Solving nonlinear matrix equations arising in Tree-Like stochastic processes

Dario A. Bini [a], Guy Latouche [b] and Beatrice Meini [a]

[a] *Dipartimento di Matematica, Università di Pisa, via Buonarroti 2, 56127 Pisa (Italy)*

[b] *Départment d'Informatique, U.L.B., C.P. 212, Boulevard du Triomphe, 1050 Bruxelles (Belgium)*

Abstract

In this paper, based on matrix structure analysis, we derive and analyze efficient algorithms to solve nonlinear matrix equations of the form $X + \sum_{1\le i\le d} A_i X^{-1} D_i = C$. This class of equations is encountered in the solution of Tree-Like stochastic processes which are a generalization of Quasi-Birth-and-Death (QBD) processes.

1 Introduction

Let m, d be positive integers and let $C, A_i, D_i \in \mathbb{R}^{m\times m}$, $i = 1, \ldots, d$, be given matrices. We consider the problem of computing a nonsingular matrix X which solves the matrix equation

$$X + \sum_{1\le i\le d} A_i X^{-1} D_i = C, \tag{1}$$

where we assume that

(1) $C = B - I$, and B is sub-stochastic;
(2) A_i and D_i have nonnegative entries;
(3) the matrices $I + C + D_i + A_1 + \cdots + A_d$, $i = 1, \ldots, d$, are stochastic.

A nonnegative matrix S is stochastic if $S\mathbf{1} = \mathbf{1}$, for $\mathbf{1} = (1, \ldots, 1)^T$, S is sub-stochastic if $S\mathbf{1} \neq \mathbf{1}$ and $S\mathbf{1} \le \mathbf{1}$, where the inequality holds componentwise.

This problem arises in the analysis of certain discrete-time bivariate Markov processes called Tree-Like processes [10]. Typical applications of Tree-Like processes have been investigated by Yeung and Sengupta [18] (single server queues with LIFO service discipline), He and Alfa [6,7] (an arriving customer interrupts the service in progress), and by Takine, Sengupta and Yeung [14] (an arriving customer is placed at the head of the queue but does not preempt the server). Van Houdt and Blondia [15] use Tree-Like processes to evaluate a medium access control protocol with an underlying stack structure. A brief general introduction is given in Latouche and Ramaswami [10, Chapter 14].

The generator matrix of a Tree-Like process has the following form

$$Q = \begin{bmatrix} C_0 & \Lambda_1 & \Lambda_2 & \dots & \Lambda_d \\ V_1 & W & 0 & \dots & 0 \\ V_2 & 0 & W & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ V_d & 0 & \dots & 0 & W \end{bmatrix}, \tag{2}$$

where $\Lambda_1, \dots, \Lambda_d$ are matrices with m rows and infinitely many columns, given by

$$\Lambda_i = \begin{bmatrix} A_i & 0 & 0 & \dots \end{bmatrix}, \qquad \text{for } 1 \le i \le d,$$

the matrices $V_1, \dots, V_d$ have m columns and infinitely many rows and are given by

$$V_i = \begin{bmatrix} D_i \\ 0 \\ 0 \\ \vdots \end{bmatrix}, \qquad \text{for } 1 \le i \le d,$$

the infinite matrix W is recursively defined by

$$W = \begin{bmatrix} C & \Lambda_1 & \Lambda_2 & \dots & \Lambda_d \\ V_1 & W & 0 & \dots & 0 \\ V_2 & 0 & W & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ V_d & 0 & \dots & 0 & W \end{bmatrix}, \tag{3}$$

and $C_0 = B_0 - I$, where B_0 is an $m \times m$ sub-stochastic matrix.

The nature of this matrix can be better understood if we introduce the set of multi-indices

$$\mathcal{S} = \bigcup_{\ell=0}^{\infty} \mathcal{S}_\ell,$$

$$\mathcal{S}_0 = \{1, 2, \dots, m\},$$

$$\mathcal{S}_\ell = \{(j_1, \dots, j_\ell; i) : 1 \le j_1, \dots, j_\ell \le d, 1 \le i \le m\}, \ \ell = 1, 2, \dots .$$

In fact, W is a linear operator defined on the linear space $\ell^1(\mathcal{S})$ made up by all the multi-index sequences $\boldsymbol{x} = (x_k)_{k\in\mathcal{S}}$, $x_k \in \mathbb{R}$, having bounded 1-norm, i.e., $||\boldsymbol{x}|| = \sum_{k\in\mathcal{S}} |x_k| < \infty$. Here, the ordering of the components of $\boldsymbol{x}$, which provides the particular structure (3) to the matrix W, is lexicographic. That is, $(j_1, \dots, j_\ell; i) < (j'_1, \dots, j'_{\ell'}; i')$ if either $\ell < \ell'$ or there exists h, $1 \le h \le \ell$ such that $j_h < j'_h$, $j_q = j'_q$ for $q = 1, \dots, h-1$, or $j_q = j'_q$ for $q = 1, \dots, \ell$ and $i < i'$. In this way the multi-index sequence $\boldsymbol{x} = (x_k)_{k\in\mathcal{S}}$ can be naturally partitioned as

$$\boldsymbol{x} = (\boldsymbol{x}^{(0)}, \boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(d)}), \tag{4}$$

where $\boldsymbol{x}^{(0)}$ is indexed by the indices of $\mathcal{S}$ of length $\ell = 0$, i.e., $\boldsymbol{x}^{(0)} = (x_i)_{i\in\mathcal{S}_0}$; the sequence $\boldsymbol{x}^{(q)}$, for $q = 1, \dots, d$ is indexed by the indices $k = (j_1, \dots, j_\ell; i) \in \mathcal{S}$ of length $\ell \ge 1$ with the first component $j_1 = q$.

As we will see in the next section, since there is a finite number of nonzero entries in each row and column of W, this infinite matrix represents a linear operator $W : \ell^1(\mathcal{S}) \to \ell^1(\mathcal{S})$ with bounded 1-norm, that is $||W|| = \sup_{||\boldsymbol{x}||=1} ||W\boldsymbol{x}|| < \infty$. The same holds for W^T.

The computation of the stationary distribution of the Markov process, i.e. the infinite nonnegative vector $\boldsymbol{\pi} = (\pi_i)_{i\in\mathcal{S}}$ such that $\boldsymbol{\pi}^{\mathrm{T}} Q = \mathbf{0}^{\mathrm{T}}$, $\sum_{i\in\mathcal{S}} \pi_i = 1$, is reduced to computing the UL factorization of (2) and this computation can be ultimately reduced to solving (1).

In this paper, we derive efficient algorithms to solve (1). We first analyze the natural fixed point iteration, already introduced in [10,17] and we perform a precise analysis of the speed of convergence based on the Perron-Frobenius theory of nonnegative matrices. Then we introduce two new improved methods. The first one consists in generating a sequence of approximations, which are computed by solving at each step d quadratic matrix equations; these equations can be solved by means of efficient algorithms, like the cyclic reduction (Bini and Meini [1,2], Latouche and Ramaswami [9,10]) which have a quadratic convergence. The second method is the application of Newton's scheme, it leads to a sequence of approximations which quadratically converges to the solution; each step of Newton's method requires the computation of the solution of a linear Sylvester-like equation of the form $Y + \sum_{1\le i\le d} H_i Y K_i = L$, for which we apply a suitable fixed point iteration. We prove the convergence of the methods, moreover we analyze and compare their convergence speeds making use once again of the Perron-Frobenius theory.

The natural fixed point iteration has a low computational cost per step, but the convergence may be very slow. The algorithm based on the solution of quadratic matrix equations generally converges faster, and is convenient for problems where the convergence of fixed point iteration is too slow. Newton's method is quadratically convergent, but its effectiveness relies on the efficient solution at each step of the Sylvester-like matrix equation.

Our paper is organized as follows. We define our notations and recall the definition of a Tree-Like process and some basic results in the next section; in particular, we recall that the solution of the nonlinear matrix equation (1) is of fundamental importance in determining the stationary distribution of the Tree-Like process. We describe in Section 3 the simple algorithm to solve that equation, based on functional iteration, and an improved algorithm in Section 4, based on the the solution of quadratic matrix equations. We apply Newton's iteration in Section 5 and we report in Section 6 on numerical experimentation to compare the three methods. We draw conclusions in Section 7.

2 The fundamental equation

Let $\{(Y_n, \phi_n) : n \geq 0\}$ be a discrete-time bivariate Markov processes in which the values of the random variables Y_n are represented by the nodes of a d-ary tree, and the random variables ϕ_n take integer values in $[1 \cdots m]$. A d-ary tree is a tree for which each node has d children.

Each node is represented by a string of integers taking values in $[1 \cdots d]$. If $J = (j_1, j_2, \ldots, j_\ell)$ is one such string, its length ℓ is denoted by $|J|$ and its children by $J + k = (j_1, j_2, \ldots, j_\ell, k)$, with $1 \leq k \leq d$. The root is represented by $\emptyset$, the empty string of length 0. Thus, the random variables (Y_n, ϕ_n) take their values in the state space $\mathcal{S}$ where

$$\mathcal{S} = \{(j_1, \ldots, j_\ell; i) : \ell \geq 0, 1 \leq j_1, \ldots, j_\ell \leq d, 1 \leq i \leq m\}.$$

The state space is partitioned into nodes

$$\mathcal{N}_J = \{(j_1, \ldots, j_\ell; i) : 1 \leq i \leq m\}$$

where $J = (j_1, j_2, \ldots, j_\ell)$. We also partition $\mathcal{S}$ into levels: for fixed ℓ,

$$\mathcal{S}_\ell = \{(j_1, \ldots, j_\ell; i) : 1 \leq j_1, \ldots, j_\ell \leq d, 1 \leq i \leq m\}$$

is the reunion of all the nodes represented by strings of length ℓ. The first, or root level, comprises the unique node

$$\mathcal{N}_\emptyset = \{i : 1 \leq i \leq m\}.$$

The possible transitions are as follows:

(1) within a node, from (J, i) to (J, i') with probability $(B_j)_{ii'}$ where j is the rightmost integer in J;
(2) within the root node, from i to i' with probability $(B_0)_{ii'}$;
(3) between a node and one of its children, from (J, i) to $(J + k, i')$ with probability $(A_k)_{ii'}$, for $1 \leq k \leq d$;
(4) between a node and its parent, from $(J + k, i)$ to (J, i') with probability $(D_k)_{ii'}$, for $1 \leq k \leq d$.

The matrices $B_0 + A_1 + \cdots + A_d$ and $B_i + D_i + A_1 + \cdots + A_d$, $i = 1, \ldots, d$, are stochastic.

