On Functional Iteration Methods for Solving M/G/1 Type Markov Chains

Paola Favati[1] and Beatrice Meini[2]
[1]Istituto di Matematica Computazionale del CNR, via S. Maria 46, 56127 Pisa, Italy; *favati@imc.pi.cnr.it*
[2]Dipartimento di Matematica, via Buonarroti 2, 56127 Pisa, Italy; *meini@dm.unipi.it*

Abstract: We consider functional iteration methods, based on the recursion $X_{n+1} = F(X_n)$, $n \geq 0$, for solving the nonlinear matrix equation $X = \sum_{i=0}^{+\infty} X^i A_i$ which arises in the numerical solution of M/G/1 type Markov chains. We propose two strategies for improving the rate of convergence of such iterative methods, based on the spectral properties of the solution G. The first strategy consists in choosing an initial approximation X_0 which shares with G some eigenvalues and the corresponding left eigenvectors; the second one relies on a relaxation technique which modifies the spectral properties of the Jacobian matrix associated with the iteration function F. Numerical results show the effectiveness of these strategies.

1 INTRODUCTION

Many queueing problems can be modeled by Markov chains of M/G/1 type [21], that is, Markov chains whose probability transition matrix has the structure

$$P^t = \begin{pmatrix} B_1 & A_0 & & \bigcirc & \\ B_2 & A_1 & A_0 & & \\ B_3 & A_2 & A_1 & A_0 & \\ \vdots & \vdots & \ddots & \ddots & \ddots \end{pmatrix},$$

where B_{i+1}, A_i, $i \geq 0$, are $k \times k$ nonnegative matrices such that $\sum_{i=1}^{+\infty} B_i$ and $\sum_{i=0}^{+\infty} A_i$ are column stochastic. It is well known that, in the case where P^t is positive recurrent, the problem of the computation of the probability invariant vector $\boldsymbol{\pi}$, i.e., the solution of the infinite system $\boldsymbol{\pi} = P^t\boldsymbol{\pi}$, $\|\boldsymbol{\pi}\|_1 = 1$, can be reduced (compare [21]) to the computation of the column stochastic matrix G, given by the minimal nonnegative solution of the nonlinear matrix equation

$$X = \sum_{i=0}^{+\infty} X^i A_i, \tag{1}$$

where X is a $k \times k$ matrix. Indeed, once G is known, the vector $\boldsymbol{\pi}$ can be computed by means of the Ramaswami formula [22]. Moreover, the knowledge of the matrix G is very important, due to its role in calculating performance measures [21].

Recently, the analysis of numerical methods for computing the solution of the matrix equation (1) has been largely developed. New quadratically convergent methods have been proposed in [2], [4], [5], [6], [7], [16], [17], [18]; classical linearly convergent methods based on functional iterations have been improved and studied in details in [10], [11], [19]. The latter methods, unlike most of the quadratically convergent methods, are very easy to implement. They are based on the recursion

$$X_{n+1} = F(X_n), \quad n \geq 0, \tag{2}$$

where

$$F(X) = K(X)(I - H(X))^{-1}, \tag{3}$$

and $H(X) = \sum_{i=0}^{+\infty} X^i H_i$, $K(X) = \sum_{i=0}^{+\infty} X^i K_i$ are such that $H_i \geq O$, $K_i \geq O$, $i \geq 0$, and $\sum_{i=0}^{+\infty} z^i A_i = z \sum_{i=0}^{+\infty} z^i H_i + \sum_{i=0}^{+\infty} z^i K_i$ (see [11], [19]).

It can be proved that, by suitably choosing the starting approximation X_0, the sequence generated by (2) linearly converges to the solution G of (1) and the rate of convergence is strongly related to the choice of X_0 and to the function F. In the case where $X_0 = 0$ in [19] it is shown that the mean asymptotic rate of convergence of $\{X_n\}$ coincides with the spectral radius of the $k^2 \times k^2$ nonnegative matrix $\widehat{R}_F$, given by the Jacobian matrix of the iteration function F computed in G. Moreover, in the case where X_0 is any column stochastic matrix, at each step n, the vector $\hat{\boldsymbol{e}}_n$, obtained by the column-wise arrangement of the entries of the matrix $G - X_n$, is orthogonal to the nonnegative left eigenvector of $\widehat{R}_F$ associated with the spectral radius. From these properties it follows that the mean asymptotic rate of convergence of the sequence X_n is less than or equal to the second largest modulus eigenvalue of $\widehat{R}_F$, if X_0 is a column stochastic matrix and $\widehat{R}_F$ is irreducible. In this way in [19] the rate of convergence is improved by suitably choosing the starting approximation X_0.

In this paper we provide a more precise analysis of the convergence speed of (2) and we propose two strategies to further improve the rate of convergence, based on the knowledge of the eigenvalues of the matrix G. These eigenvalues can be approximated by numerically computing the zeros of the function $\det(zI - \sum_{i=0}^{+\infty} A_i z^i)$ lying in the closed unit circle of the complex plane.

The first strategy is based on an extension of the results of [19] and consists in considering initial approximations X_0 that share with G more eigenvalues, and the corresponding left eigenvectors, in addition to the eigenvalue 1 and the eigenvector $\boldsymbol{e}^t = (1, \ldots, 1)$ (shared by any column stochastic matrix).

