Available online at www.sciencedirect.com
SCIENCE DIRECT®
LINEAR ALGEBRA AND ITS APPLICATIONS
ELSEVIER
Linear Algebra and its Applications 386 (2004) 67–81
www.elsevier.com/locate/laa

Algebraic Schwarz methods for the numerical solution of Markov chains

Ivo Marek [a,1], Daniel B. Szyld [b,*,2]

[a] *Czech Institute of Technology, School of Civil Engineering, Thakurova 7, 166 29 Praha 6, Czech Republic*
[b] *Department of Mathematics, Temple University (038-16), 1805 N. Broad Street, Philadelphia, PA 19122-6094, USA*

Received 12 August 2003; accepted 30 December 2003

Submitted by C. Meyer

Abstract

The convergence of additive and multiplicative Schwarz methods for computing certain characteristics of Markov chains such as stationary probability vectors and mean first passage matrices is studied. The main result is a convergence theorem for multiplicative Schwarz iterations when applied to singular systems. As a byproduct, a convergence result for alternating iterations is also obtained. It is also shown that, when the Markov chain is ergodic, additive and multiplicative Schwarz methods can be applied to the nonsingular systems that result from reducing the equations. The so-called coarse grid corrections are also studied.

AMS classification: 65F10; 65F15; 15A48

Keywords: Schwarz methods; Singular M-matrices; Markov chains; Stationary probability vectors; Iterative methods

1. Introduction

Schwarz methods are widely used nowadays in the numerical solution of partial differential equations (p.d.e.s) [23,24]. These domain decomposition methods are

* Corresponding author.
E-mail addresses: marek@ms.mff.cuni.cz (I. Marek), szyld@math.temple.edu (D.B. Szyld).
[1] This work was supported by the Grant Agency of the Czech Republic, grant no. 201/02/0595 and by a grant from the Ministry of Education of the Czech Republic, contract no. MSM 210000010.
[2] Supported in part by the National Science Foundation under grant DMS-0207525.

0024-3795/$ - see front matter
doi:10.1016/j.laa.2003.12.046

being applied in many areas of science and engineering; see, e.g., the latest proceedings of a series of conferences on this topic [8,14]. Schwarz methods are extensively used as preconditioners for Krylov subspace methods, and they are also used as iterative methods, especially for nonsymmetric problems; see, e.g., [7]. Recently, an algebraic formulation of these iterative methods has been developed and studied [2,12,21]; this algebraic formulation is reviewed in Section 3.

In this paper we explore the use of additive and multiplicative Schwarz methods (with overlap) for the solution of large sparse linear singular systems of the form

$$Ax = b. \tag{1.1}$$

Specifically, we analyze the case where the coefficient matrix $A = I - B$, where I is the identity matrix and B is a nonnegative (column) stochastic matrix, i.e., $B^{\mathrm{T}}e = e$, where $e = (1, 1, \dots, 1)^{\mathrm{T}} \in \mathbb{R}^n$. Thus A is a singular M-matrix; see Section 2 for definitions. In particular we consider the case of $b = 0$, and thus we look for the nonnegative vector v, normalized so that $v^{\mathrm{T}}e = 1$, satisfying $Av = 0$, i.e., such that $Bv = v$. This is the stationary probability distribution of the Markov chain represented by B. The standard notation for the stochastic transition probability matrix is P, and the stationary probability distribution is π such that $\pi P = \pi$ [25]. In our notation $P = B^{\mathrm{T}}$ and $\pi = v^{\mathrm{T}}$.

There is no separate treatment in the literature of Schwarz methods for singular systems in the p.d.e. context. Nevertheless the implementations derived mostly for the nonsingular case can be shown to work in the singular case as well, especially when the null space is known. This is the case, for example, when Neumann boundary conditions are present. The convergence theory developed, e.g., in [9,10], can be applied to these cases with little or no changes.

We believe that this is the first time that singular systems are analyzed using an algebraic approach to Schwarz methods (with overlap), and that Markov chains problems are studied in this context. One of our goals is to present Schwarz iterations as one more possible tool for the numerical solutions of Markov chains. In fact, multiplicative Schwarz iterations reduce to the block Gauss–Seidel method when the overlap is removed. Having the overlap has proved crucial for the fast convergence of these methods in the nonsingular case; see, e.g., [5,10]. In the singular context, having larger overlap may decrease the convergence rate of the iteration. Comparison theorems may be used to prove such decrease in convergence rate [19,20]. We mention that in [17] an analysis of multiplicative Schwarz methods without overlap was carried out using nonstationary restriction and prolongation operators. When these operators are of the type described in this paper (Section 3), i.e., constant, or stationary, those methods reduce, again, to block Gauss-Seidel.

We discuss two approaches here. Firstly, in Section 4, we show our main result: multiplicative Schwarz iterations applied directly to the $n \times n$ system (1.1) converge. Secondly, in Section 5, we consider solving a smaller nonsingular problem using the (damped) additive and multiplicative Schwarz iterations, from which the solution of (1.1) can be computed. We also show a convergent weak regular splitting for A derived from the classical Schwarz methods for the smaller system. We show that both approaches converge in the case where B is irreducible, i.e., when the Markov chain is ergodic [25]. The first approach can also be used in a more general case, when B represents a Markov chain which is free of transient states. As a byproduct of our main theorem we also obtain a convergence proof of generalized alternating direction iterations (ADI) for Markov chains.

