NUMERICAL LINEAR ALGEBRA WITH APPLICATIONS
Numer. Linear Algebra Appl. 2011; **18**:1011–1029
Published online 15 November 2011 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/nla.821

Restricted additive Schwarz methods for Markov chains

Michele Benzi*,† and Verena Kuhlemann

Department of Mathematics and Computer Science, Emory University, Atlanta, GA 30322, USA

SUMMARY

The restricted additive Schwarz method is adapted to the problem of computing the stationary probability distribution vector of large, sparse, and irreducible stochastic matrices. Inexact and two-level variants are also considered as well as acceleration by Krylov subspace methods. The convergence properties are analyzed, and extensive numerical experiments aimed at assessing the effect of varying the number of subdomains and the amount of overlap are discussed. Copyright © 2011 John Wiley & Sons, Ltd.

Received 29 October 2010; Revised 24 August 2011; Accepted 28 August 2011

KEY WORDS: iterative methods; preconditioning; Schwarz methods; Markov chains GMRES

1. INTRODUCTION

Domain decomposition methods are widely used for solving large-scale linear systems of equations arising from the discretization of PDEs on parallel computers [1]. Of particular importance among domain decomposition schemes are variants of the additive and multiplicative Schwarz methods, with or without overlap. These algorithms can be cast as stationary iterations associated with matrix splittings, making a purely algebraic analysis possible; see, for example, [2, 3]. Although there are situations where Schwarz-type methods exhibit rapid convergence, they are usually more efficient when used as preconditioners for Krylov subspace methods; see [4]. In this paper, we investigate the application of various additive Schwarz (AS) methods as solvers and as preconditioners for computing the stationary probability distribution vector of ergodic Markov chains with large and sparse transition matrices. In particular, we give the first analysis and implementation of the *restricted additive Schwarz* (RAS) method in the Markov chain context. The RAS method, first proposed in [5] as a preconditioner for nonsingular linear systems $Ax = b$, is a variant of the AS method that requires less communication and is therefore better suited for parallel implementation. Somewhat surprisingly, this method tends to exhibit convergence rates that are no worse and often better than those of the standard AS method; see [6] for an algebraic analysis in the nonsingular case. We note that RAS is the default Schwarz preconditioner for linear systems in the popular PETSc package [7]. AS methods (but not RAS) for Markov chains have been analyzed in [8, 9]. The focus of these papers is primarily theoretical, and the use of Krylov subspace acceleration is not considered. Here, we present some analysis and the results of computational experiments on realistic Markov models, including a selection of problems from the MARCA collection [10] and from certain reliability models. Our results show that RAS preconditioning (with inexact subdomain solves) is a promising approach to the solution of large, sparse Markov chain models.

*Correspondence to: Michele Benzi, Department of Mathematics and Computer Science, Emory University, Atlanta, GA 30322, USA.
†E-mail: benzi@mathcs.emory.edu

2. DEFINITIONS AND AUXILIARY RESULTS

First, we introduce some terminology and notation. If B is a matrix, we write $B \geqslant 0$ ($B > 0$) if its entries are nonnegative (positive). With $\sigma(A)$, we denote the spectrum and with $R(A)$ the range of A. For $\lambda \in \sigma(A)$, the *index* of A with respect to λ is the smallest integer k for which $R((\lambda I - A)^{k+1}) = R((\lambda I - A)^k)$. Note that this is the size of the largest Jordan block associated with the eigenvalue λ in the Jordan normal form of A. The index is denoted with $\mathrm{ind}_\lambda A$. A $n \times n$ matrix A is an *M-matrix* if $A = \beta I - B$, where I is the $n \times n$ identity, $B \geqslant 0$ and $\rho(B) \leqslant \beta$, where $\rho(B)$ is the spectral radius of B. If $\rho(B) < \beta$, then A is a nonsingular M-matrix, and $A^{-1} \geqslant 0$. Otherwise, A is singular. Recall that A is *irreducible* if the directed graph associated with A is strongly connected.

We are interested in computing the stationary probability distribution vector of finite, ergodic Markov chains. Following [11], we identify ergodic chains with irreducible ones; in particular, we allow periodic chains. The problem amounts to finding a nontrivial solution to a homogeneous system of linear equations $Ax = 0$, where A is a singular, irreducible M-matrix. It is well known that up to normalization, such solution vector x is unique [12]. This formulation applies to discrete-time as well as to continuous-time Markov chains. For a discrete-time Markov chain, we have $A = I - P^T$ where P is the row-stochastic matrix of transition probabilities. Recall that $A = M - N$ is a *splitting* if M is nonsingular. The splitting is *regular* if $M^{-1} \geqslant 0$ and $N \geqslant 0$; it is *weak regular* if $M^{-1} \geqslant 0$ and $M^{-1}N \geqslant 0$. A matrix A is *weak semipositive* if there exists a vector $x > 0$ (i.e., an entry-wise positive vector) such that $Ax \geqslant 0$; see [13]. By the Perron–Frobenius Theorem [11], it is then clear that any singular, irreducible M-matrix is weak semipositive. The standard stationary iteration associated with the splitting $A = M - N$ is of the form $x^{k+1} = Tx^k + c$, $k = 0, 1, \ldots$, where $T = M^{-1}N$, $c = M^{-1}b$, and x^0 are arbitrary. It converges for all choices of x^0 if either $\rho(T) < 1$ (in which case $\lim_{k\to\infty} T^k = 0$), or $\rho(T) = 1$, and T is *convergent*, that is, $\lim_{k\to\infty} T^k$ exists. In the cases of interest to us, T will have unit spectral radius. Such a matrix is convergent if and only if $\mathrm{ind}_1 T = 1$ and $\lambda \in \rho(T)\backslash\{1\}$ implies $|\lambda| < 1$. For $T \geqslant 0$, the latter condition can be replaced with T having positive diagonal entries [14]. Also observe that in the Markov chain context, we have $c = 0$, and that the sequence $\{x^k\}$ converges to a nontrivial solution of $Ax = 0$ only if $x^0 \notin R(M^{-1}A)$.

Finally, we will need the following result from [13, Thm. 6] in our analysis of convergence.

Theorem 2.1
Let $A = M - N$ be a weak regular splitting of $A \in \mathbf{R}^{n\times n}$, and let $T = I - M^{-1}A = M^{-1}N$. If A is weak semipositive and singular, then $\rho(T) = 1$ and $\mathrm{ind}_1 T = 1$.

In the terminology of [13], such an iteration matrix T is said to be *semiconvergent*.

3. ALGEBRAIC FORMULATION OF RESTRICTED ADDITIVE SCHWARZ

We adopt the notation and the approach developed in [8, 9]. Consider the set of indices $S = \{1, \ldots, n\}$ and let $S = \cup_{i=1}^K S_{i,0}$ be a partition of S into K disjoint, non-empty subsets. For each $S_{i,0}$ consider $S_{i,\delta}$ with $S_{i,0} \subseteq S_{i,\delta}$. Thus, $S = \cup_{i=1}^K S_{i,\delta}$ for all values of δ, with not necessarily pairwise disjoint $S_{i,\delta}$. For $\delta > 1$, this notation introduces overlap. One way to find this overlap is by considering the underlying undirected graph of A. If A does not have a symmetric pattern, we use the graph of $A + A^T$ instead. To each set of nodes $S_{i,0}$, we add all nodes with distance at most δ in the underlying graph. Here, as usual, the distance is defined as the length of the shortest path in the graph connecting a given node to any node in $S_{i,0}$. The matrix $R_{i,\delta}$ is a restriction operator from the whole space $\mathbf{R}^n$ to the subspace defined by $S_{i,\delta}$. Let $\pi_{i,\delta}$ be the matrix representation of the permutation that relabels the states in $S_{i,\delta}$ as $1, \ldots, n_{i,\delta}$, where $n_{i,\delta} = |S_{i,\delta}|$. Then $R_{i,\delta}$ is an $n_{i,\delta} \times n$ matrix given by $R_{i,\delta} = [I_{i,\delta}|0]\pi_{i,\delta}$, where $I_{i,\delta}$ is the $n_{i,\delta} \times n_{i,\delta}$ identity matrix. The matrix

$$A_{i,\delta} = R_{i,\delta} A R_{i,\delta}^T \tag{1}$$

DOI: 10.1002/nla

is the restriction of A to the subspace corresponding to $S_{i,\delta}$. Thus, $A_{i,\delta}$ is an $n_{i,\delta} \times n_{i,\delta}$ principal submatrix of A. If A is an irreducible singular M-matrix, $A_{i,\delta}$ is a nonsingular M-matrix [12]. Define $E_{i,0}$ as

$$E_{i,0} = R_{i,0}^T R_{i,0} = \pi_{i,0}^T \begin{pmatrix} I_{i,0} & 0 \\ 0 & 0 \end{pmatrix} \pi_{i,0}. \tag{2}$$

