J. SIAM NUMER. ANAL.
Vol. 3, No. 3, 1966
Printed in U.S.A.

ON ITERATIVE COMPUTATION OF GENERALIZED INVERSES AND ASSOCIATED PROJECTIONS*

ADI BEN-ISRAEL† AND DAN COHEN‡

Introduction. The generalized inverse A^+ of an arbitrary complex matrix A [9], [7] and the perpendicular projection AA^+ [8] play a sufficiently important role in matrix applications to justify the current interest and research in their computational aspects. The subject of this paper is the iterative method [2], [3]:

$$Y_0 = \alpha A^*,$$
$$Y_{k+1} = Y_k(2I - AY_k), \qquad k = 0, 1, \cdots,$$

which yields A^+ as the limit of the sequence $\{Y_k\}$, $k = 0, 1, \cdots$, when α satisfies condition (1) (or (30) below. This method, a variant of the well-known Schultz method [8], is of the 2nd order (Theorems 1, 2 below). Its relation to the iterative method [4],

$$X_0 = \alpha A^*,$$
$$X_{k+1} = X_k + \alpha(I - X_kA)A^*, \qquad k = 0, 1, \cdots,$$

is shown, in Theorem 3 below, to be:

$$Y_k = X_{2^k-1} \qquad k = 0, 1, \cdots.$$

An upper bound on $\| A^+ - Y_k \|$, and the optimal α, are given in Theorems 4 and 5.

An iterative method for computing AA^+ based on $\{Y_k\}$, $k = 0, 1, \cdots$, is: $Z_k = AY_k$, i.e.,

$$Z_0 = \alpha AA^*,$$
$$Z_{k+1} = 2Z_k - Z_k^2, \qquad k = 0, 1, \cdots.$$

The traces of Z_k, $k = 1, 2, \cdots$, are shown in Theorem 6 to be a monotone increasing sequence converging to rank A. A division free bound for rank A (Corollary 2) and a criterion for nonsingularity (Corollary 3) follow now easily.

---

* Received by the editors September 24, 1965, and in revised form January 18, 1966. This research was supported by the Office of Naval Research under Contract Nonr-1228 (10), Project NR 047-021, and by the United States Army Research Office (Durham) under Contract DA-31-124-ARO-D-322.

† Departments of Engineering Sciences and Industrial Engineering and Management Sciences, Northwestern University, Evanston, Illinois.

‡ Department of Mathematics, Technion-Israel Institute of Technology, Haifa, Israel.

Direct methods for computing AA^+ were given by Householder [8], Rosen [11], Pyle [10] and others. The correct determination of rank A is a critical factor in these methods, even more so in the direct methods for computing A^+, e.g., Golub and Kahan [6]. The iterative methods $\{Y_k\}$, $\{Z_k\}$, $k = 0, 1, \cdots$, for computing A^+ and AA^+, and the bounds for rank A, given in this paper, may consequently be of some interest.

0. Notations and preliminaries. Let A denote an $m \times n$ nonzero complex matrix, A^* its *conjugate transpose*, A^+ its *generalized inverse* [9], $R(A)$, $N(A)$ its *range* and *null space*, respectively, $r = \text{rank } A$.

Let $\lambda_1(A^*A) \geqq \lambda_2(A^*A) \geqq \cdots \geqq \lambda_n(A^*A)$ be the eigenvalues of A^*A. From rank $A = r$ it follows that $\lambda_r(A^*A) > 0$ and $\lambda_i(A^*A) = 0$ for $i = r + 1, \cdots, n$. We will use the matrix norm $\| A \| = \lambda_1^{\frac{1}{2}}(A^*A)$, which is subordinate to the Euclidean vector norm (e.g., [8, p. 44] where this matrix norm is called $\text{lub}_S(A)$).

For a subspace L of the n-dimensional complex Euclidean space E^n let P_L denote the *perpendicular projection* on L.

The following results are needed in the sequel.