We also explore the use of some of these methods to find the mean first passage matrix of a Markov chain; see Section 6. Lastly, in Section 7, we discuss the reducible case.

2. Definitions and auxiliary results

In this section we present some notation, definitions, and preliminaries. Concepts on nonnegative matrices not explicitly defined here can be found in the standard reference [4].

An $n \times n$ matrix $C = (c_{jk})$ with $c_{jk} \in \mathbb{R}$, is called nonnegative if $c_{jk} \geqslant 0$, $j, k = 1, \ldots, n$; this is denoted $C \geqslant O$. When $c_{jk} > 0$, $j, k = 1, \ldots, n$, we say that the matrix is positive and denote it by $C > O$. The same notation is used for nonnegative and positive vectors. By $\sigma(C)$ we denote the spectrum of C and by $\rho(C)$ its spectral radius. By $\mathscr{R}(C)$ and $\mathscr{N}(C)$ we denote the range and null space of C, respectively.

Let $\lambda \in \sigma(C)$ be a pole of the resolvent operator $R(\mu, C) = (\mu I - C)^{-1}$. The multiplicity of λ as a pole of $R(\mu, C)$ is called the index of C with respect to λ and denoted $\mathrm{ind}_\lambda C$. Equivalently, $k = \mathrm{ind}_\lambda C$ if it is the smallest integer for which $\mathscr{R}((\lambda I - C)^{k+1}) = \mathscr{R}((\lambda I - C)^k)$. This happens if and only if $\mathscr{R}((\lambda I - C)^k) \oplus \mathscr{N}((\lambda I - C)^k) = \mathbb{R}^n$.

Let A be an $n \times n$ matrix. A is an M-matrix if $A = \beta I - B$, B nonnegative and $\rho(B) \leqslant \beta$. A pair of matrices (M, N) is called a splitting of A if $A = M - N$ and M^{-1} exists. A splitting of a matrix A is called *of nonnegative type* if the matrix $T = M^{-1}N$ is nonnegative [18]. If, in particular, the matrices M^{-1} and N are nonnegative, the splitting is called *regular* [29]. If M^{-1} and $T = M^{-1}N$ are nonnegative, the splitting is called *weak regular* [22].

Let T be a square matrix. T is called *convergent* if $\lim_{k\to\infty} T^k$ exists and *zero-convergent*, if moreover $\lim_{k\to\infty} T^k = O$. Standard stationary iterations of the form

$$x^{k+1} = Tx^k + c, \quad k = 0, 1, \ldots, \tag{2.1}$$

converge if and only if either T is zero-convergent or, if $\rho(T) = 1$, T is convergent. A square matrix T with unit spectral radius is convergent if the following two conditions hold:

(i) if $\lambda \in \sigma(T)$ and $\lambda \neq 1$, then $|\lambda| < 1$.
(ii) $\mathrm{ind}_1 T = 1$.

When $T \geqslant O$, (i) can be replaced with T having positive diagonal entries [1].

Equivalent conditions for (ii) can be found in [27].

It is useful to write $T = Q + S$, where Q is the first term of the Laurent expansion of T, i.e., the eigenprojection onto the invariant subspace corresponding to $\lambda = 1$; see, e.g., [28]. Then $Q^2 = Q$, $QS = SQ = O$, and $1 \notin \sigma(S)$. This is called the *spectral decomposition* of T. The condition (i) above is equivalent to having $\rho(S) < 1$.

We state a very useful lemma; its proof can be found, e.g., in [6]. We note that when $\rho(T) = 1$, this lemma can be used to show condition (ii) above. To prove convergence one needs to show in addition that condition (i) also holds, or equivalently, that the diagonal entries are all positive.

Lemma 2.1. *Let T be a nonnegative square matrix such that $Tv \leqslant \alpha v$ with $v > 0$. Then $\rho(T) \leqslant \alpha$. If furthermore $\rho(T) = \alpha$, then* $\mathrm{ind}_\alpha T = 1$.

A square nonnegative matrix B is irreducible if for every pair of indices i, j there is a power $k = k(i, j)$ such that the ij entry of B^k is nonzero. This implies that in the Markov chain each state has access to every other state, i.e., the chain is ergodic [25]. The Perron–Frobenius theorem states that for $B \geqslant O$ irreducible, $\rho(B)$ is an eigenvalue, and the corresponding eigenvector is positive; see, e.g., [4].

3. Algebraic formulation of Schwarz methods

We review here the formulation and some results from [2,12]. Given an initial approximation x^0 to the solution of (1.1), the (one-level) multiplicative Schwarz method can be written as the stationary iteration (2.1), where

$$T = T_\mu = (I - P_p)(I - P_{p-1}) \cdots (I - P_1) = \prod_{i=p}^{1} (I - P_i) \tag{3.1}$$

and c is a certain vector. Here

$$P_i = R_i^{\mathrm{T}} (R_i A R_i^{\mathrm{T}})^{-1} R_i A, \tag{3.2}$$

where R_i is a matrix of dimension $n_i \times n$ with full row rank, $1 \leqslant i \leqslant p$; see, e.g., [24]. In the case of overlap we have $\sum_{i=1}^{p} n_i > n$. The additive Schwarz method for the solution of (1.1) is of the form (2.1), where

$$T = T_\theta = I - \theta \sum_{i=1}^{p} P_i = I - \theta \sum_{i=1}^{p} R_i^{\mathrm{T}} A_i^{-1} R_i A, \tag{3.3}$$

where $0 < \theta \leqslant 1$ is a damping parameter. The matrix R_i corresponds to the restriction operator from the whole space to a subset of the state space (of dimension n_i) in the domain decomposition setting, and the matrix $A_i = R_i A R_i^{\mathrm{T}}$ is the restriction of A to that subset. A solution using A_i is called a local solve, and this name carries to the purely algebraic case. For our computations we need A_i to be nonsingular (with positive diagonals). If this were not the case, one can replace this local solve with one with the shifted system $A_i + \alpha_i I$, for some positive number α_i as described later in Proposition 4.4.

