Iterative Solution of Linear Systems of Functional Equations*

A. M. Ostrowski

University of Basel, Switzerland

Submitted by Ky Fan

1. Consider a linear algebraic system,

\[ A\xi = \eta \]  \hspace{1cm} (1)

where \( A \) is a matrix \((a_{\mu\nu})\) and \( \xi \) and \( \eta \) are vectors in two \( N \)-dimensional vector spaces \( S_\xi \) and \( S_\eta \). In iterative methods of treatment of such systems often single equations and single groups of equations are used in a conveniently chosen order. We shall consider here a fixed grouping of the equations of the system (1) and of the corresponding components of the vectors \( \xi \) and \( \eta \). In order to do so systematically we shall consider the \( N \)-dimensional spaces \( S_\xi \) and \( S_\eta \) as Cartesian products of \( n \) linear vector spaces \( S_{\xi(1)}, \ldots, S_{\xi(n)} \) and \( S_{\eta(1)}, \ldots, S_{\eta(n)} \) of dimensions \( m_1, \ldots, m_n \) respectively, where of course

\[ m_1 + \ldots + m_n = N. \]  \hspace{1cm} (2)

Then the system (1) can be written as a system of \( n \) matricial equations

\[ \sum_{\nu=1}^{n} A_{\mu\nu} x_\nu = y_\mu \quad (\mu = 1, \ldots, n), \] \hspace{1cm} (3)

where generally \( A_{\mu\nu} \) is a \((m_\mu \times m_\nu)\)-matrix, \( x_\nu \) is a vector from \( S_{\xi(\nu)} \), and \( y_\mu \) a vector from \( S_{\eta(\mu)} \). The matrix \( A \) becomes then a partitioned matrix

\[ A = (A_{\mu\nu}) = \begin{bmatrix} A_{11} & \ldots & A_{1n} \\ \vdots & \ddots & \vdots \\ A_{n1} & \ldots & A_{nn} \end{bmatrix}. \] \hspace{1cm} (4)

In our treatment of the system (3) we shall consider single equations (3) as well as certain groups of these equations. So far it could

* The preparation of this paper was sponsored by the U. S. Army under Contract No. DA-11-022-ORD-2059. I am indebted for discussions to Mr. Howard Bell.
appear as if the consideration of a fixed grouping is not justified. However, the convergence conditions need not be satisfied in the general case, while they could be satisfied for a conveniently chosen grouping of variables, as we shall see in examples later. (See Secs. 9, 28.)

2. As a matter of fact, our theory does not depend at all on the fact that the $A_{\mu \nu}$ in (3) and (4) are matrices, and we shall therefore usually assume that these symbols denote general linear operators.

We shall therefore from now on assume that we have $n$ linear spaces $S_{\xi}^{(1)}, \ldots, S_{\xi}^{(n)}$, $n$ linear spaces $S_{\eta}^{(1)}, \ldots, S_{\eta}^{(n)}$ and $n^2$ linear operators $A_{\mu \nu}$ ($\mu, \nu = 1, \ldots, n$), such that the operator $A_{\mu \nu}$ is defined in $S_{\xi}^{(\nu)}$ and assumes the values from $S_{\eta}^{(\mu)}$. The elements of $S_{\xi}^{(\nu)}$ will be denoted by $x_\nu$, those of $S_{\eta}^{(\nu)}$ by $y_\nu$.

The Cartesian sums of $S_{\xi}^{(\nu)}$ resp. of $S_{\eta}^{(\nu)}$ are the spaces

$$S_{\xi} = \sum_{\nu=1}^{n} S_{\xi}^{(\nu)}, \quad S_{\eta} = \sum_{\nu=1}^{n} S_{\eta}^{(\nu)},$$

the elements of which are the "generalized vectors"

$$\xi = (x_1, \ldots, x_n), \quad \eta = (y_1, \ldots, y_n).$$

The matrix (4) is then an operator in the space $S_{\xi}$, which assumes values from the space $S_{\eta}$.

3. The spaces $S_{\xi}^{(\nu)}$ and $S_{\eta}^{(\nu)}$ are assumed to be normed and the norm of an element $x_\nu$ from $S_{\xi}^{(\nu)}$ and an element $y_\nu$ from $S_{\eta}^{(\nu)}$ will be denoted by $|x_\nu|_\phi$, $|y_\nu|_\phi$, independently of the value of $\nu$. In particular, the norms chosen for the $x_\nu$ can be quite different from those chosen for the $y_\nu$. These norms will be chosen once for all and the subscripts $\phi, \psi$ will be therefore usually dropped. $S_{\xi}$ is assumed to be complete for the norm $\phi$.

As to the operators, we will use for an operator $A$ defined in a space $S_1$ with the norm $|x|_\phi$ and assuming the values from a space $S_2$ with the norm $|x|_\phi$ the expressions

$$A_{\phi,\eta}(A) = \sup \frac{|Ax_\phi|_\phi}{|x_\phi|_\phi}, \quad \lambda_{\phi,\eta}(A) = \inf \frac{|Ax_\phi|_\phi^2}{|x_\phi|_\phi^2}$$

where again the subscripts $\psi, \phi$ will be dropped, as $\phi$ and $\psi$ are assumed fixed throughout the whole paper.

Then to the matrix (4) belongs its associated matrix

$$[A] = \begin{bmatrix}
\lambda(A_{11}) & -A(A_{12}) & \cdots & -A(A_{1n}) \\
-A(A_{21}) & \lambda(A_{22}) & \cdots & -A(A_{2n}) \\
\vdots & & \ddots & \vdots \\
-A(A_{n1}) & -A(A_{n2}) & \cdots & \lambda(A_{nn})
\end{bmatrix}. \quad (7)$$
4. The convergence conditions for the iterative methods considered in this paper, depend on the matrix \( (7) \) being a so-called \( M \)-matrix.

In Secs. 7—9 we will consider the \textit{simultaneous iteration method}; in Sec. 8 we will deduce the necessary and sufficient condition for its convergence, if the \( A_{\mu\nu} \) are matrices, and we will prove its convergence in the general case if \([A]\) is an \( M \)-matrix.

