SIAM J. MATRIX ANAL. APPL.
Vol. 13, No. 2, pp. 671–679, April 1992
© 1992 Society for Industrial and Applied Mathematics

Downloaded 01/04/13 to 132.206.27.25. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

TWO-STAGE AND MULTISPLITTING METHODS FOR THE PARALLEL SOLUTION OF LINEAR SYSTEMS*

DANIEL B. SZYLD† AND MARK T. JONES‡

Dedicated to Gene Golub on the occasion of his 60th birthday

Abstract. Two-stage and multisplitting methods for the parallel solution of linear systems are studied. A two-stage multisplitting method is presented that reduces to each of the others in particular cases. Conditions for its convergence are given. In the particular case of a multisplitting method related to block Jacobi, it is shown that it is equivalent to a two-stage method with only one inner iteration per outer iteration. A fixed number of iterations of this method, say, p, is compared with a two-stage method with p inner iterations. The asymptotic rate of convergence of the first method is faster, but, depending on the structure of the matrix and the parallel architecture, it takes more time to converge. This is illustrated with numerical experiments.

Key words. solution of linear systems, block iterative methods, parallel methods

AMS(MOS) subject classification. 65F15

1. Introduction and preliminaries.

O'Leary and White [9] introduced the multisplitting method for the parallel solution of linear systems of equations of the form

$$Ax = b,$$

where A is an $n \times n$ matrix and x and b are vectors. Lanzkron, Rose, and Szyld [6] analyzed the convergence of two-stage methods for the solution of the same system. In this paper, we study the relationship between these two methods, also analyzing rates of convergence and relative costs in specific examples. Numerical experiments that illustrate our findings are presented. Further, for illustrative purposes, an algorithm combining the two methods is given.

A matrix T is nonnegative, denoted $T \geq 0$, if it has nonnegative elements. The representation $A = M - N$ is called a splitting of A if M is nonsingular; it is called a regular splitting if $M^{-1} \geq 0$ and $N \geq 0$ [11]; and it is called a weak regular splitting if $M^{-1} \geq 0$ and $M^{-1}N \geq 0$ [10]. A splitting $A = M - N$ is convergent if $\rho(M^{-1}N) < 1$, where $\rho(T)$ denotes the spectral radius of T. Thus, if the splitting $A = M - N$ is convergent, then iterative methods of the form

$$x_{i+1} = M^{-1}Nx_i + M^{-1}b \tag{1}$$

are convergent for any initial vector x_0. It is well known that if $A^{-1} \geq 0$, every weak regular splitting of A is convergent [1]. We denote by I the identity matrix.

DEFINITION 1.1 [9]. Let A, M_k, N_k, and D_k $(k = 1, \cdots, K)$ be $n \times n$ matrices. Let $D_k \geq 0$ be diagonal. Then (M_k, N_k, D_k) $(k = 1, \cdots, K)$ is called a multisplitting of A if

* Received by the editors August 8, 1990; accepted for publication (in revised form) March 9, 1991.

† Department of Mathematics, Temple University, Philadelphia, Pennsylvania 19122-2585. The work of this author was supported by National Science Foundation grant DMS-8807338 (szyld@euclid.math.temple.edu).

‡ Mathematics and Computer Science Division, Argonne National Laboratory, 9700 South Cass Avenue, Argonne, Illinois 60439-4844 (mjones@mcs.anl.gov). This author received support from the Applied Mathematical Sciences subprogram of the Office of Energy Research, U.S. Department of Energy, under contract W-31-109-Eng-38.

(i) $A = M_k - N_k$, $k = 1, \cdots, K$, is a splitting of A;
(ii) $\sum_k D_k = I$.

Given an initial vector x_0 and a multisplitting of A, the multisplitting iterative method is defined by

$$x_{i+1} = \sum_k D_k M_k^{-1} N_k x_i + \sum_k D_k M_k^{-1} b. \tag{2}$$

The solution of the different linear systems

$$M_k y = v \tag{3}$$

in (2) can be solved by different processors in a parallel environment. This fact makes these methods particularly attractive.

O'Leary and White [9] have shown that if for $k = 1, \cdots, K$, $A = M_k - N_k$ is a weak regular splitting of A satisfying $A^{-1} \geq 0$, then the multisplitting method converges for any initial vector x_0; see also Neumann and Plemmons [8]. Examples of multisplittings can be found in these references and in White [13].

