JOURNAL OF MATHEMATICAL ANALYSIS AND APPLICATIONS 32, 274–307 (1970)

On M-Functions and Their Application to Nonlinear Gauss–Seidel Iterations and to Network Flows

WERNER C. RHEINBOLDT*

University of Maryland, College Park, Maryland 20742

Submitted by G. Birkhoff

1. INTRODUCTION

There are two important sufficient conditions for the matrix $A = (a_{ij})$ of a linear system of equations $Ax = z$ which ensure that for arbitrary z the well-known Gauss-Seidel and Jacobi iterations converge to the unique solutions of the system for any starting point. These are

$$A \text{ is symmetric and positive definite;} \tag{1.1}$$

$$\begin{aligned} &A \text{ is an } M\text{-matrix, that is } a_{ij} \leqslant 0,\ i \neq j,\ i, j = 1,..., n, \\ &\text{and } A^{-1} = (b_{ij}) \text{ exists and satisfies } b_{ij} \geqslant 0,\ i, j = 1,..., n; \end{aligned} \tag{1.2}$$

[see e.g., Varga (1962)].

For the solution of nonlinear systems of equations

$$Fx \equiv \begin{pmatrix} f_1(x_1 ,..., x_n) \\ \cdots \\ f_n(x_1 ,..., x_n) \end{pmatrix} = \begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix} \tag{1.3}$$

the Gauss–Seidel and Jacobi iterations were first generalized and analyzed by Bers (1953) in the special case of discrete analogues of mildly nonlinear elliptic boundary value problems of the form $\Delta u = \varphi(s, t, u, u_s , u_t)$. Since then, these processes were discussed by various authors, and no historical survey shall be attempted here. In particular, Schechter (1962) extended the convergence condition (1.1) to the nonlinear case by proving that when the mapping F of (1.3) has a continuous, symmetric, and uniformly positive definite (Frechet) derivative on all of R^n, then (1.3) has for any $z \in R^n$ a unique solution x^*, and for any starting point in R^n the nonlinear Gauss–Seidel iteration is well-defined and converges to x^*. On the other hand, the

* This research was completed during a stay at the Gesellschaft für Mathematik und Datenverarbeitung m.b.H., Schloß Birlinghoven/Germany. The work was in part also supported by the U.S. National Science Foundation under Grant GJ-231.

conditions, underlying convergence results given by Bers (1953), Birkhoff and Kellogg (1966), Ortega and Rheinboldt (1967), (1968), and Porsching (1968), represent, in effect, generalizations of the M-matrix condition (1.2) to particular nonlinear systems. But there appears to be as yet no general extension of the condition (1.2) to the nonlinear case as is available in Schechter's result for (1.1). Such a generalization of this linear result is presented in this article.

Following Ortega (1969), a class of so-called M-functions on R^n is considered which contains as special cases all linear mappings induced by M-matrices. In analogy to M-matrices, an M-function F is assumed to have certain monotonicity properties with respect to the variables x_i, and—in extension of the condition $A^{-1} \geqslant 0$—F is required to be inverse isotone in the sense of Collatz (1952). In Section 2, certain basic properties of M-functions are proved, and Section 3 concerns the convergence of the Gauss–Seidel and Jacobi processes for such mappings. The indicated convergence result, extending the condition (1.2), states that when F is a continuous M-function from R^n onto itself, both the Gauss–Seidel and the Jacobi process converge globally for any given $z \in R^n$.

There is a well-known close connection between M-matrices and certain types of linear network flows; in fact, related considerations probably go back to Stieltjes (1886). Similarly, there is a close connection between M-functions and nonlinear network flows. This connection is used in Sections 4 and 5 to obtain sufficient conditions for certain mappings on R^n to be M-functions. These results show in particular that the mappings, considered by Bers (1953) and Ortega and Rheinboldt (1967), (1968), are in fact M-functions, and hence that the convergence results of these authors are contained in the theorems of Section 3. Finally, Section 5 concludes with a discussion of certain boundary value problems for nonlinear network flows generalizing those considered by Birkhoff and Kellogg (1966) and Porsching (1968).

At this point, I would like to extend my special thanks to Professor J. Ortega for stimulating me to look deeper into this topic, and to Professor H. Unger and the Gesellschaft für Mathematik und Datenverarbeitung m.b.H. for providing me with the opportunity of pursuing this wotk in an atmosphere of uninterrupted quiet.

2. M-FUNCTIONS AND THEIR BASIC PROPERTIES

Throughout this paper, R^n is the n-dimensional real linear space of column vectors x with components $x_1, \ldots, x_n$, and $e^i \in R^n$, $i = 1, \ldots, n$, are the unit basis vectors with i-th component one and all others zero. On R^n the natural (or component-wise) partial ordering is written as $x \leqslant y$, and $x < y$ stands for $x_i < y_i$, $i = 1,..., n$. The space of all real $n \times n$ matrices

$$A = (a_{ij}; i, j = 1,..., n)$$

is denoted by $L(R^n)$, and the transpose of a vector $x \in R^n$ or a matrix $A \in L(R^n)$ is indicated by x^T or A^T, respectively. We use $F : D \subset R^n \to R^m$, $f_i(x)$, $i = 1,..., m$, as notation for a (nonlinear) mapping F with domain D in R^n and range $F(D)$ in R^m, and with components $f_1,..., f_m$. The following class of mappings will play an important role in the subsequent discussions:

DEFINITION 2.1. The mapping $F : D \subset R^n \to R^m$ is *isotone* (or *antitone*) (on D) if $x \leqslant y$, for any $x, y \in D$, implies that $Fx \leqslant Fy$ (or $Fx \geqslant Fy$). An isotone (or antitone) mapping F is *strictly isotone* (or *strictly antitone*) if it follows from $x < y$, for any $x, y \in D$, that $Fx < Fy$ (or $Fx > Fy$).

Collatz (1952) introduced (in a more general setting) the following converse concept:

DEFINITION 2.2. The mapping $F : D \subset R^n \to R^m$ is *inverse isotone* (on D) if

$$Fx \leqslant Fy, \qquad \text{for any} \qquad x, y \in D, \qquad \text{implies that} \qquad x \leqslant y. \tag{2.1}$$

An inverse isotone mapping F is *strictly inverse isotone* if it follows from $Fx < Fy$, for any $x, y \in D$, that $x < y$.

This terminology is justified by the following result:

LEMMA 2.3. *The mapping $F : D \subset R^n \to R^m$ is inverse isotone if and only if F is injective and $F^{-1} : F(D) \subset R^m \to R^n$ is isotone.*

Proof. If F is inverse isotone, then by (2.1), $Fx = Fy$ for any $x, y \in D$ implies that $x \leqslant y$ as well as $x \geqslant y$, and hence that $x = y$. Thus F is injective. If now $u, v \in F(D)$ and $x = F^{-1}u$, $y = F^{-1}v$, then it follows from $Fx = u \leqslant v = Fy$ that $F^{-1}u = x \leqslant y = F^{-1}v$. Conversely, if F is injective and F^{-1} isotone, then $u = Fx \leqslant Fy = v$ implies that $x = F^{-1}u \leqslant F^{-1}v = y$.

In the following three results we note some simple properties of inverse isotone mappings:

LEMMA 2.4 [Schröder (1962)]. *A continuous, inverse isotone mapping $F : D \subset R^n \to R^m$ on the open set D in R^n is strictly inverse isotone.*

Proof. Let $Fy > Fx$ for some $x, y \in D$, and set

$$\epsilon = \min_i [f_i(y) - f_i(x)] > 0.$$

Then there exists a $\delta > 0$ such that $y - te \in D$ and

$$-\epsilon e \leqslant Fy - F(y - te) \leqslant \epsilon e \qquad \text{for} \qquad |t| \leqslant \delta,$$

where $e = (1,..., 1)^T$. Hence $Fx \leqslant Fy - \epsilon e \leqslant F(y - \delta e)$ which implies that $x \leqslant y - \delta e < y$.

THEOREM 2.5. *A continuous, surjective, and inverse isotone mapping* $F : R^n \rightarrow R^n$ *is a homeomorphism from* R^n *onto itself.*

Proof. By Lemma 2.3, $F^{-1} : R^n \rightarrow R^n$ is well-defined. If $\lim_{k\to\infty} y^k = y$, then $\{y^k\}$ is bounded, and from $u \leqslant y^k \leqslant v$, $k = 0, 1,...,$ follows that $F^{-1}u \leqslant x^k = F^{-1}y^k \leqslant F^{-1}v$, $k = 0, 1,....$ Thus also $\{x^k\}$ is bounded and has at least one limit point. If $\lim_{i\to\infty} x^{k_i} = x$, then, by continuity,

$$y = \lim_{i\to\infty} y^{k_i} = \lim_{i\to\infty} Fx^{k_i} = Fx,$$

or $x = F^{-1}y$. This shows that $\{x^k\}$ has only the limit point $F^{-1}y$, and hence that F^{-1} is continuous.