In the rest of the paper we shall be concerned with the single step and group iterations and prove, as the main result of the paper, generalizing the main result of Ostrowski [3], that such an iteration always converges, for the so-called \textit{free steering}, if \([A]\) is an \( M \)-matrix. Here the case of the so-called incomplete relaxation (overrelaxation or underrelaxation) will be taken into account and bounds for the relaxation factors will be obtained which depend, in a certain sense, on "the degree of \( M \)-ness of \([A]\)".

5. We call a matrix \( M(m_{\mu\nu}) \) of order \( n \) a \textit{Minkowski matrix} if we have

\[
m_{\mu\nu} > 0, \quad m_{\mu\nu} < 0 \quad (\mu \neq \nu; \ \mu, \nu = 1, \ldots, n)
\]  

(8)

and

\[
\sum_{\nu=1}^{n} m_{\mu\nu} > 0 \quad (\mu = 1, \ldots, n)
\]  

(9)

A matrix \( M(m_{\mu\nu}) \) is called an \textit{\( M \)-matrix} if there exist \( n \) positive numbers \( \gamma_1, \gamma_2, \ldots, \gamma_n \) such that if we multiply the columns of \( M \) resp. by the \( \gamma_i \), the obtained matrix \((\gamma_i m_{\mu\nu})\) becomes a Minkowski matrix.

We shall need in this paper several results about these classes of matrices, which shall be formulated as lemmas.

6. \textbf{Lemma 1.} A necessary and sufficient condition for the matrix \( M(m_{\mu\nu}) \) satisfying (8) to be an \( M \) matrix is that the determinant of \( M \) as well as all principal minors of \( M \) are positive. If \( M \) is an \( M \) matrix, so is \( M' \). (See Ostrowski [1], pp. 74, 75.)

\textbf{Lemma 2.} The inverse matrix of an \( M \) matrix is nonnegative (has non-negative elements). If we have for an \( M \) matrix \( M(m_{\mu\nu}) \) and \( n \) real numbers \( x_{\nu} \)

\[
\sum_{\nu=1}^{n} m_{\mu\nu} x_{\nu} \leq 0 \quad (\mu = 1, \ldots, n)
\]  

(10)

then all \( x_{\nu} \) are nonpositive. (See Ostrowski [1], p. 71; [3], p. 206.)

Before formulating the following lemmata, we remind the reader that by a theorem of Perron, to any nonnegative \((n \times n)\)-matrix \( R(h_{\mu\nu}) \) belongs the so-called \textit{maximal characteristic root} of \( B, \rho \), that is, a nonnegative char-
characteristic root of $B$, such that the modulus of every other characteristic root of $B$ does not exceed $\rho$. On the other hand, if the matrix $M(m_{\nu\nu})$ satisfies (8), the matrix

$$I - (m_{\nu\nu}/m_{\nu\mu}) = \begin{pmatrix}
0 & -m_{12}/m_{11} & \cdots & -m_{1n}/m_{11} \\
-m_{21}/m_{22} & 0 & \cdots & -m_{2n}/m_{22} \\
\vdots & \vdots & \ddots & \vdots \\
-m_{n1}/m_{nn} & -m_{n2}/m_{nn} & \cdots & 0
\end{pmatrix}$$

(11)

is nonnegative.

**Lemma 3.** A necessary and sufficient condition for a matrix $M(m_{\nu\nu})$ satisfying (8) to be an $M$ matrix, is that the maximal characteristic root of the corresponding matrix (11) is $< 1$. (See Ostrowski [3], p. 182.)

Into the same connection belongs the

**Lemma 4.** If $K(k_{\mu\nu})$ is an $(n \times n)$-matrix with nonnegative elements and the maximal characteristic root $\sigma < s$, there exists an $n$-tuple $(p_1, \ldots, p_n)$ of $n$ positive $p_\nu$ such that we have

$$\sum_{\nu=1}^{n} k_{\mu\nu} p_\mu / p_\nu < s \quad (v = 1, \ldots, n).$$

(12)

If $K$ is irreducible, the positive $p_\mu$ can be found such that we have even

$$\sum_{\nu=1}^{n} k_{\mu\nu} p_\mu / p_\nu = \sigma \quad (v = 1, \ldots, n).$$

(12')

(See Ostrowski [3], p. 191.)

7. We consider first the simultaneous iteration method ["Iteration in Gesamtschritten"] going back to Jacobi, and apply it to (3). As the system (3) can be written in the form

$$A_{\mu\nu} x_\mu = y_\mu - \sum_{\nu \neq \mu} A_{\nu\nu} x_\nu \quad (\mu = 1, \ldots, n),$$

it appears reasonable to form a sequence of approximating vectors to $\xi$: $\xi_0, \xi_1, \ldots, \xi_\infty$, \ldots starting with $\xi_0$, by the iteration procedure

$$A_{\mu\nu} x_\mu^{(e+1)} = y_\mu - \sum_{\nu \neq \mu} A_{\mu\nu} x_\nu^{(e)} \quad (\mu = 1, \ldots, n),$$

(13)

where we put generally

$$\xi_\infty = (x_1^{(e)}, \ldots, x_n^{(e)}).$$

(14)
If we introduce the matrix

$$C = \sum_{\mu=1}^{n} A_{\mu} = \begin{bmatrix} A_{11} & 0 & \ldots & 0 \\ 0 & A_{22} & \ldots & 0 \\ 0 & 0 & \ldots & A_{nn} \end{bmatrix},$$

then the iteration procedure (13) can be written in the form

$$C\xi_{k+1} = \eta - (A - C)\xi_k \quad (k = 0, 1, \ldots). \quad (16)$$

In order to discuss the convergence of (14) denote by $\xi$ the solution of (1). Then from (13) it follows that

$$C(\xi_{k+1} - \xi) = -(A - C)(\xi_k - \xi),$$

$$\xi_{k+1} - \xi = (I - C^{-1} A)(\xi_k - \xi),$$

$$\xi_{k} - \xi = (I - C^{-1} A)\kappa(\xi_0 - \xi) \quad (k = 1, 2, \ldots). \quad (17)$$

8. Assume first that the $A_{\mu}$ are matrices. Then it is well known, that the necessary and sufficient condition for the convergence of the right-side vector in (17) to 0 for any choice of $\xi_0$ is that the moduli of all eigenvalues of $I - C^{-1} A$ are $< 1$. But the characteristic equation of this matrix can be written as

$$|\lambda I - I + C^{-1} A| = 0,$$

and this becomes after multiplication by $C$

$$\begin{vmatrix} \lambda A_{11} & A_{12} & \ldots & A_{1n} \\ A_{21} & \lambda A_{22} & \ldots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n1} & A_{n2} & \ldots & \lambda A_{nn} \end{vmatrix} = 0. \quad (18)$$

We see that for the convergence of our procedure it is necessary and sufficient that the moduli of all roots of (18) are $< 1$.

