ALEFELD, G.: Convergence of D. J. Evans' Implicit Matrix Inversion Process


ALEFELD, G.

On the Convergence of the Higher Order Versions of D. J. Evans' Implicit Matrix Inversion Process

Dedicated to Prof. Dr. Dr. h. c. HELMUT HEINRICH on the occasion of his 80th birthday.

Die Versionen des Evans-Verfahrens mit höherer Konvergenzordnung werden untersucht. Es werden hinreichende Bedingungen für die Konvergenz, Fehlerabschätzungen sowie Bedingungen für die Monotonie der Iterierten angegeben. Es wird außerdem gezeigt, daß diese Verfahren zur Verbesserung einer Näherung für die Pseudoinverse einer Matrix mit vollem Spaltenrang geeignet sind.

The higher order versions of the method introduced by D. J. Evans are investigated. We introduce sufficient conditions for convergence, prove some error estimations and show that under appropriate conditions the iterates are possessing monotonic behaviour. Furthermore we show that these methods can also be used to improve an approximation for the pseudoinverse of a matrix with full column rank.

In the present paper we show that Evans' method can also be used to improve an approximation for the pseudoinverse of an \((m, n)\)-matrix \(A\) which has full column rank. Furthermore we investigate the versions of Evans' method which have higher order of convergence. For these methods we present sufficient conditions for convergence and error estimations as well as statements about the monotonic behaviour of the iterates.

1. Introduction

In the recent paper [2] D. J. EVANS introduced an implicit matrix inversion process. It was demonstrated by a simple example that this method is asymptotically much faster convergent than the wellknown method of SCHULZ [3], which is usually denoted as the Hotelling-method in the English literature. In [1] the method proposed by Evans was investigated. Sufficient conditions for convergence, error estimations and statements about the monotonic behaviour of the iterates were proved.

In the present paper we show that Evans' method can also be used to improve an approximation of the pseudoinverse of an \((m, n)\)-matrix \(A\) which has full column rank. Furthermore we investigate the versions of Evans' method which have higher order of convergence. For these methods we present sufficient conditions for convergence and error estimations as well as statements about the monotonic behaviour of the iterates.

2. Preliminaries

Let there be given a real \((m, n)\)-matrix \(A\) with full column rank, that is rank \((A) = n\). It follows that \(m \geq n\). The pseudoinverse of \(A\) is defined to be the unique \((n, m)\)-matrix \(A^*\) for which the following equations (the so-called Moore-Penrose equations) hold:

\[
\begin{align*}
(a) & \quad AA^*A = A, \quad (b) \quad A^*AA^* = A^*, \\
(c) & \quad (AA^*)^T = AA^*, \quad (d) \quad (A^*A)^T = A^*A.
\end{align*}
\]

If rank\((A) = n\) and \(m = n\) hold then \(A^* = A^{-1}\). Under the above assumption rank\((A) = n\) the equation

\[
A^*A = I_n
\]

holds where \(I_n\) is the \((n, n)\)-identity matrix. This can be seen in the following manner: In general we have

\[\text{rank}(UV) \leq \min \{\text{rank}(U), \text{rank}(V)\}\]

for the rank of the product \(UV\). Hence it follows from (1), (a) that

\[\text{rank}(A) = \text{rank}(AA^*A) \leq \text{rank}(A^*A)\]

and therefore

\[\text{rank}(A^*A) \leq \text{rank}(A) \leq \text{rank}(A^*A)\]

from which rank\((A^*A) = n\) follows. Therefore the \((n, n)\)-matrix \(A^*A\) is nonsingular. Multiplying the equation

\[A^*A = A^**AA^* = A^\dagger\]

from the left by \((A^*A)^{-1}\) the assertion follows. If we set

\[X = V A^T\]

for an \((n, n)\)-matrix \(V\) and for the given matrix \(A\) then it holds that

\[X = X A A^\dagger.\]
This can be seen in the following manner: From (1), (a) and (c) it follows that
\[ A^T = A^T(A^T) = A^T(AA^T), \]
and therefore using (3),
\[ X = V A^T(AA^T) = XA^T, \]
which is (4).