LEMMA 2.6. *Let* $F : D_F \subset R^n \rightarrow R^p$ *and* $G : D_G \subset R^p \rightarrow R^m$ *be such that* $F(D_F) \subset D_G$ *and hence that* $G \circ F$ *is well-defined. If* (a) F *and* G *are inverse isotone, then also* $G \circ F$ *is inverse isotone; and if* (b) G *is isotone and* $G \circ F$ *inverse isotone, then* F *is inverse isotone.*

Proof. (a) From $G \circ Fy \geqslant G \circ Fx$, $x, y \in D_F$, follows that $Fy \geqslant Fx$ and hence that $y \geqslant x$. (b) If $Fy \geqslant Fx$, $x, y \in D_F$, then $G \circ Fy \geqslant G \circ Fx$ and therefore $y \geqslant x$.

The following two concepts were introduced by Ortega (1969); they will be central to all our considerations.

DEFINITION 2.7. The mapping $F : D \subset R^n \rightarrow R^n$ is *off-diagonally antitone* if for any $x \in R^n$ the functions

$$\varphi_{ij} : \{t \in R^1 \mid x + te^j \in D\} \rightarrow R^1, \quad \varphi_{ij}(t) = f_i(x + te^j), \quad i \neq j, \quad i, j = 1,..., n \tag{2.2}$$

are antitone. Analogously, F is *diagonally isotone* (or *strictly diagonally isotone*) if for any $x \in R^n$ the functions

$$\varphi_{ii} : \{t \in R^1 \mid x + te^i \in D\} \rightarrow R^1, \qquad \varphi_{ii}(t) = f_i(x + te^i), \qquad i = 1,..., n \tag{2.3}$$

are isotone (or strictly isotone). Finally, a strictly diagonally isotone mapping F is *surjectively diagonally isotone* if $D = R^n$ and if for any $x \in R^n$ each mapping φ_{ii} of (2.3) is surjective.

DEFINITION 2.8. The mapping $F : D \subset R^n \to R^n$ is an *M-function* if F is inverse isotone and off-diagonally antitone.

In Sections 4 and 5 we shall discuss two generically different examples of M-functions which show that such mappings occur naturally in the discretization of certain boundary value problems and also in the study of nonlinear network flows.

The definition of M-functions represents a generalization of the M-matrix concept, as the next lemma shows:

LEMMA 2.9. *A matrix $A \in L(R^n)$ is an M-matrix if and only if the induced linear mapping $A : R^n \to R^n$ is an M-function.*

The proof is a trivial consequence of Definitions 2.7 and 2.8 and of Lemma 2.3, and is omitted here.

It is well-known that all the diagonal elements of an M-matrix and of its inverse are strictly positive. The following result represents an extension of this fact to M-functions.

THEOREM 2.10. *Let $F : D \subset R^n \to R^n$ be an M-function (and hence an injection). Then F and $F^{-1} : F(D) \subset R^n \to R^n$ are strictly diagonally isotone. If $F : R^n \to R^n$ is continuous and surjective, then F and $F^{-1} : R^n \to R^n$ are surjectively diagonally isotone.*

Proof. (1) Suppose that for some $x \in R^n$; $s, t \in R^1$, $s > t$, and index i we have $x + se^i$, $x + te^i \in D$, and $f_i(x + se^i) \leqslant f_i(x + te^i)$. The off-diagonal antitonicity then implies that

$$f_j(x + se^i) \leqslant f_j(x + te^i), \qquad j \neq i, \quad j = 1,..., n,$$

or, altogether, that $F(x + se^i) \leqslant F(x + te^i)$. By the inverse isotonicity this leads to the contradiction $s \leqslant t$, which shows that F must be strictly diagonally isotone.

(2) Suppose, similarly, that for some $u \in R^n$, s, $t \in R^1$, $s > t$, and index i we have $u + se^i$, $u + te^i \in F(D)$ and $y_i \leqslant x_i$ where $y = F^{-1}(u + se^i)$, $x = F^{-1}(u + te^i)$. Lemma 2.3 implies that $y \geqslant x$, and hence that $y_i = x_i$. But then the off-diagonal antitonicity leads to the contradiction

$$f_i(y) = f_i(y_1 ,..., y_{i-1} , x_i , y_{i+1} ,..., y_n) \leqslant f_i(x) = f_i(y) - (s - t) < f_i(y)$$

which shows that F^{-1} must be strictly diagonally isotone.

(3) Let $F : R^n \to R^n$ be continuous and surjective. By part (1) of the proof, F is strictly diagonally isotone; hence it suffices to prove that for any

$x \in R^n$ the n functions φ_{ii} defined by (2.3) are surjective. For this in turn we need to show only that for any $x \in R^n$

$$\lim_{t\to\infty} f_i(x + te^i) = +\infty, \qquad i = 1,..., n \tag{2.4a}$$

and

$$\lim_{t\to-\infty} f_i(x + te^i) = -\infty, \qquad i = 1,..., n. \tag{2.4b}$$

Suppose that for some $x \in R^n$ and index i there exists a sequence $\{t_k\} \subset R^1$ for which $\lim_{k\to\infty} t_k = +\infty$ but $f_i(x + t_k e^i) \leqslant a_i < +\infty$, $k = 0, 1,...$. It is no restriction to assume that the t_k are nonnegative and monotonically increasing. Then the off-diagonal antitonicity of F implies that

$$f_j(x + t_k e^i) \leqslant f_j(x) = a_j < +\infty, \qquad j \neq i, \qquad j = 1,..., n, \qquad k = 0, 1,...,$$

and hence, altogether, that

$$F(x + t_k e^i) \leqslant a = (a_1, ..., a_n)^T, \qquad k = 0, 1,... . \tag{2.5}$$

By the surjectivity of F there is a $y \in R^n$ such that $Fy = a$, and now the inverse isotonicity leads to the contradiction $t_k \leqslant y_i - x_i < +\infty$, $k \geqslant 0$. Similarly it follows that (2.4b) holds, and therefore that F is surjectively diagonally isotone.

(4) Finally, for the proof of the surjective diagonal isotonicity of F^{-1}, suppose that there is a sequence $\{t_k\} \subset R^1$ with $\lim_{k\to\infty} t_k = +\infty$, a vector $u \in R^n$, and an index i, such that $x_i^k \leqslant a_i < +\infty$, $k = 0, 1,...$, where $x^k = F^{-1}(u + t_k e^i)$. Again, it is no restriction to assume that $\{t_k\}$ is monotonically increasing. Then by the isotonicity of F^{-1}, we have $x^k \leqslant x^{k+1}$, and thus it follows that

$$\begin{aligned} u_i + t_k = f_i(x^k) &\leqslant f_i(x_1^0, ..., x_{i-1}^0, x_i^k, x_{i+1}^0, ..., x_n^0) \\ &\leqslant f_i(x_1^0, ..., x_{i-1}^0, a_i, x_{i+1}^0, ..., x_n^0) < +\infty. \end{aligned}$$

This contradicts $\lim_{k\to\infty} t_k = +\infty$, and hence we see that for any $u \in R^n$

$$\lim_{t\to+\infty} f_i^{-1}(u + te^i) = +\infty, \qquad i = 1,..., n,$$

where the f_i^{-1} denote the components of F^{-1}. Similarly it follows that

$$\lim_{t\to-\infty} f_i^{-1}(u + te^i) = -\infty, \qquad i = 1,..., n,$$

and therefore that F^{-1} is surjectively diagonally isotone.

3. Convergence of Gauss–Seidel and Jacobi Processes

For the solution of n-dimensional equations of the form (1.3) we consider now the (underrelaxed) Gauss–Seidel iteration

$$\begin{array}{llll} \text{Solve} & f_i(x_1^{k+1},\ldots, x_{i-1}^{k+1}, x_i, x_{i+1}^k,\ldots, x_n^k) = z_i & \text{for} & x_i \\ \text{Set} & x^{k+1} = (1-\omega_k)\, x_i^k + \omega_k x_i, \quad i = 1,\ldots, n, & & k = 0, 1,\ldots, \end{array} \tag{3.1}$$

as well as the corresponding Jacobi process

$$\begin{array}{llll} \text{Solve} & f_i(x_1^k,\ldots, x_{i-1}^k, x_i, x_{i+1}^k,\ldots, x_n^k) = z_i & \text{for} & x_i \\ \text{Set} & x_i^{k+1} = (1-\omega_k)\, x_i^k + \omega_k x_i, \quad i = 1,\ldots, n, & & k = 0, 1,\ldots, \end{array} \tag{3.2}$$

where in both cases $\{\omega_k\} \subset [\epsilon, 1]$, $\epsilon > 0$, is a given sequence of relaxation factors.

In a slightly less general form the following theorem was proved by Ortega (1969); it extends results of Birkhoff and Kellogg (1966), Ortega and Rheinboldt (1968), and Porsching (1968) to a more general class of mappings.