9. We give now an example of a matrix of order 3 for which the simultaneous iteration method in the usual sense diverges, while after suitable partitioning we obtain a convergent $n$-step iteration. For the matrix

$$\begin{vmatrix} 1 & 2 & 0 \\ 2 & 1 & 2 \\ 0 & 1 & 1 \end{vmatrix}, \quad (19)$$
the convergence of the classical simultaneous iteration depends on the moduli of the roots of the equation

\[
\begin{vmatrix}
\lambda & 2 & 0 \\
2 & \lambda & 2 \\
0 & 1 & \lambda
\end{vmatrix} = 0, \quad \lambda^3 - 6\lambda = 0.
\]

Since here the maximum modulus of a root is \(\sqrt{6} > 1\), the classical simultaneous iteration is divergent. On the other hand, if (19) is partitioned taking

\[
A_{11} = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}, \quad A_{12} = \begin{pmatrix} 0 \\ 2 \end{pmatrix}, \quad A_{21} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad A_{22} = \begin{pmatrix} 1 \end{pmatrix},
\]

the corresponding equation (18) becomes

\[
\begin{vmatrix}
\lambda & 2\lambda & 0 \\
2\lambda & \lambda & 2 \\
1 & 1 & \lambda
\end{vmatrix} = 0, \quad 3\lambda(\lambda^2 - 2/3) = 0,
\]

and as here the maximal modulus of the roots of \(\sqrt{2/3} < 1\), we obtain a convergent simultaneous iteration.

10. We return now to the general case. The matrix \(I - C^{-1}A\) in (17) becomes

\[
\begin{pmatrix}
0 & A_{11}^{-1}A_{12} & \ldots & A_{11}^{-1}A_{1n} \\
\vdots & A_{n1}^{-1}A_{n1} & \ldots & 0 \\
A_{nn}^{-1}A_{nn} & A_{nn}^{-1}A_{n2} & \ldots & 0
\end{pmatrix} = (I - \delta_{\mu\nu}A_{\mu\nu}^{-1}A_{\mu\nu}). \quad (20)
\]

The upper bounds of the elements of (20) are the elements of the matrix

\[
\begin{pmatrix}
0 & A(A_{11}^{-1}A_{12}) & \ldots & A(A_{11}^{-1}A_{1n}) \\
\vdots & A(A_{n1}^{-1}A_{n1}) & \ldots & 0 \\
A(A_{nn}^{-1}A_{nn}) & A(A_{nn}^{-1}A_{n2}) & \ldots & 0
\end{pmatrix} \equiv T. \quad (21)
\]

Therefore (see Ostrowski, [5], Sec. 12) the upper bounds of the elements of \((I - C^{-1}A)^*\) are majorated by the corresponding elements of \(T^*\) and go to 0 if \(T^* \to 0\). On the other hand we have (see Ostrowski [5], Sec. 18),

\[
A(A_{\mu\nu}^{-1}A_{\mu\nu}) \leq A(A_{\mu\nu})/\lambda(A_{\mu\nu}). \quad (22)
\]

If we therefore put

\[
A(A_{\mu\nu})/\lambda(A_{\mu\nu}) = \alpha_{\mu\nu}(\mu \neq \nu), \quad \alpha_{\mu\mu} = 0 \quad (\mu = 1, \ldots, n), \quad (23)
\]
we see that our iteration procedure converges if the maximal root of the nonnegative matrix \((\alpha_{\mu})\) is < 1.

From the lemma 3 it follows now:

**Theorem 1.** The simultaneous iteration procedure defined by (13) is convergent if the matrix \([A]\) defined by (7) is an M-matrix.

11. We shall now generalize the so-called single-step and group iterations (see Ostrowski [3]) to the case of the system (3). We shall obtain in any case, starting with an initial generalized vector \(\xi_0\) arbitrarily chosen, the sequence of vectors \(\xi_\nu\) which, if the procedure is convergent, converge to a solution \(\xi\) of the equation (1). \(\xi, \eta\) and all vectors \(\xi_\nu\) are then decomposed as in (6) into their "Cartesian components," \(x_\nu, y_\nu, x_\nu^{(c)}\).

In generalizing the single-step iteration to the case considered here, we choose at the \(\kappa\)-th step, for the transition from \(\xi_\kappa\) to \(\xi_{\kappa+1}\), an index \(N_\kappa\) among the indices \(1, \ldots, n\) and change only the corresponding Cartesian component of \(\xi_\kappa\). We put therefore

\[
x_\mu^{(\kappa+1)} = x_\mu^{(\kappa)} \quad (\mu \neq N_\kappa),
\]

while in order to obtain \(x_{N_\kappa}^{(\kappa+1)}\), we use the corresponding equation (3) with \(\mu = N_\kappa\):

\[
A_{N_\kappa N_\kappa} x_{N_\kappa}^{(\kappa+1)} = y_{N_\kappa} - \sum_{\nu \neq N_\kappa} A_{N_\kappa \nu} x_\nu^{(\kappa)}.
\]

This can be written differently by introducing

\[
\delta^{(\kappa)} = x_{N_\kappa}^{(\kappa+1)} - x_{N_\kappa}^{(\kappa)}.
\]

Then we obtain

\[
A_{N_\kappa N_\kappa} \delta^{(\kappa)} = y_{N_\kappa} - \sum_{\nu = 1}^{n} A_{N_\kappa \nu} x_\nu^{(\kappa)}.
\]

12. However, in the numerical computation the vector \(\delta^{(\kappa)}\) obtained from (27) will not be used in (26) with its exact numerical value. We shall rather use instead of (26) the more general formula

\[
x_{N_\kappa}^{(\kappa+1)} = x_{N_\kappa}^{(\kappa)} + Q_\kappa \delta^{(\kappa)},
\]

where \(Q_\kappa\) is a suitable operator defined in \(S_\xi^{(N_\kappa)}\) and assuming values from the same space. Then, of course, (25) and (26) are not necessarily satisfied unless \(Q_\kappa\) is the unity operator. Equations (27) and (28) describe the generalization of the single-step iteration to the case of the system (3).
In numerical practice, the operator $Q_\kappa$ need not be taken in a very close neighborhood of the unity operator, since the numerical experience shows that a convenient choice of $Q_\kappa$ can improve the convergence very considerably.

