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]:

$$Y_0 = \alpha A^*,$$

$$Y_{k+1} = Y_k(2I - AY_k), \quad k = 0, 1, \ldots ,$$

which yields $A^+$ as the limit of the sequence $\{Y_k\}$, $k = 0, 1, \ldots$, when $\alpha$ 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],

$$X_0 = \alpha A^*,$$

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

is shown, in Theorem 3 below, to be:

$$X_k = X_{k-1} \quad k = 0, 1, \ldots .$$

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

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

$$Z_0 = \alpha AA^*,$$

$$Z_{k+1} = 2Z_k - Z_k^2, \quad k = 0, 1, \ldots .$$

The traces of $Z_k$, $k = 1, 2, \ldots$, 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 Non-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 $\lambda_1(A^*A) \geq \lambda_2(A^*A) \geq \cdots \geq \lambda_n(A^*A)$ be the eigenvalues of $A^*A$. From rank $A = r$ it follows that $\lambda_i(A^*A) > 0$ and $\lambda_i(A^*A) = 0$ for $i = r + 1, \cdots, n$. We will use the matrix norm $\|A\| = \lambda_1(A^*A)$, which is subordinate to the Euclidean vector norm (e.g., [8, p. 44] where this matrix norm is called lub($A$)).

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 $\alpha$ satisfy

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

Then the sequence

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

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

**Theorem 0.2.** Let $\alpha$ satisfy (1). Then the sequence

$$Y_0 = \alpha A^*,$$

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

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

**Proof.**

(a) The process (2) is rewritten as

$$X_{k+1} = X_k (I - \alpha AA^*) + \alpha A^*$$

$$= X_k + \alpha (I - X_k A) A^*, \quad k = 0, 1, \cdots,$$
with

\[ X_0 = \alpha A^*. \]

From (5) it follows that

\[ AA^+ - AX_{k+1} = AA^+ - AX_k - \alpha(I - AX_k)AA^*; \]

and since \( A = AA^+A, \)

\[ AA^+ - AX_{k+1} = (AA^+ - AX_k) (I - \alpha AA^*). \]

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

\[
\| P_{R(A)} - AX_{k+1} \| \leq \| I - \alpha AA^* \| \| P_{R(A)} - AX_k \| < \| P_{R(A)} - AX_k \|.
\]

(b) Similarly we verify that

\[
AA^+ - AY_{k+1} = AA^+ - AY_k - AY_k (I - AY_k)
\]

\[ = AA^+ - AY_k - AY_k (AA^+ - AY_k), \]

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

\[
AA^+ - AY_{k+1} = (AA^+ - AY_k)^2, \quad k = 0, 1, \ldots,
\]

and finally

\[
\| P_{R(A)} - AY_{k+1} \| \leq \| P_{R(A)} - AY_k \|^2, \quad k = 0, 1, \ldots.
\]

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

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

\[
\| A^+ - X_{k+1} \| \leq \| A^+ - X_k \|, \quad k = 0, 1, \ldots.
\]

(b) The process (4) satisfies:

\[
\| A^+ - Y_{k+1} \| \leq \| A \| \| A^+ - Y_k \|^2, \quad k = 0, 1, \ldots.
\]

**Proof.**

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

\[
A^+ - X_{k+1} = (A^+ - X_k) (I - \alpha AA^*), \quad k = 0, 1, \ldots,
\]

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

(b) Similarly, (14) follows from

\[
A^+ - Y_{k+1} = (A^+ - Y_k)A(A^+ - Y_k), \quad k = 0, 1, 2, \ldots,
\]
which is obtained by using the easily verified relations
\[ Y_k = A^+ A Y_k = Y_k A A^+, \quad k = 0, 1, \ldots. \]

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

**Lemma.** Let \( S \) be any square complex matrix and \( k \geq 0 \) an integer. Then
\[
\sum_{j=0}^{k+1} S(I - S)^j = SS^+[I - (I - S)^{k+1}].
\]