$E_{i,0}$ is an $n \times n$ diagonal matrix with a 1 on the diagonal for every row where $R_{i,0}^T$ has a 1, and zeroes otherwise. Thus, $\sum_{i=1}^K E_{i,0} = I$. The AS method can be given in the form of a stationary iteration, $x^{k+1} = Tx^k + c$, where

$$T = I - \sum_{i=1}^{K} R_{i,\delta}^T A_{i,\delta}^{-1} R_{i,\delta} A \tag{3}$$

and c is a certain vector. Note that for the solution of Markov chains, c will be 0. For the RAS iteration, the prolongation operator $R_{i,\delta}^T$ will be replaced by a prolongation operator that does not consider overlap. Define the "restricted" operator $\tilde{R}_{i,\delta}$ as

$$\tilde{R}_{i,\delta} = R_{i,\delta} E_{i,0}. \tag{4}$$

With this notation, the RAS method has the form of a stationary iteration, $x^{k+1} = Tx^k + c$, where $T = I - \sum_{i=1}^K \tilde{R}_{i,\delta}^T A_{i,\delta}^{-1} R_{i,\delta} A$.. We will use

$$M_{RAS,\delta}^{-1} := \sum_{i=1}^{K} \tilde{R}_{i,\delta}^T A_{i,\delta}^{-1} R_{i,\delta} \tag{5}$$

as a right preconditioner for the generalized minimal residual method (GMRES). That is, we will use GMRES to solve $AM_{RAS,\delta}^{-1} y = 0$, with $y = M_{RAS,\delta} x$. The notation $M_{RAS,\delta}^{-1}$ only makes sense if $\sum_{i=1}^K \tilde{R}_{i,\delta}^T A_{i,\delta}^{-1} R_{i,\delta}$ is a nonsingular matrix. Guaranteeing the nonsingularity of this matrix is not entirely straightforward. A simple solution would be to construct the RAS preconditioner not using the original singular matrix A, but a slightly perturbed matrix $\tilde{A} = A + \epsilon I$ for some $\epsilon > 0$. Because this matrix is guaranteed to be a nonsingular M-matrix, it follows from the general results in [6] that the corresponding RAS preconditioner is well defined, that is, nonsingular. The resulting ϵ-dependent RAS preconditioner would then be applied, of course, to the original (singular) system, $Ax = 0$. This approach, however, besides not being very elegant, requires the introduction of an additional parameter ϵ.

An alternative approach is to work with the original matrix A, but to impose conditions on the decomposition (graph partition) that will automatically produce a nonsingular RAS preconditioner, such as those specified in Theorem 3.2 next. We begin by writing

$$\pi_{i,\delta} A {\pi_{i,\delta}}^T = \begin{pmatrix} A_{i,\delta} & K_{i,\delta} \\ L_{i,\delta} & A_{\neg i,\delta} \end{pmatrix}, \tag{6}$$

where $A_{\neg i,\delta}$ is the principal submatrix of A "complementary" to $A_{i,\delta}$. Thus, $A_{\neg i,\delta}$ is a nonsingular M-matrix. Let $D_{\neg i,\delta} = \mathrm{diag}(A_{\neg i,\delta})$ and note that because $A_{\neg i,\delta}$ is a nonsingular M-matrix, $D_{\neg i,\delta}$ is nonnegative with positive diagonal entries. We construct a matrix $M_{i,\delta}$ corresponding to $R_{i,\delta}$ as follows:

$$M_{i,\delta} = \pi_{i,\delta}^T \begin{pmatrix} A_{i,\delta} & 0 \\ 0 & D_{\neg i,\delta} \end{pmatrix} \pi_{i,\delta}. \tag{7}$$

Because $A_{i,\delta}$ is nonsingular, and $D_{\neg i,\delta}$ has positive entries on the diagonal, $M_{i,\delta}$ is invertible. It has been proven by Frommer and Szyld [6] that

$$\tilde{R}_{i,\delta}^T A_{i,\delta}^{-1} R_{i,\delta} = E_{i,0} M_{i,\delta}^{-1}. \tag{8}$$

DOI: 10.1002/nla

Thus, $M_{RAS,\delta}^{-1}$ can be written as

$$M_{RAS,\delta}^{-1} = \sum_{i=1}^{K} E_{i,0} M_{i,\delta}^{-1}, \tag{9}$$

and we will use this representation in our analysis of the RAS method for Markov chains in the remainder of this section and of the next one.

We will make use of the following lemma from [15]. We denote by $F_{i,j}$ a matrix with all 0 entries except for a 1 in position (i, j).

Lemma 3.1
Assume that A is an irreducible singular M-matrix, the splittings $A = M_l - N_l$, $l = 1, \ldots K$ are regular, and that E_l, $l = 1, \ldots, K$ are nonnegative diagonal matrices such that $\sum_{l=1}^{K} E_l = I$. If there exist indices $1 \leqslant i, j \leqslant n$ such that the (i, j)-entry in N_l is nonzero for each $l = 1, \ldots, K$, then for sufficiently small $\epsilon > 0$, the splittings

$$A + \epsilon F_{i,j} = M_l - (N_l - \epsilon F_{i,j}), \quad l = 1, \ldots, K$$

are regular splittings of a nonsingular M-matrix. Thus, $M = \sum_{l=1}^{K} E_l M_l^{-1}$ is nonsingular regardless of the choice of the weighting matrices E_l.

We emphasize that in the preceding lemma, the matrices M_l are independent of ϵ. We can now state sufficient conditions that guarantee the nonsingularity of the RAS preconditioner.

Theorem 3.2
Given any set of overlapping subdomains $S_{1,\delta}, \ldots, S_{K,\delta}$, there exists a set of overlapping subdomains $\hat{S}_{1,\delta}, \ldots, \hat{S}_{K,\delta}$ such that $S_{i,0} \subseteq \hat{S}_{i,\delta} \subseteq S_{i,\delta}$ and the preconditioning matrix $M_{RAS}^{-1} = \sum_{l=1}^{K} E_{l,\delta} M_{l,\delta}^{-1}$ corresponding to $\hat{S}_{1,\delta}, \ldots, \hat{S}_{K,\delta}$ is nonsingular.

Proof
Because A is irreducible, there exist $i \in S_{1,0}$ and $j \notin S_{1,0}$ such that the (i, j)-entry of A is nonzero. Now, we construct the sets $\hat{S}_{1,\delta}, \ldots, \hat{S}_{K,\delta}$ from the sets $S_{1,\delta}, \ldots, S_{K,\delta}$ as follows: Set $\hat{S}_{1,\delta} = S_{1,\delta} - \{j\}$. Let s be the index of the set such that $j \in S_{s,0}$. Set $\hat{S}_{s,\delta} = S_{s,\delta} - \{i\}$. Also, we remove i and j from all other sets, that is, $\hat{S}_{l,\delta} = S_{l,\delta} - \{i, j\}$ for $l \in \{2, \ldots, K\} - \{s\}$. We also have to make sure that none of the subdomains coincides with the whole domain S. If there is some $\hat{S}_{l,\delta}$ that does contain everything, we simply remove any node lying in the overlapping region of this subdomain. That is, we pick $j \in \hat{S}_{l,\delta} - S_{l,0}$ at random and set $\hat{S}_{l,\delta} = \hat{S}_{l,\delta} - \{j\}$.

Consider the matrix splittings $A = M_{l,\delta} - N_{l,\delta}$ corresponding to the new subdomains. With

$$A = \pi_{l,\delta}^T \begin{pmatrix} A_{l,0} & K_{l,\delta} \\ L_{l,\delta} & A_{\neg l,\delta} \end{pmatrix} \pi_{l,\delta} \quad \text{and} \quad M_{l,\delta} = \pi_{l,\delta}^T \begin{pmatrix} A_{l,0} & 0 \\ 0 & D_{\neg l,\delta} \end{pmatrix} \pi_{l,\delta}$$

it follows that $N_{l,\delta} = \pi_{l,\delta}^T \begin{pmatrix} 0 & -K_{l,\delta} \\ -L_{l,\delta} & -A_{\neg l,\delta} + D_{\neg l,\delta} \end{pmatrix} \pi_{l,\delta}$. Because A has nonpositive off-diagonal entries, it follows that $N_{l,\delta} \geqslant 0$. Because $M_{l,\delta}^{-1} \geqslant 0$, the splittings $A = M_{l,\delta} - N_{l,\delta}$ are regular. With the updated set of overlapping subdomains $\hat{S}_{1,\delta}, \ldots, \hat{S}_{K,\delta}$, the conditions for the previous lemma are satisfied. If the (i, j)-entry of A is given by $a_{i,j}$ ($\neq 0$ by the choice of i and j), then the (i, j)-entry in $N_{l,\delta}$,$l = 1, \ldots, K$, is given by $-a_{i,j}$, because i and j are never in the same subdomain. It follows from the previous lemma that $M = \sum_{l=1}^{K} E_{l,\delta} M_{l,\delta}^{-1}$ is nonsingular. □

It is worth stressing that the conditions on the subdomains expressed in Theorem 3.2 are sufficient, but not necessary. In practice, we have not found any cases where a given overlapping set of subdomains produced a singular preconditioner, and therefore we have not needed to modify any

DOI: 10.1002/nla

subdomains so as to satisfy the theorem's conditions. On the other hand, we have been unable to prove nonsingularity of the RAS preconditioner without any additional assumptions. Because nonsingularity of the preconditioner is an essential requirement for the analysis in the next sections, we henceforth assume that the conditions in the previous theorem are satisfied.

4. PROPERTIES OF THE RESTRICTED ADDITIVE SCHWARZ SPLITTING

In this section, we analyze the convergence of the RAS iteration for irreducible singular M-matrices. In particular, in Propositions 4.2–4.3, we state sufficient conditions under which the stationary RAS iteration converges to a nontrivial solution of the linear system $Ax = 0$ for almost all initial guesses, that is, conditions under which the iteration matrix of the RAS method is convergent.