13. The single-step iteration as described by the formulas (27) and (28) can again be generalized to the case where at the $\kappa$-th step not necessarily a single component but a whole subset of Cartesian components of $\xi_\kappa$ is changed by using the corresponding set of equations (3). We shall obtain thus the generalization of the group iteration procedures to (3).

For each index $\kappa$ ($\kappa = 0, 1, \ldots$) we choose a certain subset $g_\kappa$ of the indices $1, \ldots, n$. The complementary set to $g_\kappa$, consisting of all indices $1, \ldots, n$ not contained in $g_\kappa$, will be denoted by $\bar{g}_\kappa$. The indices of $g_\kappa$ will be called active indices at the $\kappa$-th step, and the letters $\alpha, \beta$ will run through these indices. The indices contained in $\bar{g}_\kappa$ are the passive indices at the $\kappa$-th step and will usually be denoted by $\pi$.

Then the transition from $\xi_\kappa$ to $\xi_{\kappa+1}$ will be described in the simplest case, corresponding to $Q_\kappa = I$ by the formulae

\[ x^{(\kappa+1)}_\pi = x^{(\kappa)}_\pi \quad (\pi \in \bar{g}_\kappa), \]  

\[ \sum_{\beta \in g_\kappa} A_{\alpha\beta} x^{(\kappa+1)}_\beta = y_\alpha - \sum_{\pi \in \bar{g}_\kappa} A_{\beta\pi} x^{(\kappa)}_\pi \quad (\alpha \in g_\kappa). \]  

If we introduce the notation

\[ x^{(\kappa+1)}_\alpha - x^{(\kappa)}_\alpha = \delta^{(\kappa)}_\alpha \quad (\alpha \in g_\kappa), \]

the equations (30) become

\[ \sum_{\beta \in g_\kappa} A_{\alpha\beta} \delta^{(\kappa)}_\beta = y_\alpha - \sum_{r=1}^{n} A_{\alpha r} x^{(\kappa)}_r \quad (\alpha \in g_\kappa). \]  

14. In order to account for rounding-off errors and to leave open the possibility of convergence-accelerating discussions, we shall now replace (31) by

\[ x^{(\kappa+1)}_\alpha = x^{(\kappa)}_\alpha + Q^{(\kappa)}_\alpha \delta^{(\kappa)}_\alpha \quad (\alpha \in g_\kappa), \]

where $Q^{(\kappa)}_\alpha$ are conveniently chosen operators. Then we drop (30) and (31), and our group iteration is described by (32) and (33).

The "steering" of our procedure is then done by the choice of the sets $g_\kappa$ and the operators $Q^{(\kappa)}_\alpha$.

If we introduce the elements of $S^{(\kappa)}_\xi$,

\[ r^{(\kappa)}_\mu = y_\mu - \sum_{r=1}^{n} A_{\mu r} x^{(\kappa)}_r, \]  

(34)
and the generalized vector
\[ \rho^{(\kappa)} = (r_1^{(\kappa)}, \ldots, r_n^{(\kappa)}) \]
the so-called \( \kappa \)-th residual vector, then (32) becomes
\[ \sum_{\beta \in \mathcal{E}_n} A_{\alpha \beta} \theta_\beta^{(\kappa)} = r_\alpha^{(\kappa)} \quad (\alpha \in \mathcal{E}_n) \]
and the \( \kappa \)-th step of our procedure is described by (29), (36), and (33).

15. In all following discussions we shall assume that the matrix \([A]\) is an \( M \) matrix. By the lemma 3 the matrix \((a_{\mu \nu})\), where the \( a_{\mu \nu} \) are given by (23), then has a maximal characteristic root \( \rho < 1 \). Assume a number \( s \), satisfying
\[ \rho < s < 1. \]
Then by lemma 4 there exists an \( n \)-tuple of positive numbers \( p_\nu \) \( (\nu = 1, \ldots, n) \), such that we have
\[ \sum_{\mu = 1}^{n} a_{\mu \nu} p_\mu / p_\nu < s \quad (\nu = 1, \ldots, n). \]  
(38)

We can even assume, if the matrix \((a_{\mu \nu})\) is irreducible, that all left-hand side expressions in (38) have the value \( \rho \).

From the theorem 2 of Ostrowski [5] it now follows that the system (1) is satisfied in \( S_\zeta \) by a generalized vector \( \zeta \). Introducing the differences \( \xi = \zeta \) as new variables, we reduce our problem to the case that \( \zeta = 0 \). We can therefore from now on without loss of generality assume that we have
\[ \eta = 0, \quad y_\nu = 0 \quad (\nu = 1, \ldots, n). \]  
(39)
Then the solution of (1) becomes the zero vector and we shall have to discuss whether \( \xi_\kappa \rightarrow 0 \) or not.

In the further discussions we can and shall also assume that we have
\[ A_{\mu \nu} = I \quad (\mu = 1, \ldots, n). \]  
(40)
Indeed, this amounts to multiplying the \( \mu \)-th equation (3) by \( A_{\mu \nu}^{-1} \). The \( \xi_\kappa \) are not changed by this multiplication. The \( a_{\mu \nu} \) are replaced by \( A(A_{\mu \nu}^{-1} A_{\mu \nu}) \). Since we have, however (see Ostrowski [5], Sec. 18)
\[ A(A_{\mu \nu}^{-1} A_{\mu \nu}) \leq A(A_{\mu \nu}) / \lambda(A_{\mu \nu}), \]
the maximal root \( \rho \) of \((a_{\mu \nu})\) is not increased and the relations (38) remain valid.
16. Consider now the linear system

$$A^{(k)}_\alpha - \sum_{\beta \in \mathcal{E}_\alpha} \frac{p_\alpha}{p_\beta} A^{(k)}_{\alpha\beta} = p_\alpha r^{(k)}_\alpha \quad (\alpha \in \mathcal{E}_\alpha). \quad (41)$$