The second strategy relies on a relaxation technique, applied in order to modify the spectral properties of the Jacobian matrix of the iteration function F.

Let $\lambda_1 = 1, \lambda_2, \ldots, \lambda_h$ be $1 \leq h < k$ eigenvalues of the matrix G such that the corresponding left eigenvectors $\boldsymbol{w}_1^t = \boldsymbol{e}^t, \boldsymbol{w}_2^t, \ldots, \boldsymbol{w}_h^t$ are linearly independent. Suppose that the matrix X_0 satisfies

$$\boldsymbol{w}_i^t X_0 = \lambda_i \boldsymbol{w}_i^t, \ i = 1, \ldots, h, \tag{4}$$

that is, it shares with G the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_h$ and the corresponding left eigenvectors. Then we prove that the vector $\hat{\boldsymbol{e}}_n$ belongs to the subspace $\mathcal{S}_h^{\perp}$, where $\mathcal{S}_h$ is the kh-dimensional subspace generated by the vectors $\boldsymbol{e}_i \otimes \boldsymbol{w}_j$, $i = 1, \ldots, k$, $j = 1, \ldots h$, and $\boldsymbol{e}_i$ is the k-dimensional vector having the i-th entry equal to 1, and the remaining ones equal to 0. Whence, we show that the rate of convergence of the sequence X_n is bounded from above by the spectral radius of the matrix obtained by projecting $\widehat{R}_F$ onto the subspace $\mathcal{T}_h = \mathcal{S}_h^{\perp}$. In this way the rate of convergence can be substantially reduced. Once the eigenvalues $\lambda_1 = 1, \lambda_2, \ldots, \lambda_h$, and the corresponding eigenvectors $\boldsymbol{w}_1^t, \ldots, \boldsymbol{w}_h^t$ are known, a matrix X_0 satisfying (4) can simply be obtained by

$$X_0 = V Diag(\lambda_1, \ldots, \lambda_h) W^t, \tag{5}$$

where W is the $k \times h$ matrix whose columns are the vectors $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_h$, and V is the generalized inverse of W^t.

The idea to use the matrix X_0 of (5) as starting point of functional iterations, is almost similar to the idea proposed in [20], [12] for the direct approximation of the matrix G. In fact, in the case where $h = k$, the matrix V coincides with the inverse of W^t and $W^{-t} Diag(\lambda_1, \ldots, \lambda_k) W^t$ coincides with G. In practice, in the case where the matrix W is nearly singular, due to the roundoff errors generated in the computation, the matrix $\widetilde{G}$ obtained in this way can be a poor approximation of G. Our approach provides the possibility of selecting those eigenvalues and eigenvectors which bring most of the information and which are not strongly affected by rounding errors, say, the eigenvectors which are not nearly linearly dependent.

In our second approach we consider functional iterations defined by the recursion

$$X_{n+1} = F_\omega(X_n), \quad n \geq 0, \tag{6}$$

where $F_\omega(X) = (1 - \omega) X + \omega F(X)$ and ω is a relaxation parameter. Based on the relation between the convergence rate of the sequence (6) and the spectral properties of $\widehat{R}_{F_\omega}$, we may characterize the parameter ω^* which minimizes the spectral radius of the matrix obtained by projecting $\widehat{R}_{F_\omega}$ onto the subspace $\mathcal{T}_h = \mathcal{S}_h^\perp$. We prove that the parameter ω^* depends on the eigenvalues of suitable matrices Q_i, $i = 1, \ldots, k$, defined by means of the matrix G. This dependency in principle does not allow us to explicitly determine ω^*. However, this characterization allows us to devise a strategy for the dynamic choice, at each step n, of a nearly optimal value ω_n of ω. This is obtained by replacing at each step n the matrix G with the current approximation X_n, thus generating the sequence of matrices

$$X_{n+1} = (1 - \omega_n)X_n + \omega_n F(X_n), \quad n \geq 0.$$

We present numerical results which confirm the theoretical properties proved in the paper and which show the effectiveness of the proposed ideas to improve the rate of convergence.

In section 2 we analyze the rate of convergence of functional iterations (2). In section 3 we consider the relaxed sequence (6).

2 ANALYSIS OF THE RATE OF CONVERGENCE

The rate of convergence of functional iterations (2) is strongly related to the spectral properties of the starting approximation X_0. Indeed, it was observed in [15], [14] that, in many numerical experiments, the number of iterations needed to reach the required accuracy by starting with $X_0 = I$ was much less than the number of iterations needed by starting with $X_0 = 0$. This behavior has been fully explained in [19], where it is proved that the better rate of convergence is due to the fact that X_0 shares with G the eigenvalue 1 and the corresponding left eigenvector $\boldsymbol{e}^t = (1, \ldots, 1)$. We generalize this result, by considering starting approximations X_0 which share with G more eigenvalues, and corresponding left eigenvectors.

Denote by $\Lambda = \{\lambda_1, \ldots, \lambda_k\}$ the set of the eigenvalues of the matrix G, and with $\boldsymbol{w}_1^t, \ldots, \boldsymbol{w}_k^t$ the corresponding left eigenvectors, where $\lambda_1 = 1$ and $\boldsymbol{w}_1^t = \boldsymbol{e}^t$. Let Λ_h be a subset of h eigenvalues of Λ, $1 \leq h < k$, such that:

1. $\lambda_1 \in \Lambda_h$;
2. if $\lambda \in \Lambda_h$, then $\bar{\lambda} \in \Lambda_h$, where $\bar{\lambda}$ denotes the complex conjugate of λ;
3. the left eigenvectors corresponding to the eigenvalues in Λ_h are linearly independent.