Recall that the iterative method $x^{k+1} = Tx^k$ is convergent if the $\lim_{k\to\infty} T^k$ exists. As is well known (see, e.g.,[13]), a matrix T with $\rho(T) = 1$ is convergent if and only if the following conditions are both satisfied: (i) $\lambda \in \sigma(T)$ with $|\lambda| = 1$ implies $\lambda = 1$; (ii) $\mathrm{ind}_1 T = 1$; that is, $\mathrm{rank}\,(I-T) = \mathrm{rank}\,(I-T)^2$. The last condition is equivalent to T having only Jordan blocks of size 1 associated with the eigenvalue 1. Note that given an iteration matrix $T = M^{-1}N = I - M^{-1}A$ associated with a splitting $A = M - N$ with A singular, one necessarily has $1 \in \sigma(T)$ and thus $\rho(T) \geqslant 1$.

We begin our discussion of convergence by first showing that the iteration matrix of the RAS method satisfies $\rho(T) = 1$ and $\mathrm{ind}_1 T = 1$. Because of Theorem 2.1, it suffices to show that the RAS splitting is weak regular.

Theorem 4.1
Let A be an irreducible singular M-matrix, and let $M^{-1}_{RAS,\delta}$ be given by Equation (9). Then the splitting $A = M_{RAS,\delta} - N_{RAS,\delta}$ is weak regular.

Proof
With $E_{i,0} \geqslant 0$ and $M^{-1}_{i,\delta} \geqslant 0$, it follows that $M^{-1}_{RAS,\delta} = \sum_{i=1}^{K} E_{i,0} M^{-1}_{i,\delta} \geqslant 0$. Let $N_{i,\delta} = M_{i,\delta} - A$, then

$$\begin{aligned} M^{-1}_{RAS,\delta} N_{RAS,\delta} &= I - \sum_{i=1}^{K} \left(E_{i,0} M^{-1}_{i,\delta} A\right) \\ &= I - \sum_{i=1}^{K} E_{i,0} + \sum_{i=1}^{K} E_{i,0} M^{-1}_{i,\delta} N_{i,\delta} \\ &= \sum_{i=1}^{K} E_{i,0} M^{-1}_{i,\delta} N_{i,\delta}. \end{aligned}$$

The last equality holds because $\sum_{i=1}^{K} E_{i,0} = I$. Because $M^{-1}_{i,\delta} N_{i,\delta} \geqslant 0$ and $E_{i,0} \geqslant 0$, we obtain $M^{-1}_{RAS,\delta} N_{RAS,\delta} \geqslant 0$. □

The previous result, combined with Theorem 2.1, implies that the iteration matrix T of the RAS method is semiconvergent. For it to be convergent, we additionally need to show that $\lambda = 1$ is the only eigenvalue of T on the unit circle. As mentioned in section 2, this is equivalent to T having no zero diagonal entries, see [14]. In general, T may not have all positive entries along the diagonal. There are essentially two ways to enforce this condition. One approach (see, e.g., [8,9]) is to slightly modify the splittings of A so as to ensure positive diagonals. Adding a small positive constant to the diagonal entries of $M_{i,\delta}$ guarantees that all the diagonal entries of $T_{i,\delta} = M^{-1}_{i,\delta} N_{i,\delta}$ are positive [9]. The following two results show that, with this modification, the diagonal entries of the RAS iteration matrix $T = M^{-1}_{RAS,\delta} N_{RAS,\delta}$ are positive.

DOI: 10.1002/nla

Proposition 4.2
Let A be an irreducible singular M-matrix, and let $M_{RAS,\delta}$ and $M_{i,\delta}$ be defined as in Equations (9) and (7). Set $N_{i,\delta} = M_{i,\delta} - A$ and $N_{RAS,\delta} = M_{RAS,\delta} - A$. Assume that for all i, the matrix $M_{i,\delta}^{-1} N_{i,\delta}$ has positive entries on the main diagonal. Then, $T = M_{RAS,\delta}^{-1} N_{RAS,\delta}$ has positive entries on the main diagonal.

Proof
As seen in the proof of Theorem 4.1,

$$M_{RAS,\delta}^{-1} N_{RAS,\delta} = \sum_{i=1}^{K} E_{i,0} M_{i,\delta}^{-1} N_{i,\delta}.$$

For a row in which $E_{i,0}$ is 0, the diagonal entry of $E_{i,0} M_{i,\delta}^{-1} N_{i,\delta}$ is 0. For a row in which $E_{i,0}$ has a 1, the diagonal entry of $E_{i,0} M_{i,\delta}^{-1} N_{i,\delta}$ equals the diagonal entry of $M_{i,\delta}^{-1} N_{i,\delta}$, which is positive by assumption. Because $\sum_{i=1}^{K} E_{i,0} = I$, for each row, exactly one of the $E_{i,0}$ has a 1. Thus, the diagonal entries of $M_{RAS,\delta}^{-1} N_{RAS,\delta}$ are positive. □

The following proposition describes the modification that is needed to ensure that $M_{i,\delta}^{-1} N_{i,\delta}$ has positive entries along its diagonal. The result can be found in [9].

Proposition 4.3
Let $B \geqslant 0$, $B^T e = e$. Let $K > 1$ be a positive integer and let $\alpha_1, \ldots, \alpha_K$ be any positive real numbers. Let $A = I - B = M_{i,\delta} - N_{i,\delta}$, $i = 1, \ldots, K$, be defined by

$$M_{i,\delta} = \pi_{i,\delta}^T \begin{pmatrix} \alpha_i I + A_{i,\delta} & 0 \\ 0 & \alpha_i I + D_{\neg i,\delta} \end{pmatrix} \pi_{i,\delta},$$

and $N_{i,\delta} = M_{i,\delta} - A$ where $\pi_{i,\delta}$, $A_{i,\delta}$, and $D_{\neg i,\delta}$ are defined as before. Then, the splittings $A = M_{i,\delta} - N_{i,\delta}$ are regular, and the diagonals of $T_{i,\delta} = M_{i,\delta}^{-1} N_{i,\delta}$ are positive, for $i = 1, \ldots, K$.

Another way to ensure convergence was already proposed in [13]. The idea is to replace T by $T_\alpha = \alpha I + (1-\alpha)T$ for some $\alpha \in (0,1)$. Such a matrix is guaranteed to be convergent if $T \geqslant 0$, $\rho(T) = 1$ and $\mathrm{ind}_1(T) = 1$; indeed, such a T_α has positive diagonal entries.

In concluding this section, we emphasize that such modifications are needed only when the RAS iteration is used as a stationary iterative method. No modifications are necessary if Krylov subspace acceleration is used, because in this case the convergence of T is not required.

5. EXTENSION TO INEXACT SOLVES

In this section, we extend the convergence results to the case of inexact local solves. Instead of solving the linear systems $A_{i,\delta} y_i = z_i$, we want to approximate the matrices $A_{i,\delta}$ by $\hat{A}_{i,\delta}$ so that the systems $\hat{A}_{i,\delta} y_i = z_i$ are easier to solve. The diagonal modifications mentioned earlier can also be regarded as a type of inexact solve. The following propositions show that, under certain assumptions, the RAS method is also convergent in the case of inexact local solves.

Proposition 5.1
If $\hat{A}_{i,\delta}$ is a nonsingular M-matrix with $\hat{A}_{i,\delta} \geqslant A_{i,\delta}$ and

$$\hat{M}_{i,\delta} = \pi_{i,\delta}^T \begin{pmatrix} \hat{A}_{i,\delta} & 0 \\ 0 & D_{\neg i,\delta} \end{pmatrix} \pi_{i,\delta},$$

then the splittings $A = \hat{M}_{i,\delta} - \hat{N}_{i,\delta}$ are weak regular.

DOI: 10.1002/nla

Proof
First, note that $\hat{M}_{i,\delta}^{-1}$ is nonnegative, because $\hat{A}_{i,\delta}$ has a nonnegative inverse. Second, we will consider $\hat{M}_{i,\delta}^{-1}\hat{N}_{i,\delta}$ and show that it is nonnegative. Recall that A can be written as

$$A = \pi_{i,\delta}^T \begin{pmatrix} A_{i,\delta} & K_{i,\delta} \\ L_{i,\delta} & A_{\neg i,\delta} \end{pmatrix} \pi_{i,\delta},$$

and it follows that

$$\hat{M}_{i,\delta}^{-1}\hat{N}_{i,\delta} = I - \pi_{i,\delta}^T \begin{pmatrix} \hat{A}_{i,\delta}^{-1}A_{i,\delta} & \hat{A}_{i,\delta}^{-1}K_{i,\delta} \\ D_{\neg i,\delta}^{-1}L_{i,\delta} & D_{\neg i,\delta}^{-1}A_{\neg i,\delta} \end{pmatrix} \pi_{i,\delta}.$$

With $\hat{A}_{i,\delta} \geqslant A_{i,\delta}$, and the fact that both matrices are M-matrices, it follows that $\hat{A}_{i,\delta}^{-1} \leqslant A_{i,\delta}^{-1}$ and $I \geqslant \hat{A}_{i,\delta}^{-1}A_{i,\delta}$. Furthermore, $I - D_{\neg i,\delta}^{-1}A_{\neg i,\delta}$ is nonnegative, because $D_{\neg i,\delta} = \mathrm{diag}(A_{\neg i,\delta})$. With $\hat{A}_{i,\delta}^{-1}$, $D_{\neg i,\delta}^{-1} \geqslant 0$, and $K_{i,\delta}$, $L_{i,\delta} \leqslant 0$, we can conclude that $\hat{M}_{i,\delta}^{-1}\hat{N}_{i,\delta}$ is nonnegative, and that the splittings $A = \hat{M}_{i,\delta} - \hat{N}_{i,\delta}$ are weak regular. □