If the diagonal (masking) matrices D_k have only zero and one entries, the method (2) corresponds to the (parallel) block Jacobi method [11]. The splittings $A = M_k - N_k$ are then the same as $A = M - N$, since only the nonzeros corresponding to nonzeros of D_k are relevant. We are mainly concerned about this case in this paper; see also §3. We note that there is then no overlap between the groups of variables that are modified in the fractional steps of the algorithm; see (3). In §4 we make some observations for some cases when there is overlap. The dichotomy between overlap and nonoverlap is similar to that of the Schwarz additive algorithm in the context of domain decomposition methods; see, e.g., Dryja [3] or Dryja and Widlund [4] and the references given therein.

Thus, in the nonoverlap case, the processors have to be synchronized between the formation of iterates x_i and x_{i+1}. A discussion of chaotic multisplitting methods (with overlap), where synchronization is not required, can be found in [2].

Two-stage methods, also called inner/outer methods, consist of using a splitting within a splitting; i.e., given a splitting $A = M - N$, let $M = B - C$ and perform, say, p "inner" iterations [6]. Thus the resulting method is

$$x_{i+1} = (B^{-1}C)^p x_i + \sum_{j=0}^{p-1} (B^{-1}C)^j B^{-1}(Nx_i + b). \tag{4}$$

2. Two-stage multisplitting. A simple unification of the two methods described in §1 is considered where each splitting $A = M_k - N_k$ of the multisplitting method can itself be a two-stage method. The discussion in this section is general in the sense that no assumption is made on the form of the matrices D_k.

DEFINITION 2.1. Let A, M_k, B_k, C_k, N_k, and D_k be $n \times n$ matrices. Let $D_k \geq 0$ be diagonal. Then $(M_k, B_k, C_k, N_k, D_k)$ $(k = 1, \cdots, K)$ is called a two-stage multisplitting if

(i) $A = M_k - N_k$, $k = 1, \cdots, K$, is a splitting of A;
(ii) $\sum_k D_k = I$;
(iii) $M_k = B_k - C_k$, $k = 1, \cdots, K$, is a splitting of M_k.

The two-stage multisplitting algorithm is as follows:

Downloaded 01/04/13 to 132.206.27.25. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Downloaded 01/04/13 to 132.206.27.25. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Given the initial vector x_0.
For $i = 0, 1, 2, \cdots$, until convergence.
  For $k = 1$ to K
    $y_{k,0} = x_i$
    For $j = 0$ to $p - 1$
      $B_k y_{k,j+1} = C_k y_{k,j} + N_k x_i + b$
  $x_{i+1} = \sum_k D_k y_{k,p}$

Clearly, this algorithm reduces to the multisplitting method (2) when $p = 1$ and to the two-stage method (4) when $K = 1$. In the analysis of the convergence of the two-stage multisplitting algorithm, we use the following

LEMMA 2.2 [6]. *Given a nonsingular matrix A and T such that $(I - T)^{-1}$ exists, there exists a unique pair of matrices M, N, such that $T = M^{-1}N$ and $A = M - N$, where M is nonsingular. These matrices are $M_T = A(I - T)^{-1}$ and $N_T = M_T - A$. The splitting $A = M_T - N_T$ is the (unique) splitting induced by T.*

Lanzkron, Rose, and Szyld [6] have shown that if the "outer" splitting $A = M - N$ is regular and the "inner" splitting $M = B - C$ is weak regular, then the two-stage method (4) is convergent for any initial vector x_0 and for any value of p. Moreover, under these conditions, the splitting induced by the iteration matrix

$$T_p = I - (I - (B^{-1}C)^p)(I - M^{-1}N) \tag{5}$$

is a weak regular splitting. They also present an example where the outer splitting is weak regular and there is no convergence for certain values of p.