THEOREM 0.1. *Let the real* α *satisfy*

$$0 < \alpha < \frac{2}{\lambda_1(A^*A)}. \tag{1}$$

Then the sequence

$$X_k = \alpha \sum_{p=0}^{k} A^*(1 - \alpha AA^*)^p \qquad k = 0, 1, \cdots, \tag{2}$$

converges to A^+ *as* $k \to \infty$. (*See* [4].)

THEOREM 0.2. *Let* α *satisfy* (1). *Then the sequence*

$$Y_0 = \alpha A^*, \tag{3}$$

$$Y_{k+1} = Y_k(2I - AY_k), \qquad k = 0, 1, \cdots, \tag{4}$$

converges to A^+ *as* $k \to \infty$. (*See* [2], [3].)

1. On the iterative computation of A^+. In terms of the residuals $P_{R(A)} - AX_k$ and $P_{R(A)} - AY_k$ we have, as in the nonsingular case ([8, p. 94]), the following:

THEOREM 1. (a) *The process* (2) *is of the* 1*st order.* (*b*) *The process* (4) *is of the* 2*nd order.*

Proof.

(a) The process (2) is rewritten as

$$\begin{aligned} X_{k+1} &= X_k\,(I - \alpha AA^*) + \alpha A^* \\ &= X_k + \alpha\,(I - X_kA)A^*, \qquad k = 0, 1, \cdots, \end{aligned} \tag{5}$$

with

$$X_0 = \alpha A^*. \tag{6}$$

From (5) it follows that

$$AA^+ - AX_{k+1} = AA^+ - AX_k - \alpha(I - AX_k)AA^*; \tag{7}$$

and since $A = AA^+A$,

$$AA^+ - AX_{k+1} = (AA^+ - AX_k)\,(I - \alpha AA^*). \tag{8}$$

Since $\| I - \alpha AA^* \| < 1$, by (1), and $AA^+ = P_{R(A)}$, [4], it follows that:

$$\begin{aligned} \| P_{R(A)} - AX_{k+1} \| &\leqq \| I - \alpha AA^* \| \, \| P_{R(A)} - AX_k \| \\ &< \| P_{R(A)} - AX_k \|. \end{aligned} \tag{9}$$

(b) Similarly we verify that

$$\begin{aligned} AA^+ - AY_{k+1} &= AA^+ - AY_k - AY_k\,(I - AY_k) \\ &= AA^+ - AY_k - AY_k\,(AA^+ - AY_k), \end{aligned} \tag{10}$$

where $Y_k = Y_k\,AA^+$ holds because $Y_k = C_kA^*$ for some matrix C_k, [2], $k = 0, 1, \cdots$, and $A^* = A^*AA^+$, [9]. From (10) it follows that

$$AA^+ - AY_{k+1} = (AA^+ - AY_k)^2, \qquad k = 0, 1, \cdots, \tag{11}$$

and finally

$$\| P_{R(A)} - AY_{k+1} \| \leqq \| P_{R(A)} - AY_k \|^2, \qquad k = 0, 1, \cdots. \tag{12}$$

In terms of convergence to A^+, the corresponding results are given by the following theorem.

THEOREM 2. (a) *The process* (2) *satisfies*:

$$\| A^+ - X_{k+1} \| < \| A^+ - X_k \|, \qquad k = 0, 1, \cdots. \tag{13}$$

(b) *The process* (4) *satisfies*:

$$\| A^+ - Y_{k+1} \| \leqq \| A \| \, \| A^+ - Y_k \|^2, \qquad k = 0, 1, \cdots \tag{14}$$

Proof.

(a) Using (5) and $A^+AA^* = A^*$, [9], it follows that

$$A^+ - X_{k+1} = (A^+ - X_k)\,(I - \alpha AA^*), \quad k = 0, 1, \cdots, \tag{15}$$

which, because of (1), proves (13).

(b) Similarly, (14) follows from

$$A^+ - Y_{k+1} = (A^+ - Y_k)A(A^+ - Y_k), \quad k = 0, 1, 2, \cdots, \tag{16}$$

which is obtained by using the easily verified relations

$$Y_k = A^+AY_k = Y_kAA^+, \qquad k = 0, 1, \cdots .$$

To establish the relation between the processes (2) and (4) we need the following lemma.