Consider the splitting $A = \hat{M}_{RAS,\delta} - \hat{N}_{RAS,\delta}$ with $\hat{M}_{RAS,\delta} = \sum_{i=1}^{K} E_{i,0}\hat{M}_{i,\delta}^{-1}$. As seen in the proof of Theorem 4.1, $\hat{M}_{RAS,\delta}^{-1}\hat{N}_{RAS,\delta} = \sum_{i=1}^{K} E_{i,0}\hat{M}_{i,\delta}^{-1}\hat{N}_{i,\delta}$. From the proposition earlier, it follows that $A = \hat{M}_{RAS,\delta} - \hat{N}_{RAS,\delta}$ is weak regular as well. Following the idea of adding a positive value to the diagonal of each $\hat{M}_{i,\delta}$ as seen in section 4 will lead to positive diagonal entries in $\hat{M}_{RAS,\delta}^{-1}\hat{N}_{RAS,\delta}$. Alternatively, the corresponding iteration matrix can be shifted and scaled using a parameter $\alpha \in (0, 1)$, as shown at the end of section 4. With either of these modifications, the RAS method with inexact local solves associated with matrices $\hat{A}_{i,\delta} \geqslant A_{i,\delta}$ is convergent.

We are particularly interested in the case where an incomplete LU factorization is used for approximating the solution of $A_{i,\delta}y_i = z_i$. The incomplete LU (ILU) factorization was first introduced by Varga [16] and was further studied by Meijerink and van der Vorst [17]. Meijerink and van der Vorst showed that, for an M-matrix A and every zero pattern $Q \subset P_n = \{(i, j) \mid 1 \leqslant i, j \leqslant n\}$ not containing the diagonals (i, i), there is a unique lower triangular matrix L with unit diagonal, a unique upper triangular matrix U with positive diagonals, and a matrix R with

$$\begin{aligned} l_{ij} &= 0 \text{ if } (i, j) \in Q, \\ u_{ij} &= 0 \text{ if } (i, j) \in Q, \\ r_{ij} &= 0 \text{ if } (i, j) \notin Q, \end{aligned}$$

such that the splitting $A = LU - R$ is regular. Note that Q can always be chosen such that $\mathrm{diag}(R) = 0$. So, in our case, $A_{i,\delta} = L_{i,\delta}U_{i,\delta} - R_{i,\delta}$, and we use $\hat{A}_{i,\delta} = L_{i,\delta}U_{i,\delta}$ instead. Because the splitting $A_{i,\delta} = L_{i,\delta}U_{i,\delta} - R_{i,\delta}$ is regular, $R_{i,\delta}$ is nonnegative and $\hat{A}_{i,\delta} \geqslant A_{i,\delta}$. Varga showed in [18] that if A is a nonsingular M-matrix and B is a matrix satisfying

$$\begin{aligned} a_{ij} &\leqslant b_{ij} \leqslant 0, \qquad \text{for } i \neq j \text{ and} \\ 0 &< a_{ii} \leqslant b_{ii}, \end{aligned}$$

then B is also a nonsingular M-matrix. If the drop tolerance in the ILU factorization is sufficiently small, $\hat{A}_{i,\delta}$ will satisfy both of the earlier conditions. In this case, $\hat{A}_{i,\delta}$ is a nonsingular M-matrix, and the previous discussion shows that replacing $A_{i,\delta}$ with $\hat{A}_{i,\delta}$ will preserve the convergence results. Again, this is only needed if the inexact RAS method is used as a stationary iteration. If Krylov acceleration is used, the earlier conditions are not needed for convergence.

6. TWO-LEVEL METHOD

The rate of convergence of the AS method may be improved with the help of a "coarse grid" correction, leading to a two-level approach. This correction can be formed algebraically, without any

DOI: 10.1002/nla

reference to an actual grid; for simplicity, however, we will use the same terminology in use in the field of numerical PDEs. The correction can be applied in two ways, additively or multiplicatively; see for example [2, 3, 19–21].

In this section, we describe the use of a coarse grid correction for the RAS method in the case of irreducible Markov chains. Let P_0 and R_0 denote prolongation and restriction operators (respectively), and let $A_0 = R_0 A P_0$ be the resulting coarse representation of A. Let $\hat{A}_0$ denote an invertible approximation of A_0. The additive variant of the two-level method corresponds to the iteration matrix

$$T = I - \theta(P_0 \hat{A}_0^{-1} R_0 + M_{RAS}^{-1})A,$$

where $\theta > 0$ is a parameter. For the multiplicative variant, the iteration matrix is given by

$$T = (I - P_0 \hat{A}_0^{-1} R_0 A)(I - M_{RAS}^{-1} A).$$

Here, we consider the additive approach only. For use as a preconditioner, the parameter θ has no effect and can be set to 1.

Some results were obtained by using independent sets and the Schur complement. First, a maximal independent set $\mathcal{F}$ in the underlying graph of A is found. We can partition A such that

$$\pi A \pi^T = \begin{pmatrix} A_{\mathcal{F}\mathcal{F}} & A_{\mathcal{F}\mathcal{C}} \\ A_{\mathcal{C}\mathcal{F}} & A_{\mathcal{C}\mathcal{C}} \end{pmatrix}.$$

Because $\mathcal{F}$ is an independent set, $A_{\mathcal{F}\mathcal{F}}$ is a diagonal matrix. Note that $A_{\mathcal{F}\mathcal{F}}$ is nonsingular, because A has positive entries along its diagonal. We define the restriction operator

$$R_0 = [-A_{\mathcal{C}\mathcal{F}} A_{\mathcal{F}\mathcal{F}}^{-1} \quad I]$$

and the prolongation operator as

$$P_0 = \begin{bmatrix} -A_{\mathcal{F}\mathcal{F}}^{-1} A_{\mathcal{F}\mathcal{C}} \\ I \end{bmatrix}.$$

With these operators, $A_0 = R_0(\pi A \pi^T) P_0$ is given by the Schur complement

$$A_0 = A_{\mathcal{C}\mathcal{C}} - A_{\mathcal{C}\mathcal{F}} A_{\mathcal{F}\mathcal{F}}^{-1} A_{\mathcal{F}\mathcal{C}}.$$

Note that A_0 is a singular, irreducible M-matrix [22, Lemma 1], and that an ILU factorization can be used to inexactly solve systems of the form $A_0 z = y$. In other words, we set $\hat{A}_0 = \hat{L}_0 \hat{U}_0$, where $\hat{L}_0$ and $\hat{U}_0$ are incomplete factors of A_0, which can be assumed to be nonsingular; see [23] for some sufficient conditions that guarantee the existence of nonsingular ILU factors of singular M-matrices. In our experiments, the ILU factors were always nonsingular.

To reduce the size of the coarse grid correction, we apply the earlier scheme twice. That is, a maximal independent set in the underlying graph of A_0 is found, and the Schur complement of A_0 is used. We refer to Figure 1 for an example. A few numerical results using our two-level method with additive correction using the earlier independent set approach are given in section 8. A greedy algorithm was used to find the maximal independent sets [24].

For a discussion of different coarsening strategies within algebraic multi-level methods for Markov chain problems, we refer the reader to, for example, [25, 26].

7. DESCRIPTION OF THE TEST PROBLEMS

For our numerical experiments, we used the generator matrices of some Markov chain models provided in the Markov Chain Analyzer collection [10]. These matrices are infinitesimal generators of CTMCs. For our purposes, we converted them to the form $A = I - P^T$, with P row-stochastic. A corresponds to a discrete-time Markov chain, known as the *embedded Markov chain*; see [12, Chapter 1.4.3]. In Table I, the dimensions and number of nonzeroes of our selected test

DOI: 10.1002/nla

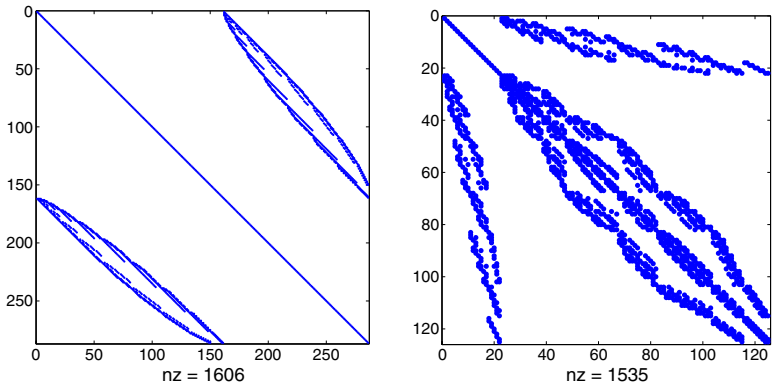

Figure 1. Nonzero patterns of matrices used to construct the coarse grid correction. Left: sparsity pattern of the matrix permuted by Luby's maximal independent set reordering. Right: sparsity pattern of the corresponding Schur complement, also permuted by Luby's maximal independent set reordering. The original matrix comes from the *ncd* family.

Table I. Properties of the generator matrices.