THEOREM 3.1. *Let $F: D \subset R^n \to R^n$ be continuous, off-diagonally antitone and strictly diagonally isotone. Suppose that for some $z \in R^n$ there exist points $x^0, y^0 \in D$ such that*

$$x^0 \leqslant y^0; \qquad J = \{x \in R^n \mid x^0 \leqslant x \leqslant y^0\} \subset D; \qquad Fx^0 \leqslant z \leqslant Fy^0. \tag{3.3}$$

Then, for $\epsilon \in (0, 1]$ and any sequence $\{\omega_k\} \subset [\epsilon, 1]$, the Gauss–Seidel iterates $\{y^k\}$ and $\{x^k\}$ given by (3.1) *and starting from y^0 and x^0, respectively, are uniquely defined and satisfy*

$$x^0 \leqslant x^k \leqslant x^{k+1} \leqslant y^{k+1} \leqslant y^k \leqslant y^0, \qquad Fx^k \leqslant z \leqslant Fy^k, \qquad k = 0, 1,\ldots \tag{3.4}$$

as well as

$$\lim_{k\to\infty} x^k = x^* \leqslant y^* = \lim_{k\to\infty} y^k, \qquad Fx^* = Fy^* = z. \tag{3.5}$$

The corresponding result holds for the Jacobi iteration (3.2).

Proof. We give here only the proof for the Gauss–Seidel process, for the Jacobi iteration it proceeds analogously and is even slightly simpler.

We proceed by induction and suppose that for some $k \geqslant 0$ and $i \geqslant 1$

$$x^0 \leqslant x^k \leqslant y^k \leqslant y^0, \qquad Fx^k \leqslant z \leqslant Fy^k, \tag{3.6}$$

$$x_j^k \leqslant x_j^{k+1} \leqslant y_j^{k+1} \leqslant y_j^k, \qquad j = 1,\ldots, i-1, \tag{3.7}$$

where for $i = 1$ the relation (3.7) is vacuous. Clearly, (3.6/7) is valid for $k = 0$ and $i = 1$. From (3.3) it follows that the functions

$$\alpha(s) = f_i(x_1^{k+1},\ldots, x_{i-1}^{k+1}, s, x_{i+1}^{k},\ldots, x_n^{k})$$
$$\beta(s) = f_i(y_1^{k+1},\ldots, y_{i-1}^{k+1}, s, y_{i+1}^{k},\ldots, y_n^{k})$$

are defined for $s \in [x_i^0, y_i^0]$. From (3.6/7) and the off-diagonal antitonicity of F we then find that

$$\beta(s) \leqslant \alpha(s), \qquad s \in [x_i^0, y_i^0] \tag{3.8}$$

and

$$\beta(x_i^k) \leqslant \alpha(x_i^k) \leqslant f_i(x^k) \leqslant z_i \leqslant f_i(y^k) \leqslant \beta(y_i^k) \leqslant \alpha(y_i^k). \tag{3.9}$$

By the continuity and strict isotonicity of α and β, (3.9) implies the existence of unique $\hat{y}_i^k$ and $\hat{x}_i^k$ for which

$$\beta(\hat{y}_i^k) = z_i = \alpha(\hat{x}_i^k), \qquad x_i^k \leqslant \hat{x}_i^k \leqslant \hat{y}_i^k \leqslant y_i^k,$$

where the relation $\hat{x}_i^k \leqslant \hat{y}_i^k$ is a consequence of (3.8). Because of $\omega_k \in [\epsilon, 1]$ we therefore have

$$\begin{aligned} y_i^k \geqslant y_i^{k+1} &= (1-\omega_k)\, y_i^k + \omega_k \hat{y}_i^k \geqslant \hat{y}_i^k \geqslant \hat{x}_i^k \geqslant x_i^{k+1} \\ &= (1-\omega_k)\, x_i^k + \omega_k \hat{x}_i^k \geqslant x_i^k \end{aligned}$$

which shows that (3.7) holds for $i = 1,\ldots, n$, and hence that

$$x^k \leqslant x^{k+1} \leqslant y^{k+1} \leqslant y^k.$$

From this it follows that

$$\begin{aligned} f_i(y^{k+1}) &\geqslant f_i(y_1^{k+1},\ldots, y_i^{k+1}, y_{i+1}^{k},\ldots, y_n^{k}) \\ &\geqslant f_i(y_1^{k+1},\ldots, y_{i-1}^{k+1}, \hat{y}_i^k, y_{i+1}^{k},\ldots, y_n^{k}) = z_i \end{aligned}$$

and similarly that

$$f_i(x^{k+1}) \leqslant f_i(x_1^{k+1},\ldots, x_{i-1}^{k+1}, \hat{x}_i^k, x_{i+1}^{k},\ldots, x_n^{k}) = z_i\,.$$

This completes the induction and hence the proof of (3.4). Clearly now the limits

$$\lim_{k\to\infty} x^k = x^* \leqslant y^* = \lim_{k\to\infty} y^k$$

exist, and

$$y^{k+1} \geqslant \hat{y}^k = -\frac{1}{\omega_k}(y^k - y^{k+1}) + y^k \geqslant -\frac{1}{\epsilon}(y^k - y^{k+1}) + y^k,$$

as well as

$$x^{k+1} \leqslant \hat{x}^k = \frac{1}{\omega_k}(x^{k+1} - x^k) + x^k \leqslant \frac{1}{\epsilon}(x^{k+1} - x^k) + x^k,$$

imply that also $\lim_{k\to\infty} \hat{x}^k = x^*$ and $\lim_{k\to\infty} \hat{y}^k = y^*$. Therefore, it follows from the definition of the Gauss–Seidel process and the continuity of F that $Fx^* = Fy^* = z$.

The following theorem provides information about the relation between the Gauss–Seidel and the Jacobi process, and about the dependence of each iteration upon the choice of relaxation factors.

THEOREM 3.2. *Suppose that the conditions of Theorem 3.1 hold and let* $\{\omega_k\}, \{\bar{\omega}_k\} \subset [\epsilon, 1]$ *be given sequences for which*

$$0 < \epsilon \leqslant \omega_k \leqslant \bar{\omega}_k \leqslant 1, \qquad k = 0, 1,... . \tag{3.10}$$

In all cases with y^0 as starting point, let $\{y^k\}$ and $\{\bar{y}^k\}$ be the Gauss–Seidel sequences formed with $\{\omega_k\}$ and $\{\bar{\omega}_k\}$ respectively, and $\{v^k\}$ and $\{\bar{v}^k\}$ the corresponding Jacobi sequences. By Theorem 3.1 these sequences are well-defined and converge monotonically to a solution of $Fx = z$ in J. More specifically, all four sequences have the same limit point $y^* \in J$, and

$$y^k \geqslant \bar{y}^k \geqslant y^*; \qquad v^k \geqslant \bar{v}^k \geqslant y^*; \qquad v^k \geqslant y^k, \qquad k = 0, 1,... . \tag{3.11}$$

Moreover, if $x \in J$ is any solution of $Fx = z$, then $x \leqslant y^*$.

The corresponding result—with all inequalities reversed—holds for the four sequences starting at x^0.

Proof. Suppose that for some $k \geqslant 0$ and $i \geqslant 1$

$$y^0 \geqslant y^k \geqslant \bar{y}^k \geqslant x, \qquad y_j^{k+1} \geqslant \bar{y}_j^{k+1} \geqslant x_j, \qquad j = 1,..., i-1; \tag{3.12}$$

this is clearly valid for $k = 0$ and $i = 1$. Then

$$\begin{aligned} &f_i\left(y_1^{k+1},..., y_{i-1}^{k+1}, \frac{1}{\omega_k}(y_i^{k+1} - y_i^k) + y_i^k, y_{i+1}^k,..., y_n^k\right) \\ &= z_i = f_i\left(\bar{y}_1^{k+1},..., \bar{y}_{i-1}^{k+1}, \frac{1}{\bar{\omega}_k}(\bar{y}_i^{k+1} - \bar{y}_i^k) + \bar{y}_i^k, \bar{y}_{i+1}^k,..., \bar{y}_n^k\right) \\ &\geqslant f_i\left(y_1^{k+1},..., y_{i-1}^{k+1}, \frac{1}{\bar{\omega}_k}(\bar{y}_1^{k+1} - \bar{y}_i^k) + \bar{y}_i^k, y_{i+1}^k,..., y_n^k\right) \end{aligned}$$

implies, by the strict diagonal isotonicity of F and by (3.4) and (3.10), that

$$\frac{1}{\omega_k}(y_i^{k+1} - y_i^k) + y_i^k \geqslant \frac{1}{\bar{\omega}_k}(\bar{y}_i^{k+1} - \bar{y}_i^k) + \bar{y}_i^k \geqslant \frac{1}{\omega_k}(\bar{y}_i^{k+1} - \bar{y}_i^k) + \bar{y}_i^k$$

and hence that $y_i^{k+1} \geqslant \bar{y}_i^{k+1}$. Moreover, from

$$f_i\left(\bar{y}_1^{k+1},..., \bar{y}_{i-1}^{k+1}, \frac{1}{\bar{\omega}_k}(\bar{y}_i^{k+1} - \bar{y}_i^k) + \bar{y}_i^k, \bar{y}_{i+1}^k,..., \bar{y}_n^k\right)$$
$$= z_i = f_i(x) \geqslant f_i(\bar{y}_1^{k+1},..., \bar{y}_{i-1}^{k+1}, x_i, \bar{y}_{i+1}^k,..., \bar{y}_n^k)$$

we find that

$$\bar{y}_i^{k+1} \geqslant \frac{1}{\bar{\omega}_k}(\bar{y}_i^{k+1} - \bar{y}_i^k) + \bar{y}_i^k \geqslant x_i .$$