LEMMA. *Let S be any square complex matrix and $k \geqq 0$ an integer. Then*

$$\sum_{j=0}^{k} S(I - S)^j = SS^+[I - (I - S)^{k+1}]. \tag{17}$$

Proof. By induction. For $k = 0, 1$, (17) holds because $S = SS^+S$. Assuming that (17) holds for k, it also holds for $k + 1$ since

$$\sum_{j=0}^{k+1} S(I - S)^j = SS^+[I - (I - S)^{k+1}] + S(I - S)^{k+1}$$
$$= SS^+[I - (I - S)^{k+2}].$$

The sought relation is that (4) is a "subprocess" of (2).

THEOREM 3.

$$Y_k = X_{2^k-1}, \qquad k = 0, 1, \cdots . \tag{18}$$

Proof. Using (4) and (3), and the remark following (10), it follows that

$$\begin{aligned} Y_k &= A^+[I - (I - AY_{k-1})^2] = A^+[I - (I - AY_{k-p})^{2^p}] \\ &= A^+[I - (I - \alpha AA^*)^{2^k}]. \end{aligned} \tag{19}$$

From (2) it follows that

$$X_{2^k-1} = \alpha \sum_{p=0}^{2^k-1} A^*(I - \alpha AA^*)^p = A^+ \sum_{p=0}^{2^k-1} (\alpha AA^*)\ (I - \alpha AA^*)^p. \tag{20}$$

Using the lemma with $S = \alpha AA^*$ and the easily verifiable fact that $\alpha AA^*(\alpha AA^*)^+ = AA^+$, we conclude that

$$X_{2^k-1} = A^+[I - (I - \alpha AA^*)^{2^k}], \tag{21}$$

which, compared with (19), proves (18).

Remark. Using Euler's identity [4],

$$(1 + x) \prod_{p=1}^{k-1} (1 + x^{2^p}) = \sum_{p=0}^{2^k-1} x^p, \qquad |x| < 1, \tag{22}$$

and Theorem 3, we obtain:

$$Y_k = \alpha A^*[I + (I - \alpha AA^*)] \prod_{p=1}^{k-1} [I + (I - \alpha AA^*)^{2^p}], \tag{23}$$

which corresponds to A_k^+ in [4, (54)].

THEOREM 4.[1]

$$\| A^+ - Y_k \| \leqq \frac{\lambda_1^{1/2}(A^*A)}{\lambda_r(A^*A)} (1 - \alpha\lambda_r(A^*A))^{2^k}, \quad k = 0, 1, \cdots . \tag{24}$$

Proof. Using Theorems 0.1 and 3 it follows that

$$\begin{aligned} A^+ - Y_k &= \alpha \sum_{p=2^k}^{\infty} A^*(I - \alpha AA^*)^p \\ &= \alpha \sum_{p=2^k}^{\infty} A^*(AA^+ - \alpha AA^*)^p, \qquad k = 0, 1, \cdots . \end{aligned} \tag{25}$$

As in [3] we verify that

$$\| AA^+ - \alpha AA^* \| = | 1 - \alpha\lambda_r(A^*A) |;$$

and therefore

$$\begin{aligned} \| A^+ - Y_k \| &\leqq \alpha \| A^* \| \sum_{p=2^k}^{\infty} \| AA^+ - \alpha AA^* \|^p \\ &\leqq \frac{\alpha \| A^* \| \| AA^+ - \alpha AA^* \|^{2^k}}{1 - \| AA^+ - \alpha AA^* \|} \\ &\leqq \frac{\lambda_1^{1/2}(A^*A)(1 - \alpha\lambda_r(A^*A))^{2^k}}{\lambda_r(A^*A)}, \qquad k = 0, 1, \cdots . \end{aligned} \tag{26}$$

Remarks. (a) This theorem corrects an error in [4, Theorem 17]. (b) As in [5] we call α_0 *optimal* if it minimizes $\| AA^+ - \alpha AA^* \|$.