Matrix	Number of rows/cols	Number of nonzeroes
ncd07	62,196	420,036
ncd09	129,766	883,246
mutex09	65,535	1,114,079
mutex12	263,950	4,031,310
tcomm16	13,671	67,381
tcomm20	17,081	84,211
twod06	66,049	263,169
twod10	263,169	1,050,625
$\text{reliab}_1(m)$	m^2	$5m^2 - 4m$
$\text{reliab}_2(m)$	m^2	$5m^2 - 4m$

matrices are shown. Each matrix is named by its family and its index in the family. The matrices from the *mutex* and *ncd* families are structurally symmetric, the matrices from the *twod* family are structurally nonsymmetric, and the matrices from the *tcomm* family are very close to being structurally symmetric. We also run some test with matrices that arise from a reliability problem. We consider a simple reliability model with two different classes of machines. We assume that each class has the same number of machines. Each machine is subject to breakdown and a subsequent repair. A state is completely specified by the ordered pair (n_1, n_2), where n_1 denotes the number of intact machines of the first class, and n_2 denotes the number of intact machines of the second class. Thus, if there are m machines in each class, the total number of possible states is $(m+1)^2$. We order these states such that state (i, j) has index $(m+1)(m-i)+m-j+1$. The times between successive breakdowns and successive repairs are both exponentially distributed. The breakdown rates of class 1 machines and class 2 machines are, respectively, λ_1 and λ_2. Similarly, the repair rates of the two classes of machines are μ_1 and μ_2. The transition rate matrix for the described reliability model is then given by Q on the next page. Here $m = 3$ machines per class where used. The diagonal elements indicated by asterisks are the negated sums of the off-diagonal elements in their corresponding rows. We note that the stationary distribution for these models is known analytically (i.e., a closed-form solution exist), which makes them well suited for testing codes and for checking the accuracy of computed solutions.

Our goal is to solve the singular system $\pi Q = 0$ subject to $\|\pi\|_1 = 1$ (the normalization condition). However, one can solve the equivalent system $-Q^T\pi^T = 0$ instead. Here, the coefficient matrix is a singular irreducible M-matrix, and the theory developed in the previous sections applies. From this point on, we let $A = -Q^T$ and $x = \pi^T$, so that we can use the notation introduced in the earlier sections. We tested two different reliability matrices. The first corresponds to a reliability

DOI: 10.1002/nla

$$Q = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \\ 10 \\ 11 \\ 12 \\ 13 \\ 14 \\ 15 \\ 16 \end{array} \begin{array}{c} \begin{array}{cccccccccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 15 & 16 \end{array} \\ \left(\begin{array}{cccccccccccccccc} * & 3\lambda_2 & & & 3\lambda_1 & & & & & & & & & & & \\ \mu_2 & * & 2\lambda_2 & & & 3\lambda_1 & & & & & & & & & & \\ & 2\mu_2 & * & \lambda_2 & & & 3\lambda_1 & & & & & & & & & \\ & & 3\mu_2 & * & & & & 3\lambda_1 & & & & & & & & \\ \mu_1 & & & & * & 3\lambda_2 & & & 2\lambda_1 & & & & & & & \\ & \mu_1 & & & \mu_2 & * & 2\lambda_2 & & & 2\lambda_1 & & & & & & \\ & & \mu_1 & & & 2\mu_2 & * & \lambda_2 & & & 2\lambda_1 & & & & & \\ & & & \mu_1 & & & 3\mu_2 & * & & & & 2\lambda_1 & & & & \\ & & & & 2\mu_1 & & & & * & 3\lambda_2 & & & \lambda_1 & & & \\ & & & & & 2\mu_1 & & & \mu_2 & * & 2\lambda_2 & & & \lambda_1 & & \\ & & & & & & 2\mu_1 & & & 2\mu_2 & * & \lambda_2 & & & \lambda_1 & \\ & & & & & & & 2\mu_1 & & & 3\mu_2 & * & & & & \lambda_1 \\ & & & & & & & & 3\mu_1 & & & & * & 3\lambda_2 & & \\ & & & & & & & & & 3\mu_1 & & & \mu_2 & * & 2\lambda_2 & \\ & & & & & & & & & & 3\mu_1 & & & 2\mu_2 & * & \lambda_2 \\ & & & & & & & & & & & 3\mu_1 & & & 3\mu_2 & * \end{array} \right) \end{array}.$$

problem with parameters $\lambda_1 = 1$, $\lambda_2 = 0.2$, $\mu_1 = 2.5$, and $\mu_2 = 6$, whereas the second corresponds to a reliability problem with parameters $\lambda_1 = 2$, $\lambda_2 = 0.9$, $\mu_1 = 0.5$, and $\mu_2 = 6$.

In Table II, we report the value of the subdominant eigenvalue (i.e., the second largest eigenvalue) of the stochastic matrices for the embedded Markov chains describing the reliability models. The fact that the gap between the dominant eigenvalue $\lambda = 1$ and the subdominant eigenvalue shrinks as the number of states is increased indicates that these problems become increasingly difficult as their size grows. In Table III, we report results for a few reliability models from the first class using GMRES(50) with the drop tolerance-based ILUTH preconditioner [12, 27]. In all cases, the drop tolerance was 10^{-3}. Note the high construction costs and very slow convergence behavior in the case of sufficiently large problems.

8. NUMERICAL EXPERIMENTS

In this section, we provide the results of our numerical experiments. The primary goal of these tests is to study the convergence behavior of RAS and of RAS-preconditioned GMRES for large Markov

Table II. Subdominant eigenvalue, reliability models.

Matrix	Subdominant eigenvalue
$\text{reliab}_1(100)$	0.9807
$\text{reliab}_1(400)$	0.9952
$\text{reliab}_1(700)$	0.9972
$\text{reliab}_1(1000)$	0.9981
$\text{reliab}_1(1200)$	0.9984
$\text{reliab}_2(100)$	0.9894
$\text{reliab}_2(400)$	0.9974
$\text{reliab}_2(700)$	0.9985
$\text{reliab}_2(1000)$	0.9989
$\text{reliab}_2(1200)$	0.9991

Table III. ILUTH results, reliability models.

Matrix	ILUTH time	Its
$\text{reliab}_1(100)$	0.23	32
$\text{reliab}_1(400)$	10.0	43
$\text{reliab}_1(700)$	70.1	50
$\text{reliab}_1(1000)$	195.0	> 500
$\text{reliab}_1(1200)$	325.0	> 500

DOI: 10.1002/nla

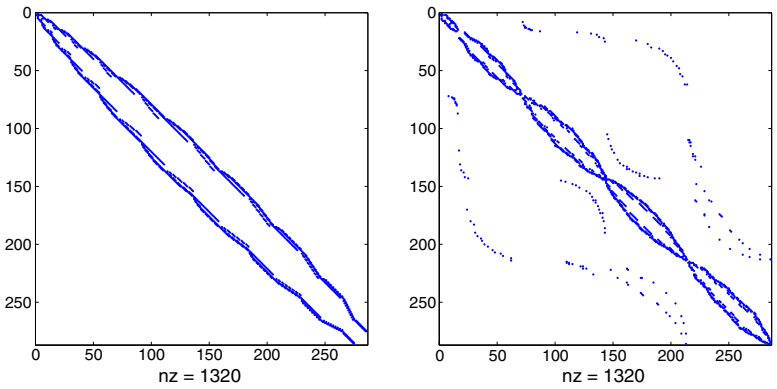

Figure 2. Left: nonzero pattern of an *ncd* matrix. Right: nonzero pattern of the same matrix when four subdomains are used, and the matrix is reordered accordingly by METIS.

Table IV. Results for GMRES with RAS preconditioner. K is the number of domains, "Constr.", the time (in seconds) needed to construct the preconditioner; "It", the number of GMRES(50) iterations needed to reduce the 2-norm of the residual below 10^{-12}; "Solve", the time (in seconds) for the GMRES(50) iterations. For each matrix, the best overall timings are in boldface. For local solves, the complete LU factorization was used.

		RAS preconditioned GMRES with					
		Small overlap			Large overlap		
Matrix	K	Constr.	Solve	It	Constr.	Solve	It
ncd07	2	99.10	17.54	28	221.35	52.36	27
	4	38.66	9.22	28	79.96	12.12	28
	8	21.96	8.43	28	30.73	9.67	28
	16	13.26	8.22	29	16.79	9.15	28
	32	7.95	8.50	30	7.67	7.58	29
	64	4.43	7.77	33	**4.49**	**7.23**	32
tcomm16	2	0.27	0.15	5	0.29	0.15	5
	4	0.28	0.29	5	0.33	0.16	5
	8	0.28	0.12	5	0.35	0.15	5
	16	0.24	0.11	5	0.31	0.13	5
	32	0.20	0.16	7	0.24	0.12	5
	64	0.17	0.19	8	**0.19**	**0.14**	6
twod06	2	16.94	2.95	12	20.78	3.27	12
	4	6.80	2.94	14	9.72	2.95	12
	8	4.44	2.83	15	6.42	3.12	14
	16	2.91	3.62	18	4.29	3.55	17
	32	**1.94**	**4.23**	21	2.87	3.64	18
	64	1.62	5.67	27	2.11	5.02	25
$\text{reliab}_1(400)$	2	71.75	21.52	13	86.27	23.72	13
	4	35.89	12.86	13	45.12	14.32	13
	8	14.82	17.39	21	26.32	14.28	14
	16	9.23	23.38	27	14.22	11.94	14
	32	6.17	20.92	29	9.14	9.25	15
	64	4.29	19.81	30	**6.19**	**8.68**	16
$\text{reliab}_2(400)$	2	50.51	21.95	18	59.22	21.02	16
	4	30.20	18.77	18	36.92	17.87	16
	8	12.99	29.33	31	20.76	16.60	17
	16	8.48	27.62	33	12.88	14.19	18
	32	5.65	33.06	39	**8.45**	**13.69**	20
	64	4.18	45.97	52	6.05	18.13	27

DOI: 10.1002/nla

chain problems as a function of algorithmic parameters like the amount of overlap and the number of subdomains, and to carry out a comparison with standard AS preconditioning. A secondary goal is to present some preliminary experiments with the two-level method. Although AS-type methods have been developed mostly for implementation on parallel computers, here we consider only a

Table V. Here, a small amount of overlap was used. K is the number of domains, "Constr.", the time (in seconds) needed to construct the preconditioner; "It", the number of GMRES(50) iterations needed to reduce the 2-norm of the residual below 10^{-12}; "Solve", the time (in seconds) for the GMRES(50) iterations. The number of GMRES(50) iterations without preconditioner is also given. For local solves, ILUT was used.