With these assumptions, it is clear that the transition graph between *nodes* is a tree, which is why we call these Tree-Like processes. In their full generality, Tree-Like processes allow a fifth type of transitions:

(5a) between a node at level ℓ and any node in the levels 0 to ℓ, for all ℓ (Yeung and Sengupta [18])

or

(5b) between a node and any of its descendants (Takine, Sengupta and Yeung [14])

but we do not consider this type of transitions in the present paper. Instead, we make the further simplifying assumption that

$$B_1 = B_2 = \cdots = B_d = B;$$

that is, we assume that, except at the root node, transition probabilities *within* a node do not depend on the specific node.

The structure of the matrix generator of this Markov process depends on the ordering of the states. There are two natural ways of ordering the nodes: level by level or lexicographically. If we enumerate the nodes level by level, then we recognize that the Markov process has the structure of a non-homogeneous QBD and we may use the general approach given in Bright and Taylor [3,4] and in [10].

It is, however, more fruitful to use the lexicographical order: $\mathcal{N}_\emptyset$ comes first, then $\mathcal{N}_1$ and all the nodes $\mathcal{N}_J$ for which the leftmost integer of J is 1, followed by $\mathcal{N}_2$ and all the nodes $\mathcal{N}_J$ for which the leftmost integer of J is 2, and so on up to $\mathcal{N}_d$ and all its descendants. With this ordering, the generator $Q = P - I$ (where P is the transition matrix of the Markov chain) is (2).

Observe that in each row and in each column of W there is a finite number of nonzero entries. From this property it follows that the 1-norm of W is finite. In order to prove this, consider a vector $\boldsymbol{x} = (x_k)_{k\in\mathcal{S}}$ such that $||\boldsymbol{x}|| = \sum_{k\in\mathcal{S}} |x_k| = 1$, and partition it as $\boldsymbol{x} = (\boldsymbol{x}^{(0)}, \boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(d)})$, as described in (4). From the definition of 1-norm of $\boldsymbol{x}$ it follows that $1 = ||\boldsymbol{x}|| = \sum_{i=0}^{d} ||\boldsymbol{x}^{(i)}||$. Now let $\boldsymbol{y} = W\boldsymbol{x}$ and partition $\boldsymbol{y}$ as $\boldsymbol{y} = (\boldsymbol{y}^{(0)}, \boldsymbol{y}^{(1)}, \ldots, \boldsymbol{y}^{(d)})$, similarly to (4). It holds

$$\begin{aligned} \boldsymbol{y}^{(0)} &= C\boldsymbol{x}^{(0)} + \textstyle\sum_{i=1}^{d} A_i \boldsymbol{x}^{(i,0)} \\ \boldsymbol{y}^{(i)} &= W\boldsymbol{x}^{(i)} + V_i \boldsymbol{x}^{(0)}, \quad i = 1, \ldots, d \end{aligned} \tag{5}$$

where $\boldsymbol{x}^{(i,0)}$ is the first block in the partitioning (4) of $\boldsymbol{x}^{(i)}$. Taking the norms in both sides of (5) we get

$$\begin{aligned} ||\boldsymbol{y}^{(0)}|| &\le \gamma(||\boldsymbol{x}^{(0)}|| + \textstyle\sum_{i=1}^{d} ||\boldsymbol{x}^{(i,0)}||) \\ ||\boldsymbol{y}^{(i)}|| &\le \gamma||\boldsymbol{x}^{(0)}|| + ||W\boldsymbol{x}^{(i)}||, \quad i = 1, \ldots, d, \end{aligned}$$

where γ is a positive constant depending on C, D_i and A_i, $i = 1, \ldots, d$. Therefore, we have

$$||W\boldsymbol{x}|| = ||\boldsymbol{y}^{(0)}|| + \cdots + ||\boldsymbol{y}^{(d)}|| \le (d+1)\gamma||\boldsymbol{x}^{(0)}|| + \gamma\sum_{i=1}^{d} ||\boldsymbol{x}^{(i,0)}|| + \sum_{i=1}^{d} ||W\boldsymbol{x}^{(i)}||$$

which recursively yields

$$||W\boldsymbol{x}|| \le (d+2)\gamma(||\boldsymbol{x}^{(0)}|| + \sum_{i=1}^{d} ||\boldsymbol{x}^{(i,0)}|| + \sum_{i,j=1}^{d} ||\boldsymbol{x}^{(i,j,0)}|| + \cdots) = \gamma(d+2)$$

where $\boldsymbol{x}^{(i,j,0)}$ is the first block component of the j-th block component of $\boldsymbol{x}^{(i)}$ in the recursive block decomposition (4). The latter inequality implies that W is a linear operator on $\ell^1(\mathcal{S})$ with bounded 1-norm.

The theorem below has been proved by different means in [18,17,10]. In particular, the proof in [10] proceeds along purely probabilistic arguments. We provide here some remarks based on the recursive structure of W in (3).

Theorem 1 *Assume that the Tree-Like process with generator (2) is positive recurrent. Denote by $\boldsymbol{\pi}_J^{\mathrm{T}}$ the sub-vector of stationary probabilities associated with the states in node $\mathcal{N}_J$.*

One has that

$$\boldsymbol{\pi}_J^{\mathrm{T}} = \boldsymbol{\pi}_\emptyset^{\mathrm{T}} R_{j_1} \cdots R_{j_\ell} \tag{6}$$

if $J=(j_1,\ldots,j_\ell)$, where

$$R_i = A_i(-S)^{-1}, \qquad \text{for } 1\le i\le d,$$

and $S=T-I$, where $T_{kk'}$ is the probability of moving from the state (J,k) to the state (J,k') at a later time, without visiting the node $\mathcal{N}_J$ or its parent in between, independently of $J\neq\emptyset$. The matrix S is the minimal solution of the nonlinear matrix equation

$$X+\sum_{i=1}^{d} A_i X^{-1} D_i = C, \tag{7}$$

with respect to the componentwise ordering. The vector $\boldsymbol{\pi}_\emptyset^{\mathrm{T}}$ is the solution of

$$\boldsymbol{\pi}_\emptyset^{\mathrm{T}}\left[C_0+\sum_{i=1}^{m} A_i(-S)^{-1}D_i\right]=\mathbf{0} \tag{8}$$

normalized by

$$\boldsymbol{\pi}_\emptyset^{\mathrm{T}}\sum_{n\ge 0}\left[\sum_{1\le i\le d} R_i\right]^n \mathbf{1}=1.$$

A first remark related to the previous theorem comes from the UL factorization of W as stated by the following

Proposition 2 *Let S be a nonsingular matrix, then matrices L, U recursively defined by*

$$U=\begin{bmatrix} S & \Lambda_1 & \Lambda_2 & \ldots & \Lambda_d\\ 0 & U & 0 & \ldots & 0\\ 0 & 0 & U & \ddots & \vdots\\ \vdots & \vdots & \ddots & \ddots & 0\\ 0 & 0 & \ldots & 0 & U\end{bmatrix},\qquad L=\begin{bmatrix} I & 0 & 0 & \ldots & 0\\ Y_1 & L & 0 & \ldots & 0\\ Y_2 & 0 & L & \ddots & \vdots\\ \vdots & \vdots & \ddots & \ddots & 0\\ Y_d & 0 & \ldots & 0 & L\end{bmatrix},$$

with $Y_i=\begin{bmatrix} S^{-1}D_i\\ O\\ \vdots\end{bmatrix}$, for $i=1,\ldots,d$, represent linear operators in $\ell^1(\mathcal{S})$ with bounded 1-norm. Moreover, S is a solution of (7) if and only if $W=UL$.

PROOF. The boundness of the operators L and U can be proved by following the same argument used for proving the boundness of W. The remaining part of the proposition can be proved by direct inspection. □

The matrices L and U have formal inverses recursively defined by

$$L^{-1} = \begin{bmatrix} I & 0 & 0 & \dots & 0 \\ -L^{-1}Y_1 & L^{-1} & 0 & \dots & 0 \\ -L^{-1}Y_2 & 0 & L^{-1} & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ -L^{-1}Y_d & 0 & \dots & 0 & L^{-1} \end{bmatrix}$$

and

$$U^{-1} = \begin{bmatrix} S^{-1} & -S^{-1}\Lambda_1 U^{-1} & -S^{-1}\Lambda_2 U^{-1} & \dots & -S^{-1}\Lambda_d U^{-1} \\ 0 & U^{-1} & 0 & \dots & 0 \\ 0 & 0 & U^{-1} & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \dots & 0 & U^{-1} \end{bmatrix},$$

respectively. From the above relations it follows that L^{-1} and U^{-1} have finite entries, however, they may have an unbounded ℓ^1-norm even under the assumptions of stochasticity and positive recurrence of the Tree-like process.

Once the matrix S is known, the stationary probability vector can be computed by using the UL factorization of W. In order to show this, we rewrite as follows the matrix Q given in (2):

$$Q = \begin{bmatrix} C_0 & \Lambda_0 \\ V_0 & W_0 \end{bmatrix},$$

and we decompose $\boldsymbol{\pi}^{\mathrm{T}}$ as $\boldsymbol{\pi}^{\mathrm{T}} = (\boldsymbol{\pi}_{\emptyset}^{\mathrm{T}}, \boldsymbol{\pi}_{\star}^{\mathrm{T}})$, where $\boldsymbol{\pi}_{*}^{\mathrm{T}}$ is the stationary probability vector for all the nodes below the root. The equation $\boldsymbol{\pi}^{\mathrm{T}}Q = \mathbf{0}$ immediately leads to

$$\boldsymbol{\pi}_{\emptyset}^{\mathrm{T}} \left[C_0 + \Lambda_0(-W_0)^{-1}V_0\right] = \mathbf{0} \tag{9}$$
$$\boldsymbol{\pi}_{\star}^{\mathrm{T}} = \boldsymbol{\pi}_{\emptyset}^{\mathrm{T}}\Lambda_0(-W_0)^{-1}. \tag{10}$$

Next we decompose $\boldsymbol{\pi}_{\star}^{\mathrm{T}}$ as $\boldsymbol{\pi}_{\star}^{\mathrm{T}} = (\boldsymbol{\pi}_{1\star}^{\mathrm{T}}, \dots, \boldsymbol{\pi}_{d\star}^{\mathrm{T}})$ where $\boldsymbol{\pi}_{i\star}^{\mathrm{T}}$ is the stationary probability sub-vector of all the nodes $\mathcal{N}_J$ for which the string J begins with the integer i; we find from (10) that

$$\boldsymbol{\pi}_{i\star}^{\mathrm{T}} = \boldsymbol{\pi}_{\emptyset}^{\mathrm{T}}\Lambda_i(-W)^{-1} = \boldsymbol{\pi}_{\emptyset}^{\mathrm{T}}\Lambda_i L^{-1}(-U)^{-1},$$

that is, $\boldsymbol{\pi}_{i\star}^{\mathrm{T}} = \boldsymbol{\pi}_{\emptyset}^{\mathrm{T}} \Lambda_i (-U)^{-1}$ and we readily find that

$$\left[\boldsymbol{\pi}_i^{\mathrm{T}} \; \boldsymbol{\pi}_{i1\star}^{\mathrm{T}} \; \ldots \; \boldsymbol{\pi}_{id\star}^{\mathrm{T}} \right] = \boldsymbol{\pi}_{\emptyset}^{\mathrm{T}} R_i \left[I \; \Lambda_1(-U)^{-1} \; \ldots \; \Lambda_d(-U)^{-1} \right]$$

where $\boldsymbol{\pi}_{ij\star}^{\mathrm{T}}$ is the stationary probability vector for all the nodes $\mathcal{N}_J$ for which the string J begins with the integers i and j.