This completes the induction and with it the proof of

$$y^0 \geqslant y^k \geqslant \bar{y}^k \geqslant x, \qquad k = 0, 1,...,$$

and therefore of

$$y^* = \lim_{k\to\infty} y^k \geqslant \lim_{k\to\infty} \bar{y}^k = \bar{y}^* \geqslant x. \tag{3.13}$$

Analogously, we can show that

$$v^0 \geqslant v^k \geqslant \bar{v}^k \geqslant x, \qquad k = 0, 1,...$$

and hence that

$$v^* = \lim_{k\to\infty} v^k \geqslant \lim_{k\to\infty} \bar{v}^k = \bar{v}^* \geqslant x. \tag{3.14}$$

Since y^* and v^* are solutions of $Fx = z$ in the set J, (3.13) and (3.14) imply that $y^* = \bar{y}^*$ and $v^* = \bar{v}^*$.

Now suppose that for some $k \geqslant 0$

$$y^0 \geqslant v^k \geqslant y^k \tag{3.15}$$

which again holds for $k = 0$. Then, using $v^k \geqslant y^k \geqslant y^{k+1}$, we find that

$$f_i\left(y_1^{k+1},..., y_{i-1}^{k+1}, \frac{1}{\omega_k}(y_i^{k+1} - y_i^k) + y_i^k, y_{i+1}^k,..., y_n^k\right)$$
$$= z_i = f_i\left(v_1^k,..., v_{i-1}^k, \frac{1}{\omega_k}(v_i^{k+1} - v_i^k) + v_i^k, v_{i+1}^k,..., v_n^k\right)$$
$$\leqslant f_i\left(y_1^{k+1},..., y_{i-1}^{k+1}, \frac{1}{\omega_k}(v_i^{k+1} - v_i^k) + v_i^k, y_{i+1}^k,..., y_n^k\right), \qquad i = 1,..., n,$$

and hence that

$$\frac{1}{\omega_k}(y_i^{k+1} - y_i^k) + y_i^k \leqslant \frac{1}{\omega_k}(v_i^{k+1} - v_i^k) + v_i^k, \qquad i = 1,..., n,$$

which implies that $y_i^{k+1} \leqslant v_i^{k+1}$, $i = 1,..., n$. Therefore (3.15) holds for all $k \geqslant 0$, and from

$$\bar{v}^* = v^* \geqslant y^* = \bar{y}^* \geqslant x$$

it finally follows that all limits are equal and not less than x.

The proof of the theorem for the lower sequences proceeds analogously.

This result shows that—as in the linear case—the Jacobi process never converges faster than the Gauss–Seidel process, and, moreover, that in this setting it is most desirable not to underrelax. The conclusion $v^k \geqslant y^k$ generalizes a result of Porsching (1968) on the comparison of the iterates of the two methods.

Every M-function is by Theorem 2.10 strictly diagonally isotone. Hence the assumptions of Theorem 3.1 about the mapping F are certainly satisfied for all continuous M-functions. Note however that they are also valid, for example, in the case of the singular matrix

$$A = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}.$$

If $F : R^n \to R^n$ is a continuous M-function, then, by the domain-invariance theorem, $F(R^n)$ is open. Hence for any $z \in F(R^n)$ there exist points $a, b \in R^n$ such that $a \leqslant z \leqslant b$ and $\{y \in R^n \mid a \leqslant y \leqslant b\} \subset F(R^n)$, and hence that $x^0 = F^{-1}a \leqslant y^0 = F^{-1}b$ satisfy the condition (3.3) of Theorem 3.1. While nonlinear M-functions need not be surjective, the important point now is that for surjective M-functions it is not only possible to find for any $z \in R^n$ suitable initial points x^0, y^0 for which (3.3) holds, but, even better, that then the processes (3.1) and (3.2) converge for any starting point. This is the content of the next theorem; it represents the desired generalization of the mentioned global convergence result for the Gauss–Seidel process when A is an M-matrix. It also extends a result of Porsching (1968) to a more general class of mappings.

THEOREM 3.3. *Let $F : R^n \to R^n$ be a continuous, surjective M-function, and $\epsilon \in (0, 1]$ a given number. Then for any $z \in R^n$, any starting point $x^0 \in R^n$, and any sequence $\{\omega_k\} \subset [\epsilon, 1]$, the Gauss–Seidel process* (3.1), *as well as the Jacobi process* (3.2), *converges to the unique solution x^* of $Fx = z$.*

Proof. We prove the result again only for the Gauss–Seidel iteration. For given x^0, $z \in R^n$ define the vectors a, b, u^0, $v^0 \in R^n$ by

$$\begin{gathered} a_i = \min(f_i(x^0), z_i), \qquad b_i = \max(f_i(x^0), z_i), \qquad i = 1,..., n \\ u^0 = F^{-1}a, \qquad v^0 = F^{-1}b. \end{gathered} \tag{3.16a}$$

Then, by the inverse isotonicity,

$$Fv^0 \geqslant z \geqslant Fu^0, \qquad v^0 \geqslant x^0 \geqslant u^0, \qquad v^0 \geqslant x^* \geqslant u^0. \tag{3.16b}$$

Let $\{v^k\}$, $\{u^k\}$, and $\{x^k\}$ denote the Gauss–Seidel sequences starting from v^0, u^0, and x^0, respectively, each of them formed with the same sequence $\{\omega_k\}$ and with z as the right-hand side. By Theorem 2.10, F is surjectively diagonally isotone, and hence the solutions $\hat{v}_i^k$, $\hat{u}_i^k$, and $\hat{x}_i^k$ of the equations

$$\left.\begin{aligned} f_i(v_1^{k+1},\ldots, v_{i-1}^{k+1}, \hat{v}_i^k, v_{i+1}^k,\ldots, v_n^k) &= z_i \\ f_i(u_1^{k+1},\ldots, u_{i-1}^{k+1}, \hat{u}_i^k, u_{i+1}^k,\ldots, u_n^k) &= z_i \\ f_i(x_1^{k+1},\ldots, x_{i-1}^{k+1}, \hat{x}_i^k, x_{i+1}^k,\ldots, x_n^k) &= z_i \end{aligned}\right\} \begin{aligned} i &= 1,\ldots, n \\ k &= 0, 1,\ldots \end{aligned}$$

exist and are unique, and, therefore, the three Gauss–Seidel sequences are well-defined. Moreover, by Theorem 3.1 we have

$$\begin{aligned} &v^0 \geqslant v^k \geqslant v^{k+1} \geqslant u^{k+1} \geqslant u^k \geqslant u^0, \qquad && k = 0, 1,\ldots \\ &Fv^k \geqslant z \geqslant Fu^k, && k = 0, 1,\ldots \\ &\lim_{k\to\infty} v^k = \lim_{k\to\infty} u^k = x^* = F^{-1}z. \end{aligned} \tag{3.17}$$

Suppose that for some $k \geqslant 0$ and $i \geqslant 1$

$$v^k \geqslant x^k \geqslant u^k, \qquad v_j^{k+1} \geqslant x_j^{k+1} \geqslant u_j^{k+1}, \qquad j = 1,\ldots, i-1, \tag{3.18}$$

which is valid for $k = 0$ and $i = 1$. Then

$$\begin{aligned} f_i(v_1^{k+1},\ldots, v_{i-1}^{k+1}, \hat{v}_i^k, v_{i+1}^k,\ldots, v_n^k) = z_i &= f_i(x_1^{k+1},\ldots, x_{i-1}^{k+1}, \hat{x}_i^k, x_{i+1}^k,\ldots, x_n^k) \\ &\geqslant f_i(v_1^{k+1},\ldots, v_{i-1}^{k+1}, \hat{x}_i^k, v_{i+1}^k,\ldots, v_n^k), \end{aligned}$$

together with the strict diagonal isotonicity of F, implies that $\hat{v}_i^k \geqslant \hat{x}_i^k$. Similarly it follows that $\hat{x}_i^k \geqslant \hat{u}_i^k$. Hence, because of $\{\omega_k\} \subset [\epsilon, 1]$, we find that

$$\begin{aligned} v_i^{k+1} = (1-\omega_k)\, v_i^k + \omega_k \hat{v}_i^k &\geqslant (1-\omega_k)\, x_i^k + \omega_k \hat{x}_i^k = x_i^{k+1} \\ &\geqslant (1-\omega_k)\, u_i^k + \omega_k \hat{u}_i^k = u_i^{k+1}. \end{aligned}$$

This completes the induction, and (3.18) and (3.17) together now imply that $\lim_{k\to\infty} x^k = x^*$.

