
93-25	August	A.A. Stoorvogel	Stabilizing solutions of the H_∞ algebraic Riccati equation
93-26	August	R. Perelaer J.K. Lenstra M. Savelsbergh F. Soumis	The Bus Driver Scheduling Problem of the Amsterdam Transit Company
93-27	August	J.M. van den Akker C.P.M. van Hoesel M.W.P. Savelsbergh	Facet Inducing Inequalities for Single-Machine Scheduling Problems
93-28	September	H.J.C. Huijberts H. Nijmeijer	Dynamic disturbance decoupling of nonlinear systems and linearization

Number	Month	Author	Title
93-29	September	L.C.G.J.M. Habets	Testing Reachability and Stabilizability of Systems over Polynomial Rings using Gröbner Bases
93-30	September	Z. Liu J.A.C. Resing	Duality and Equivalencies in Closed Tandem Queueing Networks
93-31	November	A.G. de Kok J. van der Wal	Assigning identical operators to different machines
93-32	November	M.J.A. van Eenige J.A.C. Resing J. van der Wal	A Matrix-Geometric Analysis of Queueing Systems with Periodic Service Interruptions
93-33	November	A.G. de Kok	Backorder lead time behaviour in (b, Q)-inventory models with compound renewal demand
93-34	November	J. van der Wal	The lead time shift theorem for the continuous review (s, Q) and (s, S) inventory systems
93-35	November	J.H.C.M. Verrijdt A.G. de Kok	Distribution Planning for a Divergent 2-echelon Network without Intermediate Stocks under Service Restrictions
93-36	November	A.G. Lagodimos A.G. de Kok J.H.C.M. Verrijdt	The Robustness of Multi-Echelon Service Models under Autocorrelated Demands
93-37	November	C.P.M. van Hoesel A.P.M. Wagelmans	An $O(T^3)$ algorithm for the economic lot-sizing problem with constant capacities
93-38	November	I.J.B.F. Adan J. Wessels	Analysing shortest expected delay routing for Erlang servers