			GMRES preconditioned with AS			GMRES preconditioned with RAS		
Matrix	GMRES	K	Constr.	Solve	It	Constr.	Solve	It
ncd07	>250	2	**0.52**	**4.96**	29	0.52	5.16	29
		8	0.87	4.95	29	0.87	5.32	29
		16	0.74	5.65	30	0.74	5.68	29
		32	0.77	5.68	30	0.77	5.98	30
		64	0.83	6.67	34	0.83	6.74	33
ncd09	>250	2	**1.39**	**11.42**	28	1.39	11.71	28
		8	1.86	12.75	30	1.86	13.28	30
		16	1.51	13.81	31	1.51	13.64	30
		32	1.73	14.38	31	1.73	15.38	31
		64	1.55	15.92	34	1.55	17.20	34
mutex09	21	2	5.18	0.63	6	5.18	0.66	6
		8	2.74	1.18	11	**2.74**	**1.02**	9
		16	2.71	1.44	12	2.71	1.40	11
		32	2.39	1.63	13	2.39	1.44	11
		64	2.61	1.47	14	2.61	1.36	12
mutex12	22	2	89.15	4.09	8	89.15	4.12	7
		8	32.91	7.54	12	32.91	6.46	11
		16	20.06	7.30	13	20.06	7.02	12
		32	14.16	9.15	14	14.16	8.04	13
		64	11.55	9.85	15	**11.55**	**9.84**	14
tcomm16	93	2	0.13	0.11	5	0.13	0.11	5
		8	0.14	0.09	5	0.14	0.09	5
		16	**0.12**	**0.08**	5	0.12	0.10	5
		32	0.12	0.15	7	0.12	0.16	7
		64	0.13	0.13	7	0.13	0.16	7
tcomm20	93	2	0.17	0.13	5	0.17	0.15	5
		8	0.19	0.13	5	0.19	0.15	5
		16	**0.15**	**0.11**	5	0.15	0.12	5
		32	0.15	0.12	5	0.15	0.13	5
		64	0.16	0.18	7	0.16	0.20	7
twod06	>250	2	**1.28**	**1.62**	12	1.28	1.68	12
		8	1.15	1.93	16	1.15	2.12	16
		16	0.94	2.69	19	0.94	2.93	19
		32	0.83	3.11	21	0.83	3.32	21
		64	0.79	4.52	27	0.79	4.59	27
twod10	>250	2	**6.45**	**8.29**	16	6.45	9.17	16
		8	10.47	14.95	19	10.47	11.69	18
		16	8.07	12.28	20	8.07	12.85	19
		32	4.84	20.76	26	4.84	20.23	25
		64	4.97	27.43	31	4.97	30.69	31

DOI: 10.1002/nla

single compute node. We think that it is useful to carry out such a study before investing significant parallelization effort.

The implementation was done in MATLAB 7.8.0 on a 2.13-GHz Intel Core 2 Duo Processor with 2-GB main memory. We performed a large number of tests on numerous matrices. Here, we present a selection of these results to show our overall findings. The Krylov method used was GMRES [28]. As initial vector, we used the first unit basis vector, e_1; this gave slightly better results than using a uniform probability distribution as the initial vector.

For the partitioning of our matrices, we used the METIS library [29]. In our tests, we used $K = 2$, 8, 16, 32, and 64 domains. METIS requires the matrices to be structurally symmetric. Therefore, we applied METIS to the underlying graph of $A + A^T$ for the *twod* matrices. The amount of overlap was chosen according to the distance to the domain. For a small choice of overlap, we chose all vertices within a distance of 1 in the underlying graph, that is, all vertices that are connected to a vertex in the domain. See Figure 2 for the sparsity pattern of an *ncd* matrix before and after the reordering induced by METIS. The *mutex* matrices have a large separator set, and choosing the overlap in the described way leads to domains that are close to the size of the matrix. For these matrices,

Table VI. Here, a large amount of overlap was used. K is the number of domains, "Constr.", the time (in seconds) needed to construct the preconditioner; "It", the number of GMRES(50) iterations needed to reduce the 2-norm of the residual below 10^{-12}; "Solve", the time (in seconds) for the GMRES(50) iterations. For local solves, ILUT was used.

		GMRES preconditioned with					
		AS			RAS		
Matrix	K	Constr.	Solve	It	Constr.	Solve	It
ncd07	2	**0.79**	**5.10**	28	0.79	5.34	28
	8	0.95	5.14	29	0.95	5.05	28
	16	0.81	5.26	29	0.81	5.71	29
	32	0.81	6.01	30	0.81	5.63	29
	64	0.81	6.45	34	0.81	6.34	32
ncd09	2	**1.75**	**11.49**	28	1.75	12.05	28
	8	2.12	12.17	29	2.12	12.90	29
	16	1.72	13.18	30	1.72	13.79	30
	32	1.77	14.03	31	1.77	15.18	31
	64	1.79	15.56	33	1.79	16.02	33
tcomm16	2	0.13	0.17	5	0.13	0.20	5
	8	0.17	0.10	5	0.17	0.11	5
	16	**0.13**	**0.09**	5	0.13	0.10	5
	32	0.14	0.10	5	0.14	0.11	5
	64	0.14	0.12	6	0.14	0.14	6
tcomm20	2	**0.17**	**0.13**	5	0.17	0.14	5
	8	0.22	0.13	5	0.22	0.15	5
	16	0.19	0.13	5	0.19	0.13	5
	32	0.18	0.13	5	0.18	0.14	5
	64	0.17	0.15	6	0.17	0.16	6
twod06	2	**1.57**	**1.53**	12	1.57	1.56	12
	8	1.46	1.87	15	1.46	1.85	14
	16	1.22	2.00	17	1.22	2.49	17
	32	1.05	3.10	19	1.05	2.94	19
	64	0.96	4.30	26	0.96	4.14	25
twod10	2	6.17	9.41	16	**6.17**	**9.19**	16
	8	13.46	12.16	18	13.46	12.00	18
	16	9.77	12.66	19	9.77	13.45	19
	32	6.33	17.04	23	6.33	18.25	23
	64	6.38	26.39	28	6.38	25.00	27

DOI: 10.1002/nla

we restricted the total size of a domain, that is the partition and the overlap, by $\frac{4}{3} \cdot \frac{n}{K}$, where n is the size of the matrix and K is the number of domains. For a large choice of overlap, we chose all vertices that lie within a distance of 10 in the underlying graph. We factored the blocks using an incomplete LU factorization with a symmetric reverse Cuthill–McKee reordering [30], threshold parameter $\tau = 10^{-3}$ for the *mutex* and *reliab* matrices, and $\tau = 10^{-4}$ for the other matrices. Compared with exact LU factorization, the timings and storage requirements were dramatically reduced, whereas the number of iterations remained essentially unchanged (compare Table IV with V and VI). We found that RAS always performs as well or better than the standard AS preconditioner in terms of number of iterations, whereas in terms of total runtime, the standard AS preconditioner tends to perform slightly better. This can be explained by observing that in RAS, we have to keep track of the nodes that are in the non-overlapping domains and the ones that are in the overlap. This needs some extra "bookkeeping" which might lead to RAS performing slightly slower than AS for the same number of iterations, at least with our implementation. Note, however, that AS requires more interprocessor communication, and thus we expect it to be less efficient than RAS in a parallel

Table VII. The matrix *reliab1* arises from a reliability problem with parameters $\lambda_1 = 1$, $\lambda_2 = 0.2$, $\mu_1 = 2.5$, and $\mu_2 = 6$. Here, K is the number of domains, "Constr.", the time (in seconds) needed to construct the preconditioner; "It", the number of GMRES(50) iterations needed to reduce the 2-norm of the residual below 10^{-12}; "Solve", the time (in seconds) for the GMRES(50) iterations.