THEOREM 2.3. *Let $A^{-1} \geq 0$. Let $A = M_k - N_k$ be a regular splitting, and let $M_k = B_k - C_k$ be a weak regular splitting for $k = 1, \cdots, K$. Then, the two-stage multisplitting method is convergent for any initial vector x_0 and any value of p.*

Proof. For a fixed k, and for any value of p, the splittings $A = M_k - N_k$ and $M_k = B_k - C_k$ define a (convergent) two-stage method and induce weak regular splittings $A = M_{k,p} - N_{k,p}$. The two-stage multisplitting method reduces then to the multisplitting method (2), where this collection of induced weak regular splittings is used in Definition 1.1(i), and is thus convergent for any initial vector x_0. □

We remark that the preceding discussion can be easily generalized to allow the inclusion of the nested iterative methods described in [6].

3. A comparison of some block methods. Consider a partition of the n variables as $\{1, \cdots, n\} = S = \cup_{k=1}^{K} S_k$, where S_k has n_k elements. Let the diagonal matrices $D_k \geq 0$ $(k = 1, \cdots, K)$ be such that $D_{k,i} = 1$ if $i \in S_k$, and zero otherwise. This implies in particular that $\sum_k D_k = I$. Let $A_k = D_k A D_k$, and let $\bar{A}_k$ be the $n_k \times n_k$ matrix with nonzero entries coinciding with those of A_k. If A is an M-matrix [1], [11], the splitting $A = M - N$, where $M = \sum_k A_k$, is a regular splitting. This is a standard device; and with this splitting, the method (1) is the block Jacobi method, with its inherent parallelism. Let $\bar{A}_k = \bar{B}_k - \bar{C}_k$ $(k = 1, \cdots, K)$ be weak regular splittings. Using the structure of the matrices D_k, we can "assemble" the matrices $\bar{B}_k$ and $\bar{C}_k$ in the obvious way and obtain a weak regular splitting of $M = B - C$, where $B_k = D_k B D_k$ has the same nonzeros as $\bar{B}_k$.

The two-stage method (4) with these matrices corresponds to a block method in which the same number, say, p, of "inner" iterations is performed in each block (the case of different values of p for different blocks has been studied in [6]). If $p = 1$, this two-stage method is equivalent to the multisplitting method (2), where $M_k = B_k$ and $N_k = C_k + N$ $(k = 1, \cdots, K)$. The special case of $\bar{A}_k = \bar{B}_k - \bar{C}_k$ being the splitting

corresponding to the SOR method was treated (with and without overlap) by White [12, §4].

In a parallel implementation of these methods, it is assumed that a different processor solves for the variables in different sets S_k. In the absence of communication delays between processors, p (outer) iterations of the multisplitting method take more work than one iteration of the two-stage method with p "inner" iterations. The difference is $p-1$ matrix times vector products with the matrix N; see (4). It is therefore natural to compare the convergence properties of these two methods. In the next section we study their asymptotic rate of convergence, while in §3.2 we report on some numerical experiments. We show that if T_p is the iteration matrix of the two-stage method with p "inner" iterations, then $\rho(T_p) \geq \rho(T_1)^p$, i.e., this multisplitting method (one inner iteration) is asymptotically faster. Nevertheless, we show that there are problems for which, because of the communications delay and the sparsity structure of N, $p > 1$ will give faster convergence.

3.1. Asymptotic convergence. In the comparison of the asymptotic convergence of the block methods just described, we use the following result, proofs of which can be found in [6] or [7]; conditions for strict inequalities of the spectral radius are also presented in the latter reference.

THEOREM 3.1. *Let $A = M - N = \tilde{M} - \tilde{N}$ be convergent weak regular splittings such that $\tilde{M}^{-1} \geq M^{-1}$, and let x and z be the nonnegative Frobenius eigenvectors of $T = M^{-1}N$ and $\tilde{T} = \tilde{M}^{-1}\tilde{N}$, respectively. If*

$$\tilde{N}z \geq 0 \tag{6}$$

or if $Nx \geq 0$ with $x > 0$, then $\rho(T) \geq \rho(\tilde{T})$.

We present now our main theoretical result. It applies to any two-stage method, and in particular to the case where $M = \sum_k A_k$, $A_k = D_k A D_k$, and $D_k \geq 0$ ($k = 1, \cdots, K$) are diagonal matrices with zeros and ones such that $\sum_k D_k = I$, i.e., to the multisplitting in question.