**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.**
\[
Y_k = X_{2k-1}, \quad k = 0, 1, \ldots.
\]

**Proof.** Using (4) and (3), and the remark following (10), it follows that
\[
Y_k = A^+[I - (I - AY_{k-1})^2] = A^+[I - (I - AY_{k-1})^{2\pi}]
\]
\[
= A^+[I - (I - \alpha AA^*)^{2\pi}].
\]

From (2) it follows that
\[
X_{2k-1} = \alpha \sum_{j=0}^{2k-1} A^*(I - \alpha AA^*)^j = A^+ \sum_{j=0}^{2k-1} (\alpha AA^*)^j (I - \alpha AA^*)^j.
\]

Using the lemma with \( S = \alpha AA^* \) and the easily verifiable fact that \( \alpha AA^*(\alpha AA^*)^+ = AA^+ \), we conclude that
\[
X_{2k-1} = A^+[I - (I - \alpha AA^*)^{2\pi}],
\]

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

**Remark.** Using Euler's identity [4],
\[
(1 + x) \prod_{p=1}^{k-1} (1 + x^{2\pi}) = \sum_{p=0}^{2k-1} x^p, \quad |x| < 1,
\]

and Theorem 3, we obtain:
\[
Y_k = \alpha AA^*[I + (I - \alpha AA^*)] \prod_{p=0}^{k-1} [I + (I - \alpha AA^*)^{2\pi}]
\]
which corresponds to \( A_k^+ \) in [4, (54)].
THEOREM 4.\footnote{1 Recall that $\lambda_1(A^*A)$ is the smallest nonzero (positive) eigenvalue of $A^*A$, and note that $|1 - \alpha \lambda_r(A^*A)| < 1$ since $\lambda_r(A^*A) \leq \lambda_1(A^*A)$ and (1).}

\begin{equation}
\| A^+ - Y_k \| \leq \frac{\lambda_1(A^*A)}{\lambda_r(A^*A)} (1 - \alpha \lambda_r(A^*A))^k, \quad k = 0, 1, \ldots.
\end{equation}

Proof. Using Theorems 0.1 and 3 it follows that

\begin{equation}
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, \quad k = 0, 1, \ldots.
\end{equation}

As in [3] we verify that

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

and therefore

\begin{equation}
\| A^+ - Y_k \| \leq \alpha \| A^* \| \sum_{p=2^k}^{\infty} \| AA^+ - \alpha AA^* \|^p
\leq \frac{\alpha \| A^* \| \| AA^+ - \alpha AA^* \|^k}{1 - \| AA^+ - \alpha AA^* \|}
\leq \frac{\lambda_1(A^*A)(1 - \alpha \lambda_r(A^*A))^k}{\lambda_r(A^*A)}, \quad k = 0, 1, \ldots.
\end{equation}

Remarks. (a) This theorem corrects an error in [4, Theorem 17]. (b) As in [5] we call $\alpha_0$ 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 $\alpha$ is

\begin{equation}
\alpha_0 = \frac{2}{\lambda_1(A^*A) + \lambda_r(A^*A)},
\end{equation}

for which

\begin{equation}
\| A^+ - Y_k \| \leq \frac{\lambda_1(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)^k, \quad k = 0, 1, \ldots.
\end{equation}

Proof. As in [1] the minimizing $\alpha_0$ must satisfy
\[ 1 - a \lambda_i(A^*A) = 1 - a \lambda_i(A^*A), \]
i.e., the interval \([\lambda_c(A^*A), \lambda_1(A^*A)]\) 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 \(\lambda_i(A^*A) = \lambda_i(AA^*)\), 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

\[ \lambda_i(A^*A) \leq \max_{i=1, \cdots, m} \sum_{j=1}^m |b_{ij}|. \]

Therefore (1) can be replaced by:

\[ 0 < a < \frac{2}{\max_{i=1, \cdots, m} \sum_{j=1}^m |b_{ij}|}. \]

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 \(a\) satisfy (1). Then the sequence of matrices

\[ Z_0 = aAA^*, \]
\[ Z_{k+1} = 2Z_k - Z_k^2, \quad k = 0, 1, \cdots, \]
converges to \(AA^+\) as \(k \to \infty\), and

\[ \| P_{R(A)} - Z_{k+1} \| \leq \| P_{R(A)} - Z_k \|^2 \quad k = 0, 1, \cdots. \]