Throughout, we suppose, without loss of generality, that $\Lambda_h = \{\lambda_1, \ldots, \lambda_h\}$.

Denote by $\mathcal{S}_h$ the subspace generated by the k^2–dimensional vectors $\boldsymbol{e}_i \otimes \boldsymbol{w}_j$, $i = 1, \ldots, k$, $j = 1, \ldots, h$, and let $\mathcal{T}_h = \mathcal{S}_h^\perp$. Let $S_h = I \otimes E_h$ be a $k^2 \times hk$ matrix whose columns are an orthonormal basis of $\mathcal{S}_h$, and let $T_h = I \otimes F_h$ be a $k^2 \times (k^2 - hk)$ matrix whose columns are an orthonormal basis of $\mathcal{T}_h$.

Let us denote by $\hat{\boldsymbol{e}}_n$ the vector obtained by column-wise arranging the entries of the matrix $G - X_n$. The error $\hat{\boldsymbol{e}}_{n+1}$ can be explicitly related to the error $\hat{\boldsymbol{e}}_n$, as shown in [19], by means of the recursion

$$\hat{\boldsymbol{e}}_{n+1} = \widehat{R}_n \hat{\boldsymbol{e}}_n \tag{7}$$

where

$$\widehat{R}_n = \sum_{j=0}^{+\infty} Y_{j,n} \otimes G^j$$

and

$$\begin{array}{l} Y_{0,n} = \sum_{i=1}^{+\infty} \left(X_n^{i-1} K_i (I - H(X_n))^{-1}\right)^t \\ Y_{j,n} = \sum_{i=j}^{+\infty} \left(X_n^{i-j} A_{i+1} (I - H(X_n))^{-1}\right)^t, \quad j \geq 1. \end{array} \tag{8}$$

Denote by

$$\widehat{R}_F = \sum_{j=0}^{+\infty} Y_j \otimes G^j$$

where

$$\begin{array}{l} Y_0 = \sum_{i=1}^{+\infty} \left(G^{i-1} K_i (I - H(G))^{-1}\right)^t \\ Y_j = \sum_{i=j}^{+\infty} \left(G^{i-j} A_{i+1} (I - H(G))^{-1}\right)^t, \quad j \geq 1. \end{array}$$

Observe that, if $\lim_n X_n = G$, then $\widehat{R}_F = \lim_n \widehat{R}_n$.

The mean asymptotic rate of convergence of the sequence $\hat{\boldsymbol{e}}_n$, i.e., the value

$$r = \lim_n \sqrt[n]{\|\hat{\boldsymbol{e}}_n\|},$$

where $||\cdot||$ is any vector norm, can be estimated, as shown in the following theorem:

Theorem 1. *Let X_n, $n \geq 0$, the sequence generated by $X_{n+1} = F(X_n)$, where $F(\cdot)$ is defined by (3). Let $1 \leq h < k$ and $\Lambda_h \subset \Lambda$ satisfying properties 1, 2, 3. Suppose that X_0 satisfies*

$$\boldsymbol{w}^t X_0 = \lambda \boldsymbol{w}^t,$$

for any $\lambda \in \Lambda_h$, where $\boldsymbol{w}^t$ is the corresponding left eigenvector of G. If the sequence X_n is convergent then it holds

$$r = \lim_n \sqrt[n]{\|\hat{\boldsymbol{e}}_n\|} \leq \rho(T_h^t \widehat{R}_F T_h).$$

Proof. First observe that $\hat{\boldsymbol{e}}_n \in \mathcal{T}_h$, for $n \geq 0$. In fact, for $n = 0$, this property follows from the relations $\boldsymbol{w}^t(G - X_0) = \lambda \boldsymbol{w}^t - \lambda \boldsymbol{w}^t = 0$, for any $\lambda \in \Lambda_h$; for $n \geq 1$, the property follows from (7) and from the fact that, if $\boldsymbol{u} \in \mathcal{S}_h$, then $\boldsymbol{u}^t \widehat{R}_n \in \mathcal{S}_h$, for any $n \geq 0$.

Let $\Pi = (S_h|T_h)$. From (7) it follows that

$$\Pi^t \hat{\boldsymbol{e}}_{n+1} = \begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{f}_{n+1} \end{pmatrix} = (\Pi^t \widehat{R}_n \Pi)(\Pi^t \hat{\boldsymbol{e}}_n) = (\Pi^t \widehat{R}_n \Pi) \begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{f}_n \end{pmatrix}, \tag{9}$$

where $\boldsymbol{f}_n = T_h^t \boldsymbol{e}_n$. Since $S_h^t \widehat{R}_n T_h = O$, it follows that

$$\Pi^t \widehat{R}_n \Pi = \begin{pmatrix} S_h^t \\ - \\ T_h^t \end{pmatrix} \widehat{R}_n \ (S_h|T_h) = \begin{pmatrix} V_n & O \\ T_n & Z_n \end{pmatrix}, \tag{10}$$

where $Z_n = T_h^t \widehat{R}_n T_h$. From (9) and (10), the recurrence (7) can be rewritten in the form

$$\boldsymbol{f}_{n+1} = Z_n \boldsymbol{f}_n.$$

Let $\epsilon > 0$ be fixed and let $||\cdot||_\epsilon$ be a matrix norm such that

$$||T_h^t \widehat{R}_F T_h||_\epsilon \leq \rho(T_h^t \widehat{R}_F T_h) + \epsilon.$$