By Hadamard's theorem and from (37) and (38) it follows that the determinant of (41) is \( \neq 0 \) and by lemma 2 all \( A^{(k)}_\alpha \) are nonnegative. On the other hand, by (40), (36) can be rewritten as

$$p_\alpha \delta^{(k)}_\alpha = \sum_{\beta \neq \alpha} A_{\alpha\beta} (p_\beta \delta^{(k)}_\beta) = p_\alpha r^{(k)}_\alpha$$

and if follows

$$p_\alpha |\delta^{(k)}_\alpha| - \sum_{\beta \neq \alpha} A_{\alpha\beta} |\delta^{(k)}_\beta| \leq p_\alpha |r^{(k)}_\alpha| \quad (\alpha \in \mathcal{E}_\alpha).$$

Subtracting from this term by term (41) we obtain

$$(p_\alpha |\delta^{(k)}_\alpha| - A^{(k)}_\alpha) - \sum_{\beta \neq \alpha} A_{\alpha\beta} |\delta^{(k)}_\beta| - A^{(k)}_\beta \leq 0 \quad (\alpha \in \mathcal{E}_\alpha),$$

and by lemma 2 we have

$$|p_\alpha \delta^{(k)}_\alpha| \leq A^{(k)}_\alpha \quad (\alpha \in \mathcal{E}_\alpha). \quad (42)$$

We now put for \( \kappa = 0, 1, \ldots \)

$$\tau_\kappa = \sum_{\alpha \in \mathcal{E}_\kappa} p_\alpha |r^{(k)}_\alpha|, \quad (43)$$

$$\omega_\kappa = \sum_{\nu = 1}^{n} \sum_{\gamma} p_\nu |r^{(k)}_\nu|, \quad (44)$$

$$\sigma_\kappa = \sum_{\alpha \in \mathcal{E}_\kappa} A^{(k)}_\alpha, \quad (45)$$

choose an \( \epsilon \) with

$$0 < \epsilon < \frac{1 - s}{1 + s}, \quad (46)$$
and assume that for all $\kappa$ we have
\[ A(Q_\kappa - 1) \leq \varepsilon \quad (\alpha \in g_\kappa; \quad \kappa = 0,1,\ldots). \tag{47} \]

Then we are going to prove that for each $\kappa$ we have
\[ \omega_{\kappa + 1} - \omega_{\kappa} \leq - (1 - s - \varepsilon(1 + s))\tau_{\kappa}, \tag{48} \]
\[ |\beta_\pi r^{(\kappa+1)}_\pi - \beta_\pi r^{(\kappa)}_\pi| \leq (1 + \varepsilon)\tau_{\kappa} \quad (\pi \in g_\kappa). \tag{49} \]

The proof of (48) and (49) will be given in the Sects. 17–19.

17. It follows from (41), (43), and (45), summing (41) on $\alpha$, that
\[ \sigma_{\kappa} - \tau_{\kappa} = \sum_{\pi,\beta} \frac{\beta_\pi}{\beta_\beta} \alpha_{\alpha\beta} \Delta^{(\kappa)}_\beta \tag{50} \]

and therefore
\[ \sigma_{\kappa} \geq \tau_{\kappa}. \tag{51} \]

Consider now the expression
\[ \sum_{\pi \in g_\kappa} \sum_{\alpha \in g_{\pi\kappa}} \frac{\beta_\pi}{\beta_\alpha} \alpha_{\alpha\beta} \Delta^{(\kappa)}_\beta = \sum_{\beta} \Delta^{(\kappa)}_\beta \sum_{r=1}^{n} \frac{\beta_\pi}{\beta_\beta} \alpha_{\alpha\beta} - \sum_{\beta} \Delta^{(\kappa)}_\beta \sum_{\alpha \in g_{\pi\kappa}} \frac{\beta_\pi}{\beta_\alpha} \alpha_{\alpha\beta}. \]

But here the subtrahend is, by (50), $= \sigma_{\kappa} - \tau_{\kappa}$, while the minuend, by (38) and (45), is $< s\sigma_{\kappa}$. We obtain therefore
\[ \sum_{\pi,\beta} \frac{\beta_\pi}{\beta_\alpha} \alpha_{\alpha\beta} \Delta^{(\kappa)}_\beta \leq s\sigma_{\kappa} - \sigma_{\kappa} + \tau_{\kappa}. \tag{52} \]

If, in particular, the matrix $(\alpha_{\mu\nu})$ is irreducible, the minuend has even the exact value $\rho\sigma_{\kappa}$ and the left-hand side of (52) is $= \rho\sigma_{\kappa} - \sigma_{\kappa} + \tau_{\kappa}$.

18. Writing (34) for $\kappa + 1$ and $\kappa$ and subtracting we have by (33) and (29)
\[ p_{\mu} r^{(\kappa+1)}_{\mu} - p_{\mu} r^{(\kappa)}_{\mu} = - \sum_{\alpha \in g_{\kappa}} \frac{\beta_{\mu}}{\beta_{\alpha}} A_{\alpha\mu} Q^{(\kappa)}_{\alpha} \delta^{(\kappa)}_{\alpha}, \tag{53} \]

if, in particular, the matrix $(\alpha_{\mu\nu})$ is irreducible, the minuend has even the exact value $\rho\sigma_{\kappa}$ and the left-hand side of (53) is $= \rho\sigma_{\kappa} - \sigma_{\kappa} + \tau_{\kappa}$. 
If, in particular, $\mu$ is an active index $\beta$, the first sum on the right is, by (36), $- \frac{B_\beta}{p_\beta} r^{(x)}_\beta$ and we obtain

$$p_\beta r^{(x - 1)}_\beta = - \sum_\alpha \frac{p_\alpha}{p_\beta} A_{\beta \alpha} (Q^{(x)}_\alpha - I) p_\alpha A^{(x)}_\alpha.$$  \hspace{1cm} (54)