THEOREM 3.2. *Let $A^{-1} \geq 0$. Let $A = M - N$ be a regular splitting and let $M = B - C$ be a weak regular splitting. Let T_p be the iteration matrix of the two-stage method with p "inner" iterations. Then $\rho(T_p) \geq \rho(T_1)^p$.*

Proof. By Lemma 2.2 the matrix T_p in (5) induces the splitting $A = M_{T_p} - N_{T_p}$, where

$$M_{T_p} = M(I - (B^{-1}C)^p)^{-1} = B(I - B^{-1}C)(I - (B^{-1}C)^p)^{-1}. \tag{7}$$

The iteration matrix of the multisplitting method in question is

$$H = T_1^p = [B^{-1}(C+N)]^p.$$

It induces the splitting $A = M_H - N_H$, where

$$M_H = A(I-H)^{-1} = B(I - B^{-1}(C+N))(I - (B^{-1}(C+N))^p)^{-1}. \tag{8}$$

We will use Theorem 3.1. As shown in [6, Thm. 4], condition (6) is directly satisfied for N_{T_p}. It remains only to show that $M_H^{-1} \geq M_{T_p}^{-1}$. We have that

$$M_{T_p}^{-1} = (I - (B^{-1}C)^p)(I - B^{-1}C)^{-1}B^{-1} = \sum_{i=0}^{p-1}(B^{-1}C)^i B^{-1},$$

Downloaded 01/04/13 to 132.206.27.25. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

$$M_H^{-1} = (I - (B^{-1}(C+N))^p)(I - B^{-1}(C+N))^{-1}B^{-1}$$
$$= \sum_{i=0}^{p-1}(B^{-1}(C+N))^i B^{-1},$$

where we have used the identity

$$(I - R^p)(I - R)^{-1} = \sum_{i=0}^{p-1} R^i.$$

Since $B^{-1} \geq 0$ and $N \geq 0$, $B^{-1}N \geq 0$. Therefore, $B^{-1}(C+N) \geq B^{-1}C$ and the theorem follows. □

3.2. Practical considerations. The two-stage methods with $p > 1$ do have a significant advantage on parallel architectures with nonuniform memory access times: the inner iterations can be accomplished with references exclusively to memory that is "local" to that processor. Examples of architectures that fall into this class are the BBN GP1000, the BBN TC2000, Intel hypercube computers, and the CRAY-2. Parallel computers with cache memory that is local to a processor can also take advantage of these two-stage splitting methods if the data necessary to accomplish the inner iterations can be stored in the cache. Examples of architectures that fall into this class are the Encore Multimax and Sequent Symmetry computers. The two-stage methods can gain an advantage from this data locality only if the number of outer iterations necessary to converge to a solution decreases as p increases. As we have shown in Theorem 3.2, asymptotically, the two-stage method is slower. Therefore, there must be some "optimal" number of inner iterations p, which would depend on the architecture and the matrix in question.

To briefly demonstrate the practical advantages of using $p > 1$ inner iterations on parallel processors with nonuniform memory access times, we ran two experiments on the BBN TC2000 multiprocessor. To perform the experiments, we modified a parallel block iterative code to use the block Jacobi method as the outer iteration, inducing the splitting $A = M - N$, and the point Gauss–Seidel method as the inner iteration, inducing the splitting $M = B - C$. In each experiment the number of processors varied from 1 to 32 and the number of inner iterations was increased until no further decrease in execution time was observed. The iterations were stopped when the ratio of the norms of the residual and the right-hand side was less than 10^{-6}.

In the first experiment, a matrix generated with a nine-point cross stencil on a 64×64 grid was used.[1] All the unknowns in a grid row were grouped into a block, resulting in 64 blocks of size 64×64. The right-hand side was constructed by setting the right boundary of the grid to 1s and the other boundaries to 0s. An examination of the results in Fig. 1 shows that a clear advantage is gained when p is increased. We have included in the same graph the number of outer iterations needed for convergence. Of course the number of outer iterations is independent of the number of processors.

In the second experiment, a matrix of order 1280 with a semi-bandwidth of 20 was used. The off-diagonals inside the band were -1, and the elements in the ith diagonal position were $(\sum_{j \neq i} \mid a_{ij} \mid) + 2$. The matrix was partitioned into 64 20×20

[1] Note: In order to significantly reduce the number of outer iterations as p increases, as many of the larger elements in A as possible must be included in B. To simulate this, we used the following stencil on the grid: $A(i,i) = 9.02$, $A(i,i+1) = A(i,i+2) = A(i+1,i) = A(i+2,i) = -2.24$, $A(i,i+gridsize) = A(i,i+gridsize*2) = A(i+gridsize,i) = A(i+gridsize*2,i) = -0.01$.