Since the sequence $\{Z_n\}$ converges to $T_h^t \widehat{R}_F T_h$, it can be easily verified that there exists a positive constant c such that

$$||\boldsymbol{f}_n||_\epsilon \leq c(\rho(T_h^t \widehat{R}_F T_h) + 2\epsilon)^n.$$

Whence, we have

$$r = \lim_n \sqrt[n]{||\hat{\boldsymbol{e}}_n||} = \lim_n \sqrt[n]{||\boldsymbol{f}_n||} \leq \rho(T_h^t \widehat{R}_F T_h).$$

$\square$

The above result generalizes the convergence properties stated in [19], which can be obtained in the case where $h = 1$ and X_0 is nonnegative, that is, X_0 is a column stochastic matrix.

From the above theorem it follows that, by starting with a matrix X_0 sharing with G more than one eigenvalue, the rate of convergence can be improved. In fact, the following result holds:

Theorem 2. *Under the hypotheses of Theorem 1, the set Δ of the eigenvalues of the matrix $T_h^t \widehat{R}_F T_h$ is given by*

$$\Delta = \bigcup_{i=h+1}^{k} \{\mu : \mu \text{ is eigenvalue of } Q_i\}$$

where $Q_i = \sum_{j=0}^{\infty} \lambda_i^j Y_j$, $i = 1, \ldots, k$.

Proof. Observe that the matrix $Z_F = T_h^t \widehat{R}_F T_h$ can be written as

$$Z_F = (I \otimes F_h^t) \widehat{R}_F (I \otimes F_h) = \sum_{j=0}^{\infty} Y_j \otimes F_h^t G^j F_h.$$

Whence, the matrix Z_F is similar to the matrix $\sum_{j=0}^{\infty} F_h^t G^j F_h \otimes Y_j$, which is similar to $\sum_{j=0}^{\infty} U^j \otimes Y_j$, where U is the Schur canonical form of $F_h^t G F_h$. Since the eigenvalues of $F_h^t G F_h$ are $\lambda_{h+1}, \ldots, \lambda_k$, U is an upper triangular matrix whose diagonal entries are $\lambda_{h+1}, \ldots, \lambda_k$. In this way we find that Z_F is similar to a block upper triangular matrix whose diagonal blocks are the matrices Q_i, $i = h+1, \ldots, k$. □

In particular, from the above results, it follows that

$$\rho(T_h^t \widehat{R}_F T_h) \le \rho(T_1^t \widehat{R}_F T_1)$$

for any $1 < h < k$, that is, the rate of convergence can be improved, if the starting approximation X_0 has $h > 1$ eigenvalues, and corresponding left eigenvectors, in common with G.

In the above theorems we have given conditions on the starting matrix X_0 which allow to increase the rate of convergence of the sequence $X_{n+1} = F(X_n)$, under the hypothesis that the sequence X_n is convergent. In the case where the matrix X_0 is column stochastic, the sequence is obviously convergent. In the more general case where X_0 is not nonnegative, the local convergence is guaranteed by the classical results on functional iteration methods [23], as stated by the next theorem. Let us first introduce some notations: for any $k \times k$ matrix A denote by $\boldsymbol{a} = vec(A)$ the k^2-dimensional vector obtained by column-wise arranging the entries of the matrix A; analogously, we denote $A = vec^{-1}(\boldsymbol{a})$.

Theorem 3. *Let $g_F : R^{k^2} \to R^{k^2}$ the function defined by*

$$g_F(\boldsymbol{x}) = vec(F(vec^{-1}(\boldsymbol{x}))),$$

where F is given in (3), and denote by $J_F(\boldsymbol{x})$ the $k^2 \times k^2$ Jacobian matrix associated with the function g_F. Then, $g_F(\boldsymbol{g}) = \boldsymbol{g}$ and $\widehat{R}_F = J_F(\boldsymbol{g})$, where $\boldsymbol{g} = vec(G)$.

Proof. The proof directly follows from the properties of Kronecker product, see [13]. □

Since $\rho(\widehat{R}_F) < 1$ (compare [19]), the local convergence is guaranteed.

Once the eigenvalues $\lambda_1 = 1$, λ_2, ..., λ_h, and the corresponding eigenvectors $\boldsymbol{w}_1^t$, ..., $\boldsymbol{w}_h^t$ are known, a matrix X_0 satisfying (4) can simply be obtained by

$$X_0 = V Diag(\lambda_1, \ldots, \lambda_h) W^t, \tag{11}$$

where W is the $k \times h$ matrix whose columns are the vectors $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_h$, and V is the generalized inverse of W^t.