The existence of a matrix \( V \) such that (3) holds is equivalent to the condition that the range of \( X^T \) is contained in the range of \( A: \)
\[ X = V A^T \text{ for some } (n, n) \text{-matrix } V \in R(X^T) \subset R(A) \text{ (where } R \text{ denotes the range).} \]

In order to prove this assume on the one hand that \( X = V A^T \). Then it follows for \( x \in R^n \) that \( X^T x = A V^T x \) and because of \( V^T x \in R^n \) it follows that \( R(X^T) \subset R(A) \). If on the other hand \( R(X^T) \subset R(A) \) then for every \( x \in R^n \) there exists an \( x \in R^n \) such that \( X^T x = A x \).

Choosing \( x = e_i, i = 1(1) n \), where \( e_i \) denotes the \( i \)-th unit-vector we get \( X^T = A V^T \) or \( X = V A^T \). The \( i \)-th column of \( V^T \) is a vector \( X_i \) which because of the equation \( X^T e_i = A x_i \) corresponds to \( X^T = A V^T \).

If (besides of \( \text{rank}(A) = n \)) \( m = n \) then because of \( R(A) = R^n \) it holds for all matrices \( X \) that \( R(X^T) \subset R(A) = R^n \), that is (3) is true for all matrices \( X \) with some \( V \).

We now mention some results which partly where already proved in [1].

**Theorem 1:** Assume that for the \((n, n)\)-matrix \( A = D - L - U \) (\( D \) diagonal part, \( L \) lower triangular part, \( U \) upper triangular part of \( A \)) the diagonal part \( D \) is nonsingular. Let \( s_i = 0, i = 1(1) n \), and \( S = \text{diag}(s_i) \).

a) Define the real numbers \( p_i, i = 1(1) n \), recursively by
\[ p_i = \frac{1}{|s_i|} \left( \sum_{j=1}^{i-1} \frac{|a_{ij}|}{|a_{ii}|} p_j |s_j| + \sum_{j=i+1}^{n} \frac{|a_{ij}|}{|a_{ii}|} |s_j| \right), \quad i = 1(1) n, \]
and assume that
\[ p = \max_{1 \leq i \leq n} p_i < 1. \]
Then using the norm \( ||.|| = ||S^{-1} \cdot .||_\infty \) it holds that
\[ ||(D - L)^{-1} U|| \leq p < 1. \]

b) Define the real numbers \( q_i, i = 1(1) n \), recursively by
\[ q_i = \frac{1}{|s_i|} \left( \sum_{j=1}^{i-1} \frac{|a_{ij}|}{|a_{ii}|} |s_j| + \sum_{j=i+1}^{n} \frac{|a_{ij}|}{|a_{ii}|} q_j |s_j| \right), \quad i = 1(1) n, \]
and assume that
\[ q = \max_{1 \leq i \leq n} q_i < 1. \]
Then using the norm \( ||.|| = ||S^{-1} \cdot .||_\infty \) it holds that
\[ ||(D - U)^{-1} L|| \leq q < 1. \]

**Theorem 2:** Assume that the \((n, n)\)-matrix \( A \) has nonvanishing diagonal elements and that for the numbers \( s_i \neq 0, i = 1(1) n \), and \( S = \text{diag}(s_i) \), it holds that
\[ \tau = \max_{1 \leq i \leq n} \left\{ \frac{1}{|s_i|} \sum_{j=i+1}^{n} \frac{|a_{ij}|}{|a_{ii}|} |s_j| \right\} < 1. \]
Then \( p \leq \tau < 1 \) and \( q \leq \tau < 1 \) for the numbers \( p \) and \( q \) defined in Theorem 1.

If not stated otherwise we always use the norm
\[ ||.|| = ||S^{-1} \cdot .||_\infty \]
where \( S \) is some fixed nonsingular diagonal-matrix.

We are using the following lemmata:

**Lemma 1:** Assume that \( A \) is a real \((m, n)\)-matrix and \( X \) is a real \((n, m)\)-matrix. If \( ||I_n - XA|| < 1 \) for the norm (5) then \( XA \) has nonvanishing diagonal elements.

For the case \( m = n \) the proof can be found in [1] as a part of the proof of Satz 1 of this paper. The analogous proof holds for \( m \neq n \).