We assume that the rows of R_i are rows of the $n \times n$ identity matrix I, e.g.,

$$R_i = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Formally, such a matrix R_i can be expressed as

$$R_i = [I_i | O]\pi_i \tag{3.4}$$

with I_i the identity on $\mathbb{R}^{n_i}$ and π_i a permutation matrix on $\mathbb{R}^n$. In this case, it follows that A_i is an $n_i \times n_i$ principal submatrix of A. In fact, we can write

$$\pi_i A \pi_i^{\mathrm{T}} = \begin{bmatrix} A_i & K_i \\ L_i & A_{\neg i} \end{bmatrix}, \tag{3.5}$$

where $A_{\neg i}$ is the principal submatrix of A “complementary” to A_i, i.e.,

$$A_{\neg i} = [O | I_{\neg i}] \cdot \pi_i \cdot A \cdot \pi_i^{\mathrm{T}} \cdot [O | I_{\neg i}]^{\mathrm{T}} \tag{3.6}$$

with $I_{\neg i}$ the identity on $\mathbb{R}^{n-n_i}$. Recall that if A is an M-matrix, so are its principal submatrices, and thus both A_i and $A_{\neg i}$ are M-matrices [4]. For each $i = 1, \ldots, p$, we construct diagonal matrices $E_i \in \mathbb{R}^{n \times n}$ associated with R_i from (3.4) as follows

$$E_i = R_i^{\mathrm{T}} R_i. \tag{3.7}$$

These diagonal matrices have ones on the diagonal in every row where R_i^{T} has nonzeros.

If A is an M-matrix, for each $i = 1, \ldots, p$, we construct a second set of matrices $M_i \in \mathbb{R}^{n \times n}$ associated with R_i from (3.4) as follows

$$M_i = \pi_i^{\mathrm{T}} \begin{bmatrix} A_i & O \\ O & D_{\neg i} \end{bmatrix} \pi_i, \tag{3.8}$$

where

$$D_{\neg i} = \operatorname{diag}(A_{\neg i}) \geqslant O \tag{3.9}$$

has positive entries along the diagonal and thus is invertible. Since we are assuming that A_i is invertible (or shifted so it is, as in Proposition 4.4), then we have that the matrices M_i are nonsingular.

With the definitions (3.7) and (3.8) we obtain the following equality

$$E_i M_i^{-1} = R_i^{\mathrm{T}} A_i^{-1} R_i, \quad i = 1, \ldots, p. \tag{3.10}$$

We can thus rewrite (3.1) as

$$T = T_\mu = (I - E_p M_p^{-1} A)(I - E_{p-1} M_{p-1}^{-1} A) \cdots (I - E_1 M_1^{-1} A). \tag{3.11}$$

Similarly, (3.3) can be rewritten as

$$T = T_\theta = I - \theta \sum_{i=1}^{p} E_i M_i^{-1} A. \tag{3.12}$$

This is how we interpret the multiplicative and additive Schwarz iterations.

In the context of discretizations of p.d.e.s, the use of Schwarz methods greatly benefit from the use of coarse grid corrections, and they are needed to guarantee a convergence rate independent of the mesh size [9,10,23,24]. Coarse grid corrections can be additive or multiplicative, and they have been described in the algebraic context as well [2]. Here we restrict our comments to the multiplicative corrections. To that end consider a new projection P_0 of the form (3.2) onto the "coarse space", i.e., onto a particular subset of states, usually taken in the overlap between the other set of states. Corresponding to these "coarse" states, there correspond a natural matrix R_0 as in (3.4), and $A_0 = R_0 A R_0^{\mathrm{T}}$, so that E_0 and M_0 is similarly defined as in (3.7) and (3.8). The multiplicative corrected multiplicative Schwarz iteration operator is then

$$T_{\mu c} = (I - P_0) T_\mu = (I - E_0 M_0^{-1} A) T_\mu, \tag{3.13}$$

while the multiplicative corrected additive Schwarz iteration operator, also known as hybrid II Schwarz [24], is then

$$T_{\theta c} = (I - P_0) T_\theta = (I - E_0 M_0^{-1} A) T_\theta. \tag{3.14}$$

In [2] it was shown that when A is nonsingular, $\rho(T_\mu) < 1$, and thus, the method (2.1) is convergent. Furthermore, there exists a unique splitting $A = M - N$ such that $T = T_\mu = M^{-1} N$. This splitting is a weak regular splitting. The same results hold for $T_{\mu c}$, the iteration with a "coarse grid" correction. Similar result were also shown for T_θ and $T_{\theta c}$ when $\theta < 1/q$, where q is the measure of overlap; i.e., the maximum number of nonzeros in the same rows of all E_i, $i = 1, \ldots, p$; see [2] for further details. In this paper we want explore the convergence of (2.1), using the iterations defined by (3.11)–(3.14), when A is singular.