The eigenvalues $\lambda_1, \ldots, \lambda_k$ of G are given by the zeros of the function $\varphi(z) = \det(zI - \sum_{j=0}^{+\infty} A_j z^j)$ lying in the closed unit circle of the complex plane (compare [12]), and the left eigenvectors are the solutions of the linear systems $\boldsymbol{w}_i^t(\lambda_i I - \sum_{j=0}^{+\infty} A_j \lambda_i^j) = 0$, $i = 1, \ldots, k$. The zeros of the function $\varphi(z)$ can be computed by means of any algorithm for approximating the zeros of a polynomial, for example by means of Aberth method [9], [1]; the left eigenvectors can be computed by solving the corresponding homogeneous linear systems.

In the light of theorems 1, 2, 3, choosing an initial approximation satisfying (11) can lead to a strong reduction of the number of iterations needed to reach the required accuracy.

We have verified the effectiveness of the proposed strategy on a problem arising from the modeling of a Metaring MAC protocol [3], having block size $k = 16$, for different values of the system utilization ρ. For these problems, as ρ approaches 1, functional iteration methods converge very slowly, even if the starting approximation is a column stochastic matrix. We have tested our algorithm in the cases where $\rho = 0.5, 0.8, 0.97$. We have sorted the eigenvalues in such a way that $|\lambda_1| \geq |\lambda_2| \geq \ldots \geq |\lambda_k|$, and we have chosen different subsets $\Lambda_h = \{\lambda_1, \ldots, \lambda_h\}$, where h is such that properties 1, 2, and 3 of this section are satisfied. Table 1 reports, for different values of ρ and h, the number of iterations and the CPU time (in seconds) needed by functional iterations in order to obtain an approximation X with a residual error $\|X - \sum_{i=0}^{+\infty} X^i A_i\|_1$ less than 10^{-11}. We observe that our method shows its effectiveness especially when ρ is close to 1. In particular, in the cases $\rho = 0.8$ and $\rho = 0.97$, choosing a starting approximation that shares with G only 2 eigenvalues (and corresponding left eigenvectors), leads to a substantial reduction of the number of iterations.

Figure 1 reports the logarithm (to the base 10) of the residual error for $h = 1, 2, 16$, in the case where $\rho = 0.97$. It is worth pointing out the dramatic reduction of the number of iterations when h passes from 1 to 2. In the case where $h = 16$, i.e., all the eigenvalues and corresponding eigenvectors are used, the matrix X_0 is a poor approximation of G, which can be refined by applying few steps of functional

	$\rho = 0.5$				$\rho = 0.8$				$\rho = 0.97$			
h	1	2	8	16	1	2	8	16	1	2	8	16
Iter.	76	32	24	10	204	33	20	8	2763	34	20	19
Time	5.0	2.1	1.6	0.7	15.0	2.4	1.5	0.6	203.3	2.5	1.5	1.4

Table 1: Metaring MAC Protocol

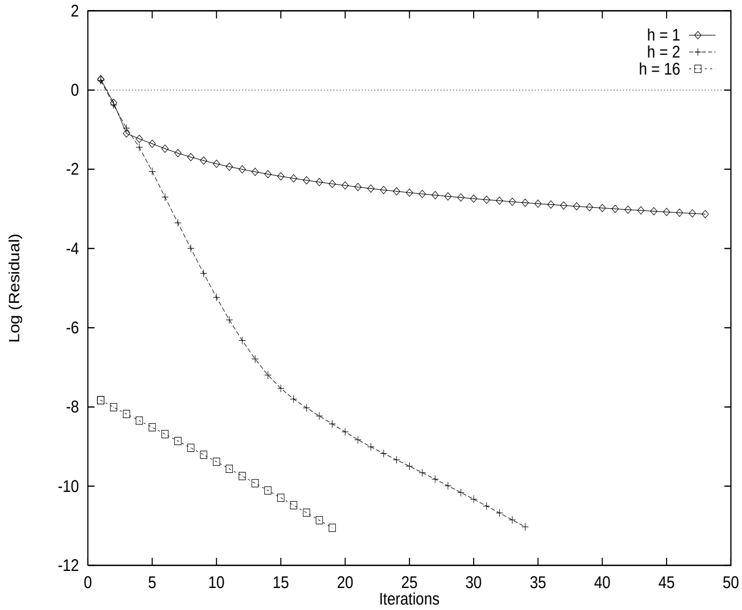

Figure 1: Metaring MAC Protocol, $\rho = 0.97$

iterations. However, from figure 1, it appears that the asymptotic rate of convergence is almost the same as in the case $h = 2$.

3 A RELAXATION TECHNIQUE

In the previous section we have observed that the rate of convergence of functional iteration methods is related to the spectral properties of the starting approximation X_0 and to the Jacobian matrix $\widehat{R}_F$ associated with the iteration function F. In this section we propose to improve the rate of convergence by applying a relaxation technique to customary functional iterations. More precisely, we consider functional iterations defined by the recursion

$$X_{n+1} = (1 - \omega)X_n + \omega F(X_n), \quad n \geq 0 \tag{12}$$

where $F(\cdot)$ is given by (3) and ω is a relaxation parameter.