The function $F(\alpha) = \| AA^+ - \alpha AA^* \|$ is convex and $F(0) = F(2/\lambda_1(A^*A)) = 1$. As in [1] it can be shown that $F(\alpha)$ has a unique minimum in the interval

$$0 < \alpha < \frac{2}{\lambda_1(A^*A)}.$$

THEOREM 5. *The optimal* α *is*

$$\alpha_0 = \frac{2}{\lambda_1(A^*A) + \lambda_r(A^*A)}, \tag{27}$$

for which

$$\| A^+ - Y_k \| \leqq \frac{\lambda_1^{1/2}(A^*A)}{\lambda_r(A^*A)} \left(\frac{\lambda_1(A^*A) - \lambda_r(A^*A)}{\lambda_1(A^*A) + \lambda_r(A^*A)} \right)^{2^k}, \quad k = 0, 1, \cdots . \tag{28}$$

Proof. As in [1] the minimizing α_0 must satisfy

[1] Recall that $\lambda_r(A^*A)$ is the smallest nonzero (positive) eigenvalue of A^*A, and note that $| 1 - \alpha\lambda_r(A^*A) | < 1$ since $\lambda_r(A^*A) \leqq \lambda_1(A^*A)$ and (1).

$$(29) \qquad 1 - \alpha\lambda_r(A^*A) = -(1 - \alpha\lambda_1(A^*A)),$$

i.e., the interval $[\lambda_r(A^*A), \lambda_1(A^*A)]$ is mapped onto an interval symmetric around the origin. Now, (29) gives (27), which yields (28) when substituted in (24).

Using well-known bounds on $\lambda_1(A^*A) = \lambda_1(AA^*)$, it is possible to replace condition (1) by another condition which is more easily checked: Writing $AA^* = (b_{ij})$, $i, j = 1, \cdots, m$, the Gershgorin theorem [8] implies that

$$\lambda_1(A^*A) \leqq \max_{i=1,\cdots,m} \sum_{j=1}^{m} |b_{ij}|.$$

Therefore (1) can be replaced by:

$$(30) \qquad 0 < \alpha < \frac{2}{\max_{i=1,\cdots,m} \sum_{j=1}^{m} |b_{ij}|}.$$

Other bounds [8] on $\lambda_1(A^*A)$ yield similar conditions.

2. On the iterative computation of AA^+. An iterative method for computing AA^+, based on the process (3) and (4), is given in the following corollary.

COROLLARY 1. *Let α satisfy* (1). *Then the sequence of matrices*

$$(31) \qquad Z_0 = \alpha AA^*,$$

$$(32) \qquad Z_{k+1} = 2Z_k - Z_k^2, \qquad k = 0, 1, \cdots,$$

converges to AA^+ as $k \to \infty$, and

$$(33) \qquad \| P_{R(A)} - Z_{k+1} \| \leqq \| P_{R(A)} - Z_k \|^2 \qquad k = 0, 1, \cdots.$$

Proof. The corollary follows from Theorems 0.2 and 1 (b) by noting that $Z_k = AY_k$, $k = 0, 1, \cdots$.

The following fact about the process (31), (32) is useful.

THEOREM 6. *The trace of Z_k is a monotone increasing function of k, $k = 1, 2, \cdots$, converging to rank A.*

Proof. From the easily verifiable fact

$$(34) \qquad Z_k = I - (I - \alpha AA^*)^{2^k}, \qquad k = 0, 1, \cdots,$$

it follows that:

$$\text{trace } Z_k = m - \text{trace } \{(I - \alpha AA^*)^{2^k}\}$$

$$(35) \qquad = m - \sum_{i=1}^{m} (1 - \alpha\,\lambda_i(AA^*))^{2^k} = m - \sum_{i=1}^{r} (1 - \alpha\lambda_i(AA^*))^{2^k}$$

$$- (m - r) = r - \sum_{i=1}^{r} (1 - \alpha\, \lambda_i(AA^*))^{2k}, \quad k = 0, 1, \cdots,$$

where the third equality in (35) follows from

$$\lambda_i\,(AA^*) = 0, \qquad i = r + 1, \cdots, m.$$

From (1) it follows that:

$$|\,1 - \alpha\, \lambda_i(AA^*)\,| < 1, \qquad i = 1, \cdots, r;$$

and from (35):

$$\text{trace } Z_{k+1} \geqq \text{trace } Z_k, \qquad k = 1, 2, \cdots, \tag{36}$$

and

$$\lim_{k\to\infty} \text{trace } Z_k = r = \text{rank } A. \tag{37}$$

Remark. For α large enough, $1 - \alpha\, \lambda_i(AA^*) < 0$ for some i. Thus it is obvious from (35) that possibly

$$\text{trace } Z_0 > \text{trace } Z_1 .$$

For a real x let $[x]$ denote the integral part of x; e.g., $[3.5] = 3$, $[-2.5] = -3$. Let $\langle x\rangle = -[-x]$; e.g., $\langle 3.5\rangle = 4$.[2]

Division free bounds on the rank and nullity of A are derived from Theorem 6.

Corollary 2. *For every integer $k \geqq 1$ and real α satisfying* (1),

$$\text{rank } A \geqq \langle \text{trace } Z_k\rangle, \tag{38}$$

$$\dim N(A^*) \leqq [\text{trace}\,\{(I - \alpha AA^*)^{2k}\}]. \tag{39}$$

Proof. Equation (38) follows from (35). Equation (39) follows from the facts that the sequence

$$(I - \alpha AA^*)^{2k} = I - Z_k, \qquad k = 0, 1, \cdots, \tag{40}$$

converges to $P_{N(A^*)}$ by Corollary 1, and the sequence of traces,

$$\{\text{trace } (I - Z_k)\}, \qquad k = 1, 2, \cdots,$$

is monotone decreasing by Theorem 6.

A consequence of the above is the following corollary.

Corollary 3. *The square matrix A is nonsingular if and only if for some integer $k \geqq 1$ and for some real $\beta > 0$*

$$\text{trace}\,\{(I - \beta\, AA^*)^{2k}\} < 1. \tag{41}$$

Proof. The proof follows from (39) by noting that a scalar $\beta > 0$ satisfies

[2] Thus $\langle x\rangle = [x] + 1$ unless x is an integer, in which case $x = [x] = \langle x\rangle$.

$$| 1 - \beta\lambda_i(AA^*) | < 1, \qquad i = 1, \cdots, r,$$

if and only if β satisfies (1).

3. Examples. The computation of A^+ by the iterative method of (3) and (4), and of AA^+ by (31) and (32), is demonstrated below. In each example, five values of α satisfying (30) were used:

$$\alpha_p = \frac{p/3}{\max_{i=1,\cdots,m} \sum_{j=1}^{m} | b_{ij} |}, \qquad p = 1, \cdots, 5.$$

The sequence of traces

$$\{\text{trace } (I - Z_k)\} = \{\text{trace } (I - \alpha AA^*)^{2^k}\}, \qquad k = 0, 1, \cdots,$$

which is monotone decreasing for $k = 1, 2, \cdots$, and converges to the nullity of A^*, indicates the rate of convergence. Computations were carried out on a PHILCO-2000.

Example 1. The matrix is:

$$A = \begin{pmatrix} 1 & 4 & 0 \\ 2 & 3 & 0 \\ 2 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}.$$

The sequence trace $(I - Z_k)$ for α_p, $p = 1, \cdots, 5$, converges to the nullity of A^* which is 1.

p	1	2	3	4	5
α_p	0.010101	0.020202	0.030303	0.040404	0.050505
k	trace $(I - Z_k)$				
0	3.646464	3.292929	2.939393	2.585858	2.232323
1	3.386287	2.959289	2.719008	2.665442	2.798592
2	3.044291	2.664607	2.498218	2.380443	2.344645
3	2.703913	2.400470	2.228713	2.111508	2.036046
4	2.412875	2.129182	1.993923	1.924015	1.882346
5	2.137676	1.930274	1.854851	1.805310	1.761924
6	1.933500	1.805974	1.721921	1.647348	1.580391
7	1.806340	1.647827	1.521131	1.419059	1.336854
8	1.648066	1.419678	1.271578	1.175610	1.113470
9	1.419988	1.176130	1.073754	1.030839	1.012875
10	1.176389	1.031022	1.005440	1.000951	1.000166
11	1.031113	1.000962	1.000029	1.000001	1.000000
12	1.000968	1.000001	1.000000	1.000000	
13	1.000000	1.000000			

