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

$$X_{n+1} = X_n(2I - AX_n) \quad (n = 0, 1, \ldots)$$

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 $E^* \hspace{1mm} n$ on the range of $A^*$, and $A^+$ the generalized inverse of $A$.

Theorem. The sequence of matrices defined by

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

Received November 9, 1964.
where $X_0$ is an $n \times m$ complex matrix satisfying

\begin{equation}
X_0 = A^*B_0 \quad \text{for some nonsingular } m \times m \text{ matrix } B_0,
\end{equation}

\begin{equation}
X_0 = C_0A^* \quad \text{for some nonsingular } n \times n \text{ matrix } C_0,
\end{equation}

\begin{equation}
\| AX_0 - P_{R(A)} \| < 1,
\end{equation}

\begin{equation}
\| X_0A - P_{R(A^*)} \| < 1,
\end{equation}

converges to the generalized inverse $A^+$ of $A$.

**Proof.** As in [3], the generalized inverse $A^+$ is characterized as the unique solution of the matrix equations,

\begin{equation}
AX = P_{R(A)},
\end{equation}

\begin{equation}
XA = P_{R(A^*)}.
\end{equation}

Thus it suffices to prove that the sequence (2) satisfies:

\begin{equation}
\lim_{n \to \infty} \| AX_n - P_{R(A)} \| = 0,
\end{equation}

\begin{equation}
\lim_{n \to \infty} \| X_nA - P_{R(A^*)} \| = 0.
\end{equation}

First we verify from (2), (3), (4) that

\begin{equation}
X_n = A^*B_n \quad \text{for } n = 0, 1, \ldots
\end{equation}

\begin{equation}
X_n = C_nA^* \quad \text{for } n = 0, 1, \ldots
\end{equation}

(where $B_n, C_n$ are recursively computed as

\begin{equation}
B_{n+1} = B_n(2P_{R(A)} - AA^*B_n),
\end{equation}

\begin{equation}
C_{n+1} = C_n(2P_{R(A^*)} - A^*AC_n),
\end{equation}

but are not used in the sequel).

Now, from (2),

\begin{equation}
P_{R(A)} - AX_{n+1} = (P_{R(A)} - AX_n)P_{R(A)} - AX_n(P_{R(A)} - AX_n);
\end{equation}

using (12), it follows that $AX_nP_{R(A)} = P_{R(A)}AX_n$.

Therefore

\begin{equation}
P_{R(A)} - AX_{n+1} = (P_{R(A)} - AX_n)^2
\end{equation}

and

\begin{equation}
\| P_{R(A)} - AX_{n+1} \| \leq \| P_{R(A)} - AX_n \|^2 \quad (n = 0, 1, \ldots),
\end{equation}

which, by (5), proves (9).

To prove (10) we write

\begin{equation}
P_{R(A^*)} - X_{n+1}A = P_{R(A^*)} - X_n(2P_{R(A)} - AX_n)A,
\end{equation}

which is rewritten, by (11), as

\begin{equation}
P_{R(A^*)} - X_{n+1}A = P_{R(A^*)} - P_{R(A^*)}X_nA - X_nA + (X_nA)^2 = (P_{R(A^*)} - X_nA)^2.
\end{equation}

\footnote{$\| \|$ is a multiplicative matrix norm.}
Thus
\[ || P_{R(A^*)} - X_{n+1}A || \leq || P_{R(A^*)} - X_nA ||^2 \quad (n = 0, 1, \ldots) \]
which, by (6), proves (10).

Remarks. (i) Similarly, the sequence defined by
\[ X_{n+1} = (2P_{R(A^*)} - X_nA)X_n \quad (n = 0, 1, \ldots), \]
with \( X_0 \) satisfying (3), (4), (5), (6), converges to \( A^+ \).

(ii) When \( A \) is nonsingular, both (2) and (16) reduce to the well-known process (1) due to Schulz [1], further studied by Dück in [4].

(iii) Conditions (5), (6) cannot be weakened as shown by:
\[
A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad P_{R(A)} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}
\]
and, taking
\[
X_0 = \begin{pmatrix} 1 \\ 0 \end{pmatrix},
\]
which satisfies (3), (4) but \( || AX_0 - P_{R(A)} || = 1 \) under the sum-of-squares norm.

(iv) The practical significance of the process proposed here is impaired by the need for knowledge of \( P_{R(A)} \). In fact, the direct computation of \( A^+ \) requires little more than the computation of \( P_{R(A)} \) and of \( P_{R(A^*)} \), and not substantially more than the computation of one alone. For any matrix \( A \) can be expressed in the form \( A = FR^* \) where the columns of \( F \) are linearly independent as are those of \( R \). Then, as shown by Householder in [5],
\[
P_{R(A)} = F(F^*F)^{-1}F^*
\]
and
\[
P_{R(A^*)} = R(R^*R)^{-1}R^*,
\]
whereas
\[
A^+ = R(R^*R)^{-1}(F^*F)^{-1}F^*.
\]
While only one of the projections \( P_{R(A)} \), \( P_{R(A^*)} \) is needed for the computation by the method proposed here, both are needed for testing (5) and (6).

(v) In the case where \( A \) is of full rank, the method proposed here is applicable. For, if rank \( A = m \), \( P_{R(A)} = I_{m \times m} \) and (2) reads:
\[ X_{n+1} = X_n(2I - AX_n). \]
In this case, \( A^+ = (AA^*)^{-1} \) and, indeed, by (11), we verify that \( X_n = A^*B_n \), where \( B_n \) converges to \( (AA^*)^{-1} \).

Similarly, if rank \( A = n \), \( P_{R(A^*)} = I_{n \times n} \) and (16) becomes
\[ X_{n+1} = (2I - X_nA)X_n. \]

Example. Let
\[
A = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 1 \end{pmatrix}
\]
and take

\[ X_0 = \frac{1}{2} A^* = \frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1 & 1 \end{pmatrix}. \]

Here, formula (17) is used to obtain:

\[ X_1 = \frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1 & 1 \end{pmatrix} \left( \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} - \frac{1}{2} \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 1 \end{pmatrix} \right) \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} \]

\[ = \frac{1}{4} \begin{pmatrix} 2 & 1 \\ 1 & 2 \\ -1 & 1 \end{pmatrix}, \]

\[ X_2 = \frac{1}{16} \begin{pmatrix} 10 & 5 \\ 5 & 10 \\ -5 & 5 \end{pmatrix}, \]

\[ X_3 = \frac{1}{256} \begin{pmatrix} 170 & 85 \\ 85 & 170 \\ -85 & 85 \end{pmatrix}, \]

etc., converging to:

\[ A^+ = \frac{1}{3} \begin{pmatrix} 2 & 1 \\ 1 & 2 \\ -1 & 1 \end{pmatrix}. \]

Technion, Israel Institute of Technology
Haifa, Israel


A Note on the Maximum Value of Determinants over the Complex Field

By C. H. Yang

The purpose of this note is to extend a theorem on determinants over the real field to the corresponding theorem over the complex field.

**Theorem.** Let \( D(n) \) be an \( n \)th order determinant with complex numbers as its entries. Then

\[
\max_{|a_{jk}| \leq K} |D(n)| = \max_{|a_{jk}| = K} |D(n)|.
\]

Received June 5, 1964. Revised December 8, 1964.