Downloaded 01/04/13 to 132.206.27.25. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

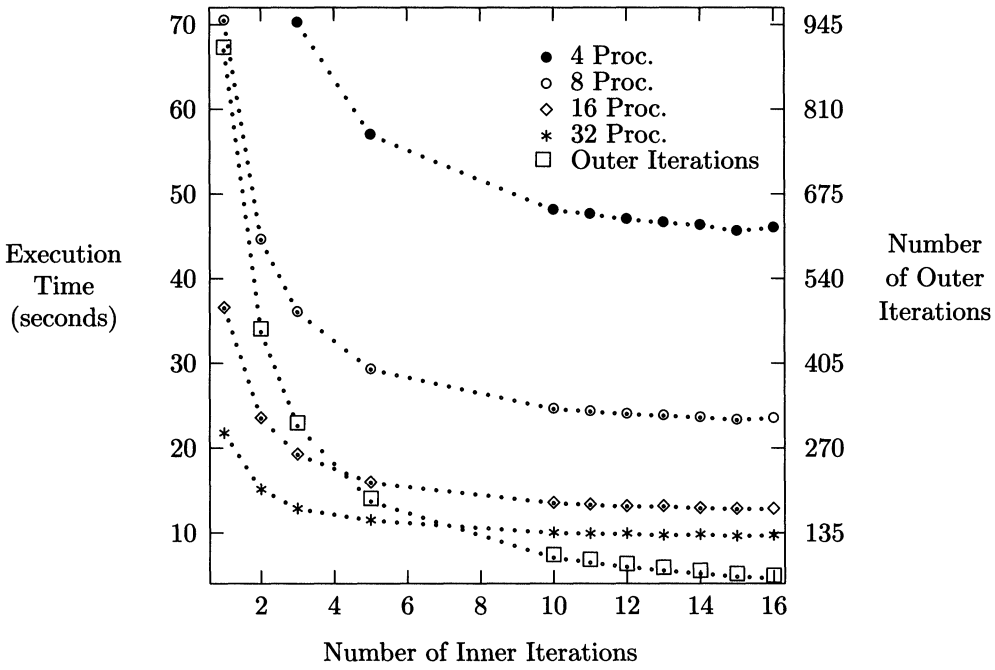

FIG. 1. *Results from experiment on grid problem.*

blocks. The right-hand side consisted of 1s in every 64th position and 0s in every other position. In this experiment, the reduction in the number of outer iterations as p increases was not large; therefore, the possible advantage to be gained from increasing p, if any, was small. The results are given in Fig. 2.

4. The effect of overlap. The asymptotic results of Theorem 3.2 apply only to a specific class of multisplittings. The same theory does not extend easily to other cases. The question arises then if the conclusions of §3.2 can also be observed in more general multisplittings. In particular, is it true that when there is overlap, $p > 1$ inner iterations is still more advantageous? Also, if this is the case, how does the order of the variables affect the efficiency of the method? This last question was studied in a recent paper by White [13].

The experiments in this section were devised to explore these questions. The matrix used corresponds to a nine-point discretization on a 256×256 grid ($n = 65536$). The coefficients associated with a row on the grid are: $-1/6$, $-1/3$, 1, $-1/3$, $-1/6$. We have considered blocks of size 2048 corresponding to 8 rows of 256 nodes, and an overlap of 256 nodes, i.e., one row of overlap.

TABLE 1
Effect of the order of the variables.

Inner Iterations	Ordering 0	Ordering 1	Ordering 2
1	320 (531)	314 (509)	339 (551)
2	211 (276)	211 (267)	223 (283)
4	165 (152)	169 (149)	174 (153)
8	162 (94)	172 (94)	172 (94)
16	207 (69)	219 (68)	222 (69)

Downloaded 01/04/13 to 132.206.27.25. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

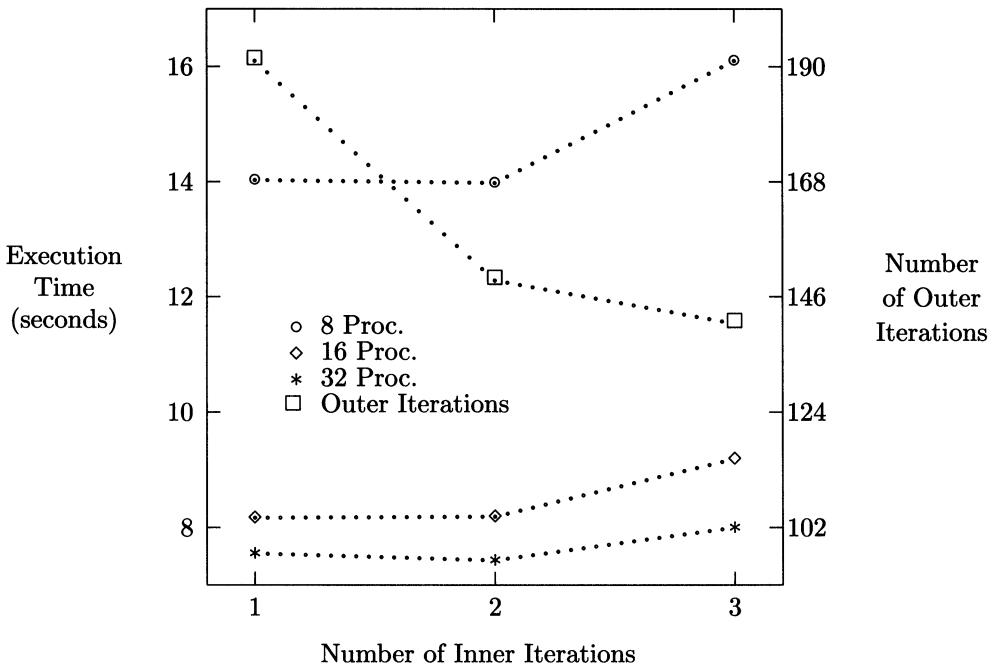

FIG. 2. *Results from experiment on band problem.*

The experiment reported in Table 1 illustrates the effect of different orderings when overlap is used. We have used 16 processors of the BBN TC2000. In each column, the time in seconds is given first and then in parentheses is the number of (outer) iterations. Ordering 0 is the natural ordering. Ordering 1 is numbering the overlapping rows first. Ordering 2 is numbering the overlapping rows last. It can be readily observed that (a) $p > 1$ can indeed give better results, as in the case described in §3, (b) fewer outer iterations are needed when numbering the overlap first, as in the case studied by White [13], and (c) the effect of the ordering diminishes dramatically as the number of inner iterations is increased. We should point out that the timings for orderings 1 and 2 can be improved by using more sophisticated code for the reordering portion of it.

In Fig. 3 we report on an experiment with the same matrix, using ordering 1, with different number of processors, with and without overlap. It is interesting to observe that for p less than the "optimal" value, the nonoverlap case is faster while the situation is reverse for p larger. At the same time, there is little difference between the two cases near the "optimal" value. For completeness we report that the number of outer iterations in the nonoverlap case are: 549, 303, 186, 135, and 114, for 1, 2, 4, 8, and 16 inner iterations, respectively. Those for the overlap case are given in Table 1.

5. Concluding remarks. We have compared, in the case of (outer) block Jacobi, a two-stage method and a corresponding multisplitting method for the parallel solution of linear systems. Based on the asymptotic rate of convergence, the first method appears to be slower. On the other hand, the savings in communication between processors makes the two-stage method more competitive for a certain range of p, the number of inner iterations. The same situation was observed in two-stage

Downloaded 01/04/13 to 132.206.27.25. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