This shows that $\boldsymbol{\pi}_i^{\mathrm{T}} = \boldsymbol{\pi}_{\emptyset}^{\mathrm{T}} R_i$ for all i, and it suffices to iterate these calculations to prove (6) for all strings J. It is a simple matter to verify that (8) and (9) are identical equations.

We should emphasize that, although we deal with discrete-time Markov chains, nevertheless our results are directly adapted to continuous-time Markov processes; in that case, the matrices $B + D_i + A_1 + \cdots + A_d$ are conservative, stable matrices, and we set $C = B$ in (1).

3 Fixed point iterations

We define the matrices $G_i = (-S)^{-1} D_i$ which have the following interpretation: $(G_i)_{kk'}$ is the probability that, starting at time 0 from the state $(i;k)$ in $\mathcal{N}_i$, the Tree-Like process eventually moves to the root node and that (k') is the first state visited there. If the process is positive recurrent, then G_i is stochastic for all i.

With this definition, we may rewrite (7) as a system of coupled equations:

$$S = C + \sum_{1 \le i \le d} A_i G_i, \tag{11}$$

$$G_i = (-S)^{-1} D_i, \qquad \text{for } 1 \le i \le d, \tag{12}$$

from which we may obtain S by fixed point iterations, as stated in the following theorem (see Latouche and Ramaswami [10, Section 14.3] and Yeung and Alfa [17, Section 8] for the proof).

Theorem 3 *The sequences $\{S_n : n \ge 0\}$ and $\{G_{i,n} : n \ge 0\}$, $1 \le i \le d$, defined by*

$$S_n = C + \sum_{1 \le i \le d} A_i G_{i,n}, \tag{13}$$

$$G_{i,n+1} = (-S_n)^{-1} D_i, \qquad \textit{for } 1 \le i \le d,\ n \ge 0, \tag{14}$$

with $G_{1,0} = \ldots = G_{d,0} = 0$, monotonically converge to S and G_i, respectively.

It is useful to note that $G_{i,n}$ is the matrix of first passage probabilities from $\mathcal{N}_i$ to $\mathcal{N}_\emptyset$ in a truncated process where transitions are not allowed beyond the level $\mathcal{S}_n$. Furthermore, it is shown in [17] that S is minimal in the following sense (recall that $S = T - I$ where T is a sub-stochastic matrix).

Lemma 4 *The matrix $T = S + I$ is the minimal nonnegative solution of the equation*

$$X = B + \sum_{1 \le i \le d} A_i (I - X)^{-1} D_i.$$

Armed with these, we may define another converging sequence.

Theorem 5 *The sequences $\{\bar{S}_n : n \ge 0\}$ and $\{\bar{G}_{i,n} : n \ge 0\}$, $1 \le i \le d$, defined by (13, 14) with $\bar{G}_{1,0} = \ldots = \bar{G}_{d,0} = I$ converge to S and G_i respectively. Moreover, $\bar{G}_{i,n}$ is stochastic for all i and n.*

PROOF. The proof proceeds along the same lines as the proof of [10, Theorem 8.3.1] and we briefly outline it here.

First, we consider a Tree-Like process on the finite set of levels $\mathcal{N}_\emptyset \cup \mathcal{S}_1 \cup \cdots \cup \mathcal{S}_n$. We keep the same transition probabilities as in the original process except at the last level $\mathcal{S}_n$: if $\mathcal{N}_J$ is a leaf, that is, if J is of length $|J| = n$, it is not possible to move from $(J;k)$ to any of the states $(J + i; k')$, with $1 \le i \le d$; instead, the transition probability to (J, k') is now equal to $(B + A_1 + \cdots + A_d)_{kk'}$.

By repeating verbatim the argument on [10, Page 180] we show that the new Tree-Like process is irreducible, for every $n \ge 1$. Moreover, one readily verifies that $\bar{G}_{i,n}$ is the matrix of first passage probabilities from $\mathcal{N}_i$ to $\mathcal{N}_\emptyset$ in this new process, from which we conclude that $\bar{S}_n$ is nonsingular and that $\bar{G}_{i,n}$ is a stochastic matrix for all n and i.

Since the set of stochastic matrices of order m is compact, the sequence $\{(\bar{G}_{1,n}, \ldots, \bar{G}_{d,n}) : n \ge 0\}$ has at least one accumulation point. We denote by $(G_1^*, \ldots, G_d^*)$ one such point and we choose a subset of indices $\{n_1, n_2, \ldots\}$ such that $\lim_{t\to\infty} \bar{G}_{i,n_t} = G_i^*$. By Theorem 3 we have that $\lim_{t\to\infty} G_{i,n_t} = G_i$. Note that both G_i^* and G_i are stochastic matrices.

Furthermore, $\bar{G}_{i,0} \ge G_{i,0}$ and we easily show by induction that $\bar{G}_{i,n} \ge G_{i,n}$ for all n, so that $G_i^* \ge G_i$. Since $G_i^* \mathbf{1} = G_i \mathbf{1} = \mathbf{1}$, necessarily $G_i^* = G_i$ and the sequence $\{(\bar{G}_{1,n}, \ldots, \bar{G}_{d,n}) : n \ge 0\}$ has only one accumulation point, which proves the theorem. $\Box$

Both the sequences $\{S_n\}_n$ and $\{\bar{S}_n\}_n$ are linearly convergent, and each step of the fixed point iteration requires one $m \times m$ matrix inversion and $2d$ matrix products. The sequence $\{\bar{S}_n\}_n$ converges faster than $\{S_n\}_n$ and we estimate the asymptotic rate of convergence in Theorem 7 below. Before doing so, however, we need to introduce the following notations: $\mathrm{vec}(A)$ is the mn-dimensional vector obtained by arranging column-wise the entries of the $m\times n$ matrix A; $C = A\otimes B$ is the matrix having block entries $C_{i,j} = a_{i,j}B$, where $A = (a_{i,j})_{i,j}$. We make use of the fact that $Y = AXB$ if and only if $\mathrm{vec}(Y) = (B^{\mathrm{T}}\otimes A)\mathrm{vec}(X)$, where A, B, X, Y are matrices of compatible size.

The following result relates the error at two subsequent steps:

Theorem 6 *Let $\{S_n : n \geq 0\}$ be the sequence defined by* (13)-(14)*, with arbitrary initial approximations $G_{i,0}$, $i = 1, \dots, d$, and define $\boldsymbol{e}_n = \mathrm{vec}(E_n)$, where $E_n = S - S_n$. The following relation holds:*

$$\boldsymbol{e}_{n+1} = \left(\sum_{1\leq i\leq d} ((-S_n)^{-1}D_i)^{\mathrm{T}} \otimes R_i\right) \boldsymbol{e}_n, \quad n \geq 0. \tag{15}$$

PROOF. By subtracting (13) from (11) and by observing that

$$G_i - G_{i,n+1} = S^{-1}E_nS_n^{-1}D_i, \quad i = 1, \dots, d, \quad n \geq 0,$$

we obtain that

$$E_{n+1} = \sum_{1\leq i\leq d} R_iE_n(-S_n)^{-1}D_i, \quad n \geq 0,$$

from which (15) immediately follows. $\square$

In the following, $\rho(A)$ denotes the spectral radius of the matrix A and $\lambda(A)$ denotes the set of its eigenvalues. From Theorem 6 we derive the following estimate of the mean asymptotic rate of convergence:

Theorem 7 *The eigenvalues of the matrix $R = \sum_{1\leq i\leq d} R_i$ are such that*

$$\lambda(R) \subset \lambda\left(\sum_{1\leq i\leq d} G_i^{\mathrm{T}} \otimes R_i\right).$$

Moreover, if R has a positive left eigenvector, corresponding to its spectral radius, then

$$\rho(R) = \rho\left(\sum_{1\leq i\leq d} G_i^{\mathrm{T}} \otimes R_i\right)$$

and the sequences $\{S_n : n \geq 0\}$, $\{\bar{S}_n : n \geq 0\}$ defined in (13)-(14)*, obtained with $G_{1,0} = \ldots = G_{d,0} = 0$ and $\bar{G}_{1,0} = \ldots = \bar{G}_{d,0} = I$, respectively, are such that*

$$\lim_{n\to\infty} \| \boldsymbol{e}_n \|^{1/n} = \rho(R), \tag{16}$$

$$\lim_{n\to\infty} \| \bar{\boldsymbol{e}}_n \|^{1/n} \leq \max \left\{ |\sigma| : \sigma \in \lambda \left(\sum_{1\leq i\leq d} G_i^{\mathrm{T}} \otimes R_i \right) \setminus \lambda(R) \right\} \tag{17}$$

for any vector norm $||\cdot||$, *where* $\boldsymbol{e}_n = \mathrm{vec}(S - S_n)$, $\bar{\boldsymbol{e}}_n = \mathrm{vec}(S - \bar{S}_n)$.