For any $A \in L(R^n)$ with $a_{ij} \leqslant 0$, $i \neq j$, it is well-known [see e.g., Varga (1962)] that A is an M-matrix if and only if A has a strictly positive diagonal and the Jacobi process converges globally. The following theorem extends this result to M-functions.

THEOREM 3.4. *Let $F : R^n \to R^n$ be continuous and off-diagonally antitone. Then the following three statements are equivalent:*

(a) *F is a surjective M-function;*

(b) *F is injective and surjectively diagonally isotone, and for any $x^0, z \in R^n$ the Jacobi sequence $\{x^k\}$ given by* (3.2) (*with $\omega_k \equiv 1$*) *converges;*

(c) *F is injective and surjectively diagonally isotone, and for any $x^0, z \in R^n$ the Gauss–Seidel sequence $\{x^k\}$ given by* (3.1) (*with $\omega_k \equiv 1$*) *converges.*

Proof. (a) $\Rightarrow$ (b) and (a) $\Rightarrow$ (c) are direct consequences of Lemma 2.3 and Theorems 2.10 and 3.3.

(b) $\Rightarrow$ (a). Since F is surjectively diagonally isotone, the Jacobi sequence $\{x^k\}$ for any $z \in R^n$ and $x^0 \in R^n$ exists and is uniquely defined. Because of the continuity of F, $\{x^k\}$ converges to a solution of $Fx = z$, and, since z was arbitrary, it follows that F is surjective. Let now $Fx \leqslant Fy$ for some $x, y \in R^n$, and consider for $z = Fy$ the Jacobi sequence $\{x^k\}$ starting from $x^0 = x$. By induction we find that

$$x^k \leqslant x^{k+1}, \qquad Fx^k \leqslant z, \qquad k = 0, 1, \ldots . \tag{3.19}$$

In fact, if $Fx^k \leqslant z$ for some $k \geqslant 0$, then

$$z_i = f_i(x_1{}^k, \ldots, x_{i-1}^k, x_i^{k+1}, x_{i+1}^k, \ldots, x_n{}^k) \geqslant f_i(x^k), \qquad i = 1, \ldots, n,$$

shows that $x_i^{k+1} \geqslant x_i{}^k$, $i = 1, \ldots, n$, and hence that

$$z_i = f_i(x_1{}^k, \ldots, x_{i-1}^k, x_i^{k+1}, x_{i+1}^k, \ldots, x_n{}^k) \geqslant f_i(x^{k+1}), \qquad i = 1, \ldots, n.$$

From (3.19) it follows, because of the injectivity of F, that

$$x = x^0 \leqslant \lim_{k \to \infty} x^k = F^{-1}z = y$$

and, therefore, that F is inverse isotone.

(c) $\Rightarrow$ (a). The proof proceeds analogously to that of (b) $\Rightarrow$ (a).

As the one-dimensional example $Fx = \exp(x)$ shows, the surjective diagonal isotonicity of F in (b) and (c) cannot be reduced to strict diagonal isotonicity. But it is conjectured that—as in the linear case—the injectivity assumption in these statements is not needed.

Any principal submatrix of an M-matrix is again an M-matrix. We use Theorem 3.4 to extend this fact to surjective M-functions.

THEOREM 3.5. *Let $F : R^n \to R^n$ be a continuous, surjective M-function. For some permutation $(m_1, ..., m_n)$ of $(1, ..., n)$, and with given $1 \leqslant p < n$ and fixed numbers $c_{p+1}, ..., c_n$, define*

$$G : R^p \to R^p, \qquad g_i(x_1, ..., x_p) = f_{m_i}\left(\sum_{j=1}^{p} x_j e^{m_j} + \sum_{j=p+1}^{n} c_j e^{m_j}\right), \qquad j = 1, ..., p.$$

Then G is again a continuous, surjective M-function.

Proof. It suffices to prove the statement for $p = n - 1$, since the general case will then follow by repeated application of the special result. For ease of notation, let $(m_1, ..., m_n)$ be the identity permutation, that is, let

$$G : R^{n-1} \to R^{n-1}, \qquad g_i(\bar{x}) = f_i(x_1, ..., x_{n-1}, c_n), \qquad i = 1, ..., n-1,$$

where c_n is fixed and we denote vectors of R^{n-1} by $\bar{x} = (x_1, ..., x_{n-1})^T$. Clearly G is again continuous, off-diagonally antitone, and surjectively diagonally isotone. In order to prove the injectivity of G, suppose that $G\bar{x} = G\bar{y}$ for some $\bar{x}, \bar{y} \in R^{n-1}$. We may then assume that, say,

$$f_n(x_1, ..., x_{n-1}, c_n) < f_n(y_1, ..., y_{n-1}, c_n),$$

since in the case of equality it would follow from the injectivity of F that $\bar{x} = \bar{y}$. By the surjective diagonal isotonicity of F, there exists a number $d_n > c_n$ such that

$$f_n(y_1, ..., y_{n-1}, c_n) = f_n(x_1, ..., x_{n-1}, d_n) \tag{3.19a}$$

and hence that

$$f_i(y_1, ..., y_{n-1}, c_n) = f_i(x_1, ..., x_{n-1}, c_n) \geqslant f_i(x_1, ..., x_{n-1}, d_n), \qquad i = 1, ..., n-1, \tag{3.19b}$$

By the inverse isotonicity of F, (3.19a/b) together imply that $\bar{x} \leqslant \bar{y}$ and $d_n \leqslant c_n$, which contradicts $d_n > c_n$. Thus G must be injective.

Consider now the (unrelaxed) Jacobi sequence $\{\bar{x}^k\}$ for G starting from some $\bar{x}^0 \in R^{n-1}$ and formed with the right hand side $\bar{z} \in R^{n-1}$. Clearly, $\{\bar{x}^k\}$ is uniquely defined. By the surjective diagonal isotonicity of F^{-1} there exists a unique z_n such that

$$c_n = f_n^{-1}(z_1, ..., z_{n-1}, z_n) \tag{3.20}$$

where f_n^{-1} denotes the n-th component of F^{-1}. Set $z = (z_1, ..., z_{n-1}, z_n)^T$ and $x^* = F^{-1}z$, then (3.20) implies that $x_n^* = c_n$. Now let

$$x^0 = (x_1^0, ..., x_{n-1}^0, c_n)^T$$

and define the vectors a, b, u^0, $v^0 \in R^n$ by (3.16a). Then (3.16b) holds and the (unrelaxed) Jacobi sequences $\{v^k\}$, $\{u^k\}$ for F with right-hand side z and with v^0 and u^0 as starting vectors satisfy (3.17). Suppose that for some $k \geqslant 0$

$$v_i^k \geqslant x_i^k \geqslant u_i^k, \qquad i = 1,\dots, n-1, \qquad v_n^k \geqslant c_n \geqslant u_n^k$$

which is certainly valid for $k = 0$. Then

$$f_i(v_1^k,\dots, v_i^{k+1},\dots, v_{n-1}^k, v_n^k) = z_i = f_i(x_1^k,\dots, x_i^{k+1},\dots, x_{n-1}^k, c_n)$$
$$\geqslant f_i(v_1^k,\dots, v_{i-1}^k, x_i^{k+1}, v_{i+1}^k,\dots, v_{n-1}^k, v_n^k),$$
$$i = 1,\dots, n-1$$

implies that $v_i^{k+1} \geqslant x_i^{k+1}$, $i = 1,\dots, n-1$, while it follows from

$$f_n(v_1^k,\dots, v_{n-1}^k, v_n^{k+1}) = z_n = f_n(x_1^*,\dots, x_{n-1}^*, c_n) \geqslant f_n(v_1^k,\dots, v_{n-1}^k, c_n)$$

that $v_n^{k+1} \geqslant c_n$. Similarly, we find that $u_i^{k+1} \leqslant x_i^{k+1}$, $i = 1,\dots, n-1$, and $u_n^{k+1} \leqslant c_n$. Thus

$$\lim_{k\to\infty} v_i^k = \lim_{k\to\infty} x_i^k = \lim_{k\to\infty} u_i^k = x_i^*, \qquad i = 1,\dots, n-1,$$

and it follows from Theorem 3.4 that G is a surjective M-function.

This generalizes a result of Duffin (1948) for mappings of the form

$$F : R^n \to R^n,$$
$$f_i(x) = g_i(x_i - x_1,\dots, x_i - x_{i-1}, x_i, x_i - x_{i+1},\dots, x_i - x_n), \qquad i = 1,\dots, n, \tag{3.21}$$

with isotone g_i.

We end this Section with a result which shows that the surjectivity condition needed in Theorem 3.3 is equivalent to another, in many cases more easily verifiable, condition.