**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

\[ Z_k = I - (I - aAA^*)^k, \quad k = 0, 1, \cdots, \]
it follows that:

\[ \text{trace } Z_k = m - \text{trace } [(I - aAA^*)^k] \]

\[ = m - \sum_{i=1}^m (1 - a \lambda_i(AA^*))^{2k} = m - \sum_{i=1}^m (1 - a \lambda_i(AA^*))^{2k} \]
\[-(m - r) = r - \sum_{k=1}^{r} (1 - \alpha \lambda_i(AA^*)^{*k})^k, \quad k = 0, 1, \ldots,\]

where the third equality in (35) follows from
\[
\lambda_i(AA^*) = 0, \quad i = r + 1, \ldots, m.
\]

From (1) it follows that:
\[
|1 - \alpha \lambda_i(AA^*)| < 1, \quad i = 1, \ldots, r;
\]

and from (35):
\[
\text{trace } Z_{k+1} \geq \text{trace } Z_k, \quad k = 1, 2, \ldots,
\]

and
\[
\lim_{k \to \infty} \text{trace } Z_k = r = \text{rank } A.
\]

Remark. For \(\alpha\) 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 \((x) = -[-x]; \text{ e.g., } (3.5) = 4.2\).

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

**Corollary 2.** For every integer \(k \geq 1\) and real \(\alpha\) satisfying (1),
\[
(38) \quad \text{rank } A \geq \text{trace } Z_k,
\]

\[
(39) \quad \dim N(A^*) \leq \lfloor \text{trace } ((I - \alpha AA^*)^{*k}) \rfloor.
\]

**Proof.** Equation (38) follows from (35). Equation (39) follows from the facts that the sequence
\[
(40) \quad (I - \alpha AA^*)^{*k} = I - Z_k, \quad k = 0, 1, \ldots,
\]

converges to \(P_{N(A^*)}\) by Corollary 1, and the sequence of traces,
\[
\lfloor \text{trace } (I - Z_k) \rfloor, \quad k = 1, 2, \ldots,
\]

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 \(k \geq 1\) and for some real \(\beta > 0\)
\[
(41) \quad \text{trace } ((I - \beta AA^*)^{*k}) < 1.
\]

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

\[^{\text{Thus } (x) = [x] + 1 \text{ unless } x \text{ is an integer, in which case } x = [x] = (x).}\]
ITERATIVE COMPUTATION

\[ |1 - \beta \lambda_i(AA^*)| < 1, \quad i = 1, \ldots, r, \]

if and only if \( \beta \) 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 \( \alpha \) satisfying (30) were used:

\[
\alpha_p = \frac{p/3}{\max_{i=1,\ldots,m} \sum_{j=1}^m |b_{ij}|}, \quad p = 1, \ldots, 5.
\]

The sequence of traces

\[
\{\text{trace} (I - Z_k)\} = \{\text{trace} (I - \alpha AA^*)^k\}, \quad k = 0, 1, \ldots,
\]

which is monotone decreasing for \( k = 1, 2, \ldots, \) 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 \( \text{trace} (I - Z_k) \) for \( \alpha_p, \ p = 1, \ldots, 5 \), converges to the nullity of \( A^* \) which is 1.