PROOF. We proceed in a manner similar to Meini [11] and define the orthogonal $m^2 \times m^2$ matrix $\Pi = [\Pi_1|\Pi_2]$, with $\Pi_1 = (1/\sqrt{m})\mathbf{1} \otimes I \in \mathbb{R}^{m^2\times m}$, and $\Pi_2 \in \mathbb{R}^{m^2\times(m^2-m)}$. Since $\mathbf{1}^{\mathrm{T}} G_i^{\mathrm{T}} = \mathbf{1}^{\mathrm{T}}$, we have that

$$\Pi_1^{\mathrm{T}} \left(\sum_{1\leq i\leq d} G_i^{\mathrm{T}} \otimes R_i \right) \Pi_2 = 0$$

and

$$\Pi_1^{\mathrm{T}} \left(\sum_{1\leq i\leq d} G_i^{\mathrm{T}} \otimes R_i \right) \Pi_1 = R,$$

that is,

$$\Pi^{\mathrm{T}} \left(\sum_{1\leq i\leq d} G_i^{\mathrm{T}} \otimes R_i \right) \Pi = \begin{bmatrix} R & 0 \\ H & K \end{bmatrix}, \tag{18}$$

where $H = \Pi_2^{\mathrm{T}} \left(\sum_{1\leq i\leq d} G_i^{\mathrm{T}} \otimes R_i \right) \Pi_1$ and $K = \Pi_2^{\mathrm{T}} \left(\sum_{1\leq i\leq d} G_i^{\mathrm{T}} \otimes R_i \right) \Pi_2$. Thus, it is clear that $\lambda(R) \subset \lambda \left(\sum_{1\leq i\leq d} G_i^{\mathrm{T}} \otimes R_i \right)$.

Let $\boldsymbol{v}^{\mathrm{T}} > 0$ be a left eigenvector of R corresponding to $\rho(R)$. Since G_i, $i = 1, \ldots, d$, are stochastic, it follows that

$$(\mathbf{1}^{\mathrm{T}} \otimes \boldsymbol{v}^{\mathrm{T}}) \left(\sum_{1\leq i\leq d} G_i^{\mathrm{T}} \otimes R_i \right) = \rho(R)(\mathbf{1}^{\mathrm{T}} \otimes \boldsymbol{v}^{\mathrm{T}}), \tag{19}$$

that is, $\mathbf{1}^{\mathrm{T}} \otimes \boldsymbol{v}^{\mathrm{T}}$ is a left eigenvector corresponding to the eigenvalue $\rho(R)$. Since $\sum_{1\leq i\leq d} G_i^{\mathrm{T}} \otimes R_i$ is a nonnegative matrix and $\mathbf{1}^{\mathrm{T}} \otimes \boldsymbol{v}^{\mathrm{T}}$ is a positive vector, it follows that $\rho \left(\sum_{1\leq i\leq d} G_i^{\mathrm{T}} \otimes R_i \right) = \rho(R)$ (see [16]).

The relation (16) is proved as follows. Since $\boldsymbol{v}$ has positive components, we may define the vector norm $||\cdot||$ such that $||\boldsymbol{x}|| = (\mathbf{1}^{\mathrm{T}} \otimes \boldsymbol{v}^{\mathrm{T}})|\boldsymbol{x}|$, where $\boldsymbol{x} \in \mathbb{R}^{m^2}$ and $|\boldsymbol{x}| = (|x_i|)_i$. Since the vectors $\boldsymbol{e}_n$ are nonnegative and since the sequences $\{(-S_n)^{-1} D_i\}_n$ converge monotonically to G_i, we obtain from (15) that

$$\begin{aligned} \| \boldsymbol{e}_n \| = (\mathbf{1}^{\mathrm{T}} \otimes \boldsymbol{v}^{\mathrm{T}}) \boldsymbol{e}_n &\leq (\mathbf{1}^{\mathrm{T}} \otimes \boldsymbol{v}^{\mathrm{T}}) \left(\sum_{1\leq i\leq d} G_i^{\mathrm{T}} \otimes R_i \right) \boldsymbol{e}_{n-1} \\ &= \rho(R)(\mathbf{1}^{\mathrm{T}} \otimes \boldsymbol{v}^{\mathrm{T}}) \boldsymbol{e}_{n-1} = \rho(R) ||\boldsymbol{e}_{n-1}||. \end{aligned}$$

Hence $\lim_{n\to\infty} \| \boldsymbol{e}_n \|^{1/n} \leq \rho(R)$. Let us now show the opposite inequality. From (15) one has

$$\begin{aligned}
&\| \boldsymbol{e}_n \| = (\mathbf{1}^{\mathrm{T}} \otimes \boldsymbol{v}^{\mathrm{T}})\boldsymbol{e}_n = \\
&(\mathbf{1}^{\mathrm{T}} \otimes \boldsymbol{v}^{\mathrm{T}}) \left(\left(\sum_{1\leq i\leq d} G_i^{\mathrm{T}} \otimes R_i \right) - \left(\sum_{1\leq i\leq d} (G_i - (-S_{n-1})^{-1} D_i)^{\mathrm{T}} \otimes R_i \right) \right) \boldsymbol{e}_{n-1} = \\
&\rho(R) ||\boldsymbol{e}_{n-1}|| - (\mathbf{1}^{\mathrm{T}} \otimes \boldsymbol{v}^{\mathrm{T}}) \left(\sum_{1\leq i\leq d} (G_i - (-S_{n-1})^{-1} D_i)^{\mathrm{T}} \otimes R_i \right) \boldsymbol{e}_{n-1}.
\end{aligned}$$

Since the sequences $\{(-S_n)^{-1} D_i\}_n$ converge monotonically to G_i, for any $\epsilon > 0$ there exists an integer n_0 such that

$$G_i - (-S_n)^{-1} D_i \leq \epsilon(\mathbf{1}\ \mathbf{1}^{\mathrm{T}})/m$$

for any $i = 1, \ldots, d$ and for any $n \geq n_0$. Thus, we obtain that for any $n \geq n_0$

$$\begin{aligned}
&\| \boldsymbol{e}_n \| \geq \rho(R) ||\boldsymbol{e}_{n-1}|| - (\epsilon/m)(\mathbf{1}^{\mathrm{T}} \otimes \boldsymbol{v}^{\mathrm{T}}) \left(\sum_{1\leq i\leq d} (\mathbf{1}\ \mathbf{1}^{\mathrm{T}}) \otimes R_i \right) \boldsymbol{e}_{n-1} = \\
&(1-\epsilon)\rho(R) ||\boldsymbol{e}_{n-1}||,
\end{aligned}$$

so that $\lim_{n\to\infty} \| \boldsymbol{e}_n \|^{1/n} \geq \rho(R)(1-\epsilon)$. Since ϵ is arbitrary, we deduce that $\lim_{n\to\infty} \| \boldsymbol{e}_n \|^{1/n} = \rho(R)$, and such equality holds for any vector norm for the equivalence of the norms.

We finally prove (17). Since the matrices $\bar{G}_{i,n}$ are stochastic for any i and n, it can be easily shown by induction on n that $\Pi_1^{\mathrm{T}} \boldsymbol{e}_n = \mathbf{0}$ for all n, and thus $\Pi^{\mathrm{T}} \boldsymbol{e}_n = \left[\begin{array}{c} \mathbf{0} \\ \hline \boldsymbol{h}_n \end{array} \right]$, where $\boldsymbol{h}_n = \Pi_2^{\mathrm{T}} \boldsymbol{e}_n$. Since

$$\begin{aligned}
\Pi^{\mathrm{T}} \boldsymbol{e}_n &= \Pi^{\mathrm{T}} \left(\sum_{1\leq i\leq d} ((-S_{n-1})^{-1} D_i)^{\mathrm{T}} \otimes R_i \right) \Pi (\Pi^{\mathrm{T}} \boldsymbol{e}_{n-1}) \\
&= \begin{bmatrix} R & 0 \\ H_{n-1} & K_{n-1} \end{bmatrix} (\Pi^{\mathrm{T}} \boldsymbol{e}_{n-1}),
\end{aligned}$$

where

$$H_n = \Pi_2^{\mathrm{T}} \left(\sum_{1\leq i\leq d} ((-S_n)^{-1} D_i)^{\mathrm{T}} \otimes R_i \right) \Pi_1$$

and

$$K_n = \Pi_2^{\mathrm{T}} \left(\sum_{1 \le i \le d} ((-S_n)^{-1} D_i)^{\mathrm{T}} \otimes R_i \right) \Pi_2,$$

one has $\boldsymbol{h}_n = K_{n-1} \boldsymbol{h}_{n-1}$, from which we deduce (17) by applying the same techniques as in [11]. $\square$

From the above theorem it follows that the convergence of the sequence obtained with $G_{i,0} = I$ is generally faster than the one obtained with $G_{i,0} = 0$, when the Perron-Frobenius eigenvector $\boldsymbol{v}^{\mathrm{T}}$ of R is strictly positive.

Remark 8 *The property that the Perron-Frobenius eigenvector $\boldsymbol{v}^{\mathrm{T}}$ of R is strictly positive is related to irreducibility characteristics of the stochastic process under study and does not seem to be a very restrictive assumption. For instance, for $d = 1$, if $A_1 + B + D_1$ is irreducible then $\boldsymbol{v}^{\mathrm{T}} > 0$ (Neuts [12, Lemma 1.3.2]). For a general Tree-Like process, such simple sufficient conditions are not as readily available.*

4 Reduction to quadratic equations

By writing (12) as $D_i + SG_i = 0$ and replacing S by the right-hand side of (11), we obtain the system

$$D_i + (C + \sum_{\substack{1 \le j \le d \\ j \ne i}} A_j G_j) G_i + A_i G_i^2 = 0, \tag{20}$$

for $i = 1, \ldots, d$. If we define F_i as

$$F_i = C + \sum_{\substack{1 \le j \le d \\ j \ne i}} A_j G_j, \tag{21}$$

then (20) becomes

$$D_i + F_i G_i + A_i G_i^2 = 0, \quad 1 \le i \le d. \tag{22}$$

For each i, we recognize this as the equation which defines the matrix of first passage probabilities to lower levels in a QBD characterized by the transition matrices A_i, F_i and D_i. Those quadratic equations may be solved by the cyclic reduction method (Bini and Meini [1,2] and Latouche and Ramaswami [9,10]) which converges quadratically. Thus, we may determine the matrices G_i, $1 \le i \le d$, and the matrix S by the following iterative procedure: we define sequences of matrices $\{G_{i,n} : 1 \le i \le d, n \ge 0\}$ such that, for $n \ge 1$ and for each i, $G_{i,n}$ is the minimal nonnegative solution of the matrix quadratic equation

$$D_i + F_{i,n} G_{i,n} + A_i G_{i,n}^2 = 0, \tag{23}$$

where

$$F_{i,n} = C + \sum_{1 \le j \le i-1} A_j G_{j,n} + \sum_{i+1 \le j \le d} A_j G_{j,n-1}. \tag{24}$$

Theorem 9 *The sequences $\{G'_{i,n} : n \ge 0\}$ obtained from (23, 24), starting with $G'_{1,0} = G'_{2,0} = \cdots = G'_{d,0} = 0$, monotonically converge to G_i, for $1 \le i \le d$. The sequences $\{\tilde{G}_{i,n} : n \ge 0\}$ obtained from (23, 24), starting with $\tilde{G}_{1,0} = \tilde{G}_{2,0} = \cdots = \tilde{G}_{d,0} = I$, converge to G_i, for $1 \le i \le d$. Moreover, $\tilde{G}_{i,n}$ is stochastic for all i and n.*

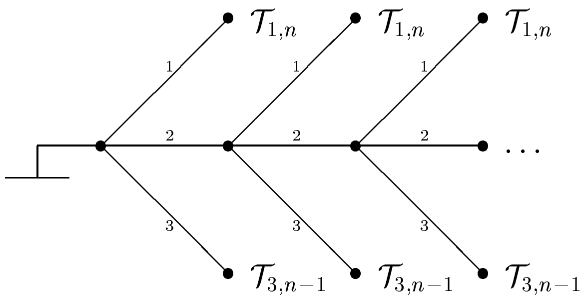

Fig. 1. A pictorial representation of the tree $\mathcal{T}_{2,n}$ for $d = 3$.