		RAS Preconditioned GMRES with					
		Small overlap			Large overlap		
Matrix	K	Constr.	Solve	It	Constr.	Solve	It
$\text{reliab}_1(100)$	2	0.14	0.37	15	**0.16**	**0.23**	10
	4	0.14	0.34	15	0.17	0.24	10
	8	0.15	0.31	15	0.17	0.24	10
	16	0.11	0.49	20	0.14	0.35	14
	32	0.11	0.72	24	0.13	0.60	21
	64	0.19	0.81	27	0.18	0.72	24
$\text{reliab}_1(400)$	2	**4.51**	**4.47**	13	4.76	4.50	13
	4	5.23	4.85	13	5.74	5.14	13
	8	3.62	10.38	22	4.63	5.55	14
	16	3.17	15.04	27	4.08	6.52	15
	32	2.83	17.95	29	3.51	7.17	16
	64	2.52	17.86	30	3.04	7.26	17
$\text{reliab}_1(700)$	2	21.69	53.00	26	21.75	17.52	16
	4	18.98	41.59	26	20.05	18.16	16
	8	13.33	54.13	30	15.47	19.29	16
	16	12.65	54.92	31	14.76	21.88	17
	32	9.45	50.95	30	12.06	22.88	16
	64	7.39	41.42	27	**9.59**	**19.64**	16
$\text{reliab}_1(1000)$	2	112.94	53.47	17	116.71	42.47	17
	4	75.26	42.75	17	65.91	55.57	17
	8	45.66	47.74	19	48.00	45.74	18
	16	27.23	80.44	27	**29.62**	**42.28**	18
	32	21.46	196.66	36	26.32	48.84	19
	64	19.31	251.73	36	23.82	60.82	19
$\text{reliab}_1(1200)$	2	154.85	161.76	19	160.28	157.36	19
	4	112.29	61.02	18	108.95	71.54	18
	8	73.51	224.24	33	77.90	62.63	19
	16	50.25	227.65	32	56.04	113.21	20
	32	38.74	228.14	37	**43.84**	**73.57**	20
	64	26.60	143.57	30	32.89	94.01	20

DOI: 10.1002/nla

implementation. A selection of our results is presented in the following tables. Results in bold indicate the best total runtime, that is, the time to construct the preconditioner plus the time taken by the Krylov method. Runtimes and iteration counts for the Markov Chain Analyzer examples, using AS and RAS with small and large overlaps, are given in Tables V and VI, respectively. A few results for the reliability problems are shown in Tables VII and VIII. In Tables IX and X, results using a two-level RAS preconditioner are given.

Our results indicate many favorable properties of the RAS preconditioner for the numerical solution of Markov chains. In particular, we focused on the rate of convergence depending on the size of the problem, the number of domains used, and the amount of overlap chosen. First of all, the time required to construct the preconditioner, when incomplete LU factorizations are used, is very small. Also note that, with an increased number of domains, the size of the submatrices that require an incomplete LU factorization decreases, and thus the construction time for the preconditioner decreases. This can be best observed in the largest example of the first reliability problem ($\text{reliab}_1(1200)$, see Table VII). Here the time to construct the preconditioner can be cut from 155 s for two domains to 27 s for 64 domains. Furthermore, the construction expenses mainly consist of

Table VIII. The matrix *reliab2* arises from a reliability problem with parameters $\lambda_1 = 2$, $\lambda_2 = 0.9$, $\mu_1 = 0.5$, and $\mu_2 = 6$. Here, K is the number of domains, "Constr.", the time (in seconds) needed to construct the preconditioner; "It", the number of GMRES(50) iterations needed to reduce the 2-norm of the residual below 10^{-12}; "Solve", the time (in seconds) for the GMRES(50) iterations.

		RAS Preconditioned GMRES with					
		Small overlap			Large overlap		
Matrix	K	Constr.	Solve	It	Constr.	Solve	It
$\text{reliab}_2(100)$	2	0.20	0.37	14	0.24	0.34	12
	4	**0.18**	**0.32**	14	0.23	0.33	12
	8	0.17	0.47	20	0.20	0.40	14
	16	0.17	1.20	30	0.17	0.80	22
	32	0.16	1.28	35	0.20	1.04	28
	64	0.14	1.78	42	0.15	1.71	40
$\text{reliab}_2(400)$	2	7.54	10.24	19	8.27	6.83	17
	4	5.81	7.93	19	6.65	7.27	17
	8	4.21	18.85	31	5.33	8.09	18
	16	3.83	21.21	33	**4.82**	**7.89**	18
	32	2.96	31.28	39	4.11	10.19	20
	64	2.96	43.42	52	3.54	14.49	27
$\text{reliab}_2(700)$	2	50.86	50.88	25	48.79	25.99	20
	4	22.25	35.34	25	23.65	24.83	20
	8	12.41	69.96	38	**13.75**	**28.50**	22
	16	13.76	72.80	37	15.81	29.15	22
	32	10.08	121.10	49	13.62	31.67	23
	64	9.05	223.73	92	12.19	46.52	28
$\text{reliab}_2(1000)$	2	102.53	111.52	25	102.22	68.41	23
	4	80.11	72.52	25	82.25	64.44	23
	8	45.23	77.08	26	48.13	67.96	24
	16	36.72	425.31	88	41.34	76.41	26
	32	25.78	187.05	44	**30.74**	**76.68**	26
	64	20.82	272.93	69	28.14	106.39	28
$\text{reliab}_2(1200)$	2	90.33	126.93	26	92.78	98.79	25
	4	123.10	206.22	26	125.40	130.95	25
	8	81.95	146.24	27	87.95	128.44	25
	16	58.11	238.04	41	62.35	133.43	27
	32	44.42	245.89	43	**49.78**	**124.00**	27
	64	34.60	454.97	81	42.56	148.94	30

DOI: 10.1002/nla

Table IX. A few results with two-level methods are shown. Here, K is the number of domains, "Constr.", the time (in seconds) needed to construct the preconditioner; "It", the number of GMRES(50) iterations needed to reduce the 2-norm of the residual below 10^{-12}; "Solve", the time (in seconds) for the GMRES(50) iterations.

		GMRES preconditioned with two-level RAS					
		Small overlap			Large overlap		
Matrix	K	Constr.	Solve	It	Constr.	Solve	It
ncd07	2	**2.31**	**3.01**	19	2.47	3.02	19
	4	2.45	3.13	19	2.65	3.36	19
	8	2.70	3.27	19	2.80	3.24	19
	16	2.57	3.95	19	2.65	4.03	20
	32	2.57	3.84	20	2.68	3.76	20
	64	2.69	3.68	21	2.57	3.24	20
ncd09	2	**5.22**	**7.27**	19	5.57	7.09	19
	4	5.48	7.32	19	5.61	7.16	19
	8	5.55	7.46	20	5.88	7.93	20
	16	5.36	8.43	21	5.59	8.85	21
	32	5.63	9.28	21	5.69	9.30	21
	64	5.89	9.82	22	6.44	9.56	22
tcomm16	2	0.43	0.25	7	0.39	0.20	7
	4	0.39	0.18	7	0.41	0.21	7
	8	0.41	0.17	7	0.45	0.23	7
	16	**0.37**	**0.17**	7	0.45	0.19	7
	32	0.46	0.21	7	0.45	0.22	7
	64	0.43	0.24	8	0.47	0.24	7
tcomm20	2	0.58	0.34	7	0.57	0.30	7
	4	0.60	0.33	7	0.61	0.30	7
	8	0.59	0.30	7	0.64	0.31	7
	16	**0.55**	**0.27**	7	0.61	0.28	7
	32	**0.54**	**0.28**	7	0.58	0.28	7
	64	0.56	0.35	8	0.58	0.30	7
twod06	2	6.35	1.88	10	5.97	1.98	10
	4	**3.19**	**1.79**	11	3.47	1.69	10
	8	3.09	2.44	12	3.56	1.95	11
	16	2.85	2.66	14	3.19	2.77	14
	32	2.78	3.21	16	3.08	2.94	15
	64	2.71	4.46	21	3.08	4.29	20
twod10	2	22.10	8.47	12	**16.65**	**8.21**	12
	4	20.34	8.77	12	21.45	8.62	12
	8	21.54	11.02	14	24.72	10.69	13
	16	18.31	11.40	14	20.17	11.75	14
	32	15.37	17.09	18	17.31	15.84	17
	64	15.58	24.79	22	17.50	22.51	20

the work needed to compute the incomplete LU factorizations as the time spent to find the non-overlapping domains with the METIS software is almost negligible. The task of computing several incomplete LU factorizations of different submatrices is trivial to parallelize, and we expect an even better construction time when used in a parallel setting. Another observation that can be made from our results is that there is only a slight increase in the number of iterations when the size of the problem is increased. This property is particularly desirable for preconditioners. In some cases, the number of iterations is even decreased as the size of the problem grows larger (see for example Table VII). The choice of the number of domains also has an influence on the convergence. As the number of domains increases so does the number of iterations. This behavior is not surprising, because the propagation of information across the computational domain takes longer for an

DOI: 10.1002/nla

Table X. A few results with two-level methods are shown. Here, K is the number of domains, "Constr.", the time (in seconds) needed to construct the preconditioner; "It", the number of GMRES(50) iterations needed to reduce the 2-norm of the residual below 10^{-12}; "Solve", the time (in seconds) for the GMRES(50) iterations.