It can be readily observed that, in this case, the error $\hat{\boldsymbol{e}}_{n+1}$ is related to the error $\hat{\boldsymbol{e}}_n$ by means of the recursion

$$\hat{\boldsymbol{e}}_{n+1} = ((1-\omega)I + \omega\widehat{R}_n)\hat{\boldsymbol{e}}_n, \ n \geq 0. \tag{13}$$

From the above equation, it follows that the rate of convergence of the relaxed sequence is related to the parameter ω, as stated by the following theorem:

Theorem 4. *Let $\omega > 0$ a relaxation parameter. Let $1 \leq h < k$ and $\Lambda_h \subset \Lambda$ satisfying properties 1, 2, 3. Suppose that X_0 satisfies*

$$\boldsymbol{w}^t X_0 = \lambda \boldsymbol{w}^t,$$

for any $\lambda \in \Lambda_h$, where $\boldsymbol{w}^t$ is the corresponding left eigenvector of G. If the sequence X_n, $n \geq 0$ of (12) is convergent, then it holds

$$r = \lim_n \sqrt[n]{\|\hat{\boldsymbol{e}}_n\|} \leq \rho((1-\omega)I + \omega T_h^t \widehat{R}_F T_h).$$

Proof. The proof can be carried out by following the same lines of the proof of Theorem 2. □

The above result suggests us to determine a value ω^* of ω for which the upper bound $\rho((1-\omega)I + \omega T_h^t \widehat{R}_F T_h)$ to the rate of convergence is minimum.

Observe that, if $\omega = 1$ the local convergence of the sequence X_n is guaranteed, as pointed out in the previous section, since $\rho(T_h^t \widehat{R}_F T_h) \leq \rho(\widehat{R}_F) < 1$. By choosing $\omega = \omega^*$ the local convergence is still guaranteed since $\rho((1-\omega^*)I + \omega^* T_h^t \widehat{R}_F T_h) \leq \rho(T_h^t \widehat{R}_F T_h)$.

Hereafter we use the notations

$$W_\omega = (1-\omega)I + \omega T_h^t \widehat{R}_F T_h$$

and

$$f_\mu(x) = \frac{2(\mathrm{Re}\ x - \mathrm{Re}\ \mu)}{2(\mathrm{Re}\ x - \mathrm{Re}\ \mu) + |\mu|^2 - |x|^2}.$$

In the following theorem we characterize the value ω^* for which the upper bound to the asymptotic rate of convergence is minimum :

Theorem 5. *Let $X_n = (1-\omega)X_{n-1} + \omega F(X_{n-1})$, for $n \geq 1$. Let $1 \leq h < k$ and $\Lambda_h \subset \Lambda$ satisfying properties 1, 2, 3 and suppose that X_0 satisfies the hypothesis of theorem 4. Let Δ be the set of the eigenvalues of the matrix $T_h^t \widehat{R}_F T_h$ and $\Gamma = \{\alpha \in$*

$\Delta : \text{Re } \alpha = \max_{\lambda\in\Delta} \text{Re } \lambda\}$. *Then the parameter* ω^* *such that* $\rho(W_{\omega^*}) = \min_{\omega>0}\rho(W_\omega)$ *is given by*

$$\omega^* = \min\left\{ f_\mu(\nu), \frac{1-\text{Re } \mu}{1-2\text{Re } \mu + |\mu|^2}\right\}$$

where:

- $\mu \in \Gamma$ *is such that* $\text{Im } \mu = \max_{\alpha\in\Gamma} \text{Im } \alpha$
- $\nu \in \Delta - \Gamma$ *is such that* $f_\mu(\nu) = \min_{\lambda\in\Delta-\Gamma} f_\mu(\lambda)$.

Proof. Let us estimate the spectral radius of $\rho(W_\omega)$. It can be easily verified that the eigenvalues of the matrix W_ω are given by the set

$$\{1-\omega+\omega\lambda, \lambda \in \Delta\}.$$

Whence, we have

$$\rho(W_\omega)^2 = \max_{\lambda\in\Delta} p_\lambda(\omega),$$

where

$$p_\lambda(\omega) = |1-\omega+\omega\lambda|^2 = \omega^2(1+|\lambda|^2 - 2\text{Re } \lambda) - 2\omega(1-\text{Re } \lambda) + 1.$$

The function $p_\lambda(\omega)$ belongs to a sheaf of $k^2 - hk$ parabolas with center in $(0,1)$ and vertices in

$$(x_\lambda, y_\lambda) = \left(\frac{1-\text{Re } \lambda}{1+|\lambda|^2-2\text{Re } \lambda}, \frac{\text{Im }^2\lambda}{1+|\lambda|^2-2\text{Re } \lambda}\right), \lambda \in \Delta.$$

Since $\rho(T_h^t \widehat{R}_F T_h) < 1$, for any $\lambda \in \Delta$ we have $\text{Re } \lambda < 1$, whence $x_\lambda > 0$ and $0 < y_\lambda < 1$. Moreover we have $p_\lambda(0) = 1$ and $p'_\lambda(0) = 2(\text{Re } \lambda - 1)$. Hence, for ω in a right neighborhood of zero, the value $\max_{\lambda\in\Delta} p_\lambda(\omega)$ is reached for $\mu \in \Gamma$ such that $\text{Im } \mu = \max_{\alpha\in\Gamma} \text{Im } \alpha$. It can be observed that

$$\min_{\omega>0} \; (\max_{\lambda\in\Delta} p_\lambda(\omega)) = p_\mu(\omega^*)$$

where

$$\omega^* = \min(x_\mu, \widehat{\omega}),$$

with

$$\widehat{\omega} = \min\{\omega > 0 : p_\lambda(\omega) = p_\mu(\omega), \lambda \neq \mu\}$$

