On Optimal Alternating Direction Parameters for Singular Matrices

By R. B. Kellogg and J. Spanier

1. Introduction. One of the most widely used iterative methods for approximating solutions of partial difference equations in two space dimensions is the method of alternating directions, considered originally by Peaceman and Rachford [6]. Application of this method involves the selection of certain auxiliary numbers, called acceleration parameters, which are chosen to enhance the convergence of the iterative process. Extremely favorable convergence rates may be obtained for this method, but only when the acceleration parameters are properly chosen.

The matrix problem to be solved may be written in the form

$$(H + V)x = y_0, \tag{1}$$

where H and V are symmetric, positive semi-definite matrices and $H + V$ is also symmetric and positive semi-definite. A necessary and sufficient condition for a solution of (1) to exist is that y_0 be orthogonal to the null space of $H + V$. Of course, if this null space is empty, $H + V$ is nonsingular and a unique solution x exists but in some applications this is not the case. For example, solving Laplace's equation with Neumann boundary conditions by finite difference methods leads to an equation (1) in which $H + V$ is singular.

The alternating direction method applied to (1) may be defined by

$$\begin{aligned} x_{k+(1/2)} &= -(H + a_{k+1}D)^{-1}[(V - a_{k+1}D)x_k - y_0], \\ x_{k+1} &= -(V + a_{k+1}D)^{-1}[(H - a_{k+1}D)x_{k+(1/2)} - y_0], \end{aligned} \tag{2}$$

where x_0 is an arbitrary initial vector, D is a symmetric positive definite normalizing matrix, usually defined as in [3], [8], or [12], and the a_k are the acceleration parameters. It has been shown [4] that there exist parameter sequences for which the iterates x_k converge to a solution of (1), even when $H + V$ is singular. However, all algorithms known to the authors for generating acceleration parameters which result in rapid convergence of the iterates x_k degenerate to the absurd choice $a_k = 0$. This paper shows how useful acceleration parameters may be chosen for singular problems.

2. Preliminaries. We shall make use of the notation and some of the results of [4] to set the stage for our analysis. We begin with the following assumptions:

(a) The vector y_0 is orthogonal to $\eta(H + V)$ = the null space of $H + V$.

(b) The sequence $\{a_k\}$ is monotone and cyclic; i.e., $a_1 \geqq a_2 \geqq \cdots \geqq a_t > 0$, $a_k = a_{k+t}$.

The integer t will be referred to as the cycle length.

Received January 11, 1965.

Now define matrices T_k and Z by

$$T_k = (V + a_kD)^{-1}(H - a_kD)(H + a_kD)^{-1}(V - a_kD),$$
$$Z = T_tT_{t-1} \cdots T_1,$$

so that Z is the iteration matrix after one cycle of t iterations. As in [4], let

$$H_1 = D^{-1/2}HD^{-1/2}, \qquad V_1 = D^{-1/2}VD^{-1/2},$$
$$T_k' = D^{1/2}T_kD^{-1/2} = (V_1 + a_kI)^{-1}(H_1 - a_kI)(H_1 + a_kI)^{-1}(V_1 - a_kI),$$
$$Z_1 = D^{1/2}ZD^{-1/2} = T_t'T_{t-1}' \cdots T_1',$$

and let E denote the orthogonal projection on $\eta(H_1 + V_1)$.

The following results are established in [4] and will be useful in our analysis:

(T1) The matrix Z_1 coincides with E on $\eta(H_1 + V_1)$.

(T2) Let α, β be any positive numbers with $\alpha \leqq \beta$. Then there exists a positive integer t_0 such that any parameter sequence S of the form

$$S = \{a_1, \cdots, a_{t_0}\}, \qquad \beta \geqq a_1 \geqq \cdots \geqq a_{t_0} \geqq \alpha,\ a_k = a_{k+t_0},$$

will cause the powers Z_1^n to converge to the projection E and the powers Z^n to converge to $D^{-1/2}ED^{1/2}$, a projection on $\eta(H + V)$.

(T3) With the same assumptions as in (T2), the sequence $\{x_k\}$ converges to a solution of (1).