**Lemma 2:** Assume that \( A \) is a real \((m, n)\)-matrix and \( X \) is a real \((n, m)\)-matrix. If \( ||I_n - XA|| < 1 \) for the norm (5) then
\[ ||I_n - D^{-1}XA|| = ||D^{-1}(L + U)|| < 1 \]
where \( XA = D - L - U. \)

For the case \( m = n \) the proof can be found in [1] as a part of the proof of Satz 1 of this paper. The analogous proof holds for \( m \neq n \).
Lemma 3. Assume that \( A \) is a real \((m,n)\)-matrix and \( X \) is a real \((n,m)\)-matrix. If \( AX = D - L - U \) with nonsingular diagonal part \( D \) then if
\[
||I_n - D^{-1}XA|| = ||D^{-1}(L + U)|| < 1
\]
for the norm (5) it always holds that
\[
||I_n - D^{-1}XA|| \leq ||I_n - XA||.
\]
Furthermore if all diagonal elements of \( XA \) are different from one then (6) holds with the strict \(<\>-sign.

Proof: By assumption
\[
\frac{1}{|s_i|} \left( \sum_{j=1}^{n} \left| \frac{[s_j]}{d_{ij}} + \frac{u_{ij}}{d_{ij}} \right| \right) < 1,
\]
where we have set \( D = \text{diag}(d_{ii}), L = (l_{ij}), U = (u_{ij}). \)

If \( d_{ii} \neq 1 \) then multiplying this inequality by \( 1 - d_{ii} \neq 0 \) and using \( 1 - |d_{ii}| \leq 1 - d_{ii} \) it follows that
\[
\frac{1}{|s_i|} \left( \sum_{j=1}^{n} \left| \frac{[s_j]}{d_{ij}} + \frac{u_{ij}}{d_{ij}} \right| \right) < 1 - d_{ii} + \frac{1}{|s_i|} \left( \sum_{j=1}^{n} \left| [s_j] + |u_{ij}| \right| s_i \right).
\]
If \( d_{ii} \neq 1 \), \( i = 1(1)n \), then it follows that
\[
||I_n - D^{-1}XA|| < ||I_n - XA||. If at least one \( d_{ii} = 1 \) then we only can conclude that ||\( I_n - D^{-1}XA|| \leq ||I_n - XA||.
\]

3. Methods

Let there be given an \((m,n)\)-matrix \( A \) with full column rank, that is rank \((A) = n \leq m). The \((n,n)\)-matrix \( X_1 \) is assumed to be an approximation for the pseudoinverse \( A^\dagger \) of \( A \). We are forming the product \( X_1A \) which is an \((n,n)\)-matrix and consider the splitting
\[
X_1A = D_k - L_k - U_k
\]
where \( D_k \) is the diagonal part and \( L_k \) and \( U_k \) denote the strictly lower and upper triangular parts of \( X_1A \), respectively. If \( D_k \) is nonsingular then we define
\[
\tilde{L}_k = D_k^{-1}L_k, \quad \tilde{U}_k = D_k^{-1}U_k
\]
and
\[
F_k = \tilde{L}_k \tilde{U}_k (I_n - \tilde{L}_k)^{-1} (I_n - \tilde{U}_k), \quad G_k = (I_n - \tilde{L}_k)(I_n - \tilde{U}_k).
\]
It then holds that
\[
D_k^{-1}X_1A = (I_n - F_k) G_k.
\]
In order to improve \( X_k \) we consider for some nonnegative integer \( r \) the iteration method
\[
X_{k+1} = G_k^{-1}(I_n + F_k + \ldots + F_k^r)D_k^{-1}X_k, \quad k = 1, 2, \ldots
\]
For the special case \( m = n \) and for \( r = 0 \) this method was proposed by EVANS in [2].

Theorem 3 (Convergence statements): Let \( A \) be an \((m,n)\)-matrix with full column rank. Assume that for some \((n,n)\)-matrix \( V_1 \) it holds that
\[
X_1 = V_1A^\dagger \quad \text{and} \quad ||I_n - X_1A|| < 1
\]
where the norm is used. Then method (11) is well-defined and it holds that \( \lim_{k \to \infty} X_k = A^\dagger. \)

For the sequence of the residuals \( I_n - X_kA \) it holds, using the norm (5), that
\[
||I_n - X_{k+1}|| \leq ||I_n - D_k^{-1}X_kA||^{[r_k+1]a} \leq ||I_n - X_kA||^{[r_k+1]a}. \quad (12)
\]