that is

$$\widehat{\omega} = \min_{\lambda\in\Delta-\Gamma} f_\mu(\lambda).$$

$\square$

From theorem 2 it follows that the entries of the set Δ, and consequently, the asymptotic optimal relaxation parameter ω^*, cannot be evaluated without knowing the matrix G. We overcome this problem by dynamically computing an optimal relaxation parameter by minimizing, at each step, the spectral radius of a suitable matrix which relates the error of two successive approximations. Observe that, if X_0 satisfies the hypothesis of Theorem 4, then the error vector $\hat{\boldsymbol{e}}_n$ belongs to the linear space $\mathcal{T}_h$ for any n. Whence, by premultiplying both sides by the matrix Π^t as in the proof of Theorem 1, the recurrences (13) can be rewritten in the form

$$T_h^t \hat{\boldsymbol{e}}_{n+1} = ((1-\omega)I + \omega T_h^t \widehat{R}_n T_h))(T_h^t \hat{\boldsymbol{e}}_n).$$

Based on these considerations, we generate the sequence of matrices

$$X_{n+1} = (1-\omega_n)X_n + \omega_n F(X_n), \ \ n \geq 0,$$

where the value of ω_n is such that

$$\rho(W_{n,\omega_n}) = \min_{\omega>0} \rho(W_{n,\omega}),$$

and $W_{n,\omega} = (1-\omega)I + \omega T_h^t \widehat{R}_n T_h$. Indeed, the following results hold:

Theorem 6. *Let* $X_n = (1-\omega)X_{n-1} + \omega F(X_{n-1})$. *Let* $1 \leq h < k$ *and* $\Lambda_h \subset \Lambda$ *satisfying properties 1, 2, 3 and suppose that* X_0 *satisfies the hypothesis of theorem 4. Let* Δ_n *be the set of the eigenvalues of the matrix* $T_h^t \widehat{R}_n T_h$ *and* $\Gamma_n = \{\alpha \in \Delta_n : \mathrm{Re}\ \alpha = \max_{\lambda \in \Delta_n} \mathrm{Re}\ \lambda\}$. *If* $\rho(T_h^t \widehat{R}_n T_h) < 1$ *then the parameter* ω_n *such that* $\rho(W_{n,\omega_n}) = \min_{\omega>0} \rho(W_{n,\omega})$ *is given by*

$$\omega_n = \min\left\{ f_\mu(\nu), \frac{1-\mathrm{Re}\ \mu}{1-2\mathrm{Re}\ \mu + |\mu|^2} \right\}$$

where:

- $\mu \in \Gamma_n$ *is such that* $\mathrm{Im}\ \mu = \max_{\alpha \in \Gamma_n} \mathrm{Im}\ \alpha$

- $\nu \in \Delta_n - \Gamma_n$ *is such that* $f_\mu(\nu) = \min_{\lambda \in \Delta_n - \Gamma_n} f_\mu(\lambda)$.

Proof. The proof can be carried out analogously to the proof of Theorem 5. □

In the case where $\rho(T_h^t \widehat{R}_n T_h) \geq 1$ we set $\omega_n = 1$, i.e. we apply the standard functional iteration formulas; however, since the non relaxed sequence converges to the matrix G, after a finite number of steps the condition $\rho(W_n) < 1$ is satisfied.

Observe that the optimal parameter ω_n depends on the eigenvalues of the matrix $T_h^t \widehat{R}_n T_h$, that can be computed by only knowing the approximation X_n and the eigenvalues of the matrix G, as stated by the following result, which is analogous to Theorem 2:

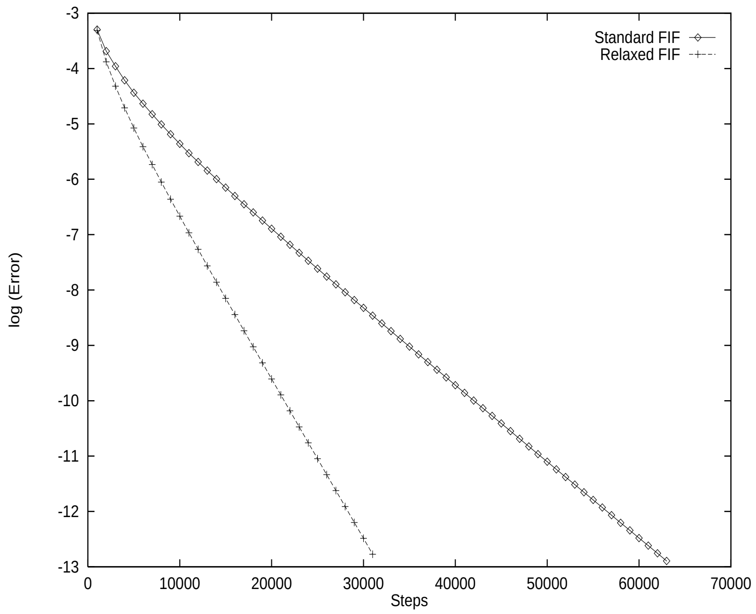

Figure 2: Teletraffic system

Theorem 7. *Let*

$$Q_{i,n} = \sum_{j=0}^{+\infty} \lambda_i^j Y_{j,n}, \ \ i = 1, \ldots, k,$$

where the matrices $Y_{j,n}$ are defined in (8) and λ_i, $i = 1, \ldots, k$, are the eigenvalues of the matrix G. Under the hypothesis of Theorem 1 the set Δ_n of the eigenvalues of the matrix $T_h^t \widehat{R}_n T_h$ is given by

$$\Delta_n = \bigcup_{i=h+1,\ldots,k} \{\mu : \mu \text{ is eigenvalue of } Q_{i,n}\}$$

From the many numerical experiments that we performed, we observed that the sequence ω_n rapidly converges to the asymptotic optimal relaxation parameter ω^* of Theorem 5. This suggested us to stop the computation of the dynamical relaxation parameter ω_n when the difference between the parameters obtained at two subsequent steps is smaller than a positive constant σ, depending on the used machine precision.