If we now use (47) and (42), it follows from (53), replacing there $\mu$ by a passive index $\pi$,

$$\left| \frac{p_\pi}{p_\beta} r^{(x - 1)}_\pi \right| - \left| \frac{p_\pi}{p_\beta} r^{(x)}_\pi \right| \leq \left| \frac{p_\pi}{p_\beta} r^{(x - 1)}_\pi - p_\pi r^{(x)}_\pi \right| \leq (1 + \varepsilon) \sum_\alpha \frac{p_\alpha}{p_\pi} \alpha_{\pi \alpha} A^{(x)}_\alpha$$

and summing this over all passive indices $\pi$,

$$\sum_\pi \left| \frac{p_\pi}{p_\beta} r^{(x - 1)}_\pi \right| - \sum_\pi \left| \frac{p_\pi}{p_\beta} r^{(x)}_\pi \right| = \sum_\pi \left| \frac{p_\pi}{p_\beta} r^{(x - 1)}_\pi \right| - \omega_\pi \tau_\pi \leq$$

$$\sum_\pi \left| \frac{p_\pi}{p_\beta} r^{(x - 1)}_\pi \right| - \frac{p_\pi}{p_\beta} r^{(x)}_\pi \right| \leq (1 + \varepsilon) \sum_{\pi, \alpha} \frac{p_\alpha}{p_\pi} \alpha_{\pi \alpha} A^{(x)}_\alpha.$$  \hspace{1cm} (55)

On the other hand, from (54), (47), and (42) we have, as $\alpha_{\beta \mu} = 0$,

$$\left| \frac{p_\beta}{p_\beta} r^{(x + 1)}_\beta \right| \leq \varepsilon \sum_\alpha \frac{p_\alpha}{p_\beta} \alpha_{\beta \alpha} A^{(x)}_\alpha + \varepsilon A^{(x)}_\beta,$$

and summing this over all active indices $\beta$,

$$\sum_\beta \left| \frac{p_\beta}{p_\beta} r^{(x + 1)}_\beta \right| \leq \varepsilon \sum_{\beta, \alpha} \frac{p_\alpha}{p_\beta} \alpha_{\beta \alpha} A^{(x)}_\alpha + \varepsilon A^{(x)}_\beta.$$  \hspace{1cm} (56)

Adding (55) and (56), we have by (44)

$$\omega_{\pi + 1} - \omega_\pi + \tau_\pi \leq \sum_{\pi, \alpha} \frac{p_\alpha}{p_\pi} \alpha_{\pi \alpha} A^{(x)}_\alpha + \varepsilon \sum_{\mu = 1}^n \frac{p_\mu}{p_\pi} \alpha_{\mu \pi} A^{(x)}_\alpha + \varepsilon A^{(x)}_\pi.$$  \hspace{1cm} (57)

19. In (57) we use for the first term on the right the estimate (52). As for the second term on the right, we have from (38) and (45)

$$\varepsilon \sum_{\mu = 1}^n \sum_\pi \frac{p_\mu}{p_\pi} \alpha_{\mu \pi} A^{(x)}_\pi = \varepsilon \sum_\pi A^{(x)}_\pi \sum_{\mu = 1}^n \frac{p_\mu}{p_\pi} \alpha_{\mu \pi} \leq \varepsilon \sigma_\pi,$$
and we have from (57) using (51)
\[ \omega_{k+1} - \omega_k + \tau_k \leq (s\sigma_k - \sigma_k + \tau_k) + \varepsilon s\sigma_k + \varepsilon \sigma_k = \tau_k - (1 - \varepsilon - s(1 + \varepsilon))\sigma_k, \]
and therefore finally
\[ \omega_{k+1} - \omega_k \leq - (1 - s - \varepsilon(1 + s))\sigma_k. \quad (58) \]

By (51) and (46), (48) follows now immediately.

On the other hand, from the last inequality in (55) we have by (52)
\[ \left| p_{n+1}^{(k)} - p_n^{(k)} \right| \leq (1 + \varepsilon) \sum_{\pi,\alpha} p_{n+1}^{(k)} \alpha_{\pi,\alpha} A_{\alpha}^{(k)} \leq (1 + \varepsilon) (s\sigma_k - \sigma_k + \tau_k), \]
and (49) follows now from \( s < 1. \)

20. It is now easy to show that our iteration procedure converges if the \( Q_k^{(s)} \) satisfy (47), (46). Indeed, then the right-hand bound in (48) can be written, putting \( p = 1 - s - \varepsilon(1 + s) > 0, \) as \( -pt_k, \) and (48) becomes
\[ \omega_{k+1} - \omega_k \leq - pt_k. \quad (59) \]
We see that the sequence \( \omega_k \) is monotonically decreasing and tends therefore to a nonnegative limit \( \omega, \)
\[ \omega_k \downarrow \omega \geq 0. \quad (60) \]

The infinite series \( \sum_{\sigma=0}^{\infty} (\omega_\sigma - \omega_{\sigma+1}) \) is therefore convergent, and as we have from (59)
\[ \tau_{\kappa} \leq \frac{1}{p} (\omega_{\kappa} - \omega_{\kappa+1}), \quad (61) \]
the series
\[ \sum_{\sigma=0}^{\infty} \tau_\sigma \quad (62) \]
is convergent. We have therefore in the sequence
\[ \lambda_\kappa = \sum_{\sigma=\kappa}^{\infty} \tau_\sigma \quad (\kappa = 0,1,\ldots) \quad (63) \]
a sequence monotonically decreasing to zero.

21. In order to obtain further estimates for \( \omega_k \) consider the passive indices at the \( \kappa \)-th step, \( \tau_1,\ldots,\tau_\kappa. \)
We shall now assume that each of these indices becomes active at one of the following steps, and we denote for the general $\tau$ by $k$ the smallest index $> \kappa$ for which $\tau$ becomes active. We have then

$$|p_\tau r^{(\kappa)} - p_\tau r^{(k)}| \leq \sum_{\sigma = \kappa}^{k-1} |p_\tau r^{(\sigma)} - p_\tau r^{(\sigma+1)}|.$$  

Observe that here for each $\sigma$ in the right-hand sum $\tau$ is a passive index at the $\sigma$-th step. We have therefore by (49) and (63)

$$|p_\tau r^{(\kappa)} - p_\tau r^{(k)}| \leq (1 + \epsilon) \sum_{\sigma = \kappa}^{\infty} \tau_\sigma = (1 + \epsilon) \lambda_\kappa,$$

$$|p_\tau r^{(\kappa)}| \leq (1 + \epsilon) \lambda_\kappa + |p_\tau r^{(k)}| \leq (1 + \epsilon) \lambda_\kappa + \tau_k \leq (2 + \epsilon) \lambda_\kappa,$$

$$\sum_{\pi} |p_\pi r^{(\kappa)}| \leq (n - 1) (2 + \epsilon) \lambda_\kappa,$$

$$\omega_\kappa = \tau_\kappa + \sum_{\pi \in \mathcal{G}_\kappa} |p_\pi r^{(\kappa)}| \leq n(2 + \epsilon) \lambda_\kappa.$$

(64)

We see that $\omega_\kappa \to 0$ as $\kappa \to \infty$ and $\omega$ in (60) is 0.