Proof: From the assumption \( ||I_n - X_1A|| < 1 \) it follows, using Lemma 1, that \( X_1A \) has nonvanishing diagonal elements. Hence \( X_1A \) can be computed using (11). Multiplying (11) from the right by \( A \) it follows that
\[
X_2A = G_1^{-1}(I_n + F_1 + \ldots + F_1^r)D_1^{-1}X_1A = G_1^{-1}(I_n + F_1 + \ldots + F_1^r)(I_n - \tilde{L}_1)(I_n - \tilde{U}_1)^{-1} = G_1^{-1}(I_n + F_1 + \ldots + F_1^r)(I_n - \tilde{L}_1)(I_n - \tilde{U}_1)^{-1} G_1 = G_1^{-1}F_1^{-1}G_1,
\]
If \( X_k - X_{k+1}A = G_k^{-1}(I_n + F_k + \ldots + F_k^r)D_k^{-1}X_k \), \( k = 1, 2, \ldots \)

Since the matrices \( \tilde{L}_k \) and \( (I - \tilde{L}_k)^{-1} \) commute we obtain using the definition of \( G_k \) and \( F_k \)
\[
I_n - X_kA = (G_k^{-1}F_kG_k)^{-1} = (I_n - \tilde{L}_k)^{-1}(I_n - \tilde{L}_k)^{-1} \tilde{L}_k \tilde{U}_k)^{-1} = [(I_n - \tilde{U}_k)^{-1} (I_n - \tilde{L}_k)^{-1} \tilde{L}_k (I_n - \tilde{U}_k)^{-1} \tilde{L}_k)^{-1} \tilde{U}_k]^{-1} \tilde{U}_k \tilde{U}_k^{-1}.
\]
Using Lemma 2 it follows from the assumption \( ||I_n - X_1A|| < 1 \) that \( ||I_n - D_k^{-1}X_kA|| < 1 \). From Theorem 2 and Lemma 3 it therefore follows that
\[
||I_n - \tilde{U}_k^{-1} \tilde{L}_k|| \leq ||I_n - D_k^{-1}X_kA|| \leq ||I_n - X_kA||
\]
and
\[ \| (I_n - \tilde{L}_n)^{-1} \tilde{U}_n \| \leq \| I_n - D_1^{-1} X_1 A \| \leq \| I_n - X_1 A \| \]
where in both lines the second \( \leq \)-sign can be replaced by the \( < \)-sign if all diagonal elements of \( X_1 A \) are not equal to one.

Therefore we get
\[ \| I_n - X_2 A \| \leq \| I_n - D_1^{-1} X_2 A \|^{2(r+1)} \leq \| I_n - X_2 A \|^{2(r+1)} < 1. \]

Therefore for \( X_2 \) the same conditions hold as we have assumed for \( X_1 \) and \( X_3 \) can be computed.

It follows that
\[ \| I_n - X_3 A \| < I_n - X_2 A \|^{2(r+1)} < 1. \]

Therefore for \( X_2 \) the same conditions hold as we have assumed for \( X_1 \) and \( X_3 \) can be computed.

It follows that
\[ \| I_n - X_2 A \| < I_n - X_2 A \|^{2(r+1)} < 1. \]

Therefore for \( X_2 \) the same conditions hold as we have assumed for \( X_1 \) and \( X_3 \) can be computed.

From the assumption \( X_2 = V_2 A^T \) and the iteration method (11) it follows by mathematical induction that
\[ X_{k+1} = V_{k+1} A^T \]
where \( V_{k+1} \) is the \( (n, n) \)-matrix
\[ V_{k+1} = G_k^{-1} (I_n + F_k + \ldots + F_k^r) D_k^{-1} V_k. \]

Because of (3) and (4) it follows that \( X_{k+1} = X_{k+1} A^T \).

Therefore, multiplying the equation
\[ I_n X_{k+1} A = (G_k^{-1} F_k G_k)^{r+1} \]
from the right by \( A \) it follows that
\[ A^T - X_{k+1} = (G_k^{-1} F_k G_k)^{r+1} A^T \]
from which the assertion \( \lim_{k \to \infty} X_k = A^T \) follows.