PROOF. We define a sequence of trees $\{\mathcal{T}_{i,n} : n \ge 0, 1 \le i \le d\}$ as follows (see Figure 1 for a graphical description). The trees $\mathcal{T}_{i,0}$ comprise a single node. For $n \ge 1$, the root of $\mathcal{T}_{i,n}$ has d children; each child with index $j \le i-1$ is the root of a tree which is isomorphic to $\mathcal{T}_{j,n}$; each child with index $j \ge i+1$ is the root of a tree which is isomorphic to $\mathcal{T}_{j,n-1}$; the child with index i is the root of a tree isomorphic to $\mathcal{T}_{i,n}$ itself. The transition probabilities are the same as in the original Tree-Like process, except at the leaf nodes where two cases are considered.

In the first case, for every node $\mathcal{N}_J$ which has no child, the transition probability from (J,k) to (J,k') is $B_{kk'}$. One proves by induction that for all $n \ge 1$ and all i, $G'_{i,n}$ is the matrix of first passage probabilities from $\mathcal{N}_i$ to $\mathcal{N}_\emptyset$ in $\mathcal{T}_{i,n}$. The detailed proof is similar to that of Theorem 3.

In the second case, for every node $\mathcal{N}_J$ which has no child, the transition probability from (J,k) to (J,k') is $(B + A_1 + \cdots + A_d)_{kk'}$. One proves by induction that $\tilde{G}_{i,n}$ is the matrix of first passage probabilities from $\mathcal{N}_i$ to $\mathcal{N}_\emptyset$ in $\mathcal{T}_{i,n}$. Then we repeat the argument of Theorem 5. $\square$

Let us now analyze the convergence speed of the sequences $\{\tilde{G}_{i,n}\}_n$ and $\{G'_{i,n}\}_n$. Define the matrices $E_{i,n} = G_i - G_{i,n}$, $i = 1, \ldots, d$, representing the errors at step n, and the corresponding vectors $\boldsymbol{e}_{i,n} = \operatorname{vec}(E_{i,n})$. The errors at two subsequent steps are related as stated in the following theorem.

Theorem 10 *Let $\{G_{i,n}\}_n$, $i = 1, \ldots, d$, be the sequences generated by means of (23),(24), with arbitrary initial approximations $G_{i,0}$, $i = 1, \ldots, d$. Then, at each step n we have that*

$$\boldsymbol{e}_{i,n} - \sum_{1\le j\le i} \left(G_{i,n}^{\mathrm{T}} \otimes (-S)^{-1} A_j\right) \boldsymbol{e}_{j,n} = \sum_{i+1\le j\le d} \left(G_{i,n}^{\mathrm{T}} \otimes (-S)^{-1} A_j\right) \boldsymbol{e}_{j,n-1}, \quad (25)$$

for $i = 1, \ldots, d$.

PROOF. From (22, 23) we obtain that

$$F_i G_i - F_{i,n} G_{i,n} + A_i (G_i^2 - G_{i,n}^2) = 0$$

for all i and n. By replacing in the latter equation F_i and $F_{i,n}$ respectively by the expressions from (21) and (24), and $G_i^2 - G_{i,n}^2$ by $G_i E_{i,n} + E_{i,n} G_{i,n}$, we obtain that

$$\begin{aligned} CE_{i,n} + \sum_{1\le j\le i-1} A_j (G_j G_i - G_{j,n} G_{i,n}) + \sum_{i+1\le j\le d} A_j (G_j G_i - G_{j,n-1} G_{i,n}) + \\ + A_i G_i E_{i,n} + A_i E_{i,n} G_{i,n} = 0, \end{aligned}$$

from which, using

$$G_j G_i - G_{j,n} G_{i,n} = G_j E_{i,n} + E_{j,n} G_{i,n}$$

and

$$G_j G_i - G_{j,n-1} G_{i,n} = G_j E_{i,n} + E_{j,n-1} G_{i,n},$$

we arrive at

$$\left(C + \sum_{1\le j\le d} A_j G_j\right) E_{i,n} + \sum_{1\le j\le i} A_j E_{j,n} G_{i,n} + \sum_{i+1\le j\le d} A_j E_{j,n-1} G_{i,n} = 0.$$

We obtain (25) after multiplying this equation on the left by S^{-1}. $\square$

The equation (25) may also be written as

$$\boldsymbol{f}_n = (I - H_n)^{-1} K_n \boldsymbol{f}_{n-1} \quad (26)$$

where $\boldsymbol{f}_n$ is the dm^2-dimensional vector made up by the vectors $\boldsymbol{e}_{i,n}$ ($i = 1, \ldots, d$), H_n is the $d \times d$ block lower triangular matrix with block entries

$$(H_n)_{i,j} = \begin{cases} G_{i,n}^{\mathrm{T}} \otimes \tilde{R}_j & i = 1, \ldots, d, \quad j = 1, \ldots, i \\ 0 & \text{otherwise,} \end{cases} \quad (27)$$

K_n is the $d \times d$ block strictly upper triangular matrix with block entries

$$(K_n)_{i,j} = \begin{cases} G_{i,n}^{\mathrm{T}} \otimes \tilde{R}_j & i = 1, \dots, d, \quad j = i+1, \dots, d \\ 0 & \text{otherwise}, \end{cases} \tag{28}$$

and $\tilde{R}_i = (-S)^{-1} A_i = S^{-1} R_i S$, $1 \leq i \leq d$.

Since $H_n + K_n = D_{1,n} Z D_{2,n}$ where

$$\begin{aligned} D_{1,n} &= \mathrm{Diag}\left(G_{1,n}^{\mathrm{T}} \otimes I, G_{2,n}^{\mathrm{T}} \otimes I, \dots, G_{d,n}^{\mathrm{T}} \otimes I\right), \\ D_{2,n} &= \mathrm{Diag}\left(I \otimes \tilde{R}_1, I \otimes \tilde{R}_2, \dots, I \otimes \tilde{R}_d\right), \end{aligned}$$

and

$$Z = \begin{bmatrix} I \\ I \\ \vdots \\ I \end{bmatrix} \begin{bmatrix} I \; I \; \dots \; I \end{bmatrix},$$

we conclude that $H_n + K_n$ has $(d-1)m^2$ eigenvalues equal to 0, and the remaining m^2 eigenvalues are the eigenvalues of $\sum_{1 \leq i \leq d} G_{i,n}^{\mathrm{T}} \otimes \tilde{R}_i$. Since $R_i = S \tilde{R}_i S^{-1}$, these are the eigenvalues of $\sum_{1 \leq i \leq d} G_{i,n}^{\mathrm{T}} \otimes R_i$.

Let H and K respectively be the $d \times d$ block lower and strictly upper triangular matrices with block entries

$$(H)_{i,j} = \begin{cases} G_i^{\mathrm{T}} \otimes \tilde{R}_j & i = 1, \dots, d, \quad j = 1, \dots, i \\ 0 & \text{otherwise}, \end{cases} \tag{29}$$

$$(K)_{i,j} = \begin{cases} G_i^{\mathrm{T}} \otimes \tilde{R}_j & i = 1, \dots, d, \quad j = i+1, \dots, d \\ 0 & \text{otherwise}. \end{cases} \tag{30}$$

The matrices H and K are the limits of $\{H_n\}_n$ and $\{K_n\}_n$, respectively, when $G_{i,0}$, $i = 1, \dots, d$ are the null matrices, or the identity matrices. As for $H_n + K_n$, $H + K$ has $(d-1)m^2$ eigenvalues equal to 0, and the remaining m^2 eigenvalues are the eigenvalues of $\sum_{1 \leq i \leq d} G_i^{\mathrm{T}} \otimes R_i$. By Theorem 7, we have $\lambda(R) \subset \lambda(H + K)$ and, if the Perron-Frobenius eigenvector of R is strictly positive, then $\rho(H+K) = \rho(R) < 1$. Since $H + K$ is nonnegative, $M = I - H - K$ is a nonsingular M-matrix and $N = (I - H)^{-1} K$ is the iteration matrix obtained by means of a regular splitting applied to M. From the Perron-Frobenius theory (see [16]) it follows that $\rho(N) \leq \rho(H + K)$, and thus $\rho(N) \leq \rho(R)$. This inequality allows us to prove that the iteration defined by (23, 24) with

$G'_{i,0} = 0$ is generally faster than the iteration (13, 14) starting with $G_{i,0} = 0$, $i = 1, \ldots, d$. This result is reported in the next theorem which shows also that the sequences $\{\tilde{G}_{i,n}\}_n$, $i = 1, \ldots, d$, converge faster than the sequences $\{G'_{i,n}\}_n$, $i = 1, \ldots, d$.