We have tested the relaxation technique on the family of teletraffic systems, modeled as a continuous-time QBD Markov chain and defined by blocks of dimension 24, described in [8].

Figure 2 reports, for one of these problems, the logarithm (to the base 10) of the residual error $\|X_n - \sum_{i=0}^{+\infty} X_n^i A_i\|_1$ of the approximations obtained at step n, with the relaxed and standard functional iteration methods, starting with $X_0 = I$.

For this problem, the asymptotic optimal relaxation parameter is $\omega^* = 2.086$. The time needed for the computation of the relaxation parameter does not significantly affect the overall computational time of the relaxed method. The ratio between the number of iterations (and the computational time) of the standard and relaxed method, needed to reach a residual error less than 10^{-12}, is about 2.

4 CONCLUDING REMARKS

In this paper we have proposed two strategies which allow to improve the rate of convergence of functional iterations. They are based on a different use of spectral properties of the matrix G and of related matrices (e.g., $\widehat{R}_F$, W_ω). For this reason, in general we cannot establish when a strategy is more effective than the other one. In the applications they can also be applied simultaneously, as proposed in Section 3.

REFERENCES

[1] O. Aberth. Iteration methods for finding all zeros of a polynomial simultaneously. *Math. Comp.*, 27, 1973.

[2] N. Akar and K. Sohraby. An invariant subspace approach in M/G/1 and G/M/1 type Markov chains. *Commun. Statist. Stochastic Models*, 13:381–416, 1997.

[3] G. Anastasi, L. Lenzini, and B. Meini. Performance evaluation of a worst case model of the MetaRing MAC protocol with global fairness. *Performance Evaluation*, 29:127–151, 1997.

[4] D. Bini and B. Meini. On cyclic reduction applied to a class of Toeplitz-like matrices arising in queueing problems. In W. J. Stewart, editor, *Computations with Markov Chains*, pages 21–38. Kluwer Academic Publisher, 1995.

[5] D. 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.

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

[7] D. Bini and B. Meini. Inverting block Toeplitz matrices in block Hessenberg form by means of displacement operators: application to queueing problems. *Linear Algebra Appl.*, 272:1–16, 1998.

[8] J. N. Daigle and D. M. Lucantoni. Queueing systems having phase-dependent arrival and service rates. In W. J. Stewart, editor, *Numerical solution of Markov Chains*, pages 161–202. Marcel Dekker, New York, 1991.

[9] L. W. Ehrilch. A modified Newton method for polynomials. *Comm. A.C.M.*, 10:107–108, 1967.

[10] P. Favati and B. Meini. On functional iteration methods for solving nonlinear matrix equations arising in queueing problems. Submitted for publication, 1997.

[11] P. Favati and B. Meini. Relaxed functional iteration techniques for the numerical solution of M/G/1 type Markov chains. *BIT*, 38, 1998. In printing.

[12] H. R. Gail, S. L. Hantler, and B. A. Taylor. Solutions of the basic matrix equation for M/G/1 and G/M/1 type Markov chains. *Commun. Statist. Stochastic Models*, 10:1–43, 1994.

[13] A. Graham. *Kronecker Products and Matrix Calculus with Applications.* Ellis Horwood Limited, Chichester, 1981.

[14] G. Latouche. Algorithms for evaluating the matrix G in Markov chains of PH/G/1 type. Technical report, Bellcore, 1992.

[15] G. Latouche. Algorithms for infinite Markov chains with repeating columns. In C.D. Meyer and R.J. Plemmons, editors, *Linear Algebra, Queueing Models and Markov Chains*, pages 231–265. Springer-Verlag, New York, 1993.

[16] G. Latouche. Newton's iteration for non-linear equations in Markov chains. *IMA J. of Numerical Analysis*, 14:583–598, 1994.

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

[18] G. Latouche and G.W. Stewart. Numerical methods for M/G/1 type queues. In W. J. Stewart, editor, *Computations with Markov Chains*, pages 571–581. Kluwer Academic Publishers, 1996.

[19] 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.

[20] G. R. Murthy, M. Kim, and E. J. Coyle. Equilibrium analysis of skip free Markov chains: nonlinear matrix equations. *Commun. Statist. Stochastic Models*, 7:547–571, 1991.

[21] M. F. Neuts. *Structured Stochastic Matrices of M/G/1 Type and Their Applications.* Marcel Dekk., New York, 1989.

[22] V. Ramaswami. A stable recursion for the steady state vector in Markov chains of M/G/1 type. *Commun. Statist. Stochastic Models*, 4:183–188, 1988.

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