In the case of a continuous isotone function $\varphi : R^1 \to R^1$ the assumptions

$$\lim_{t\to+\infty} \varphi(t) = +\infty, \qquad \lim_{t\to-\infty} \varphi(t) = -\infty \tag{3.22}$$

are necessary and sufficient for the surjectivity of φ. An analogous result holds for continuous M-functions; in order to phrase it, it will be convenient to introduce the following concept.

DEFINITION 3.6. (a) For any sequence $\{x^k\} \subset R^n$ we write

$$\lim_{k\to\infty} x^k = +\infty \qquad (\text{or} \lim_{k\to\infty} x^k = -\infty)$$

if

$$\lim_{k\to\infty} x_i^k = +\infty \qquad (\text{or} \lim_{k\to\infty} x_i^k = -\infty)$$

for at least one index i.

(b) The mapping $F : R^n \to R^n$ is order-coercive if for any sequence $\{x^k\} \subset R^n$

$$\begin{gathered} x^k \leqslant x^{k+1}, \qquad k = 0, 1,..., \qquad \lim_{k\to\infty} x^k = +\infty \\ \text{implies that} \qquad \lim_{k\to\infty} Fx^k = +\infty \end{gathered} \tag{3.23a}$$

and

$$\begin{gathered} x^k \geqslant x^{k+1}, \qquad k = 0, 1,..., \qquad \lim_{k\to\infty} x^k = -\infty \\ \text{implies that} \qquad \lim_{k\to\infty} Fx^k = -\infty. \end{gathered} \tag{3.23b}$$

Note that $\lim_{k\to\infty} x^k = +\infty$ for some $\{x^k\} \subset R^n$ does, in general, not exclude that $\lim_{k\to\infty} x^k = -\infty$, and vice versa, unless the sequence is bounded to one side.

THEOREM 3.7. *Let $F : R^n \to R^n$ be a continuous M-function. Then F is surjective if and only if F is order-coercive.*

Proof. Let F be surjective and hence, by Lemma 2.3, bijective. If $\{x^k\} \subset R^n$ is any sequence, then it follows from $Fx^k \leqslant a$, $k = 0, 1,...$, that $x^k \leqslant F^{-1}a$, $k = 0, 1,...$, and hence that the implication (3.23a) is valid. Similarly, we see that (3.23b) holds, and therefore that F is order-coercive.

Conversely, let now F be order-coercive. Then Theorem 3.1 will ensure that $Fx = z$ has a solution for every $z \in R^n$ if only we can prove that for every $z \in R^n$ there exist points $x^0, y^0 \in R^n$ for which (3.3) holds. We shall construct such points by means of the Jacobi process (3.2) with $\omega_k = 1$ for all $k \geqslant 0$.

Observe first that F is surjectively diagonally isotone. In fact, part (3) of the proof of Theorem 2.10 carries over verbatim except that in this case the order-coercivity is used to conclude from (2.5) that $\{t_k\}$ is bounded.

Let now $z \in R^n$ be given, and with arbitrary points $u^0, v^0 \in R^n$ define the vectors z'', $z' \in R^n$ by

$$z_i'' = \max(f_i(u^0), z_i), \qquad z_i' = \min(f_i(v^0), z_i), \qquad i = 1,..., n.$$

By the surjective diagonal isotonicity of F, the Jacobi sequences $\{u^k\}$ and $\{v^k\}$ starting from u^0 and v^0 and satisfying

$$\begin{aligned} f_i(u_1^{\,k},..., u_{i-1}^k, u_i^{k+1}, u_{i+1}^k,..., u_n^{\,k}) &= z_i'' \\ f_i(v_1^{\,k},..., v_{i-1}^k, v_i^{k+1}, v_{i+1}^k,..., v_n^{\,k}) &= z_i' \end{aligned} \qquad i = 1,..., n; \quad k = 0, 1,...$$

exist and are uniquely defined. Moreover, we have

$$u^k \leqslant u^{k+1}, \qquad Fu^k \leqslant z'', \qquad k = 0, 1,... \tag{3.24a}$$

and

$$v^k \geqslant v^{k+1}, \qquad Fv^k \geqslant z', \qquad k = 0, 1,... . \tag{3.24b}$$

In fact, clearly $Fu^0 \leqslant z''$, and, if $Fu^k \leqslant z''$ for some $k \geqslant 0$, then

$$f_i(u_1^{\,k},..., u_{i-1}^k, u_i^{k+1}, u_{i+1}^k,..., u_n^{\,k}) = z_i'' \geqslant f_i(u^k), \qquad i = 1,..., n,$$

and thus $u_i^{k+1} \geqslant u_i^{\,k}$, $i = 1,..., n$. Hence, by the off-diagonal antitonicity,

$$z_i'' = f_i(u_1^{\,k},..., u_{i-1}^k, u_i^{k+1}, u_{i+1}^k,..., u_n^{\,k}) \geqslant f_i(u^{k+1}), \qquad i = 1,..., n.$$

Similarly we can prove (3.24b).

From the order-coercivity it now follows that $\{u^k\}$ and $\{v^k\}$ must remain bounded above and below, respectively. Thus $\lim_{k\to\infty} u^k = y^0$ and $\lim_{k\to\infty} v^k = x^0$ exist, and by the continuity of F we have $Fy^0 = z'' \geqslant z$ and $Fx^0 = z' \leqslant z$. Moreover, the inverse isotonicity implies that $x^0 \leqslant y^0$, which completes the proof.

The idea of using the Jacobi process (3.2) to construct points x^0, y^0 for which (3.3) holds is due to Birkhoff and Kellogg (1966).

The simple one-dimensional example $f : R^1 \to R^1$, $f(t) = t - 1$ for $t < 0$, and $f(t) = t$ for $t \geqslant 0$ shows that when F is not continuous, order-coercivity does not necessarily imply surjectivity.

4. The Response Condition for Inverse Isotonicity

The global convergence result of Theorem 3.3 raises the question when a mapping $F : R^n \to R^n$ is a surjective M-function. In order to formulate necessary and sufficient conditions for this, some network-theoretical language will be useful.

We consider here only finite directed ***networks***[1] $\Omega = (N, A)$ consisting of

[1] In view of possible confusions with the analytic concept of the graph of a function, we shall use here the term network rather than graph.

the set of n *nodes* $N = \{1, 2, \ldots, n\}$ together with some set $\Lambda \subset N \times N$ of (directed) *links* connecting certain of these nodes. More specifically, there is a link from $i \in N$ to $j \in N$ exactly if the element $(i, j) \in N \times N$ is contained in Λ. There shall never be any loops in Ω, that is, Λ shall not contain elements of the form $(i, i) \in N \times N$. A (directed) *path* from i to j is a sequence of links in Λ of the form $(i, i_1), (i_1, i_2), \ldots, (i_m, j)$, and the network is *connected* if any two nodes are connected by some path.

DEFINITION 4.1. (a) The *associated network* $\Omega_F = (N, \Lambda_F)$ of the mapping $F : R^n \to R^n$ consists of the set of nodes $N = \{1, 2, \ldots, n\}$ and the set of links

$$\Lambda_F = \{(i, j) \in N \times N \mid i \neq j, \text{ and, for some } x \in R^n, \varphi_{ij} \text{ is not constant}\}$$

where the functions φ_{ij} are defined by (2.2).

(b) Let $F : R^n \to R^n$ be off-diagonally antitone, then a link $(i, j) \in \Lambda_F$ of Ω_F is *strict* (or *surjective*) if, for any $x \in R^n$, φ_{ij} is strictly antitone (or strictly antitone and surjective). A *path* in Ω_F is *strict* (or *surjective*) if all its links are strict (or surjective), and a node $i \in N$ is *strictly* (or *surjectively*) *connected* to a node $j \in N$ if there exists a strict (or surjective) path from i to j.

In terms of network flows this notation can be interpreted as follows: The variables $x_1, \ldots, x_n$ are *state variables* associated with the n nodes of Ω_F, and the value $f_i(x_1, \ldots, x_n)$ is the *efflux from the node i* when the network is in state x. The node $i \in N$ is connected to the node $j \in N$ if and only if for some state x a certain change in the state x_j of j produces some change in the efflux f_i from the node i. If F is off-diagonally antitone, then for any linked nodes $i, j \in N$ an increase (or decrease) of the state x_j of j produces the reverse effect in the efflux from i. This is exactly the expected situation in a linear network in which the x_i are potentials and the flow from i to j is proportional to the potential difference $x_i - x_j$.

In the linear case $A \in L(R^n)$, all links of the associated network are surjective, and Ω_A is simply the (loopless) directed graph which has A as an adjacency matrix.

In order to answer the question when an off-diagonally antitone function $F : R^n \to R^n$ is inverse isotone, we shall always require two different conditions. The first concerns connectivity properties of the associated network, while the second contains information either about the response shown by the efflux from a certain node to some change of the state of all nodes, (response condition), or about the influence of a state-change of one node upon the efflux from all nodes, (influence condition). As an illustrative example for this, we quote here the following linear result, [see Schröder (1961) and also Collatz (1964)].