From this point on we assume that t is fixed and that k is a multiple of t, so that our discussion is aimed at methods in which cycles of t iterations are always completed. If x is any solution of (1), since x is transformed into itself by (2), we have

$$\begin{aligned} x - x_k &= Z^{k/t}(x - x_0) \\ &= Z^n(x - x_0), \qquad n = k/t \text{ an integer.} \end{aligned} \tag{3}$$

It is easily seen from (3) and (T2) that, when the x_k converge to a solution of (1), they converge to that solution w whose projection on $\eta(H + V)$ satisfies

$$D^{-1/2}ED^{1/2}w = D^{-1/2}ED^{1/2}x_0.$$

Setting $x = w$ in (3), we have

$$\begin{aligned} w - x_k &= Z^n(w - x_0) \\ &= (Z^n - D^{-1/2}ED^{1/2})(w - x_0) \\ &= (Z - D^{-1/2}ED^{1/2})^n(w - x_0). \end{aligned} \tag{4}$$

It is common practice (see, e.g., [9, p. 67]) in dealing with iterative methods such as (2) to define the rate of convergence in terms of the spectral radius of the iteration matrix. Thus, if the spectral radius is used as a criterion, equation (4) reveals that we should seek to minimize

$$\rho(Z - D^{-1/2}ED^{1/2}) = \sup \lambda(Z - D^{-1/2}ED^{1/2}),$$

the supremum being taken over all eigenvalues of $Z - D^{-1/2}ED^{1/2}$. But

$$Z - D^{-1/2}ED^{1/2} = D^{-1/2}(Z_1 - E)D^{1/2}$$

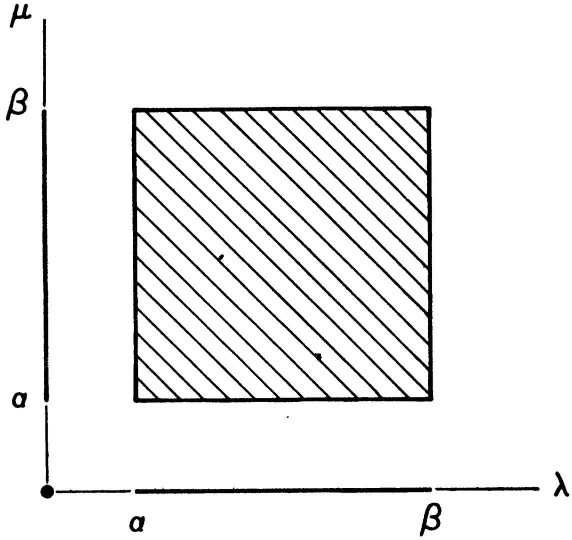

FIG. 1

so that it suffices to minimize $\rho(Z_1 - E)$. This minimization will be performed in the next section for a restricted class of pairs of matrices (H_1, V_1).

3. Commutative Analysis. In this section we will show how to select acceleration parameters which minimize $\rho(Z_1 - E)$, assuming that $H_1V_1 = V_1H_1$. We will also show how to apply these acceleration parameters in the noncommutative case.

The assumption of commutativity, while necessary for our analysis, is rarely satisfied in practice. Nevertheless, parameters based on the commutative assumption have worked well in a variety of nonsingular problems [1]. Our own experience with singular problems is along the same lines. A variety of heat conduction problems have been solved using the program HOT [7] in which both matrices H and V were singular. In some of these problems, $H + V$ was also singular. Observed convergence rates, using the parameters suggested by our theorem, were substantially the same as those observed in nonsingular problems.

Let α and β be given with $0 < \alpha < \beta$. Let λ and μ be real variables and let R be the point set in the (λ, μ) plane defined by

$$R = (0, 0) \cup \{(\lambda, 0): \alpha \leqq \lambda \leqq \beta\} \cup \{(0, \mu): \alpha \leqq \mu \leqq \beta\} \cup \{(\lambda, \mu): \alpha \leqq \lambda, \mu \leqq \beta\}.$$

The set R is shown in Figure 1. Let $\mathcal{C}$ be the collection of all pairs (H_1, V_1) of symmetric $N \times N$ matrices which commute, $H_1V_1 = V_1H_1$, and which satisfy the condition that if λ and μ are eigenvalues of H_1 and V_1, respectively, having the same eigenvector, then $(\lambda, \mu) \in R$.