		RAS Preconditioned GMRES with					
		Small overlap			Large overlap		
Matrix	K	Constr.	Solve	It	Constr.	Solve	It
$\text{reliab}_1(100)$	2	0.37	0.40	14	**0.35**	**0.31**	13
	4	0.34	0.34	14	0.36	0.32	13
	8	0.34	0.34	14	0.37	0.35	13
	16	0.33	0.40	16	0.34	0.38	14
	32	0.42	0.53	18	0.36	0.53	17
	64	0.41	0.58	19	0.35	0.52	18
$\text{reliab}_1(400)$	2	**8.23**	**6.64**	15	8.36	6.68	15
	4	8.55	6.88	15	8.78	6.95	15
	8	8.05	9.00	18	8.63	7.09	15
	16	7.75	11.00	20	8.31	8.04	16
	32	7.94	12.24	21	8.20	8.38	16
	64	7.58	11.85	21	8.05	8.87	17
$\text{reliab}_1(700)$	2	26.88	28.69	20	27.32	22.05	17
	4	24.18	28.64	20	24.65	22.44	17
	8	23.80	31.04	21	24.69	22.68	17
	16	23.02	33.56	22	24.12	22.97	17
	32	22.15	31.12	21	**23.63**	**23.25**	17
	64	21.70	29.53	20	23.15	23.79	17
$\text{reliab}_2(100)$	2	0.41	0.45	14	0.44	0.41	14
	4	**0.38**	**0.39**	14	0.43	0.42	14
	8	0.43	0.42	16	0.40	0.44	15
	16	0.40	0.65	21	0.44	0.61	19
	32	0.40	0.73	22	0.39	0.72	21
	64	0.42	1.00	26	0.47	1.16	25
$\text{reliab}_2(400)$	2	10.40	9.28	19	10.60	9.21	19
	4	**10.26**	**9.13**	19	10.46	9.26	19
	8	9.70	11.19	22	10.15	10.08	20
	16	9.25	12.41	23	10.10	9.77	19
	32	9.28	15.03	25	9.91	12.63	21
	64	9.67	19.29	29	9.91	15.25	24
$\text{reliab}_2(700)$	2	49.69	42.19	24	48.96	38.35	23
	4	44.34	40.44	24	45.58	37.59	23
	8	41.97	67.55	29	43.40	43.06	25
	16	40.99	48.30	27	**42.00**	**41.73**	24
	32	39.96	57.47	29	41.61	45.48	25
	64	39.81	66.60	32	41.42	51.54	27

increased number of subdomains. Note, however that, for several matrices, the increase in the number of iterations is moderate. As mentioned earlier, the construction time is faster if more domains are used. In most cases, the time saved during the construction is smaller than the increase in the time required by the Krylov method, and thus the overall time needed increases with the number of domains. This behavior seems to be reversed for some very large problems. Here, the overall time needed is reduced for a large number of domains, because the decreased construction time outweighs the larger time needed to solve the linear system (see larger problems in both reliability models, Tables VII and VIII). Another aspect that is worth mentioning is the effect of the amount of overlap on the convergence. Whereas larger overlap leads to an increase in construction time, the number of iterations is decreased. The most notable reduction can be seen in the second reliability

DOI: 10.1002/nla

model. Here, in one case, the number of iterations could be cut to 28 with an increase of overlap. For a small choice of overlap, 92 iterations were required. In many cases, the reduction in the number of iterations induced a reduction in the overall solution time.

Finally, we comment on the results obtained with the two-level method. The number of iterations could be reduced (except for the *tcomm* examples) with our two-level approach. For example, for problems coming from the *ncd* family, the number of iterations could be reduced from about 30 to about 20. It can also be seen that the properties mentioned before are more pronounced in the two-level method. That is, there is an even slower increase in the number of iterations as the problem size increases. Also, the increase in the number of iterations with the increase in the number of domains is reduced. Unfortunately, the increased construction time in the two-level method almost outweighs the time saved during the iterative solve. In most cases, we could only observe a very small reduction in the overall time needed, if any. In our case, a better choice of overlap seems to be more effective than a two-level method. We should mention that a number of attempts were made to find more cost-effective coarsening strategies, but we were unable to obtain better results.

9. SUMMARY AND CONCLUSIONS

In this paper, we have extended the RAS method to the computation of the stationary vector of large, sparse Markov chains. Our results suggest that, when combined with GMRES acceleration and inexact solves, RAS is a promising approach for the solution of Markov chains with large, sparse transition matrices. Although primarily designed with parallel computing in mind, for sufficiently large problems the proposed technique is found to be superior to standard approaches (like ILU preconditioning) even in a sequential implementation. Future work will focus on parallel implementation and multi-level extensions.

ACKNOWLEDGEMENTS

We would like to thank Daniel Szyld and an anonymous referee for the useful comments and suggestions, which resulted in several improvements.

REFERENCES

1. Toselli A, Widlund O. *Domain Decomposition Methods – Algorithms and Theory*, Springer Series in Computational Mathematics. Springer-Verlag: Berlin/Heidelberg, 2005.
2. Frommer A, Szyld DB. Weighted max norms, splittings, and overlapping additive Schwarz iterations. *Numerische Mathematik* 1999; **83**:259–278.
3. Benzi M, Frommer A, Nabben R, Szyld DB. Algebraic theory of multiplicative Schwarz methods. *Numerische Mathematik* 2001; **89**:605–639.
4. Saad Y. *Iterative Methods for Sparse Linear Systems*. Society for Industrial and Applied Mathematics: Philadelphia, 2003.
5. Cai X-C, Sarkis M. A restricted additive Schwarz preconditioner for general sparse linear systems. *SIAM Journal on Scientific Computing* 1999; **21**:792–797.
6. Frommer A, Szyld DB. An algebraic convergence theory for restricted additive Schwarz methods using weighted max norms. *SIAM Journal on Numerical Analysis* 2000; **39**:463–479.
7. Balay S, Gropp W, McInnes L, Smith B. *et al*. The Portable, Extensible Toolkit for Scientific Computing, Version 2.0.13, 1996. http://www.mcs.anl.gov/petsc, code and documentation.
8. Bru R, Pedroche F, Szyld DB. Additive Schwarz iterations for Markov chains. *SIAM Journal on Matrix Analysis and Applications* 2005; **27**:445–458.
9. Marek I, Szyld DB. Algebraic Schwarz methods for the numerical solution of Markov chains. *Linear Algebra and its Applications* 2004; **386**:67–81.
10. Stewart WJ. MARCA Models: A collection of Markov chain models. Available at http://www4.ncsu.edu/~billy/MARCA_Models/MARCA_Models.html.
11. Berman A, Plemmons RJ. *Nonnegative Matrices in the Mathematical Sciences*. Academic Press: New York, 1979.
12. Stewart WJ. *Introduction to the Numerical Solution of Markov Chains*. Princeton University Press: Princeton, 1994.
13. Neumann M, Plemmons RJ. Convergent nonnegative matrices and iterative methods for consistent linear systems. *Numerische Mathematik* 1978; **31**:265–279.
14. Alefeld G, Schneider N. On square roots of M-matrices. *Linear Algebra and its Applications* 1982; **42**:119–132.

DOI: 10.1002/nla

15. Kavanagh JP, Neumann M. Consistency and convergence of the parallel multisplitting method for singular M-matrices. *SIAM Journal on Matrix Analysis and Applications* 1989; **10**:210–218.
16. Varga RS. Factorization and normalized iterative methods. In *Boundary Problems in Differential Equations*, Langer RE (ed.). University of Wisconsin Press: Madison, 1960; 121–142.
17. Meijerink JA, van der Vorst HA. An iterative solution method for linear systems of which the coefficient matrix is a symmetric M-matrix. *Mathematics of Computation* 1977; **31**:148–162.
18. Varga RS. *Matrix Iterative Analysis*. Prentice-Hall: Englewood Cliffs, 1962.
19. Chan TF, Mathew TP. Domain decomposition methods. *Acta Numerica* 1994; **3**:61–143.
20. Quarteroni A, Valli A. *Domain Decomposition Methods for Partial Differential Equations*, Oxford Science Publications. Clarendon Press: Oxford, 1999.
21. Smith BF, Björstad PE, Gropp WD. *Domain Decomposition: Parallel Multilevel Methods for Elliptic Partial Differential Equations*. Cambridge University Press: Cambridge - New York - Melbourne, 1996.
22. Benzi M, Tůma M. A parallel solver for large-scale Markov chains. *Applied Numerical Mathematics* 2002; **41**:135–153.
23. Buoni J J. Incomplete factorization of singular *M*-matrices. *SIAM Journal on Algebraic and Discrete Methods* 1986; **7**:193–198.
24. Luby M. A simple parallel algorithm for the maximal independent set problem. *SIAM Journal on Computing* 1986; **15**:1036–1053.
25. De Sterck H, Manteuffel TA, McCormick SF, Miller K, Pearson J, Ruge J, Sanders G. Smoothed aggregation multigrid for Markov chains. *SIAM Journal on Scientific Computing* 2010; **32**:40–61.
26. Virnik E. An algebraic multigrid preconditioner for a class of singular *M*-matrices. *SIAM Journal on Scientific Computing* 2007; **29**:1982–1991.
27. Philippe B, Saad Y, Stewart WJ. Numerical methods in Markov chain modeling. *Operations Research* 1992; **40**:1156–1179.
28. Saad Y, Schultz M. GMRES: a generalized minimum residual algorithm for solving nonsymmetric linear systems. *SIAM Journal on Scientific and Statistical Computing* 1986; **7**:856–869.
29. Karypis G, Kumar V. MeTiS: A Software Package for Partitioning Unstructured Graphs, Partitioning Meshes, and Computing Fill-Reducing Orderings of Sparse Matrices, Version 4.0, September 1998. University of Minnesota, Department of Computer Science and Army HPC Research Center, Minneapolis, MN.
30. Cuthill E, McKee J. Reducing the bandwidth of sparse symmetric matrices. *Proceedings of the ACM National Conference* 1969; **24**:157–172.

DOI: 10.1002/nla