We add some remarks:

Assume that \( m = n \). Then the second order Schulz-method
\[ Y_{k+1} = Y_k + (I_n - Y_k A) Y_k, \quad k = 1, 2, \ldots, \]
is convergent to \( A^{-1} \) if \( \varphi(I_n - Y_1 A) < 1 \) where \( \varphi \) denotes the spectral radius. \( \varphi(I_n - Y_1 A) < 1 \) holds, for example, for \( Y_1 = X_1 \) if \( \| I_n - X_1 A \| < 1 \). A simple computation shows that
\[ \| I_n - Y_{k+1} A \| \leq \| I_n - X_1 A \|^{2^k}. \]

On the other hand it follows from (12) for \( r = 0 \) that
\[ \| I_n - X_{k+1} A \| \leq \| I_n - D_1^{-1} X_1 A \|^{2^k} \leq \| I_n - X_1 A \|^{2^k}. \]

As mentioned in the proof of Theorem 3 the second \( \leq \)-sign can be replaced by the strict \( < \)-sign if all diagonal elements of \( X_1 A \) are different from one. This means that we have a better estimation for the residual \( \| I_n - X_{k+1} A \| \) compared with the estimation of the residual \( \| I_n - X_{k+1} A \| \).

In this sense the method of Evans is faster than the method of Schulz.

Theorem 4 (Monotone Convergence): Assume that the \( (m, n) \)-matrix \( A \) has full column rank and that for some \( (n, m) \)-matrix \( X_k \geq 0 \) it holds that \( I_n - X_k A \geq 0 \) where \( X_k = \bar{V}_k A^T \) for some \( (n, n) \)-matrix \( V_k \). If \( A^T \geq 0 \) and if each row of \( X_k \) contains at least one nonvanishing entry then (11) is well-defined and it holds that
\[ 0 \leq X_k \leq X_k^* \leq \ldots \leq X_k \leq X_{k+1} \leq \ldots \leq A^T, \]
that is \( \lim_{k \to \infty} X_k = X^* \leq A^T \) if \( \psi(I_n - X_k A) < 1 \) where \( \psi \) denotes the spectral radius then \( \lim_{k \to \infty} X_k = A^T \).

Proof: We first show that under the assumptions \( I_n - X_k A \geq 0, X_k \geq 0, X_k = \bar{V}_k A^T \) for some \( k \geq 0 \) it holds that \( 0 < d_k \leq 1 \) for the diagonal elements of \( X_k A \). This can be seen as follows: From the assumption \( I_n - X_k A \)
\[ = I_n - D_k + L_k + U_k \geq 0 \]
and therefore that
\[ X_k A = D_k - L_k - U_k \leq D_k. \]

Since by assumption \( A^T \geq 0 \) it follows that
\[ X_k A A^T \leq D_k A^T. \]
By (3) and (4) it follows from the hypothesis $X_k = V_k A^\top$ that $X_k = X_k A A^\top$. Therefore the last inequality can be written as

$$X_k \leq D_k A A^\top.$$  

Set $A^\top = (b_{ij})$, $X_k = (x_{ij})$ and consider a fixed row $i$. Then $x_{ij} \leq d_i b_{ij}$, $1 \leq j \leq m$, and therefore $d_i \geq 0$. If $d_i = 0$ then $x_{ij} = 0$ for $1 \leq j \leq m$ which contradicts the assumption. Therefore $X_{k+1}$ can be computed using (11). We now show that the assumptions $X_k = V_k A^\top$ and $I_n - X_k A \geq 0$ imply that $X_{k+1} \geq X_k$.

Since $D_k \geq 0$, $L_k \geq 0$, $U_k \geq 0$, $X_k \geq 0$ it follows that $F_k \leq 0$ and therefore using (11) that

$$X_{k+1} = G_k^{-1}(I_n + F_k + \ldots + F_{k-1}) D_k^{-1}X_k = (I_n + \tilde{U}_k + \ldots + \tilde{U}_{k-1}) (I_n + \tilde{L}_k + \ldots + \tilde{L}_{k-1}^{-1}) (I_n + F_k + \ldots + F_{k-1}) D_k^{-1}X_k =$$

$$= D_k^{-1}X_k + \text{nonnegative terms} \geq D_k^{-1}X_k \geq X_k.$$  

As in the proof of Theorem 3 one shows that

$$I_n - X_{k+1}A^\top = (G_k^{-1} F_k G_k)^{r+1} = [(I_n - \tilde{U}_k)^{-1} (I_n - \tilde{L}_k)^{-1} \tilde{L}_k \tilde{U}_k]^{r+1}$$