22. If we now isolate in (34) for $y_\mu = 0$ the term $x_\mu^{(\kappa)}$ we have

$$p_\mu |x_\mu^{(\kappa)}| \leq p_\mu |r^{(\kappa)}| + \sum_{r=1}^{n} \frac{p_\mu}{p_r} \alpha_{\mu\tau} x_\tau^{(\kappa)} \quad (\mu = 1, \ldots, n),$$

and summing over $\mu$ by (38) and (64)

$$\sum_{\mu=1}^{n} p_\mu |x_\mu^{(\kappa)}| \leq \omega_\kappa + s \sum_{r=1}^{n} p_r |x_r^{(\kappa)}|,$$

$$(1 - s) \sum_{\mu=1}^{n} p_\mu |x_\mu^{(\kappa)}| \leq \omega_\kappa \leq n(2 + \epsilon) \lambda_\kappa,$$

$$\sum_{\mu=1}^{n} p_\mu |x_\mu^{(\kappa)}| \leq \frac{\omega_\kappa}{1 - s} \leq \frac{(2 + \epsilon)n}{1 - s} \lambda_\kappa.$$

(65)
We see that \( \xi_\kappa \to 0 \), and our procedure is convergent. The formula (65) gives at each step a measure for the error implied in \( \xi_\kappa \) in terms of \( \omega_\kappa \), formed for the corresponding residual vector \( p^{(\kappa)} \).

If we put \( \max_{\mu, \nu} p_\mu / p_\nu = \gamma \), we have

\[
\frac{\sum_{\mu} \xi_\mu^{(\kappa)}}{\sum_{\mu} r_\mu^{(\kappa)}} \leq \frac{1}{1 - s} \sum_{\mu = 1}^{n} \frac{r_\mu^{(\kappa)}}{r_\mu^{(\kappa)}}.
\]

and from (65)

\[
\sum_{\mu = 1}^{n} \xi_\mu^{(\kappa)} \leq \frac{\gamma}{1 - s} \sum_{\mu = 1}^{n} r_\mu^{(\kappa)}.
\]

The value of the sum \( \sum_{\mu} r_\mu^{(\kappa)} \) is easily estimated at every step of the computation. If we have a suitable estimate of \( \gamma \), we have at each step also a good estimate of error made by using \( \xi_\kappa \) instead of \( \xi \).

23. In practical computation, the steering will usually be chosen according to the relative values of \( \tau_\kappa \) and \( \omega_\kappa \). One possibility is of course to choose the active indices \( \alpha \) at the \( \kappa \)-th step in such a way that \( \tau_\kappa \) is as large as possible, as long as the solution of the corresponding system (30) does not become too complicated.

This approach appears to be technically the best, although, as a matter of fact, it may even slow down the convergence in the long run.

In any case, if for a positive constant \( \delta \), at each step \( \tau_\kappa \) is chosen in such a way that we have

\[
\tau_\kappa \geq \delta \omega_\kappa,
\]

this is certainly possible if \( \delta \leq 1/n \), then it follows from (59) that

\[
\omega_{\kappa + 1} \leq (1 - \delta \omega) \omega_\kappa.
\]

and we see that the \( \omega_\kappa \) and with them the \( \xi_\kappa \) converge at least as the terms of a suitable geometric series. The convergence is then linear.

24. We can now formulate our results in the following

**Theorem.** Consider a linear system (3) in which each \( A_{\mu \nu} \) is a linear operator defined in a normed complete linear space \( S_\xi^{(\nu)} \) and assuming values from a normed linear space \( S_\eta^{(\nu)} \). Assume that each \( A_{\mu \nu} \) establishes a one-to-one correspondence between \( S_\xi^{(\nu)} \) and \( S_\eta^{(\nu)} \) and that the matrix \( [A] \) is an \( M \)-matrix. Denote the maximal characteristic root of \( (\alpha_{\mu \nu}) \) by \( \sigma \), and by \( \varepsilon \) a positive number satisfying

\[
0 < \varepsilon < \frac{1 - \sigma}{1 + \sigma}.
\]

\[
\]
Consider then the iteration procedure defined by (29), (32), and (33), where the linear operators \( Q_n^{(s)} \) satisfy the condition (47), and assume that for the chosen steering of the iteration each index \( v \) among 1, \ldots, \( n \) is used infinitely often as an active index in the sense of Sec. 13.

Then the sequence of the approximating vectors \( \xi_\nu \) is convergent to the solution \( \xi_0 \) of (1).

If, further, \( \tau_\nu, \omega_\nu \) are defined by (43) and (44) and the steering is such that at every step we have (66) with a convenient positive \( \delta \), then the convergence of the \( \xi_\nu \) is at least linear in the sense that for a convenient \( \theta, 0 < \theta < 1, \) we have

\[
\xi_\nu - \xi = O(\theta^\nu), \quad 0 < \theta < 1 \ (\kappa \to \infty).
\] (69)

To prove this theorem, observe that if the condition (68) is satisfied, then for any \( s > \sigma \) and sufficiently close to \( \sigma \) we have also \( \varepsilon < (1 - s)/(1 + s) \). But then the condition (46) is also satisfied and the assertion of the theorem follows from the results of the Sections 22 and 23.

25. In the practical applications of our error estimates the difficulty arises that usually the value of \( \sigma \) is not known and an estimate of the quotients of the \( \rho_\mu \) satisfying (38) is missing.