Consider a set of positive acceleration parameters $a_1, \cdots, a_t$. As has been seen, the spectral radius $\rho(Z_1 - E)$ is a measure of the rate of convergence of the iterations (2) using these acceleration parameters. Therefore, the function

$$r(a_1, \cdots, a_t) = \sup \{\rho(Z_1 - E): (H_1, V_1) \in \mathcal{C}\}$$

is a measure of the least favorable rate of convergence that would be encountered in using the parameters $a_1, \cdots, a_t$ on pairs of matrices drawn from the class $\mathfrak{C}$.

The following theorem gives a set of parameters that makes this least favorable convergence rate as good as possible. To state the theorem, we introduce a function $f(a_1, \cdots, a_t)$ defined by

$$f(a_1, \cdots, a_t) = \max_{\alpha \leqq s \leqq \beta} \prod_{j=1}^{t} \frac{|a_j - s|}{|a_j + s|}. \tag{5}$$

Let $\tilde{a}_1, \cdots, \tilde{a}_t$ be the unique (in descending order) set of parameters which minimize $f(a_1, \cdots, a_t)$. The existence and uniqueness of this minimizing set of parameters is proved in [10].

THEOREM. *If $a_1, \cdots, a_t$ is a nonnegative set of parameters, then* $r(\tilde{a}_1, \cdots, \tilde{a}_t) \leqq r(a_1, \cdots, a_t) = f(a_1, \cdots, a_t)$.

Proof. Let $(H_1, V_1) \in \mathfrak{C}$ and let $a_1, \cdots, a_t$ be a nonnegative set of parameters. Since H_1 and V_1 are symmetric and commute, there is an orthogonal basis of common eigenvectors of H_1 and V_1. If w is an element of the basis, and if $H_1 w = \lambda w$, $V_1 w = \mu w$, then w is an eigenvector of Z_1 with eigenvalue

$$\nu = \prod_{j=1}^{t} \frac{(a_j - \lambda)(a_j - \mu)}{(a_j + \lambda)(a_j + \mu)}.$$

E is the orthogonal projection on the subspace generated by the basis vectors w such that $\lambda + \mu = 0$. If $\lambda + \mu = 0$, then $\lambda = \mu = 0$ and $\nu = 1$. Hence, w is an eigenvector of $Z_1 - E$ with eigenvalue 0 if $\lambda = \mu = 0$ and with eigenvalue ν otherwise. Since $(\lambda, \mu) \in R$, it is easily seen that $|\nu| \leqq f(a_1, \cdots, a_t)$ when $\lambda + \mu > 0$. Hence, $\rho(Z_1 - E) \leqq f(a_1, \cdots, a_t)$, and, therefore,

$$r(a_1, \cdots, a_t) \leqq f(a_1, \cdots, a_t). \tag{6}$$

By compactness, the maximum in (5) must be obtained for some $s \in [\alpha, \beta]$. Let $H_1 = sI$, $V_1 = 0$. Then $(H_1, V_1) \in \mathfrak{C}$ and $\rho(Z_1 - E) = f(a_1, \cdots, a_t)$. Hence, using (6),

$$r(a_1, \cdots, a_t) = f(a_1, \cdots, a_t).$$

The rest of the theorem follows from the definition of $\tilde{a}_1, \cdots, \tilde{a}_t$.

The parameters $\tilde{a}_1, \cdots, \tilde{a}_t$ may be considered optimal parameters for matrix pairs in the class $\mathfrak{C}$. These parameters are determined by the numbers α, β, and t. There is in the literature a variety of methods for computing them [10, 11] or approximating them [2, 12].

If this theorem is compared with equation (7.42) of [9], it is seen that, using the parameters $\tilde{a}_1, \cdots, \tilde{a}_t$, a nonsingular problem may be expected to converge twice as fast as a singular problem with the same parameter bounds α and β.