which implies that $I_n - X_{k+1}A^\top \geq 0$ and therefore since $A^\top \geq 0$

$$A^\top - X_{k+1}A A^\top \geq 0.$$  

As in the proof of Theorem 3 it follows from $X_2 = V_2 A^\top$ that $X_{k+2} = V_{k+1} A^\top$ and therefore that $X_{k+2} A A^\top = X_{k+1}$. We therefore have $A^\top \geq X_{k+1}$. Since $0 \leq X_k \leq X_{k+1}$ it follows that $X_{k+1}$ has in each row at least one entry different from zero. Therefore (11) is well defined and it holds that

$$0 \leq X_1 \leq X_2 \leq \ldots \leq X_k \leq X_{k+1} \leq \ldots \leq A^\top$$

from which lim $X_k = X^*$ $\leq A^\top$ follows. If $g(I_n - X_k A) < 1$ then there exists a nonsingular diagonal matrix $S$ such that $||I_n - X_k A|| < 1$ for the norm (5). Hence lim $X_k = A^\top$ follows from Theorem 3.

4. Error estimations and computational amount

For simplicity we consider only the special case $m = n$ in this section. This implies that $A^\top = A^{-1}$. Using (10) and (11) we arrive at

$$A^{-1} - X_{k+1} = (I_n - D_k^{-1}X_k A) (A^{-1} - X_{k+1}) = D_k^{-1}X_k A (A^{-1} - X_{k+1}) =$$

$$= D_k^{-1}X_k - D_k^{-1}X_k A X_{k+1} = D_k^{-1}X_k - (I_n - F_k) G_k G_k^{-1} (I_n + F_k + \ldots + F_{k-1}) D_k^{-1}X_k =$$

$$= D_k^{-1}X_k - (I - F_k^{-1}-1) D_k^{-1}X_k = F_k^{-1} D_k^{-1}X_k.$$  

For the norm (5) the inequality $|A| \leq |B|$ implies that $||A|| \leq ||B||$. Assuming $||I_n - D_k^{-1}X_k A|| < 1$ and using besides of this

$$||D_k^{-1}I_n| | \leq ||D_k^{-1}I_n + D_k^{-1}U_k| | \leq ||I_n - D_k^{-1}X_k A||,$$

$$||D_k^{-1}U_k| | \leq ||D_k^{-1}I_n + D_k^{-1}U_k| | \leq ||I_n - D_k^{-1}X_k A||$$

we get

$$||A^{-1} - X_{k+1}|| \leq \frac{1}{1 - ||I_n - D_k^{-1}X_k A||} \frac{||D_k^{-1}I_n||^{r+1}||D_k^{-1}U_k||^{r+1}}{||I_n - D_k^{-1}X_k A||} \leq$$

$$\leq \frac{1}{1 - ||I_n - D_k^{-1}X_k A||} \frac{||I_n - D_k^{-1}X_k A||^{(2r+1)}}{||I_n - D_k^{-1}X_k A||^{(2r+1)}} \frac{||D_k^{-1}X_k||}{||I_n - D_k^{-1}X_k A||}.$$  

Since by Lemma 2 $||I_n - X_k A|| < 1$ implies that $||I_n - D_k^{-1}X_k A|| < 1$ we have the following result.

Theorem 6 (Error estimation): Assume that for some $(n, n)$-matrix $A$ there exists an $(n, n)$-matrix $X_k$ such that for the norm (5) $||I_n - X_k A|| < 1$. Then for the iterates computed by (11) it follows that

$$||A^{-1} - X_{k+1}|| \leq \frac{1}{1 - ||I_n - D_k^{-1}X_k A||} \frac{||D_k^{-1}I_n||^{r+1}||D_k^{-1}U_k||^{r+1}}{||I_n - D_k^{-1}X_k A||} \leq$$

$$\leq \frac{1}{1 - ||I_n - D_k^{-1}X_k A||} \frac{||I_n - D_k^{-1}X_k A||^{(2r+1)}}{||I_n - D_k^{-1}X_k A||^{(2r+1)}} \frac{||D_k^{-1}X_k||}{||I_n - X_k A||^{(2r+1)}}.$$  

Proof: For the proof we only have to note that $||I_n - X_k A|| < 1$ implies $||I_n - X_k A|| < 1$ for all $k \geq 1$. This was already proved as a part of the proof of Theorem 3.