<table>
<thead>
<tr>
<th>( p )</th>
<th>1 ( \alpha_p )</th>
<th>2 ( \alpha_p )</th>
<th>3 ( \alpha_p )</th>
<th>4 ( \alpha_p )</th>
<th>5 ( \alpha_p )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( k )</td>
<td>trace ( (I - Z_k) )</td>
<td>trace ( (I - Z_k) )</td>
<td>trace ( (I - Z_k) )</td>
<td>trace ( (I - Z_k) )</td>
<td>trace ( (I - Z_k) )</td>
</tr>
<tr>
<td>0</td>
<td>3.646464</td>
<td>3.293293</td>
<td>2.939393</td>
<td>2.585858</td>
<td>2.323232</td>
</tr>
<tr>
<td>1</td>
<td>3.385287</td>
<td>2.959287</td>
<td>2.719008</td>
<td>2.605442</td>
<td>2.798592</td>
</tr>
<tr>
<td>2</td>
<td>3.044291</td>
<td>2.604007</td>
<td>2.498218</td>
<td>2.390443</td>
<td>2.344645</td>
</tr>
<tr>
<td>3</td>
<td>2.703913</td>
<td>2.400470</td>
<td>2.228713</td>
<td>2.115086</td>
<td>2.036046</td>
</tr>
<tr>
<td>4</td>
<td>2.412875</td>
<td>2.129152</td>
<td>1.993923</td>
<td>1.924015</td>
<td>1.882346</td>
</tr>
<tr>
<td>5</td>
<td>2.137675</td>
<td>1.930274</td>
<td>1.854851</td>
<td>1.805310</td>
<td>1.761924</td>
</tr>
<tr>
<td>6</td>
<td>1.933500</td>
<td>1.805974</td>
<td>1.721921</td>
<td>1.647348</td>
<td>1.580391</td>
</tr>
<tr>
<td>7</td>
<td>1.806340</td>
<td>1.647827</td>
<td>1.521131</td>
<td>1.419059</td>
<td>1.336854</td>
</tr>
<tr>
<td>8</td>
<td>1.648066</td>
<td>1.419678</td>
<td>1.271578</td>
<td>1.175010</td>
<td>1.113470</td>
</tr>
<tr>
<td>9</td>
<td>1.419988</td>
<td>1.176130</td>
<td>1.073754</td>
<td>1.030383</td>
<td>1.012875</td>
</tr>
<tr>
<td>10</td>
<td>1.176389</td>
<td>1.030222</td>
<td>1.005440</td>
<td>1.000000</td>
<td>1.000000</td>
</tr>
<tr>
<td>11</td>
<td>1.031113</td>
<td>1.000562</td>
<td>1.000000</td>
<td>1.000000</td>
<td>1.000000</td>
</tr>
<tr>
<td>12</td>
<td>1.000668</td>
<td>1.000001</td>
<td>1.000000</td>
<td>1.000000</td>
<td>1.000000</td>
</tr>
<tr>
<td>13</td>
<td>1.000000</td>
<td>1.000000</td>
<td>1.000000</td>
<td>1.000000</td>
<td>1.000000</td>
</tr>
</tbody>
</table>
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

$$A = (a_{ij}) = \frac{1}{i}, \quad i, j = 1, \ldots, 10.$$ 

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

<table>
<thead>
<tr>
<th>$k$</th>
<th>$\text{trace}(I - Z_k)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>9.333333</td>
</tr>
<tr>
<td>1</td>
<td>9.111111</td>
</tr>
<tr>
<td>2</td>
<td>9.012345</td>
</tr>
<tr>
<td>3</td>
<td>9.00182</td>
</tr>
<tr>
<td>4</td>
<td>9.00000</td>
</tr>
</tbody>
</table>

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

**Example 3.** The matrix is the $10 \times 10$ Hilbert matrix

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

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 $\{\text{trace}(I - Z_k), k = 0, 1, \ldots\}$ converges to the nullity of $A$ which is 0.

<table>
<thead>
<tr>
<th>$k$</th>
<th>$\text{trace}(I - Z_k)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>9.432031463</td>
</tr>
<tr>
<td>1</td>
<td>9.163480102</td>
</tr>
<tr>
<td>10</td>
<td>7.790923364</td>
</tr>
<tr>
<td>20</td>
<td>6.298501575</td>
</tr>
<tr>
<td>30</td>
<td>4.615961308</td>
</tr>
<tr>
<td>35</td>
<td>0.358589858</td>
</tr>
<tr>
<td>36</td>
<td>0.036953381</td>
</tr>
<tr>
<td>37</td>
<td>0.000712251</td>
</tr>
<tr>
<td>38</td>
<td>0.000000448</td>
</tr>
<tr>
<td>39</td>
<td>$0.20042748 \times 10^{-13}$</td>
</tr>
</tbody>
</table>
The elements of \((AY^n - I)\) 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