As to the value of \( \sigma \), by a theorem of Frobenius, \( \sigma \) lies between the greatest and smallest of the sums \( \sum_{r=1}^n \alpha_{\mu r} \). Since \( \sigma \) does not change if the matrix \( (\alpha_{\mu r}) \) is transformed by a positive diagonal matrix, \( \sigma \) lies also between the greatest and the smallest of the sums

\[
\sum_{r=1}^n \alpha_{\mu r} \rho_\mu | \rho_r | \quad \text{for any set of positive } \rho_\mu. \tag{70}
\]

If the matrix \( (\alpha_{\mu r}) \) is irreducible, there exists a set of positive \( \rho_\mu \) for which all sums (70) have the same value \( \sigma \), namely the set of the eigenvector components corresponding to \( \sigma \). Otherwise, sets of positive \( \rho_\mu \) can be found, for which the greatest and the smallest of the sums (70) differ by an arbitrarily small quantity. Therefore \( \sigma \) can usually be enclosed between narrow limits in choosing conveniently the \( \rho_\mu \), unless an exact computation of the maximal root of \( (\alpha_{\mu r}) \) is justified by the nature of the problem.

26. As to the quotients of the \( \rho_\mu \), we can, if the matrix \( (\alpha_{\mu r}) \) is positive or at least irreducible, use some results about the quotients of the components of the (left sided) eigenvector of \( (\alpha_{\mu r}) \) correspondent to \( \sigma \) entering into (12').

Put

\[
M = \max_{\mu, \nu} \alpha_{\mu r}; \quad m = \min_{\mu, \nu} \alpha_{\mu r}. \tag{71}
\]
\[ \kappa = \min \alpha_{\mu \nu} \quad (\alpha_{\mu \nu} > 0, \mu \neq \nu); \quad (72) \]

\[ r = \min_{\mu} \sum_{\nu} \alpha_{\mu \nu}. \quad (73) \]

Then we have (Ostrowski [6], formulas (10), (14), and (33) replacing \( \kappa_1, M_1 \) by 0, \( \kappa_2 \) by \( m \) and \( M_2 \) by \( M \), if all \( \alpha_{\mu \nu} (\mu \neq \nu) \) are positive and all \( \alpha_{\mu \mu} \) vanish,

\[ \gamma \leq \frac{M}{m}; \quad \gamma \leq \frac{M}{m} \frac{\sigma + m}{\sigma + M}; \quad \gamma \leq \frac{\sigma - r + m}{m}. \quad (74) \]

If we have \( \alpha_{\mu \nu} \geq 0 \) and \( (\alpha_{\mu \nu}) \) is irreducible, we have (Schneider [1])

\[ \gamma \leq \left( \frac{\sigma}{\kappa} \right)^{n-1}. \quad (75) \]

An improvement of (75) can be obtained in the following way, applying the results of Ostrowski [6] to the matrix \( A' \). Put

\[ \kappa^{(\mu)} = \min_{\nu} \alpha_{\nu \mu} \quad (\alpha_{\nu \mu} > 0) \quad (\mu = 1, 2, \ldots, n), \quad (76) \]

then we have

\[ \gamma \leq \prod_{\kappa^{(\mu)} < \sigma} \frac{\sigma}{\kappa^{(\mu)}}. \quad (77) \]

27. If the matrix \( A \) is reducible the above results do not apply. However, as we are only interested in the inequalities (38), the following method can be still used. Fill up the zeros in the matrix \( A \) by a positive constant \( \varepsilon \). The matrix \( A_1 \) thus obtained has a maximal root \( \sigma' > \sigma \) which is still < \( s \) if \( \varepsilon \) is chosen sufficiently small. Then the corresponding component of the left sided eigenvector of \( A_1 \) can be then taken as the \( p_\mu \) in (38). As to the value of \( \varepsilon \), it can be obtained using the results of Ostrowski [4]. If we choose \( \varepsilon \) in any case \( \leq M = \max \alpha_{\mu \nu} \), the expression \( \delta \) defined by (4) l.c. is \( < (n^2 \varepsilon)/M \); then it follows easily from the estimate (5) of Ostrowski [4] that we have

\[ \sigma' - \sigma < (n + 2) M^{1/n} \leq (n + 2) \left( \frac{n^2 \varepsilon}{M} \right)^{1/n}. \]

Therefore the sufficient condition for \( \sigma' \) to be < \( s \) is

\[ \varepsilon < \frac{(s - \rho)^n}{(n + 2)^{n+2}} M^{-n+1}. \quad (78) \]
We obtain then an estimate for $\gamma$, for instance, from the first inequality (74) replacing there $m$ by

$$\min \left( m, \frac{(s - \rho)^n}{(n + \rho)^n + 2} M^{-n+1} \right).$$

28. Consider now again the case where $A_{\mu\nu}$ are matrices and the operator $A$ the partitioned matrix (4) of total order $N$. Then it could be asked whether, if (7) is an $M$ matrix, then already for the matrix $A$, as an $(N \times N)$-matrix, the corresponding condition is satisfied. This however is not generally the case. Consider indeed the matrix

\[
\begin{pmatrix}
1 & 3 & 1 \\
3 & 1 & 1 \\
\cdots & \cdots & \cdots \\
-1 & -1 & 3
\end{pmatrix}
\] (79)

For this matrix, considered as a nonpartitioned matrix of order 3, our condition is not valid, since the corresponding associated matrix

\[
\begin{pmatrix}
1 & -3 & -1 \\
-3 & 1 & -1 \\
-1 & -1 & 3
\end{pmatrix}
\]

is not an $M$-matrix. (We have $\frac{1}{-s} = -8 < 0$). On the other hand, if (79) is considered as a partitioned matrix with the partitioning as indicated and $n = 2$, we have

\[
A_{11} = \begin{pmatrix} 1 & 3 \\ 3 & 1 \end{pmatrix}, \quad A_{11}^{-1} = \frac{1}{8} \begin{pmatrix} -1 & 3 \\ 3 & -1 \end{pmatrix}, \quad A_{22} = 3, \quad A_{11}^{-1} A_{12} = \begin{pmatrix} \frac{1}{4} \\ \frac{1}{4} \end{pmatrix},
\]

$A_{22}^{-1} A_{21} = (-\frac{3}{3} - \frac{3}{4})$, and using Euclidean norms, obtain for the corresponding matrix $T$ in (21)

\[
\begin{pmatrix}
0 & \sqrt{2}/4 \\
\sqrt{2}/3 & 0
\end{pmatrix}
\]

with the maximal root $\sqrt{1/6} < 1$.

---

1 See the detailed treatment of such norms in Ostrowski [5], Chap. I.
2 These bounds are extensively discussed in Ostrowski [5], Chap. II.
References