From Theorem 6 it follows that the order of convergence of (11) is (at least) $2(r + 1)$.

We are now discussing the computational amount necessary to perform one step of the method (11). For simplicity we again assume that $m = n$. As it is general use we only count the so-called point operations (multiplications and divisions) and neglect lower order powers of $n$.  

1. The computation of $X_kA$ needs $n^3$ operations.

2. Computation of $F_k$: Writing the equation (9) as

$$F_k(I_n - D_k^{-1}L_k)(I_n - D_k^{-1}U_k) = D_k^{-1}L_kD_k^{-1}U_k,$$

we get $F_k$ by first solving the equation

$$S_k(I_n - D_k^{-1}U_k) = D_k^{-1}L_k,$$

for $S_k$ and subsequently solving the equation

$$F_k(I_n - D_k^{-1}L_k) = S_k,$$

for $F_k$. Hence we have to solve two matrix equations with triangular matrices as coefficient matrices. This needs altogether $n^3$ operations if the right hand sides are known. For getting the right hand sides we have to compute the product $D_k^{-1}L_kD_k^{-1}U_k$. For this one has to perform $\frac{1}{2}n^3$ operations. Therefore the computation of $F_k$ needs $\frac{3}{2}n^3$ operations.

3. Computation of $I_n + F_k + \ldots + F_r$ using Horner's scheme needs $r-1$ matrix multiplications that is $n^3(r-1)$ operations.

4. Computation of $(I_n + F_k + \ldots + F_r)D_k^{-1}X_k$ needs one matrix multiplication that is $n^3$ operations.

5. Computation of $X_{k+1}$: Writing (11) as

$$G_kX_{k+1} = (I_n + F_k + \ldots + F_r)D_k^{-1}X_k =: R_k,$$

then $X_{k+1}$ can be computed by solving

$$(I_n - D_k^{-1}L_k)Z_k = R_k,$$

for $Z_k$ and subsequently solving

$$(I_n - D_k^{-1}U_k)X_{k+1} = Z_k,$$

This needs $n^3$ operations.

Adding the operations from steps 1. to 5. we have the result that for $r > 0$

$$n^3 + \frac{3}{2}n^3 + n^3(r-1) + n^3 + n^3 = n^3 \left( \frac{15}{2} + r \right)$$

operations are necessary to perform one step of (11).

Note that for $r = 0$ steps 2), 3) and 4) are not performed. Therefore one step of (11) needs $2n^3$ operations if $r = 0$.

We now compare the amount of work that necessary to perform one step of the higher order version of the Schulz method: If $r$ is a nonnegative integer then the method

$$Y_{k+1} = \left[ I_n + (I_n - Y_kA) + \ldots + (I_n - Y_kA)^{r+1} \right]Y_k,$$

is convergent if $\rho(I_n - Y_kA) < 1$. The order of convergence is $r+2$.

The computational amount of work necessary to perform one step of this method is as follows:

1. Computation of $Y_kA$ needs $n^3$ operations.

2. The term $[\ldots]$ can be computed by using Horner's scheme. This needs $r \cdot n^3$ operations.

3. Multiplying the term $[\ldots]$ by $Y_k$ needs $n^3$ operations.

Therefore altogether one step of (13) needs

$$n^3 + rn^3 + n^3 = n^3(r+2)$$

operations. This result also holds for $r = 0$.

In the following Table we list for some values of $r$ the amount of work necessary to perform one step of (11) and (13). Furthermore the Table contains the order of convergence for these values of $r$.

<table>
<thead>
<tr>
<th>$r$</th>
<th>(13) Order of Convergence</th>
<th>(11) Order of Convergence</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>$2n^3$</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>$3n^3$</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>$4n^3$</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>$5n^3$</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>$6n^3$</td>
<td>6</td>
</tr>
<tr>
<td>5</td>
<td>$7n^3$</td>
<td>7</td>
</tr>
<tr>
<td>6</td>
<td>$8n^3$</td>
<td>8</td>
</tr>
</tbody>
</table>

References


Received October 10, 1983

Address: Prof. Dr. G. ALEFELD, Institut für Angew. Mathematik, Universität Karlsruhe, Kaiserstr. 12, D-7500 Karlsruhe, F.R.G.