Theorem 11 *One has*

$$\lim_{n\to\infty} ||\boldsymbol{f}'_n||^{1/n} = \rho((I-H)^{-1}K)$$

for any vector norm $\|\cdot\|$, and

$$\lambda\left((I-\tilde{H})^{-1}\tilde{K}\right) \subset \lambda\left((I-H)^{-1}K\right),$$

where $\boldsymbol{f}'_n$ is the dm^2-dimensional vector made up by the vectors $\boldsymbol{e}'_{i,n} = \mathrm{vec}(G_i - G'_{i,n})$, $i = 1, \ldots, d$,

$$\tilde{H} = \begin{bmatrix} \tilde{R}_1 & & & 0 \\ \tilde{R}_1 & \tilde{R}_2 & & \\ \vdots & \vdots & \ddots & \\ \tilde{R}_1 & \tilde{R}_2 & \ldots & \tilde{R}_d \end{bmatrix} \quad and \quad \tilde{K} = \begin{bmatrix} 0 & \tilde{R}_2 & \tilde{R}_3 & \ldots & \tilde{R}_d \\ & 0 & \tilde{R}_3 & \ldots & \tilde{R}_d \\ & & \ddots & \ddots & \vdots \\ & & & 0 & \tilde{R}_d \\ 0 & & & & 0 \end{bmatrix}.$$

Moreover, if the Perron-Frobenius eigenvector of $(I-\tilde{H})^{-1}\tilde{K}$ is strictly positive, then one has that

$$\rho\left((I-\tilde{H})^{-1}\tilde{K}\right) = \rho\left((I-H)^{-1}K\right)$$

and

$$\lim_{n\to\infty} \|\tilde{\boldsymbol{f}}_n\|^{1/n} \le \max\left\{|\sigma| : \sigma \in \lambda\left((I-H)^{-1}K\right) \setminus \lambda\left((I-\tilde{H})^{-1}\tilde{K}\right)\right\}$$

for any vector norm $\|\cdot\|$, where $\tilde{\boldsymbol{f}}_n$ is the dm^2-dimensional vector made up by the vectors $\tilde{\boldsymbol{e}}_{i,n} = \mathrm{vec}(G_i - \tilde{G}_{i,n})$, $i = 1, \ldots, d$.

PROOF. Concerning the convergence of $\{\boldsymbol{f}'_n\}_n$, from the monotonicity of the sequences $\{G'_{i,n}\}_n$, from (27), (28) and (26) we deduce that

$$\boldsymbol{f}'_n \le (I-H)^{-1}K\boldsymbol{f}'_{n-1} \le ((I-H)^{-1}K)^n \boldsymbol{f}'_0$$

whence $\lim_{n\to\infty} ||\boldsymbol{f}'_n||^{1/n} \le \rho((I-H)^{-1}K)$. For the opposite inequality observe that

$$\boldsymbol{f}'_n = (I-H)^{-1}K\boldsymbol{f}'_{n-1} - ((I-H)^{-1}K - (I-H_n)^{-1}K_n)\boldsymbol{f}'_{n-1}. \tag{31}$$

Let E be any nonnegative matrix such that $(I-H)^{-1}K-\epsilon E \geq 0$ for any $\epsilon > 0$ in a suitable neighborhood $\mathcal{U}$ of 0. Then for any positive $\epsilon \in \mathcal{U}$ there exists n_0 such that for any $n \geq n_0$ we have $(I-H)^{-1}K-(I-H_n)^{-1}K_n \leq \epsilon E$. From the latter inequality and from (31) we deduce that $\boldsymbol{f}'_n \geq ((I-H)^{-1}K-\epsilon E)\boldsymbol{f}'_{n-1} \geq ((I-H)^{-1}K-\epsilon E)^n \boldsymbol{f}'_0$, hence $\lim_{n\to\infty} ||\boldsymbol{f}'_n||^{1/n} \geq \rho((I-H)^{-1}K-\epsilon E)$. For the arbitrariness of ϵ we obtain $\lim_{n\to\infty} ||\boldsymbol{f}'_n||^{1/n} = \rho((I-H)^{-1}K)$. Concerning the remaining part of the theorem, let us define the $(dm^2)\times(dm^2)$ orthogonal matrix $\tilde{\Pi} = [\tilde{\Pi}_1|\tilde{\Pi}_2]$, where $\tilde{\Pi}_1 = I_d \otimes \Pi_1$, $\tilde{\Pi}_2 = I_d \otimes \Pi_2$, I_d is the $d \times d$ identity matrix, and $\Pi = [\Pi_1|\Pi_2]$ is the $(m^2)\times(m^2)$ orthogonal matrix defined in the proof of Theorem 7. Since $\tilde{G}_i$ is stochastic for any i, from (29, 30) it follows that $\tilde{\Pi}_1^{\mathrm{T}} H = \tilde{H}\tilde{\Pi}_1^{\mathrm{T}}$, and that $\tilde{\Pi}_1^{\mathrm{T}} K = \tilde{K}\tilde{\Pi}_1^{\mathrm{T}}$. Thus, we have that

$$\tilde{\Pi}^{\mathrm{T}}(I-H)^{-1}K\tilde{\Pi} = \begin{bmatrix} (I-\tilde{H})^{-1}\tilde{K} & 0 \\ T_1 & T_2 \end{bmatrix},$$

where $T_1 = \tilde{\Pi}_2^{\mathrm{T}}(I-H)^{-1}K\tilde{\Pi}_1$ and $T_2 = \tilde{\Pi}_2^{\mathrm{T}}(I-H)^{-1}K\tilde{\Pi}_2$, and we conclude that $\lambda\left((I-\tilde{H})^{-1}\tilde{K}\right) \subset \lambda\left((I-H)^{-1}K\right)$. If $\boldsymbol{u}$ is a positive Perron-Frobenius right eigenvector of $(I-\tilde{H})^{-1}\tilde{K}$, that is, $(I-\tilde{H})^{-1}\tilde{K}\boldsymbol{u} = \rho\boldsymbol{u}$ then $(I-H)^{-1}K\boldsymbol{v} = \rho\boldsymbol{v}$, where $\boldsymbol{v} = \tilde{\Pi}_1\boldsymbol{u}$. Since $\boldsymbol{v}$ is positive and $(I-H)^{-1}K$ is nonnegative, ρ is the spectral radius of $(I-H)^{-1}K$ [16]. Moreover, since $\tilde{G}_{i,n}$ is stochastic for any i and n, we have that $\tilde{\Pi}_1^{\mathrm{T}}\tilde{\boldsymbol{f}}_n = \boldsymbol{0}$ for any n. Thus, as in the proof of Theorem 7, $\lim_{n\to\infty} \|\tilde{\boldsymbol{f}}_n\|^{1/n} \leq \rho(T_2)$. $\square$

Unfortunately, the result above does not allow us to conclude that, when one starts with $G_{i,0} = I$, $i = 1, \ldots, d$, the iteration defined by (23, 24) is generally faster than the iteration (13, 14). Nevertheless, we conjecture that the procedure described in this section does require fewer iterations than the ones in Section 3. Our argument is illustrated in Figure 2 where we show how the tree fills up in the case where $d = 2$ during the first two iterations of (23, 24). We represent at the top the first 8 levels of a binary tree; underneath are the same levels for the trees $\mathcal{T}_{1,1}$, $\mathcal{T}_{2,1}$, $\mathcal{T}_{1,2}$ and $\mathcal{T}_{2,2}$. By contrast, only the first 2 levels are filled after the first two iterations of (13, 14).

Concerning the computational cost, the most expensive part at each iteration is the solution of the d quadratic equations (23), the computation of the coefficients $F_{i,n}$, $1 \leq i \leq d$, in (23) requires only d matrix products. The quadratic matrix equations can be solved by means of the cyclic reduction algorithm, which is quadratically convergent, and which costs one matrix inversion and 6 matrix products per step.

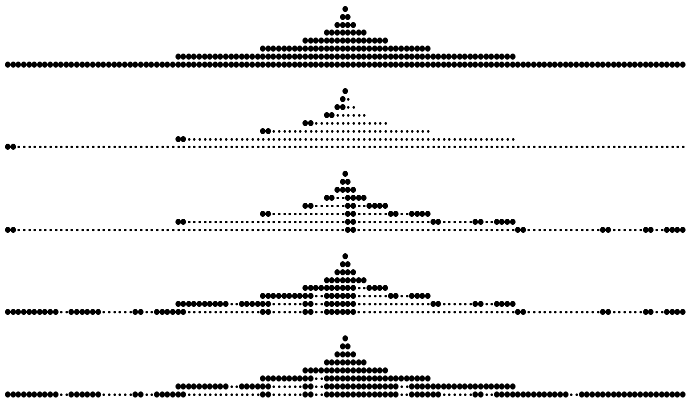

Fig. 2. Pictorial representation of the filling up of the nodes of the Tree-Like process, when $d = 2$. The first 8 levels are depicted, for the whole tree and for the trees $\mathcal{T}_{1,1}$, $\mathcal{T}_{2,1}$, $\mathcal{T}_{1,2}$ and $\mathcal{T}_{2,2}$.

5 Newton's iteration

Define the matrices

$$G = \begin{bmatrix} G_1 \\ \vdots \\ G_d \end{bmatrix}, \quad D = \begin{bmatrix} D_1 \\ \vdots \\ D_d \end{bmatrix}, \quad A = \begin{bmatrix} A_1 \cdots A_d \end{bmatrix},$$

and the four matrix operators

$$\begin{aligned}
\mathcal{I} &: \mathbb{R}^{md\times m} \to \mathbb{R}^{md\times m} : \mathcal{I}X = X, \\
\mathcal{M} &: \mathbb{R}^{md\times m} \to \mathbb{R}^{m\times m} \ \ : \mathcal{M}X = C + AX, \\
\mathcal{L} &: \mathbb{R}^{m\times m} \to \mathbb{R}^{md\times m} \ \ : \mathcal{L}V = [I_d \otimes (-V)^{-1}]D, \\
\mathcal{F} &: \mathbb{R}^{md\times m} \to \mathbb{R}^{md\times m} : \mathcal{F}X = (\mathcal{I} - \mathcal{L}\mathcal{M})X.
\end{aligned}$$

The first is the identity operator. The operator $\mathcal{M}$ is defined for matrices X in $< 0, G >$, that is, for matrices such that $0 \leq X \leq G$, and the operator $\mathcal{L}$ is defined for matrices V in $< C, S >$.

It is not difficult to verify that if $C \leq V \leq S$, then $0 \leq -C^{-1} \leq -V^{-1} \leq -S^{-1}$, so that $\mathcal{L}V$ is in $< 0, G >$; similarly, it is clear that if X is in $< 0, G >$, then $\mathcal{M}X$ is in $< C, S >$. We eventually conclude that G is the minimal nonnegative solution of $\mathcal{F}G = 0$.

Newton's method yields the sequence of matrices

$$\widehat{G}_{n+1} = \widehat{G}_n - \mathcal{F}'(\widehat{G}_n)^{-1}\mathcal{F}\widehat{G}_n, \tag{32}$$

with $\widehat{G}_0 = 0$. As we show below, the chain rule applies, so that

$$\mathcal{F}'(X) = \mathcal{I} - \mathcal{L}'(\mathcal{M}X)\mathcal{M}'(X)$$

with

$$\begin{aligned} \mathcal{M}'(X)H &= AH \\ \mathcal{L}'(V)K &= \left[I \otimes V^{-1}KV^{-1}\right] D. \end{aligned}$$

To see that the chain rule applies, one needs to repeat nearly verbatim the argument in Latouche [8]. We only repeat the salient steps here.