In the noncommutative case, $H_1 V_1 \neq V_1 H_1$, nothing is known about the nature of an optimal set of parameters. Nevertheless one can, with proper interpretation, apply the optimal commutative parameters $\tilde{a}_1, \cdots, \tilde{a}_t$ in the noncommutative case. It suffices to observe that α and β are, respectively, lower and upper bounds for the least positive and the largest eigenvalues of the matrices H_1 and V_1, $(H_1, V_1) \in \mathfrak{C}$. If $H_1 V_1 \neq V_1 H_1$ we may still construct such lower and upper bounds α and β and from these obtain parameters $\tilde{a}_1, \cdots, \tilde{a}_t$ for use in (2). The upper bound

β may be obtained from Gershgorin's theorem. A method of obtaining lower bounds for the least positive eigenvalue of a certain type matrix is discussed in [5].

Bettis Atomic Power Laboratory
Westinghouse Electric Corporation
West Mifflin, Pennsylvania

1. G. BIRKHOFF, R. S. VARGA & D. YOUNG, *Alternating Direction Implicit Methods*, Advances in Computers, Vol. 3, Academic Press, New York, 1962.
2. C. DE BOOR & J. R. RICE, "Chebyshev approximation by a $\Pi(x - r_i)/(x + s_i)$ and application to ADI iteration," *J. Soc. Indust. Appl. Math.*, v. 11, 1963, pp. 159–169. MR **28** #466.
3. J. DOUGLAS, "Alternating direction methods for three space variables," *Numer. Math.*, v. 4, 1962, pp. 41–63. MR **24** #B2122.
4. J. DOUGLAS & C. PEARCY, "On convergence of alternating direction procedures in the presence of singular operators," *Numer. Math.*, v. 5, 1963, pp. 175–184. MR **27** #4384.
5. W. H. GUILINGER & R. B. KELLOGG, "Eigenvalue lower bounds," submitted to the *Bettis Technical Review*.
6. 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.
7. R. B. SMITH & J. SPANIER, *HOT-1: A Two Dimensional Steady-State Heat Conduction Program for the Philco-2000*, WAPD-TM-465, Bettis Atomic Power Laboratory, Pittsburgh, Pa., 1964.
8. J. SPANIER & W. H. GUILINGER, "Matrix conditioning for alternating direction methods," *Bettis Technical Review*, WAPD-BT-29, 1963, pp. 69–79.
9. R. S. VARGA, *Matrix Iterative Analysis*, Prentice-Hall, Englewood Cliffs, N. J., 1962. MR **28** #1725.
10. E. L. WACHSPRESS, "Optimum alternating-direction-implicit iteration parameters for a model problem," *J. Soc. Indust. Appl. Math.*, v. 10, 1962, pp. 339–350. MR **27** #921.
11. E. L. WACHSPRESS, "Extended application of alternating direction implicit iteration model problem theory," *J. Soc. Indust. Appl. Math.*, v. 11, 1963, pp. 994–1016.
12. 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.

An Iterative Method for Computing the Generalized Inverse of an Arbitrary Matrix

By Adi Ben-Israel

Abstract. The iterative process, $X_{n+1} = X_n(2I - AX_n)$, for computing A^{-1}, is generalized to obtain the generalized inverse.

An iterative method for inverting a matrix, due to Schulz [1], is based on the convergence of the sequence of matrices, defined recursively by

$$X_{n+1} = X_n(2I - AX_n) \qquad (n = 0, 1, \cdots) \tag{1}$$

to the inverse A^{-1} of A, whenever X_0 approximates A^{-1}. In this note the process (1) is generalized to yield a sequence of matrices converging to A^+, the generalized inverse of A [2].

Let A denote an $m \times n$ complex matrix, A^* its conjugate transpose, $P_{R(A)}$ the perpendicular projection of E^m on the range of A, $P_{R(A^*)}$ the perpendicular projection of E^n on the range of A^*, and A^+ the generalized inverse of A.

THEOREM. *The sequence of matrices defined by*

$$X_{n+1} = X_n(2P_{R(A)} - AX_n) \qquad (n = 0, 1, \cdots), \tag{2}$$

Received November 9, 1964.