LEMMA 4.2. *Let $A \in L(R^n)$ be such that $a_{ij} \leqslant 0$ for $i \neq j$, $i, j = 1,..., n$, and $u > 0$ a vector for which*

$$Au = v \geqslant 0, \qquad v \neq 0.$$

Assume further that for any $i \neq j$ there exists a sequence of nonzero elements $a_{ii_1}, a_{i_1 i_2}, ..., a_{i_m j}$. Then A is an M-matrix.

In terms of networks we have here indeed the mentioned two conditions, namely, (a) the associated network is connected, and, (b) if the state vector x is changed according to the linear "test-function" $x + tu$, the efflux from any given node i,

$$f_i(x + tu) = \sum_{j=1}^{n} a_{ij}x_j + v_i t, \tag{4.1}$$

is an isotone function of t. Evidently, (b) is a response condition.

In the case of nonlinear mappings it will be useful to allow also nonlinear test-functions:

THEOREM 4.3. *Let $F : R^n \to R^n$ be off-diagonally antitone, and suppose that there is a continuous, bijective, and isotone "test-function" $H : R^n \to R^n$ such that for any $x \in R^n$ the mapping*

$$P : R^1 \to R^n, p_i(t) = f_i(H(x + te)), \qquad i = 1,..., n, \tag{4.2}$$

with $e = (1, 1,..., 1)^T$, is isotone. Assume further that every node $i \in N$ of Ω_F is strictly connected to a node $l = l(i)$ at which, for any $x \in R^n$, p_l is strictly isotone. Then F is inverse isotone and hence an M-function.

Proof. Let $Fx \leqslant Fy$ for some $x, y \in R^n$, *and set* $u = H^{-1}x$, $v = H^{-1}y$. Since $v + te \geqslant u \geqslant v - te$, and hence $H(v + te) \geqslant Hu \geqslant H(v - te)$, for all sufficiently large t, and since H is bijective, it follows that

$$+\infty > t_0 = \inf\{t \in R^1 \mid H(v + te) \geqslant Hu\} > -\infty.$$

Moreover, the continuity of H implies that $H(v + t_0 e) \geqslant Hu$ and that the set

$$N_0 = \{i \in N \mid h_i(v + t_0 e) = h_i(u)\}$$

is not empty. If $t_0 \leqslant 0$, then $x = Hu \leqslant H(v + t_0 e) \leqslant Hv = y$; suppose, therefore, that $t_0 > 0$. By the connectivity assumption, there exists for any $i \in N$ at least one strict link $(i, j) \in \Lambda_F$ to some other node $j \in N$, unless p_i is strictly isotone for each $x \in R^n$. Assume first that $(i, j) \in \Lambda_F$ is strict and that

$i \in N_0$. If $j \notin N_0$, then, using $h_j(v + t_0 e) > h_j(u)$, $h_i(v + t_0 e) = h_i(u)$, and the strictness of the link, we find that

$$\begin{aligned} &f_i(H(v + t_0 e)) \\ &\quad < f_i(h_1(v + t_0 e),\dots, h_{j-1}(v + t_0 e), h_j(u), h_{j+1}(v + t_0 e),\dots, h_n(v + t_0 e)) \\ &\quad \leqslant f_i(Hu) = f_i(x) \leqslant f_i(y) = f_i(Hv) \leqslant f_i(H(v + t_0 e)) \end{aligned}$$

which is a contradiction. Hence $j \in N_0$, and by assumption, there is at least one node $i \in N_0$ such that p_i is strictly isotone for each $x \in R^n$. But then $h_i(v + t_0 e) = h_i(u)$ together with $t_0 > 0$ leads to the contradiction

$$\begin{aligned} &f_i(H(v + t_0 e)) \\ &\quad = f_i(h_1(v + t_0 e),\dots, h_{i-1}(v + t_0 e), h_i(u), h_{i+1}(v + t_0 e),\dots, h_n(v + t_0 e)) \\ &\quad \leqslant f_i(Hu) = f_i(x) \leqslant f_i(y) = f_i(Hv) < f_i(H(v + t_0 e)). \end{aligned}$$

Thus we must have $t_0 \leqslant 0$ and the proof is complete.

For the special case $H = I$ and for mappings of the form (3.21) with isotone g_i, this result was first proved by Duffin (1948) by means of a reduction to the linear case. Schröder (1962) proved a related result for diagonal matrices H and strictly isotone functions P. Later Schröder (1966) also considered certain nonlinear test-functions in connection with the inverse isotonicity of some problems for quasi-linear differential equations.

It should be noted that, in general, the strictness assumptions in the connectivity condition of Theorem 4.3 cannot be relaxed. In fact, the two-dimensional examples

$$Fx = \begin{pmatrix} \alpha(x_1 - x_2) \\ x_2 \end{pmatrix}, \qquad Gx = \begin{pmatrix} x_1 \\ \alpha(x_2) \end{pmatrix},$$

$$\alpha(t) = \begin{cases} t - 1 & \text{for} \quad t \geqslant 1 \\ 0 & \text{for} \quad -1 \leqslant t \leqslant 1 \\ t + 1 & \text{for} \quad t \leqslant -1 \end{cases}$$

satisfy the conditions of the theorem with $H = I$, except that in Ω_F the link $(1, 2)$ is not strict, and that for G the function p_2 is not strictly isotone. Both functions are not inverse isotone.

As an application of Theorem 4.3 we obtain the following linear result:

THEOREM 4.4. *Let $A \in L(R^n)$ be such that $a_{ij} \leqslant 0$, $i \neq j$, $i, j = 1,\dots, n$. Then the following three statements are equivalent:*

(a) *A is an M-matrix*

(b) *There exists a vector $u > 0$ such that $Au > 0$.*

(c) *There exists a vector* $u > 0$ *such that* $Au = v \geqslant 0$ *and that for every* i *there is a sequence of nonzero elements* $a_{ii_1}, a_{i_1i_2}, \ldots, a_{i_ml}$ *ending at an index* $l = l(i)$ *for which* $v_l > 0$.

Proof. (a) $\Rightarrow$ (b). Let $A^{-1} = (b_{ij})$, then $b_{ij} \geqslant 0$ for all i, j and, by Theorem 2.10, we have $b_{ii} > 0$, $i = 1, \ldots, n$. Thus

$$u = A^{-1}e = \left(\sum_{j=1}^{n} b_{1j}, \ldots, \sum_{j=1}^{n} b_{nj}\right)^T \geqslant (b_{11}, \ldots, b_{nn})^T > 0$$

and $Au = e > 0$.

(b) $\Rightarrow$ (c). Trivial.

(c) $\Rightarrow$ (a). Evidently, with $H = \mathrm{diag}(u_1, \ldots, u_n)$ the mapping

$$P(t) = AH(x + te) = AHx + tv$$

is isotone in t for any $x \in R^n$. This together with the sequence assumption ensures, by Theorem 4.3, that A is an M-function and thus an M-matrix.

For $u = e$ the implication (c) $\Rightarrow$ (a) was proved by Duffin (1948). The implication (b) $\Rightarrow$ (a) appears to be due to Ky Fan (1958), see also Schröder (1961) where several related literature references are given. Evidently, Theorem 4.4 is a genuine extension of Lemma 4.2.

Theorem 4.4 provides a simple proof of the well-known result that when A is an M-matrix and $B \geqslant 0$ has the property that $A + B$ still has only non-positive off-diagonal elements, then $A + B$ is again an M-matrix. In fact, by Theorem 4.4(b) there is a vector $u > 0$ for which $Au > 0$; hence $(A + B)u \geqslant Au > 0$ implies by the same theorem that $A + B$ is an M-matrix.

The same result carries over to the surjective nonlinear case.

THEOREM 4.5. *Let* $F : R^n \to R^n$ *be a continuous, surjective M-function, and* $G : R^n \to R^n$ *a continuous, isotone mapping such that* $\hat{F} = F + G$ *is still off-diagonally antitone. Then* $\hat{F}$ *is again a continuous, surjective M-function.*

Proof. Evidently, $H = F^{-1}$ is a continuous, bijective and isotone mapping, and every component of

$$\begin{aligned}\hat{P}(t) &= \hat{F}(H(x + te)) = F(H(x + te)) + G(H(x + te)) \\ &= x + te + G(H(x + te))\end{aligned}$$

is strictly isotone in t for any $x \in R^n$. Thus, by Theorem 4.3, $\hat{F}$ is an M-function. Suppose now that $\{x^k\} \subset R^n$ is such that $x^k \leqslant x^{k+1}$, $k = 0, 1, \ldots$. Then $Fx^k + Gx^0 \leqslant \hat{F}x^k \leqslant a$ implies that $x^k \leqslant F^{-1}(a - Gx^0)$, $k = 0, 1, \ldots,$ and hence that (3.23a) holds. Similarly, we see that (3.23b) is valid, and hence that F is order-coercive. By Theorem 3.7, F is therefore surjective.