The sequence (4) converges to the generalized inverse

$$A^+ = \begin{pmatrix} -0.6 & 0.8 & 0 & 0 \\ 0.4 & -0.2 & 0 & 0 \\ 1.2 & -1.6 & 1 & 0 \end{pmatrix};$$

and the sequence (32) converges to

$$AA^+ = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Example 2. The matrix is

$$A = (a_{ij}) = \tfrac{1}{10}, \qquad i, j = 1, \cdots, 10.$$

For $\alpha_2 = 0.666667$ the sequence of traces is:

k	trace $(I - Z_k)$
0	9.333333
1	9.111111
2	9.012345
3	9.000152
4	9.000000

And the sequence (4) converges to $A^+ = A = AA^+$.

Example 3. The matrix is the 10×10 Hilbert matrix

$$A = (a_{ij}) = \left(\frac{1}{i + j - 1}\right), \qquad i, j = 1, \cdots, 10.$$

As expected, the convergence is very slow. About 40 iterations are needed for (4) to converge to the inverse of A. For $\alpha_3 = 0.178152$ the sequence of traces $\{\text{trace } (I - Z_k)\}$, $k = 0, 1, \cdots$, converges to the nullity of A which is 0.

k	trace $(I - Z_k)$
0	9.432031463
1	9.163480102
10	7.790923364
20	6.298591575
30	4.615991308
35	0.358689858
36	0.036953381
37	0.000712251
38	0.000000448
39	$0.200492748 \times 10^{-12}$

The elements of $(AY_{39} - I)$ are all smaller, in absolute value, than 10^{-12}.

Acknowledgment. The authors gratefully acknowledge the computing facilities made available to them by the SWOPE Foundation.

REFERENCES

[1] M. ALTMAN, *An optimum cubically convergent iterative method of inverting a linear bounded operator in Hilbert space*, Pacific J. Math., 10 (1960), pp. 1107–1113.

[2] A. BEN-ISRAEL, *An iterative method for computing the generalized inverse of an arbitrary matrix*, Math. Comp., 19 (1965), p. 452.

[3] ———, *A note on an iterative method for generalized inversion of matrices*, Ibid., 20 (1966), pp. 439–440.

[4] A. BEN-ISRAEL AND A. CHARNES, *Contributions to the theory of generalized inverses*, J. Soc. Indust. Appl. Math., 11 (1963), pp. 667–699.

[5] A. BEN-ISRAEL AND Y. IJIRI, *A report on the machine computation of the generalized inverse of an arbitrary matrix*, O.N.R. Research Memo. No. 110, Carnegie Institute of Technology, Pittsburgh, 1963.

[6] G. GOLUB AND W. KAHAN, *Calculating the singular values and pseudo-inverse of a matrix*, this Journal, 2 (1965), pp. 205–224.

[7] T. N. E. GREVILLE, *The pseudo-inverse of a rectangular or singular matrix and its applications to the solution of linear equations*, SIAM Rev., 1 (1959), pp. 38–43.

[8] A. S. HOUSEHOLDER, *The Theory of Matrices in Numerical Analysis*, Blaisdell, New York, 1964.

[9] R. PENROSE, *A generalized inverse for matrices*, Proc. Cambridge Philos. Soc., 51 (1955), pp. 406–413.

[10] L. D. PYLE, *Generalized inverse computations using the gradient projection method*, J. Assoc. Comput. Mach., 11 (1964), pp. 422–428.

[11] J. B. ROSEN, *The gradient projection method for nonlinear programming, Part I: Linear constraints*, J. Soc. Indust. Appl. Math., 8 (1960), pp. 181–217.