4. Convergence of multiplicative Schwarz

In this section we assume that the stationary probability distribution is positive, i.e., $\pi^{\mathrm{T}} = v > 0$. This is the case when $B = P^{\mathrm{T}}$ is irreducible, but also in other cases, e.g., when P represents a Markov chain free of transient states [25]. If in addition we require that the diagonals of the iteration matrices are positive, (using Proposition 4.4) we show in the next theorem that the matrix (3.11) is convergent, i.e., that the multiplicative Schwarz iterations are convergent.

The argument used in the proof of this theorem can also be applied to the matrix

$$\widehat{T} = T_p \cdots T_1 = \prod_{i=p}^{1} T_i, \tag{4.1}$$

where $T_i = M_i^{-1} N_i$ and $A = M_i - N_i$, $i = 1, \ldots, p$, and we therefore include it in the same result. This product corresponds to p alternating iterations, i.e., p intermediate steps of the form (2.1). This is a generalization of the case of $p = 2$ which include the classical SSOR and ADI methods; see, e.g., [3,30].

Theorem 4.1. *Let $A = I - B$, where B is an $n \times n$ column stochastic matrix such that $Bv = v$ with $v > 0$. Let $p > 1$ be a positive integer and $A = M_i - N_i$ be splittings of nonnegative type such that the diagonals of $T_i = M_i^{-1} N_i$, $i = 1, \ldots, p$, are positive. Then* (3.11) *and* (4.1) *are convergent matrices. Furthermore, there is a splitting of nonnegative type*

$$A = M - N \tag{4.2}$$

such that $T = M^{-1} N$, and the matrix T possesses the following properties:

$$T = Q + S, \quad Q^2 = Q, \quad QS = SQ = O, \quad \rho(S) < 1, \tag{4.3}$$

and

$$AQ = O. \tag{4.4}$$

The existence of a splitting of nonnegative type, and properties (4.3) *and* (4.4) *also hold for $\widehat{T}$.*

Proof. We begin with the matrix $\widehat{T}$. Let $v > 0$ be such that $Bv = v$, i.e., $Av = 0$. For each splittings of $A = M_i - N_i$, we then have that $M_i v = N_i v$. This implies that $\widehat{T} v = v$, and by Lemma 2.1 we have that $\rho(\widehat{T}) = 1$ and that the index is 1. To show that $\widehat{T}$ is convergent, we show that its diagonal is positive. This follows from the fact that each of the diagonals of the nonnegative matrices T_i is positive.

We follow a similar logic for the multiplicative Schwarz iteration matrix (3.11). Since $Av = 0$, $Tv = v$, and thus $\rho(T) = 1$ and $\text{ind}_1 T = 1$. Each factor in (3.11) can be written as

$$I - E_i + E_i(I - M_i^{-1}A) = I - E_i + E_i M_i^{-1} N_i,$$

and since $O \leqslant E_i \leqslant I$ and $M_i^{-1}N_i \geqslant O$, each factor is nonnegative. For a row in which E_i is zero, the diagonal entry in this factor has value one. For a row in which E_i has value one, the diagonal entry in this factor is the positive diagonal entry of $M_i^{-1}N_i$. Thus, again, we have a product of nonnegative matrices, each having positive diagonals, implying that the product T has positive diagonal entries, and therefore it is convergent.

The rest of the proof applies equally to T and $\widehat{T}$, we only detail it for T. The matrix T being convergent implies the spectral decomposition (4.3), where Q is the spectral projection onto the eigenspace of T corresponding to $\rho(T) = 1$. Furthermore since $T \geqslant O$, $Q = \lim_{k\to\infty} T^k \geqslant O$.

We show now that $\mathscr{N}(I - T) = \mathscr{N}(A)$. According to construction of T, $\mathscr{N}(A) \subset \mathscr{N}(I - T)$. Any element of $y \in \mathscr{N}(I - T)$ which does not belong to $\mathscr{N}(A)$ has to have a form $y = Ax$ for some x and $y \neq 0$. Since $Q \geqslant O$, we have that $y \geqslant 0$. On the other hand $y^{\mathrm{T}}e = x^{\mathrm{T}}A^{\mathrm{T}}e = 0$, a contradiction. Since we then have that $\mathscr{N}(I - T) = \mathscr{N}(A)$, the existence of a splitting of the form (4.2) follows from Theorem 2.1 of [3]. The fact that $T \geqslant O$ indicates that this splitting is of nonnegative type.