As a simple application, consider a mapping

$$F : R^n \to R^n, \qquad Fx = Ax + Gx \tag{4.3}$$

where $A \in L(R^n)$ is an M-matrix and G is isotone and a so-called *diagonal* mapping, that is a function of the form

$$G : R^n \to R^n, \qquad g_i(x) = g_i(x_i), \qquad i = 1,..., n. \tag{4.4}$$

For continuous G, Theorem 4.5 implies that F is a continuous, surjective M-function. Interestingly, in this case, F is already an M-function even if G is not continuous. In fact, if $u > 0$ is selected such that $v = Au > 0$, then, with $H = \operatorname{diag}(u_1 ,..., u_n)$, the function

$$P(t) = F(H(x + te)) = AHx + tv + G(Hx + tu)$$

is strictly isotone in t for all $x \in R^n$, and thus, by Theorem 4.3, F is an M-function. Mappings of the type (4.3/4) arise naturally in the discretization of mildly nonlinear elliptic boundary value problems such as $\Delta u = \varphi(u)$, $x \in \mathscr{D}^0$; $u = g$, $x \in \dot{\mathscr{D}}$.

If we restrict the class of test-functions permissible in Theorem 4.3 and if we strengthen, at the same time, the strictness assumptions in that theorem by including surjectivity, then we obtain a sufficient condition for a continuous, off-diagonally antitone mapping to be a surjective M-function. Note here that a continuous, diagonal mapping (in the sense of (4.4)) is a surjective M-function if and only if each component is strictly isotone and surjective.

THEOREM 4.6. *Let $F : R^n \to R^n$ be continuous and off-diagonally antitone, and suppose that there is a continuous, diagonal, and surjective M-function $H : R^n \to R^n$ such that, for any $x \in R^n$, the mapping P of (4.2) is isotone. Assume further that every node i of Ω_F is surjectively connected to a node $l = l(i)$ such that, for any $x \in R^n$, the component p_l of P is strictly isotone and surjective. Then F is a surjective M-function.*

Proof. We prove the statement first for the special case $H = I$, that is for $P(t) = F(x + te)$. From Theorem 4.3 it follows that F is an M-function, and hence, by Theorem 3.7, it suffices to show that F is order-coercive.

Let $\{x^k\} \subset R^n$ be such that $x^k \leqslant x^{k+1}$, $k = 0, 1,...$, as well as $\lim_{k\to\infty} x^k = +\infty$, and suppose that $Fx^k \leqslant a \in R^n$ for all $k \geqslant 0$. By selecting, if necessary, a suitable subsequence, (and denoting it again by $\{x^k\}$), we can ensure that $x^k \geqslant 0$ and $x^k_{i_0} = \max_i x_i^k$ for all $k \geqslant 0$, where i_0 is a fixed index.

By assumption, the set

$$N_\infty = \{i \in N \mid \lim_{k\to\infty} x_i^k = +\infty\}$$

is not empty, and, as a consequence, N_∞ contains the subset

$$N_\infty' = \{i \in N \mid x^k - x_i^k e \leqslant \beta_i e,\ k = 0, 1,\ldots, \beta_i < +\infty\}.$$

In fact, from $i \in N_\infty' \cap (N \sim N_\infty)$ it would follow that

$$x^k \leqslant (x_i^k + \beta_i)\, e \leqslant (\alpha_i + \beta_i)\, e, \qquad k = 0, 1,\ldots$$

which implies that N_∞ is empty. Evidently, N_∞' is not empty, since by construction $i_0 \in N_\infty'$ with $\beta_{i_0} = 0$. Note that always $\beta_i \geqslant 0$.

For any $i \in N_\infty'$ either p_i is strictly isotone and surjective, or there exists a surjective link $(i, j) \in \Lambda_F$ to some other node $j \in N$. In the latter case, by the isotonicity of p_i, we have for all $k \geqslant 0$

$$a_i \geqslant f_i(x^k) \geqslant f_i(x^k - (x_i^k + \beta_i)\, e) \geqslant f_i(-\beta_i e^i + (x_j^k - x_i^k - \beta_i)\, e^j) \tag{4.5}$$

and from the surjectivity of the link it follows that

$$\sup\{t \in R^1 \mid f_i(-\beta_i e^i - te^j) \leqslant a_i\} = \beta_j < +\infty.$$

Hence, (4.5) implies that $0 \leqslant -(x_j^k - x_i^k - \beta_i) \leqslant \beta_j$, or

$$x^k \leqslant (x_i^k + \beta_i)\, e \leqslant (x_j^k + \beta_j)\, e, \qquad k = 0, 1,\ldots,$$

that is, $j \in N_\infty'$.

As a consequence, there must be at least one node $i \in N_\infty'$ such that p_i is strictly isotone and surjective. Then

$$a_i \geqslant f_i(x^k) \geqslant f_i(x^k - \beta_i e) \geqslant f_i(-\beta_i e^i + x_i^k e), \qquad k = 0, 1,\ldots, \tag{4.6}$$

and, since p_i is surjective,

$$\sup\{t \in R^1 \mid f_i(-\beta_i e^i + te) \leqslant a_i\} = \beta < +\infty.$$

Thus it follows from (4.6) that $x_i^k \leqslant \beta$ for all $k \geqslant 0$, which contradicts $i \in N_\infty$. Hence, we must necessarily have $\lim_{k\to\infty} Fx^k = +\infty$.

Analogously, it follows for any sequence $\{x^k\} \subset R^n$ with $x^k \geqslant x^{k+1}$, $k = 0, 1,\ldots,$ and $\lim_{k\to\infty} x^k = -\infty$, that necessarily $\lim_{k\to\infty} Fx^k = -\infty$. Altogether, therefore, F is order-coercive and the proof for the special case $H = I$ is complete.

For the proof of the general case, set

$$\hat{F} : R^n \to R^n, \qquad \hat{f}_i(x) = f_i(h_1(x_1), \ldots, h_n(x_n)), \qquad i = 1, \ldots, n$$

and

$$\hat{P} : R^1 \to R^n, \qquad \hat{P}(t) = \hat{F}(x + te) = F(H(x + te)).$$

Evidently, $\hat{F}$ is continuous and off-diagonally antitone, $\hat{P}$ is isotone for any $x \in R^n$, and, by the surjective diagonal isotonicity of H, every surjective link of Ω_F is also surjective in $\Omega_{\hat{F}}$. Hence $\hat{F}$ satisfies the statement of the theorem with $H = I$, and by the proof for this special case, $\hat{F}$ is surjective. But then the same is true for $F = \hat{F} \circ H^{-1}$.

For mappings of the form (3.21) and for $H = I$ the result was again first proved by Duffin (1948) and our proof follows his proof-idea.

As an application of this theorem, consider the two-point boundary value problem

$$u'' = \varphi(t, u, u'), \qquad 0 < t < 1; \qquad u(0) = \alpha, \qquad u(1) = \beta, \tag{4.7}$$

where φ is continuous on

$$S = \{(t, u, p)^T \in R^3 \mid 0 < t < 1, -\infty < u, p < +\infty\}.$$

Assume further that φ has continuous partial derivatives on S with respect to its second and third argument, and that with some $\gamma > 0$

$$\left.\begin{array}{r} \partial_2\varphi(t, u, p) \geqslant 0 \\ |\partial_3\varphi(t, u, p)| \leqslant \gamma \end{array}\right\} \qquad \text{for all} \qquad (t, u, p)^T \in S. \qquad \begin{array}{r}(4.8a)\\(4.8b)\end{array}$$

With $s_j = jh$, $j = 0, 1, \ldots, n + 1$, $h = 1/(n + 1)$, we introduce the simple discrete analogue $Fx = 0$ of (4.7) where $F : R^n \to R^n$ has the components

$$f_i(x) = 2x_i - x_{i-1} - x_{i+1} + h^2\varphi\left(s_i, x_i, \frac{1}{2h}(x_{i+1} - x_{i-1})\right),$$
$$i = 1, \ldots, n; \qquad x_0 = \alpha, \qquad x_{n+1} = \beta.$$

Evidently, F is continuous on all of R^n. For fixed $h \in (0, 2/\gamma)$, (4.8b) implies that each f_i is an antitone function of x_{i-1} and of x_{i+1}, and hence that F is off-diagonally antitone. Set

$$P : R^1 \to R^n, \qquad p_i(t) = f_i(x + te), \qquad i = 1, \ldots, n,$$

then, by (4.8a), the components $p_2, \ldots, p_{n-1}$ are certainly isotone, while it follows from (4.8a/b) together with $h \in (0,2/\gamma)$ that

$$p_1(t) = 2x_1 - \alpha - x_2 + t + h^2\varphi\left(s_1, x_1 + t, \frac{1}{2h}(x_2 + t - \alpha)\right)$$

is strictly isotone and surjective, and, similarly, that the same is true for p_n. In Ω_F every node $i = 2,..., n - 1$ is linked to its "neighbors" $i - 1$ and $i + 1$, and thus every i is connected to 1 as well as to n. Finally, again by (4.8b) and $h \in (0, 2/\gamma)$, it follows that each function $f_i(x + te^j)$, $| i - j | = 1$, is strictly antitone and surjective in t, and hence that every link of Ω_F is surjective. