

On the Effects of Scaling of the Peaceman-Rachford Method

By Olof B. Widlund*

Abstract. The alternating direction method of Peaceman and Rachford is considered for elliptic difference schemes of second order and with two independent variables. An earlier result of the author's on the rapid convergence of multi-parameter noncommutative problems is described and a connection is established between that result and theorems on optimal scaling of band matrices. Simple algorithms to decrease the condition number and increase the rate of convergence are discussed.

1. Introduction. In this paper we shall consider the alternating direction implicit (ADI) method of Peaceman and Rachford [12] when applied to difference approximations to elliptic problems with two independent variables. It is known that this method is often quite powerful, especially when different acceleration parameters are used in the different iteration steps. Usually, these parameters are chosen in a cyclic way. We shall assume that this is the case and denote the cycle length by m .

It has been proved that the method always converges when $m = 1$, but for the potentially much more powerful multi-parameter case the theory is still not satisfactory. Indeed, there seems to be little hope that there will ever be a very general convergence theory because of the fact that divergence has been observed in numerical experiments.

Under certain additional restrictions on the problem, we can theoretically explain the full power of the method. Thus, there exists a very satisfactory theory in the case when the two matrices, corresponding to the different independent variables, commute. Cf. Varga [15] or Wachspress [16]. The commutativity condition is however very limiting because, as was shown by Birkhoff and Varga [1], it imposes severe restrictions on the coefficients as well as on the region. The region thus has to be rectangular. In fact all problems giving rise to commutative problems can be handled by separation-of-variables techniques. It is of interest to note that for separable problems there now exist faster methods than the ADI or SOR methods. Cf. Hockney [10] and Buzbee, Golub and Nielson [2] for methods which are in fact very efficient computer implementations of the separation-of-variables idea.

We shall now make a short survey of results for the noncommutative case. (Cf. Wachspress [16] for more details.) One of the more interesting results follows from an observation by Guilinger [8]. It is thus possible to prove the convergence of

Received March 5, 1970.

AMS 1970 subject classifications. Primary 65F10, 65F35, 65N20; Secondary 15A12.

Key words and phrases. Alternating direction implicit methods, Peaceman-Rachford method, elliptic of second order, multi-parameter, noncommutative, separation-of-variables, condition number, rate of convergence, optimal scaling.

* This work was sponsored in part by the U. S. Atomic Energy Commission, Contract AT(30-1)-1480, at the Courant Institute of Mathematical Sciences, New York University; and in part by the Office of Naval Research, Contract ONR-3656(23), at the University of California at Berkeley.

Copyright © 1971, American Mathematical Society

the ADI method for a difference approximation to Laplace's equation on convex regions for any set of positive parameters. However, the result fails to explain the very rapid convergence which has been observed in many applications. Pearcy [13] has shown that any given problem can be made to converge by choosing m sufficiently large. Practical experience also indicates that a choice of a long cycle is a cure for divergence. Pearcy's technique is however not refined enough to give realistic bounds for the rate of convergence or to explain in any satisfactory way the marked difference between one- and multi-parameter ADI.

A class of noncommutative problems was treated in an earlier paper, Widlund [18]. A recipe was given for the choice of parameters to ensure a rate of convergence which, for small enough mesh sizes, is as large as those of comparable commutative cases. In order to make the proof work, severe restrictions had to be imposed. We were thus only able to treat Dirichlet problems on rectangular regions for equations with sufficiently smooth coefficients. Furthermore, we had to let the iteration parameters depend on the independent variables in a special way. Numerical experiments strongly indicate that such a choice of parameters improves the performance of the algorithm, even in much more general situations than those covered by the theory, changing divergent or slowly convergent problems into rapidly convergent ones. As will be explained in Section 2, the use of parameters depending on the space variables is equivalent to a scaling of the matrices of the problems by a diagonal matrix. Frequently, one can look upon the scaling procedure as a device which will change the original problem into one which is closer to one which can be treated by separation-of-variables techniques.

We shall now outline an idea of the proof of the main result in Widlund [18] in order to be able to describe the results of the present paper. Denote by S_m the matrix which maps the initial error vector into the error vector after a full cycle of m iterations. Our objective is to give a good bound for the spectral radius of this matrix. Such a bound can be given in terms of the norm of a matrix S'_m which is similar to our original matrix S_m . (We will use the spectral norm as our matrix norm throughout this paper.) After a choice of a similarity transformation, we write S'_m as a sum of a principal matrix P_m and a remainder matrix R_m . The matrix P_m is the product of two matrices each of which corresponds to operations in one space direction only. The norm of P_m can therefore be estimated as if we were dealing with two one-dimensional problems. In commutative cases $R_m \equiv 0$ and, by an appropriate choice of a similarity transformation, the same is true for $m = 1$ as well. Under the assumptions mentioned above we were able to give a good enough bound for the norm of R_m . In the present paper, we shall, instead, concentrate our attention on minimizing the norm of P_m by an appropriate scaling. It will be seen that we will end up with exactly the same recommendation as in Widlund [18], where we were concerned primarily with the norm of R_m .

The idea of scaling ADI problems is far from new. Cf. Wachspress [16], Wachspress and Habetler [17], Douglas [4] and Gunn [9] for various ideas and results.

It would be most interesting if something conclusive could be said in a comparison between the successive overrelaxation methods and an optimally scaled one-parameter Peaceman-Rachford algorithm, (in which case the remainder matrix vanishes). No general theorem seems to be true which would rank one method ahead of the other. However, numerical experiments seem to indicate that in cases, when

the coefficients of the elliptic problem vary very much in magnitude, well scaled ADI methods give quite good rates of convergence while the successive overrelaxation methods are very time consuming.

We end this section with a warning. The scaling which is recommended in this paper is based on considerations about the norm of P_m only. A scaling can turn a commutative problem into a noncommutative one and such a problem might conceivably give rise to a divergent ADI algorithm. However, such a situation appears to be quite unlikely in problems arising from physics or other applications. We also note in this context that it is quite simple to decide whether or not a problem is commutative.

Acknowledgement. The author wants to express his gratitude to Professor Gene Golub. Not only did he observe a relation between the results in Widlund [18] and those of Forsythe and Straus [5], an observation which started this investigation, but he also gave generously of his time discussing the details of this paper.

2. Presentation of the Algorithm and Earlier Results. Let $A_h u = f$ be a system of linear equations which has arisen from a difference approximation to a second-order selfadjoint elliptic equation with two independent variables and no mixed derivatives. The subscript h is a parameter which goes to zero with the mesh size of the problem. We assume throughout that A_h is the sum of two matrices H and V . These are symmetric positive-definite sparse matrices. In order to get an efficient algorithm we also assume that $H + D$ and $V + D$ can be inverted rapidly for any choice of a diagonal matrix D with positive elements. We shall refrain from going into details about how to split A_h into the sum of H and V and only mention that appropriate splittings are often suggested by the original problem and by efficiency considerations. The matrices H and V typically have band structure or, as in the case of periodic boundary conditions, almost band structure with only a few nonzero elements in each row. The corresponding linear system can therefore be rapidly solved with the help of a Cholesky or LU decomposition or in important special cases by odd/even reduction (cf. Buzbee, Golub and Nielson [2]). It is well known that these procedures can be made numerically stable.

We refer to Varga [15] for a description of how one sets up difference approximations to elliptic problems. It could be mentioned that we can always assure the symmetry of our matrices by choosing a method of discretization based on a variational formulation.

For a chosen splitting we thus write our system of linear equations as

$$(H + V)u = f.$$

The ADI algorithm is defined as follows: Given some initial approximation u_1 , compute u_{n+1} , $n = 1, 2, \dots$, by

$$\begin{aligned} (\omega_{nH} D^2 + H)u_{n+1/2} &= (\omega_{nH} D^2 - V)u_n + f, \\ (\omega_{nV} D^2 + V)u_{n+1} &= (\omega_{nV} D^2 - H)u_{n+1/2} + f. \end{aligned}$$

D is a diagonal matrix with strictly positive elements. The iteration parameters ω_{nH} and ω_{nV} are chosen in a cyclic way, i.e.

$$\omega_{nH} = \omega_{lH}, \quad \omega_{nV} = \omega_{lV} \quad \text{for } n = l \bmod m.$$

We want to choose D and $\omega = \{\omega_{nH}, \omega_{nV}\}$, $n = 1, 2, \dots, m$, so that the rate of convergence is as large as possible. Our choice can be restricted so that $\omega_{nH}D^2 + H$ and $\omega_{nV}D^2 + V$ are positive definite.

Denote by S_m the matrix which maps the initial error into the error after a full cycle of iterations. Clearly,

$$S_m = \prod_{n=1}^m (\omega_{nV}D^2 + V)^{-1}(\omega_{nV}D^2 - H)(\omega_{nH}D^2 + H)^{-1}(\omega_{nH}D^2 - V)$$

if we adopt the convention

$$\prod_{i=1}^l A_i = A_l A_{l-1} \cdots A_1.$$

Introduce the matrices

$$H_D = D^{-1}HD^{-1}, \quad V_D = D^{-1}VD^{-1}.$$

It is easy to show that

$$DS_mD^{-1} = \prod_{n=1}^m (\omega_{nV}I + V_D)^{-1}(\omega_{nV}I - H_D)(\omega_{nH}I + H_D)^{-1}(\omega_{nH}I - V_D).$$

Thus S_m is similar to the error matrix which we get by applying the algorithm with $D = I$ to the scaled problem $(H_D + V_D)v = g$. Let T be some appropriate transformation matrix. (Cf. Section 1.) Define S'_m by $S'_m = TDS_mD^{-1}T^{-1}$. The spectral radii of S_m and S'_m are clearly the same. The principal matrix P_m and the remainder matrix R_m are defined by

$$P_m = \prod_{n=1}^m (\omega_{nH}I + H_D)^{-1}(\omega_{nV}I - H_D) \prod_{n=1}^m (\omega_{nV}I + V_D)^{-1}(\omega_{nH}I - V_D)$$

and

$$R_m = S'_m - P_m.$$

We refer to Widlund [18] for a good choice of T and D which enables us to give a useful bound for the norm of R_m for the five-point difference approximation to a class of problems with smooth coefficients. A further restriction in that paper namely that $\omega_{nH} = \omega_{nV}$ can be removed. Our earlier results thus hold true in particular for parameters chosen optimally as in Wachspress [16], cf. Section 4. In our earlier paper we gave a set of easily computable parameters, such that

$$\|P_m\| \leq 1 - C_1 h^{1/m}$$

and

$$\|R_m\| \leq C_2 h^{3/2m},$$

provided D^2 is chosen to be equal to the diagonal of H or V . C_1 and C_2 are strictly positive constants. An immediate consequence is

$$\rho(S_m) \leq 1 - C_1 h^{1/m} + C_2 h^{3/2m}$$

and rapid convergence for small mesh sizes. ($\rho(A)$ = spectral radius of A .) Here h is a meshparameter such that the order of our matrix A_h is a const h^{-2} .

We shall now concentrate on making $\|P_m\|$ small by an appropriate choice of D and ω . We assume that accurate lower and upper bounds are available for the spectra of H_D and V_D but that nothing useful is known about the location of the eigenvalues inside these intervals. In practice such bounds might be computed with the help of a Sturm sequence or a symmetric QR subroutine. Cf. Kahan and Varah [11]. Repeating standard argument, we get

$$\begin{aligned} \|P_m\| &\leq \left\| \prod_{n=1}^m (\omega_{nH} I + H_D)^{-1} (\omega_{nV} I - H_D) \right\| \\ &\quad \times \left\| \prod_{n=1}^m (\omega_{nV} I + V_D)^{-1} (\omega_{nH} I - V_D) \right\| \\ &\leq \max_{\lambda \in [\lambda_l, \lambda^u]; \mu \in [\mu_l, \mu^u]} \left| \prod_{n=1}^m \left(\frac{\omega_{nV} - \lambda}{\omega_{nH} + \lambda} \right) \left(\frac{\omega_{nH} - \mu}{\omega_{nV} + \mu} \right) \right| \\ &= \phi_m(\omega, \lambda_l, \lambda^u, \mu_l, \mu^u). \end{aligned}$$

Here λ_l is the smallest eigenvalue of H_D or an accurate lower bound of it, λ^u the largest eigenvalue of H_D or an upper bound thereof and μ_l and μ^u are the corresponding bounds for V_D . The problem of minimizing $\phi_m(\omega, \lambda_l, \lambda^u, \mu_l, \mu^u)$ with respect to ω has been solved completely. Cf. Wachspres [16].

By the homogeneity of ϕ_m the minimum of ϕ_m with respect to ω will depend only on three parameters, $\kappa(H_D) = \lambda^u/\lambda_l$, $\kappa(V_D) = \mu^u/\mu_l$, and $\Lambda = \lambda^u/\mu^u$. Two of them, $\kappa(H_D)$ and $\kappa(V_D)$, are equal to or very close to the spectral condition numbers of the matrices H_D and V_D . Denote by

$$\psi_m(\kappa(H_D), \kappa(V_D), \Lambda) = \min_{\omega} \phi_m(\omega, \lambda_l, \lambda^u, \mu_l, \mu^u).$$

It is the purpose of our study to show that an appropriate scaling D of our problem will give values for the arguments of ψ_m such that we are quite close to the best possible estimate for $\|P_m\|$.

3. On the Eigenvalues of Certain Matrices. Let $B_D = D^{-1}BD^{-1}$ be symmetric and positive definite and denote by $\kappa(B_D)$ its spectral condition number defined as the ratio between the largest and smallest eigenvalues of the matrix. The diagonal matrix D is strictly positive. We say that B_D is best conditioned if $\kappa(B_D) \leq \kappa(B_{D'})$ for all such diagonal matrices D' . The following result follows immediately from Forsythe and Straus [5] and Golub [7].

LEMMA 3.1. *Let B have Young's property A. Then B_D is best conditioned if its diagonal elements are equal. Furthermore, the value of the diagonal elements is equal to the average of the largest and smallest eigenvalues of B_D .*

All tridiagonal matrices satisfy property A. This is also true for many matrices which correspond to one-dimensional differential operators with periodic boundary conditions.

For more general symmetric positive-definite sparse matrices, we can use the following interesting result by van der Sluis [14].

LEMMA 3.2. *Let B have at most q nonzero elements in any row. Then the condition number of the best conditioned matrix is at most a factor q smaller than that of a matrix B_D which has constant diagonal elements.*

van der Sluis also proved that this kind of diagonal scaling can never improve the condition number of a matrix by more than a factor $n =$ order of the matrix. That result, however, seems to be of less interest to us because we are primarily interested in quite large matrices.

It is clear that the scaling strategy suggested by these results is very attractive because of its simplicity. In many cases we are dealing with elliptic problems where the coefficients of the second order terms are the same. The diagonals of H_D and V_D will then typically be almost equal. The following bound on possible improvement of the condition numbers is then of interest.

LEMMA 3.3. *Let the maximum ratio of any two diagonal elements of B be γ . Then the condition number cannot be improved by more than a factor γ by making the diagonal elements equal.*

The proof follows directly from Courant's min max principle.

These results will be of interest in our attempts to make ψ_m small. It will thus be shown in Section 4 that ψ_m will decrease when $\kappa(H_D)$ and $\kappa(V_D)$ both decrease provided Λ is kept fixed. This naturally leads to the question whether Λ can be expected to change only slightly under diagonal scaling. We shall now describe a result in that direction for the important special case of tridiagonal matrices.

LEMMA 3.4. *Let B_h be a class of symmetric tridiagonal matrices the order n_h of which increases with decreasing h . For simplicity let $h \cdot n_h = 1$. The elements of B_h are such that*

$$(B_h)_{i,i} = f(ih, h) > 0, \quad i = 1, 2, \dots, n_h,$$

$$(B_h)_{i,i+1} = -\frac{1}{2}f(ih, h)(1 + o(1)), \quad i = 1, 2, \dots, n_h - 1,$$

where $f(x, h)$ depends smoothly on h and is piecewise smooth with respect to $x \in [0, 1]$. The largest eigenvalue λ^u of B_h will then satisfy

$$\lambda^u = 2 \max_{x \in [0, 1]} f(x, 0)(1 + o(1)).$$

Proof. Geršgorin's theorem immediately gives that $\lambda^u \leq 2 \max f(x, 0)(1 + o(1))$. In order to get a reverse inequality we use the variational formulation of the largest eigenvalue. An appropriate trial vector can easily be constructed. First, choose for a given $\epsilon > 0$ a subinterval such that $f(x, h)$ is smooth and $f(x, 0) \geq \max_{x \in [0, 1]} f(x, 0) - \epsilon$. Let all vector components corresponding to x values outside this interval be zero and let the rest of them be ± 1 alternatingly. It is easy to see that this will result in a good enough lower bound for λ^u .

Remark. This proof contains elements quite similar to the arguments which are used in derivation of asymptotic expressions for the eigenvalues of two-dimensional elliptic problems. Cf. Courant, Hilbert [3] or Garabedian [6]. Indeed, interesting information about the larger eigenvalues of difference equations can be obtained by such an approach. It could also be remarked that the method of transformations of variables which gives the asymptotic behavior of the eigenvalues for continuous two-point boundary-value problems does not seem to have any useful discrete analogue.

Possible applications of Lemma 3.4 should now be obvious. Thus consider a problem for which the principal part has the form $-\partial_x a(x, y)\partial_x u - \partial_y a(x, y)\partial_y u$, $a(x, y)$ piecewise smooth. Set up the standard five-point difference approximation

and make a natural splitting of the matrix. Lemma 3.4 then implies that Λ will change only slightly under a scaling which is such that the elements of the scaling matrix are values of a piecewise continuous function in two variables evaluated at the corresponding mesh points. Many problems in physics lead to elliptic problems the principal parts of which have this simple form provided Cartesian coordinates are used.

Remark. One might believe that if a scaling decreases the condition number of H_D and V_D then the solution of the tridiagonal systems of equations in the ADI scheme might become more accurate. However, typically, if we use Gaussian elimination we do not want to do any pivoting of positive-definite matrices and, with an unchanged pivoting strategy, the quality of the solution will not be improved by scaling.

4. Recommended Scalings of the Matrices. We shall begin this section by reviewing one aspect of the theory for the optimal choice of ω . The components of ω can be computed easily in the case when $m =$ power of 2; in the general case, there exist accurate approximate formulas. We shall rely heavily on Wachspress' [16] presentation of the theory in this section. One can compute the value of $\psi_m(\kappa(H_D), \kappa(V_D), \Lambda)$, for given $\lambda_l, \lambda^u, \mu_l$ and μ^u , in terms of a parameter $k', 0 < k' < 1$, defined by

$$k' = 1/((m' + 1) + (m'(m' + 2))^{1/2}),$$

where

$$m' = 2(\lambda^u - \lambda_l)(\mu^u - \mu_l)/(\lambda^u + \mu^u)(\lambda_l + \mu_l).$$

For any cycle length m , there exists a strictly positive constant C_m such that

$$\psi_m = ((1 - C_m(k')^{1/2m})/(1 + C_m(k')^{1/2m}))^2,$$

when the mesh size goes to zero. Our problem is therefore reduced to the study of the value of k' , which measures how well conditioned our problem is with respect to the ADI algorithm.

LEMMA 4.1. *Let λ^u/μ^u be fixed. Then k' increases if both λ^u/λ_l and μ^u/μ_l decrease.*

To prove this lemma we rewrite m' as

$$(4.1) \quad m' = 2 \frac{\frac{\lambda^u}{\mu^u} \left(1 - \frac{\lambda_l}{\lambda^u}\right) \left(1 - \frac{\mu_l}{\mu^u}\right)}{\left(\frac{\lambda^u}{\mu^u} \times \frac{\lambda_l}{\lambda^u} + \frac{\mu_l}{\mu^u}\right) \left(\frac{\lambda^u}{\mu^u} + 1\right)}$$

and note that a decreasing m' will increase the value of k' .

We can now give an explicit recommendation for a scaling, when the principal part of the differential operator has the form $-\partial_x a(x, y) \partial_x u - \partial_y a(x, y) \partial_y u$ and the matrices H_D and V_D are similar, via permutations, to tridiagonal matrices. It follows immediately from Lemmas 3.1, 3.3, 3.4 and 4.1 that a good choice for D^2 is the diagonal of H or V .

Remark. This scaling is close to the one suggested by Douglas [4]. He also gave the scaling an interpretation in terms of parabolic equations. To any positive-definite elliptic problem one can associate many parabolic equations whose steady-state

solution is equal to the solution of the elliptic problem. Each scaling corresponds to a particular choice of a parabolic problem.

Several scaling strategies now suggest themselves for the general problem. One, which for smooth enough $a(x, y)$ and small mesh sizes would lead to an almost identical algorithm in the special case above, amounts to choosing D^2 equal to the diagonal of $H + V$. There are, however, some objections to such a strategy. On the basis of our earlier results, Widlund [18], and numerical experiments, it seems as if an attempt should be made to scale the original problem so that the modified problem could "almost" be treated by a separation-of-variables technique, provided the region happened to be appropriate. Choosing D^2 equal to the diagonal of $H + V$ in the case of a diffusion problem written in polar coordinates would thus be quite unnatural. In that case, it seems much more natural to choose D^2 equal to the diagonal of the matrix corresponding to the derivatives with respect to the angle. By the results of Section 3, such a scaling would either minimize the condition number of one of our matrices or at least make it quite well conditioned. Such a strategy also coincides with the one suggested by the analysis in Widlund [18].

The recommended strategy would therefore be to choose D^2 equal to the diagonal of H or V . We cannot support this choice as strongly in the general case as in the special case above, because λ^u/μ^u might vary a great deal. However, examining the formula (4.1) again, we see that not only can we compare the values of k' for a scaled and an unscaled case, if we have eigenvalue bounds available, but we can also give an upper bound for m' , and thus a lower bound for k' , in terms of the condition numbers of H_D and V_D alone. Varying λ^u/μ^u , we thus find that

$$m' \leq 2 \frac{\left(1 - \frac{\lambda_l}{\lambda^u}\right) \left(1 - \frac{\mu_l}{\mu^u}\right)}{\left(\left(\frac{\lambda_l}{\lambda^u}\right)^{1/2} + \left(\frac{\mu_l}{\mu^u}\right)^{1/2}\right)^2},$$

with equality only for $\lambda^u/\mu^u = ((\mu_l/\mu^u)/(\lambda_l/\lambda^u))^{1/2}$. This formula should give a realistic estimate of the size of m' in many cases. A more accurate comparison can of course be given if we have information about the size of λ^u/μ^u and how it changes under scaling. It follows from Lemma 3.4 and its proof that a Geršgorin estimate often gives an accurate bound for the largest eigenvalues. It is clear from formula (4.1) that, if we can show that λ^u/μ^u does not change very much and at least one of the condition numbers decreases a lot, then we can be assured of a larger value for k' .