With this splitting, using (4.3) the following identity holds $AQ = M(I - T)Q = O$, so we also have (4.4). □

Proposition 4.2. *Let the hypothesis of Theorem* 4.1 *hold. In addition, assume that each of the splittings* $A = M_i - N_i$ *is weak regular*, $i = 1, \ldots, p$, *then, the induced splitting* $A = M - N$ *is also weak regular.*

Proof. All we need to show is that $M^{-1} \geqslant O$, where M is such that $T = I - M^{-1}A$. We have that T is as in (3.11). We define Q_i, $i = 1, \ldots, p$, the matrix such that

$$I - Q_i A = \prod_{k=i}^{1} (I - E_k M_k^{-1} A).$$

Thus $Q_1 = E_1 M_1^{-1}$, and $Q_p = M^{-1}$. We show by induction that $Q_i \geqslant O$, $i = 1, \ldots, p$. Since $E_i \geqslant O$, and $M_i^{-1} \geqslant O$, $i = 1, \ldots, p$, we have in particular that $Q_1 \geqslant O$. We also have that $I - E_i M_i^{-1} A \geqslant O$, $i = 1, \ldots, p$. The proof is based on the following recursive formula:

$$Q_i = (I - E_i M_i^{-1} A) Q_{i-1} + E_i M_i^{-1}, \quad i = 2, \ldots, p, \tag{4.5}$$

from where it follows that if $Q_{i-1} \geqslant O$, we also have $Q_i \geqslant O$. The recursion follows from the following identity by a regularization argument using $A + \varepsilon I = M_i + \varepsilon I - N_i$ and the corresponding limit regime as $\varepsilon \to 0$:

$$\begin{aligned} I - Q_i A &= (I - E_i M_i^{-1} A)(I - Q_{i-1} A) \\ &= I - E_i M_i^{-1} A - Q_{i-1} A + E_i M_i^{-1} A Q_{i-1} A \\ &= I - \big((I - E_i M_i^{-1} A) Q_{i-1} + E_i M_i^{-1}\big) A. \qquad \square \end{aligned}$$

We mention that a recursion similar to (4.5) was used in [21] in the context of nonsingular A.

Corollary 4.3. *Theorem* 4.1 *and Proposition* 4.2 *apply verbatim to the case of "coarse grid" correction, by considering the additional splitting* $A = M_0 - N_0$, *with* $T_0 = M_0^{-1} N_0$ *having positive diagonals, so that* $T_{\mu c}$ *of* (3.13) *is convergent, and it induces a splitting of the appropriate type.*

An example of splittings that lead to iteration matrices satisfying the hypotheses of Theorem 4.1, Proposition 4.2 and Corollary 4.3 is described in the following proposition requiring no proof. It provides a possible modification to the local solves to guarantee that the iteration matrix defined by (3.8) does not have positive diagonals.

Proposition 4.4. *Let* $B \geqslant O$, $B^{\mathrm{T}} e = e$. *Let* $p > 1$ *be a positive integer. Let* $\alpha_1, \ldots, \alpha_p$, *be any positive real numbers. Let* $A = I - B = M_i - N_i$, $i = 0, \ldots, p$, *be defined by*

$$M_i = \pi_i^{\mathrm{T}} \begin{bmatrix} \alpha_i I + A_i & 0 \\ 0 & \alpha_i I + D_{\neg i} \end{bmatrix} \pi_i \tag{4.6}$$

and $N_i = M_i - A$, *where* π_i, A_i *and* $D_{\neg i}$ *are defined in* (3.4), (3.5) *and* (3.8), (3.9). *Then, the splittings are regular, and the diagonals of* $T_i = M_i^{-1} N_i$ *are positive,* $i = 0, \ldots, p$.

It follows that when implementing multiplicative Schwarz with the splitting (4.6) instead of solving a local problem (or the coarse problem) with a coefficient matrix A_i one needs to solve a local problem with the coefficient matrix $(\alpha_i I + A_i)$.

With these splittings, we can now define the iteration matrices (3.11) and (3.13) which by Theorem 4.1 and Corollary 4.3 are convergent, and we can thus find the stationary probability distribution v (or $\pi = v^{\mathrm{T}}$) by using the iteration (2.1) (with $c = 0$).

In [3] it was shown that in the case of two alternating iterations, i.e., for the matrix $\widehat{T} = T_1 T_2$, one needs a compatibility condition to hold in order to guarantee the existence of an induced splitting $A = M - N$ with $M^{-1} N = \widehat{T}$. This condition is that $M_1 + M_2 - A$ be nonsingular. In the case of the splitting (4.6), this condition is satisfied. Similarly, for any value of p we have that the matrix $M_1 + \cdots + M_p - A$ is also nonsingular.

Let $\gamma = \max\{|\lambda|, \lambda \in \sigma(T), \lambda \neq 1\}$. The fact that T is convergent implies that $\gamma < 1$; see, e.g., [4]. Therefore Theorem 4.1 indicates that for multiplicative Schwarz, $\sigma(M^{-1}A) = \sigma(I - T)$ has zero as an isolated eigenvalue with index 1, and the rest of the spectrum is contained in a ball with center 1 and radius γ. Furthermore, the smaller γ is, the smaller this ball around 1 is. This configuration of the spectrum often gives good convergence properties to Krylov subspace methods preconditioned with multiplicative Schwarz.

5. Stationary probability distribution

We assume in this section that $B = P^{\mathrm{T}}$ is irreducible, i.e., that the Markov chain is ergodic. In this case, every principal submatrix of order $n - 1$ of matrix $A = I - B$, is a nonsingular M-matrix [4]. One approach used with direct methods is to "remove an equation", i.e., to reduce the problem of finding the stationary probability distribution v to a linear system with a nonsingular $(n - 1) \times (n - 1)$ matrix [25]. We take this approach here for the Schwarz iterations (2.1).