Lemma 12 *The operators $\mathcal{M}$ and $\mathcal{M}'$ are Lipschitz-continuous on $< 0, G >$ and $\mathcal{M}'(X)$ is the Fréchet derivative of $\mathcal{M}$. The norm of $\mathcal{M}'$ is uniformly bounded on $< 0, G >$.*

PROOF. The proof directly follows that of [8, Lemma 4.1]; it is much simpler here because $[\mathcal{M}'(X) - \mathcal{M}'(Y)]H \equiv 0$. $\square$

Lemma 13 *The operator $\mathcal{L}'$ is Lipschitz-continuous and uniformly bounded in $< C, S >$.*

PROOF. We use the l_∞-norm for matrices.

If V and W are in $< C, S >$ and if $\| K \| = 1$, then

$$\begin{aligned} &\| [\mathcal{L}'(V) - \mathcal{L}'(W)]K \| \\ &= \| [I \otimes V^{-1}KV^{-1} - I \otimes W^{-1}KW^{-1}]D \| \\ &\leq \| D \| \| V^{-1}KV^{-1} - W^{-1}KW^{-1} \| \\ &\leq \| D \| \, (\| (V^{-1} - W^{-1})KV^{-1} \| + \| W^{-1}K(V^{-1} - W^{-1}) \|) \\ &\leq \| D \| \| W^{-1} \| \| V^{-1} \| \| V - W \| \, (\| V^{-1} \| + \| W^{-1} \|) \\ &\leq 2 \, \| D \| \| S^{-1} \|^3 \| V - W \| \end{aligned}$$

since $0 \leq -V^{-1}, -W^{-1} \leq -S^{-1}$. The remainder of the proof is identical to that of [8, Lemma 4.2]. $\square$

Since both $\mathcal{M}'$ and $\mathcal{L}'$ are Fréchet-derivatives, the chain-rule applies, $\mathcal{F}'(X) = \mathcal{I} - \mathcal{L}'(\mathcal{M}X)\mathcal{M}'(X)$ is the Fréchet-derivative of $\mathcal{F}$ at X by [13, Proposition 3.1.7] and Newton's sequence (32) may be written as

$$\widehat{G}_{n+1} = \widehat{G}_n - Z_n \tag{33}$$

with

$$Z_n = \left[\mathcal{I} - \mathcal{L}'(\mathcal{M}\widehat{G}_n)\mathcal{M}'(\widehat{G}_n)\right]^{-1} (\mathcal{I} - \mathcal{L}\mathcal{M})\, \widehat{G}_n.$$

If we define

$$\widehat{S}_n = \mathcal{M}\widehat{G}_n = C + A\widehat{G}_n,$$

we may write that Z_n is the solution of

$$\left[\mathcal{I} - \mathcal{L}'(\widehat{S}_n)\mathcal{M}'(\widehat{G}_n)\right] Z_n = \widehat{G}_n - \mathcal{L}\mathcal{M}\widehat{G}_n$$

and this may be written after a few algebraic manipulations as

$$Z_n - \left[I \otimes \widehat{S}_n^{-1} A Z_n \widehat{S}_n^{-1}\right] D = \widehat{G}_n + \left[I \otimes \widehat{S}_n^{-1}\right] D. \tag{34}$$

If we premultiply this equation by A and if we define

$$Y_n = A Z_n, \tag{35}$$

we find that Y_n is a solution of the equation

$$Y_n + \sum_{1 \le i \le d} A_i \widehat{S}_n^{-1} Y_n (-\widehat{S}_n)^{-1} D_i = L_n, \tag{36}$$

where

$$L_n = \widehat{S}_n - C + \sum_{1 \le i \le d} A_i \widehat{S}_n^{-1} D_i.$$

Now, with $AZ_n = A\widehat{G}_n - A\widehat{G}_{n+1} = \widehat{S}_n - \widehat{S}_{n+1}$, we have

$$\widehat{S}_{n+1} = \widehat{S}_n - Y_n.$$

Finally, we have by (33, 34, 35) that

$$\widehat{G}_{n+1} = -\left[I \otimes \widehat{S}_n^{-1} Y_n \widehat{S}_n^{-1}\right] D - \left[I \otimes \widehat{S}_n^{-1}\right] D,$$

so that

$$\widehat{G}_{i,n+1} = \widehat{S}_n^{-1}(Y_n(-\widehat{S}_n)^{-1} D_i - D_i).$$

In summary, Newton's method generates $d+1$ sequences $\{\widehat{S}_n \;:\; n \ge 0\}$, $\{\widehat{G}_{i,n} \;:\; n \ge 0\}$, $i = 1, \ldots, d$, recursively defined by

$$\begin{aligned}
&\widehat{S}_{n+1} = \widehat{S}_n - Y_n, \\
&\widehat{G}_{i,n+1} = \widehat{S}_n^{-1}(Y_n(-\widehat{S}_n)^{-1} D_i - D_i),\ i = 1, \ldots, d,\ n \ge 0, \\
&\widehat{S}_0 = C,
\end{aligned}$$

where Y_n solves (36). The above sequences quadratically converge to S and G_i, $i = 1, \ldots, d$, respectively.

The applicability of this method relies on the efficient solution of the linear matrix equation (36). Indeed, the computation of L_n and of the matrix coefficients in (36) requires $3d$ matrix products, while the computation of $\widehat{S}_{n+1}$ and $\widehat{G}_{i,n+1}$, $1 \le i \le d$, requires $2d$ more matrix products.

Equation (36) constitutes a linear system of m^2 equations in m^2 unknowns. The customary solution algorithms like Gaussian elimination would cost $O(m^6)$ ops. There are efficient direct methods in the literature for solving matrix equations of the form $H_1YK_1 + H_2YK_2 = L$, which are known as Sylvester equations; these methods consist in performing a Hessenberg-Schur orthogonal transformation of the matrix coefficients H_1, H_2, K_1, K_2, and in solving a quasi-triangular linear system for the overall cost $O(m^3)$ (see [5]). Unfortunately, the more general case $\sum_{1\le i\le r} H_iYK_i = L$, where $r > 2$, cannot be solved with these methods.

To solve (36), we have applied two kinds of fixed point iterations. The first fixed point iteration consists in generating the sequence $\{Y_{n,h} : h \ge 0\}$ defined by

$$Y_{n,h+1} = L_n + \sum_{1\le i\le d} (A_i\widehat{S}_n^{-1})Y_{n,h}(\widehat{S}_n^{-1}D_i), \quad h \ge 0, \tag{37}$$

with $Y_{n,0} = L_n$. The second one generates the sequence $\{Y_{n,h} : h \ge 0\}$ defined by

$$Y_{n,h+1} - (A_r\widehat{S}_n^{-1})Y_{n,h+1}(\widehat{S}_n^{-1}D_r) = L_n + \sum_{\substack{1\le i\le d\\ i\ne r}} (A_i\widehat{S}_n^{-1})Y_{n,h}(\widehat{S}_n^{-1}D_i), \quad h \ge 0, \tag{38}$$

with $Y_{n,0} = L_n$, and where r is such that $1 \le r \le d$.

Concerning the choice of r we observe that (36) can be viewed as a linear system with matrix $I - B_n$, $B_n = \sum \hat{S}_n^{-1}D_i \otimes A_iS_n^{-1}$. If $B_n \ge 0$ and $\rho(B_n) < 1$ then $I - B_n$ is a nonsingular M-matrix, therefore, for the properties of M-matrices [16], the more convenient choice for r is the one for which $||A_r\widehat{S}_n^{-1}|| \cdot ||\widehat{S}^{-1}D_r||$ is maximum for a given norm $||\cdot||$.

From the numerical experiments it seems that the functional iteration (38) is generally faster than (37), but each step requires the solution of a Sylvester matrix equation. In order to solve it, we can apply the algorithm of [5], where the Hessenberg-Schur orthogonal transformation of the matrix coefficients is done once and for all. In principle, the convergence of the sequences (37) and (38) is not guaranteed; in fact, we have found an example where (38) diverges.

6 Numerical experimentation

We have implemented in Fortran 90 the fixed point iteration (FPI), the algorithm based on the reduction to quadratic equations (QE) and Newton's method (NM). In the QE method we have solved the quadratic matrix equations by using the cyclic reduction algorithm. Concerning NM, we have solved (36) by applying the fixed point iteration (38), with $r = 1$. The programs have been run on a Pentium III at 500 MHz.

We have tested the algorithms on the following problem.

We consider a system similar to the M/M/1 queue in a random environment (see [10, Example 9.2.2]). The service rate is constant and the arrival rate depends on the state of a Markovian environmental process. Each arrival has a label which indicates which direction is taken in the tree.

The parameters are as follows: the number of children is $d = 2$, the size of the blocks is m, $D_1 = \alpha I$, $D_2 = D_1$, $A_1 = \mathrm{Diag}(a, b, \ldots, b)$, $A_2 = bI$, $C = T - D_1 - A_1 - A_2$, $a = \rho f m$, $b = \rho(1-f)m/(2m-1)$, and $T = (t_{i,j})_{i,j=1,m}$, $t_{i,i} = -1$, for $i = 1, \ldots, m$, $t_{i,i+1} = 1$, for $i = 1, \ldots, m-1$, $t_{m,1} = 1$, $t_{i,j} = 0$ elsewhere.

This system is stable if ρ (the stationary arrival rate) is less than α (the service rate). The parameter f, which ranges from 0 to 1, measures the fraction of arrivals which occur in phase 1; when f is small, the system behaves nearly like an ordinary M/M/1 queue, when f is large, the arrivals are very bursty: in phase 1 the system is strongly driven to higher levels in the direction of the children labeled "1", in all the other phases the system is strongly driven toward the root of the tree. The burstiness is higher for large values of m.

We give in Table 1, for different values of the size m of the matrices, the number of iterations and the residual error $\| \widetilde{S} - C + A_1 \widetilde{S}^{-1} D_1 + A_2 \widetilde{S}^{-1} D_2 \|_1$, where $\widetilde{S}$ is the approximation provided by each of the algorithms. For the QE method we report in parenthesis the largest number of iterations needed by cyclic reduction to solve the matrix quadratic equations. For NM we report in parenthesis the largest number of iterations needed by the fixed point iteration (38). We observe that all the algorithms provide very accurate approximations. The QE and NM algorithms seem insensitive, in terms of number of iterations, to the growth of the size m; the number of iterations of FPI grows with m.