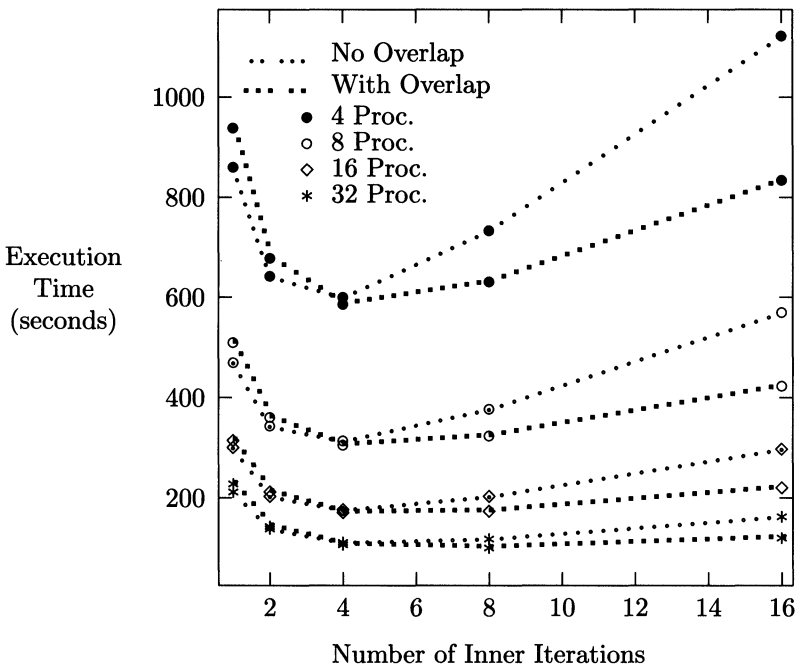

FIG. 3. *Time in seconds with and without overlap.*

multisplittings with overlap.

Acknowledgments. We extend thanks to Julio César Díaz and Michael Neumann for asking us about the relationship between two-stage and multisplitting methods, to Paul Lanzkron for an interesting discussion about his computational experience in [5], and to James Wilkes for making available his code for parallel block iterative methods. We also thank a referee for useful comments which led us to write §4.

REFERENCES

[1] A. BERMAN AND R. J. PLEMMONS, *Nonnegative Matrices in the Mathematical Sciences*, Academic Press, New York, 1979.

[2] R. BRU, L. ELSNER, AND M. NEUMANN, *Models of parallel chaotic iteration methods*, Linear Algebra Appl., 103 (1988), pp. 175–192.

[3] M. DRYJA, *An additive Schwarz algorithm for two- and three-dimensional finite element problems*, in Second International Symposium on Domain Decomposition Methods for Partial Differential Equations, T. Chan, R. Glowinski, G. Meurant, J. Pèriaux, and O. Widlund, eds., Society for Industrial and Applied Mathematics, Philadelphia, PA, 1989, pp. 168–172.

[4] M. DRYJA AND O. B. WIDLUND, *Some domain decomposition algorithms for elliptic problems*, in Iterative Methods for Large Linear Systems, Academic Press, San Diego, CA, 1989, pp. 273–291 (Proceeding of the Conference on Iterative Methods for Large Linear Systems held in Austin, TX, October 19–21, 1988, to celebrate the sixty-fifth birthday of David M. Young, Jr.).

[5] P. J. LANZKRON, *Nested iterative methods for linear systems*, Ph.D. thesis, Department of Computer Science, Duke University, Durham, NC, August 1989.

[6] P. J. LANZKRON, D. J. ROSE, AND D. B. SZYLD, *Convergence of nested classical iterative methods for linear systems*, Numer. Math., 58 (1991), pp. 685–702.

[7] I. MAREK AND D. B. SZYLD, *Comparison theorems for weak splittings of bounded operators*, Numer. Math., 58 (1990), pp. 387–397.

Downloaded 01/04/13 to 132.206.27.25. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

[8] M. NEUMANN AND R. J. PLEMMONS, *Convergence of parallel multisplitting methods for M-matrices*, Linear Algebra Appl., 88-89 (1987), pp. 559–574.
[9] D. P. O'LEARY AND R. E. WHITE, *Multi-splittings of matrices and parallel solution of linear systems*, SIAM J. Algebraic Discrete Meth., 6 (1985), pp. 630–640.
[10] J. M. ORTEGA AND W. C. RHEINBOLDT, *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York, London, 1970.
[11] R. S. VARGA, *Matrix Iterative Analysis*, Prentice-Hall, Englewood Cliffs, NJ, 1962.
[12] R. E. WHITE, *Multisplittings and parallel iterative methods*, Comput. Meth. Appl. Mech. Engrg., 64 (1987), pp. 567–577.
[13] ———, *Multisplitting of a symmetric positive definite matrix*, SIAM J. Matrix Anal. Appl., 11 (1990), pp. 69–82.

Downloaded 01/04/13 to 132.206.27.25. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php