Without loss of generality we assume that the "removed" equation is the last one, and partition A (and B) as follows

$$A = \begin{bmatrix} \tilde{A} & -g \\ -d^{\mathrm{T}} & 1 - b_{nn} \end{bmatrix}, \tag{5.1}$$

where $\tilde{A}$ is $(n - 1) \times (n - 1)$, $g, d \in \mathbb{R}^{n-1}$, $g^{\mathrm{T}} = (b_{1n}, \ldots, b_{n-1,n})$, and $d^{\mathrm{T}} = (b_{n1}, \ldots, b_{n,n-1})$. Let $\tilde{v} \in \mathbb{R}^{n-1}$, $\tilde{v}^{\mathrm{T}} = (v_1, \ldots, v_{n-1})$, then our problem $Av = 0$ reduces to

$$\tilde{A}\tilde{v} = v_n g. \tag{5.2}$$

Since we have one degree of freedom, as with direct methods, one fixes the value of v_n, say $v_n = 1$ and once (5.2) is solved, the entries are renormalized so that $e^{\mathrm{T}}v = 1$.

Since $\tilde{A}$ is a nonsingular M-matrix, we can use the multiplicative Schwarz iterations (2.1), or the additive ones with $\theta < 1/q$, as described in Section 3 on the "reduced" system (5.2). One can also use the iterations with the "coarse grid" correction (3.13) and (3.14). All these are convergent. Furthermore, all results from [2,12] apply here, e.g., the possible improvement of the rate of convergence when the overlap increases.

In the rest of this section we present a convergent weak regular splitting of A. It is constructed from the splitting of $\tilde{A}$ induced by either the additive or multiplicative Schwarz iterations (with or without the "coarse grid" correction). Let $\tilde{A} = \tilde{M} - \tilde{N}$ be such that $\tilde{M}^{-1}\tilde{N} = \tilde{T}$, the matrix of the form (3.11) or that of the form (3.12) with $\theta < 1/q$, when applied to $\tilde{A}$. Recall that these are weak regular splittings and they are uniquely determined.

Proposition 5.1. *Let*

$$M = \begin{bmatrix} \tilde{M} & 0 \\ -d^{\mathrm{T}} & 1 - b_{nn} \end{bmatrix}, \quad N = \begin{bmatrix} \tilde{N} & g \\ 0 & 0 \end{bmatrix}. \tag{5.3}$$

Then, $A = M - N$ *is a weak regular splitting and* $T = M^{-1}N$ *is convergent*.

Proof. It is obvious from the fact that $\tilde{A} = \tilde{M} - \tilde{N}$ and from the form of A in (5.1) that $A = M - N$. Since $\rho(\tilde{M}^{-1}\tilde{N}) < 1$, the iterative process $\tilde{M}\tilde{v}^{k+1} = \tilde{N}\tilde{v}^k + g$, $k = 0, 1, \ldots,$ is convergent for any $\tilde{v}^0$; cf. (2.1). This is equivalent to the convergence of the following iteration

$$\begin{bmatrix} \tilde{M} & 0 \end{bmatrix} \begin{bmatrix} \tilde{v}^{k+1} \\ 1 \end{bmatrix} = \begin{bmatrix} \tilde{N} & g \end{bmatrix} \begin{bmatrix} \tilde{v}^{k} \\ 1 \end{bmatrix},$$

for $k = 0, 1, \ldots$; cf. [26]. Let $v^{\mathrm{T}} = (\tilde{v}^{\mathrm{T}}, 1)$ be the limit of this iterative process, i.e., such that $Av = 0$. We now "add back" formally the last equation and have the equivalent iteration

$$\begin{bmatrix} \tilde{M} & 0 \\ -d^{\mathrm{T}} & 1 - b_{nn} \end{bmatrix} \begin{bmatrix} \tilde{v}^{k+1} \\ 1 \end{bmatrix} = \begin{bmatrix} \tilde{N} & g \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \tilde{v}^{k} \\ 1 \end{bmatrix} + \begin{bmatrix} 0 \\ \epsilon_k \end{bmatrix},$$

where $\epsilon_k = -d^{\mathrm{T}}\tilde{v}^{k+1} + (1 - b_{nn})$, for $k = 0, 1, \ldots$, which is thus convergent. Since $Av = 0$, we have that $\lim_{k\to\infty} \epsilon_k = 0$, and thus, $\lim_{k\to\infty} T^k$ exists.

A simple calculation shows that

$$M^{-1} = \begin{bmatrix} \tilde{M}^{-1} & 0 \\ \frac{1}{1-b_{nn}} d^{\mathrm{T}}\tilde{M}^{-1} & \frac{1}{1-b_{nn}} \end{bmatrix} \geqslant 0.$$

We can explicitly compute T and obtain

$$M^{-1}N = \begin{bmatrix} \tilde{M}^{-1}\tilde{N} & \tilde{M}^{-1}g \\ \frac{1}{1-b_{nn}} d^{\mathrm{T}}\tilde{M}^{-1}\tilde{N} & \frac{1}{1-b_{nn}} d^{\mathrm{T}}\tilde{M}^{-1}g \end{bmatrix}, \tag{5.4}$$

which is nonnegative since $\tilde{M}^{-1} \geqslant O$ and $\tilde{M}^{-1}\tilde{N} \geqslant O$. □

We remark that the splitting (5.3) is not the only one which can produce the iteration matrix (5.4); see [3].