We report in Table 2 the number of iterations and the residual error, for different values of the parameter f. It is interesting to observe the different behavior of the number of iterations for the 3 methods. The number of iterations of FPI grows with f. The number of iterations of QE decreases, as f grows, and the number of inner iterations needed to solve the quadratic matrix equations is

	FPI		QE		NM	
m	Steps	Residual	Steps	Residual	Steps	Residual
10	125	5.7e-15	10 (11)	4.7e-16	10 (33)	3.5e-16
20	280	7.6e-15	10 (12)	1.0e-15	10 (29)	9.2e-16
30	459	6.9e-15	9 (12)	1.5e-15	10 (28)	9.1e-16
40	649	8.7e-15	9 (13)	1.5e-15	10 (28)	7.9e-16
50	860	9.1e-15	9 (13)	2.0e-15	10 (28)	9.0e-16
60	1065	8.9e-15	9 (13)	2.7e-15	10 (28)	8.9e-16
70	1279	9.0e-15	9 (14)	2.3e-15	10 (28)	6.8e-16
80	1494	9.8e-15	9 (14)	1.8e-15	10 (27)	7.9e-16
90	1693	8.7e-15	9 (14)	2.3e-15	10 (27)	7.9e-16
100	1927	1.0e-14	9 (14)	2.5e-15	11 (27)	8.9e-16

Table 1
$\alpha = 2,\ \rho = 1.8,\ f = 0.8$

	FPI		QE		NM	
f	Steps	Residual	Steps	Residual	Steps	Residual
0.1	269	7.8e-15	25 (10)	3.6e-15	10 (101)	9.3e-16
0.2	445	8.0e-15	21 (11)	2.8e-15	10 (82)	5.4e-16
0.3	643	9.5e-15	17 (12)	2.6e-15	10 (69)	4.2e-16
0.4	873	9.1e-15	15 (12)	2.2e-15	10 (60)	3.7e-16
0.5	1133	9.7e-15	13 (13)	3.4e-15	10 (53)	5.8e-16
0.6	1379	9.0e-15	11 (13)	3.5e-15	10 (46)	6.7e-16
0.7	1664	9.7e-15	10 (14)	2.3e-15	10 (38)	1.0e-15
0.8	1927	1.0e-14	9 (14)	2.5e-15	10 (27)	8.9e-16
0.9	2153	1.0e-14	8 (15)	1.8e-15	10 (21)	8.9e-16
0.99	2411	9.5e-15	6 (15)	1.5e-15	11 (9)	1.0e-15

Table 2
$\alpha = 2,\ \rho = 1.8,\ m = 100$

almost constant. The number of iterations of NM is almost constant, as f grows, while the number of inner iterations of (38) decreases.

We give in Table 3 the CPU time for different values of the size m and of the parameter f. In the QE and NM algorithms the CPU time strongly depends on the number of inner iterations to solve the quadratic equations (23) and

$f = 0.8$				$m = 100$			
m	FPI	QE	NM	f	FPI	QE	NM
10	0.05	0.04	0.1	0.1	16	42	296
20	0.2	0.2	0.9	0.2	27	36	263
30	0.8	0.5	3	0.3	38	31	238
40	2	1	7	0.4	52	28	212
50	6	2	13	0.5	68	24	189
60	14	4	22	0.6	82	20	164
70	27	6	34	0.7	100	20	150
80	45	9	50	0.8	117	18	105
90	75	13	72	0.9	128	16	95
100	117	18	105	0.99	145	11	46

Table 3
$\alpha = 2$, $\rho = 1.8$, CPU time (sec.)

the linear equation (36), respectively.

In Figure 3, we represent the CPU time in function of the parameter f. It is interesting to observe the opposite monotonicity of FPI, with respect to QE and NM (which have the same behavior). The algorithm FPI is convenient when the number of iterations is not too large; otherwise QE is more convenient. QE and NM have the same behavior in terms of monotonicity.

7 Conclusions

We have compared three different algorithms for solving the nonlinear matrix equation (7). The first algorithm (FPI) is described in Section 3; it is the natural fixed point iteration and it is already used in [10,17]. The second algorithm (QE) is proposed in Section 4, it requires at each step the solution of d quadratic matrix equations. Finally, the last algorithm (NM), proposed in Section 5, consists in applying Newton's scheme.

The procedure FPI is linearly convergent and has a low computational cost per step; unfortunately, for difficult problems, the convergence is so slow that the algorithm is not applicable. The QE method generates a sequence that still converges linearly to the solution, but the convergence is generally faster than FPI; moreover the quadratic matrix equations can be solved by using efficient algorithms, like cyclic reduction.

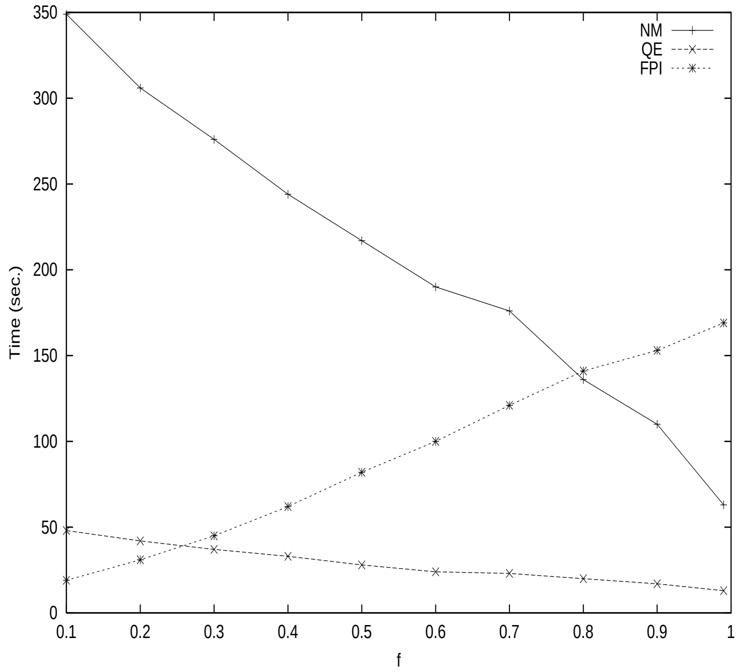

Fig. 3. $\alpha = 2$, $\rho = 1.8$, $m = 100$, CPU time

The NM algorithm has a quadratic convergence; the drawback is that it requires at each step the solution of the generalized Sylvester equation (36); in order to solve that equation we have applied two different fixed point iterations, which could converge slowly, or might even diverge. The study of effective algorithms for the solution of that equation is under investigation.

Acknowledgements

Guy Latouche is grateful to the Italian "Istituto Nazionale di Alta Matematica" for its financial support, which permitted him to visit the University of Pisa in February, 2000. The authors wish to thank two anonymous referees whose suggestions allowed to improve the presentation of the paper and to simplify the proof of Lemma 13.

References

[1] D. A. Bini and B. Meini. On the solution of a nonlinear matrix equation arising in queueing problems. *SIAM J. Matrix Anal. Appl.*, 17:906–926, 1996.

[2] D. A. Bini and B. Meini. Improved cyclic reduction for solving queueing problems. *Numerical Algorithms*, 15:57–74, 1997.

[3] L. Bright and P. G. Taylor. Calculating the equilibrium distribution in level dependent quasi-birth-and-death processes. *Comm. Statist. Stochastic Models*, 11(3):497–525, 1995.

[4] L. Bright and P. G. Taylor. Equilibrium distributions for level-dependent quasi-birth-and-death processes. In S. R. Chakravarthy and A. S. Alfa, editors, *Matrix-Analytic Methods in Stochastic Models*, pages 359–375. Marcel Dekker, New York, 1997.

[5] J. D. Gardiner, A. J. Laub, J. J. Amato, and C. B. Moler. Solution of the Sylvester matrix equation $AXB^T + CXD^T = E$. *ACM Trans. Math. Software*, 18:223–231, 1992.

[6] Q.-M. He and A. S. Alfa. The MMAP[K]/PH[K]/1 queues with a last-come-first-served preemptive service discipline. *Queueing Systems*, 29:269–291, 1998.

[7] Q.-M. He and A. S. Alfa. The discrete time MMAP[K]/PH[K]/1/LCFS-GPR queue and its variants. In G. Latouche and P. Taylor, editors, *Advances in Algorithmic Methods for Stochastic Models – Proceedings of the 3rd International Conference on Matrix-Analytic Methods*, pages 167–190. Notable Publications Inc, NJ, July 2000.

[8] G. Latouche. Newton's iteration for nonlinear equations in Markov chains. *IMA J. Numer. Anal.*, 14:583–598, 1994.

[9] G. Latouche and V. Ramaswami. A logarithmic reduction algorithm for Quasi-Birth-Death processes. *J. Appl. Probability*, 30:650–674, 1993.

[10] G. Latouche and V. Ramaswami. *Introduction to Matrix Analytic Methods in Stochastic Modeling.* ASA-SIAM Series on Statistics and Applied Probability 5. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1999.

[11] B. Meini. New convergence results on functional iteration techniques for the numerical solution of M/G/1 type Markov chains. *Numer. Math.*, 78:39–58, 1997.

[12] M. F. Neuts. *Matrix-Geometric Solutions in Stochastic Models. An Algorithmic Approach.* The Johns Hopkins University Press, Baltimore, MD, 1981.

[13] J. M. Ortega and W. C. Rheinboldt. *Iterative Solution of Nonlinear Equations in Several Variables.* Academic Press, New York, 1970.

[14] T. Takine, B. Sengupta, and R. W. Yeung. A generalization of the matrix M/G/1 paradigm for Markov chains with a tree structure. *Comm. Statist. Stochastic Models*, 11:411–421, 1995.

[15] B. Van Houdt and C. Blondia. Stability and performance of stack algorithms for random access communication modeled as a tree structured QBD Markov chain. *Comm. Statist. Stochastic Models*, 17:247–270, 2001.

[16] R. S. Varga. *Matrix Iterative Analysis.* Prentice Hall, New Jersey, 1962.

[17] R. W. Yeung and A. S. Alfa. The quasi-birth-death type Markov chain with a tree structure. *Comm. Statist. Stochastic Models*, 15:639–659, 1999.

[18] R. W. Yeung and B. Sengupta. Matrix product-form solutions for Markov chains with a tree structure. *Adv. in Appl. Probab.*, 26:965–987, 1994.