New York University
 Courant Institute of Mathematical Sciences
 251 Mercer Street
 New York, New York 10012

1. G. BIRKHOFF & R. S. VARGA, "Implicit alternating direction methods," *Trans. Amer. Math. Soc.*, v. 92, 1959, pp. 13–24. MR 21 #4549.

2. B. L. BUZBEE, G. H. GOLUB & C. W. NIELSON, *The Method of Odd/Even Reduction and Factorization with Application to Poisson's Equation*, Stanford Computer Science Department Report, 1969.

3. R. COURANT & D. HILBERT, *Methods of Mathematical Physics*. Vol. I, Interscience, New York, 1953. MR 13, 800.

4. J. DOUGLAS, JR., "Alternating direction methods for three space variables," *Numer. Math.*, v. 4, 1962, pp. 41–63. MR 24 #B2122.
5. G. E. FORSYTHE & E. G. STRAUS, "On best conditioned matrices," *Proc. Amer. Math. Soc.*, v. 6, 1955, pp. 340–345. MR 16, 1054.
6. P. R. GARABEDIAN, *Partial Differential Equations*, Wiley, New York, 1964. MR 28 #5247.
7. G. H. GOLUB, "Comparison of the variance of minimum variance and weighted least squares regression coefficients," *Ann. Math. Statist.*, v. 34, 1963, pp. 984–991. MR 27 #5336.
8. W. H. GUILINGER, JR., "The Peaceman-Rachford method for small mesh increments," *J. Math. Anal. Appl.*, v. 11, 1965, pp. 261–277. MR 32 #607.
9. J. E. GUNN, "On the two-stage iterative method of Douglas for mildly nonlinear elliptic difference equations," *Numer. Math.*, v. 6, 1964, pp. 243–249. MR 29 #6637.
10. R. W. HOCKNEY, "A fast direct solution of Poisson's equation using Fourier analysis," *J. Assoc. Comput. Mach.*, v. 12, 1965, pp. 95–113. MR 35 #3913.
11. W. KAHAN & J. VARAH, *Two Working Algorithms for the Eigenvalues of a Symmetric Tridiagonal Matrix*, Stanford Computer Science Department Report, 1966.
12. D. W. PEACEMAN & H. H. RACHFORD, JR., "The numerical solution of parabolic and elliptic differential equations," *J. Soc. Indust. Appl. Math.*, v. 3, 1955, pp. 28–41. MR 17, 196.
13. C. M. PEARCY, "On the convergence of alternating direction procedures," *Numer. Math.*, v. 4, 1962, pp. 172–176. MR 26 #3206.
14. A. VAN DER SLUIS, "Condition numbers and equilibration of matrices." (To appear.)
15. R. S. VARGA, *Matrix Iterative Analysis*, Prentice-Hall, Englewood Cliffs, N. J., 1962. MR 28 #1725.
16. E. L. WACHSPRESS, *Iterative Solution of Elliptic Systems, and Applications to the Neutron Diffusion Equations of Reactor Physics*, Prentice-Hall, Englewood Cliffs, N. J., 1966. MR 38 #2965.
17. E. L. WACHSPRESS & G. J. HABETLER, "An alternating-direction-implicit iteration technique," *J. Soc. Indust. Appl. Math.*, v. 8, 1960, pp. 403–424. MR 22 #5132.
18. O. B. WIDLUND, "On the rate of convergence of an alternating direction implicit method in a noncommutative case," *Math. Comp.*, v. 20, 1966, pp. 500–515. MR 37 #7104.