6. Mean first passage matrix

In Markov chain modeling, in addition to the stationary probability vectors, it is often important to obtain the moment matrices [25]. They are defined as follows

$$F = (f_{jk}), \quad f_{jk} = \sum_{m=1}^{\infty} f_{jk}^{(m)}, \quad j, k = 1, \ldots, n,$$
$$M = (m_{jk}), \quad m_{jk} = \sum_{m=1}^{\infty} m f_{jk}^{(m)}, \quad j, k = 1, \ldots, n,$$
$$W = (w_{jk}), \quad w_{jk} = \sum_{m=1}^{\infty} m^2 f_{jk}^{(m)}, \quad j, k = 1, \ldots, n,$$

where $f_{jk}^{(m)}$ denotes the probability that first return to state j occurs exactly m steps after leaving from state k. These matrices depend directly on the stochastic matrix $P = B^{\mathrm{T}}$ representing the Markov chain, but we do not denote this dependency explicitly.

In computations the following formulas can be used [16],

$$(I - B^{\mathrm{T}})F = B^{\mathrm{T}}(I - F_D), \quad F_D = \mathrm{diag}(F), \tag{6.1}$$
$$(I - B^{\mathrm{T}})M = F - B^{\mathrm{T}} M_D, \quad M_D = \mathrm{diag}(M), \tag{6.2}$$
$$(I - B^{\mathrm{T}})W = 2M - F - B^{\mathrm{T}} W_D, \quad W_D = \mathrm{diag}(W). \tag{6.3}$$

We mention that the expressions (6.1)–(6.3) derived in [16] are valid for any transition matrix $B^{\mathrm{T}} = P$, including the reducible case. Naturally, these formulas reduce to the well-known relations when $B^{\mathrm{T}} = P$ is irreducible; see, e.g., [25, pp. 9–10].

We assume in the rest of this section that $B = P^{\mathrm{T}}$ is irreducible. The reducible case is treated in the next section.

As example of moment matrices belonging to a Markov chain let us consider the first moment matrix M called *mean first passage matrix* [25]. This $n \times n$ matrix is under the irreducibility assumption a solution of the following matrix equation

$$M = E + B^{\mathrm{T}}[M - \mathrm{diag}(M)], \tag{6.4}$$

where $E = \begin{bmatrix} 1 & \cdot & \cdot & \cdot & 1 \\ 1 & \cdot & \cdot & \cdot & 1 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & \cdot & \cdot & \cdot & 1 \end{bmatrix} = ee^{\mathrm{T}}$, cf. [15].

It is easy to see that the linear system (6.4) defining the matrix M can be equivalently written as n linear systems, one for each of the columns of M, denoted as $M^{(j)}$, i.e.,

$$(I - B^{\mathrm{T}})M^{(j)} = Ee_j - B^{\mathrm{T}} e_j e_j^{\mathrm{T}} M e_j,$$

where e_j is the jth standard vector. Since

$$m_{jj}e_j = e_j e_j^{\mathrm{T}} M e_j = e_j e_j^{\mathrm{T}} M^{(j)},$$

the resulting systems read

$$\left[I - B^{\mathrm{T}}(I - e_j e_j^{\mathrm{T}})\right] M^{(j)} = e, \quad j = 1, \ldots, n. \tag{6.5}$$

Due to irreducibility of B^{T} the coefficient matrix of each system (6.5) is nonsingular. Consequently, this form of the defining system is computationally very suitable. In particular, additive and multiplicative Schwarz iterative methods as described in Section 3 apply in the form described in [2].

7. The reducible case

We consider here the general case, where $B = P^{\mathrm{T}}$ might not be irreducible. There is a permutation matrix H such that the symmetric permutation of B is lower block-triangular [13, p. 341], and in fact it has the following form (the so-called Romanovskij canonical form)

$$HBH^{\mathrm{T}} = \begin{bmatrix} G_0 & O & \cdots & O \\ G_1 & C_1 & \cdots & O \\ \vdots & \vdots & \ddots & \vdots \\ G_p & O & \cdots & C_p \end{bmatrix}, \tag{7.1}$$

where $\lim_{k\to\infty} G_0^k = O$ and C_i is an irreducible and stochastic matrix, $i = 1, \ldots, p$. There are efficient algorithms to compute the permutation matrix H, and thus, the form (7.1). For example, Tarjan's algorithm has almost linear complexity and good software is available for it [11].

Solving linear systems with the matrix B, or $P = B^{\mathrm{T}}$, reduces then to solving systems with each of the diagonal blocks of (7.1). This consideration applies equally to the computation of the stationary probability vectors as to the moment matrices F, M, W, etc. This can be accomplished using the techniques described in Sections 4 and 5 for irreducible stochastic matrices. In particular, this means that once the Romanovskij canonical form (7.1) of B is known, the moment matrices can be computed by applying Schwarz methods to each block or more precisely to each block column separately in the manner shown in Section 6.

8. Concluding remarks

We have described several computational approaches to the numerical solution of Markov chains using (additive and) multiplicative Schwarz methods. Our main result is a new convergence theorem for multiplicative Schwarz iterations for singular systems. In the irreducible case, we have also exploited the fact that any principal matrix is nonsingular. In the reducible case, one can reduce the problem to several irreducible smaller problems. The attractive properties of multiplicative Schwarz iterations such as the monotonicity of the convergence rate with respect to changes of the number of diagonal blocks, their sizes, etc., presented in [2] carry over naturally to the problems treated here.

Acknowledgements

We thank Michele Benzi, Olof Widlund, and the referees for their comments and suggestions which helped improve our presentation.

References

[1] G. Alefeld, N. Schneider, On square roots of M-matrices, Linear Algebra Appl. 42 (1982) 119–132.
[2] M. Benzi, A. Frommer, R. Nabben, D.B. Szyld, Algebraic theory of multiplicative Schwarz methods, Numer. Math. 89 (2001) 605–639.
[3] M. Benzi, D.B. Szyld, Existence and uniqueness of splittings for stationary iterative methods with applications to alternating methods, Numer. Math. 76 (1997) 309–321.
[4] A. Berman, R.J. Plemmons, Nonnegative Matrices in the Mathematical Sciences, third ed., Academic Press, New York, 1979, Reprinted by SIAM, Philadelphia, 1994 (updated).
[5] P.E. Bjørstad, O.B. Widlund, To overlap or not to overlap: a note on a domain decomposition method for elliptic problems, SIAM J. Sci. Stat. Comput. 10 (1989) 1053–1061.
[6] E. Bohl, I. Marek, A model of amplification, J. Comput. Appl. Math. 63 (1995) 27–47.
[7] T.F. Chan, T.P. Mathew, Domain decomposition methods, Acta Numer. (1994) 61–143.
[8] N. Debit, M. Garbey, R. Hoppe, J. Périaux, D. Keyes, Y. Kuznetzov (Eds.), Proceedings of the Thirteenth International Conference on Domain Decomposition Methods in Lyon, France, October 2000, CIMNE, UPS, Barcelona, 2001. See also http://www.ddm.org.
[9] M. Dryja, B. Smith, O.B. Widlund, Schwarz analysis of iterative substructuring algorithms for elliptic problems in three dimensions, SIAM J. Numer. Anal. 31 (1994) 1662–1694.
[10] M. Dryja, O.B. Widlund, Domain decomposition algorithms with small overlap, SIAM J. Sci. Stat. Comput. 15 (1994) 604–620.
[11] I.S. Duff, J.K. Reid, An implementation of Tarjan's algorithm for the block triangularization of a matrix, ACM Trans. Math. Software 4 (1978) 147–337.
[12] A. Frommer, D.B. Szyld, Weighted max norms, splittings, and overlapping additive Schwarz iterations, Numer. Math. 83 (1999) 259–278.
[13] F.R. Gantmacher, Applications of the Theory of Matrices, Interscience, New York, 1959.
[14] I. Herrera, D. Keyes, O. Widlund, R. Yates, Proceedings of the Fourteenth International Conference on Domain Decomposition Methods, Cocoyoc, Mexico, January 2002. National Autonomous University of Mexico (UNAM), Mexico City, Mexico, 2003. See also http://www.ddm.org.
[15] J.G. Kemeny, J.L. Snell, Finite Markov Chains, van Nostrand, Princeton, NY, 1960, Reprinted, Springer, New York, 1976.
[16] Š. Klapka, Markov chain modeling in railways safety techniques, Ph.D. Thesis, Charles University, Faculty of Mathematics and Physics, Prague, 2002.

[17] U. Krieger, Numerical solution of large finite Markov chains by algebraic multigrid techniques, in: W.J. Stewart (Ed.), Computations with Markov Chains, Kluwer Academic, Boston, 1995, pp. 403–424.

[18] I. Marek, Frobenius theory of positive operators, comparison theorems and applications, SIAM J. Appl. Math. 19 (1970) 608–628.

[19] I. Marek, D.B. Szyld, Comparison theorems for the convergence factor of iterative methods for singular matrices, Linear Algebra Appl. 316 (2000) 67–87.

[20] I. Marek, D.B. Szyld, Comparison of convergence of general stationary iterative methods for singular matrices, SIAM J. Matrix Anal. Appl. 24 (2002) 68–77.

[21] R. Nabben, Comparisons between additive and multiplicative Schwarz iterations in domain decomposition methods, Numer. Math. 95 (2003) 145–162.

[22] J.M. Ortega, W. Rheinboldt, Iterative Solution of Nonlinear Equations in Several Variables, Academic Press, New York, San Francisco, London, 1970, Reprinted by SIAM, Philadelphia, 2000.

[23] A. Quarteroni, A. Valli, Domain Decomposition Methods for Partial Differential Equations, Oxford Science Publications, Clarendon Press, Oxford, 1999.

[24] B.F. Smith, P.E. Bjørstad, W.D. Gropp, Domain Decomposition: Parallel Multilevel Methods for Elliptic Partial Differential Equations, Cambridge University Press, Cambridge, New York, Melbourne, 1996.

[25] W.J. Stewart, Introduction to the Numerical Solution of Markov Chains, Princeton University Press, Princeton, NJ, 1994.

[26] J. Stoer, R. Bulirsch, Introduction to Numerical Analysis, third ed., Springer Verlag, New York, 2002.

[27] D.B. Szyld, Equivalence of convergence conditions for iterative methods for singular equations, Numer. Linear Algebra Appl. 1 (1994) 151–154.

[28] A.E. Taylor, D.C. Lay, Introduction to Functional Analysis, second ed., John Wiley and Sons, New York, 1980.

[29] R.S. Varga, Matrix Iterative Analysis, Prentice-Hall, Englewood Cliffs, New Jersey, 1962, Second ed., revised and expanded, Springer, Berlin, Heidelberg, New York, 2000.

[30] D.M. Young, Iterative Solution of Large Linear Systems, Academic Press, New York, 1971.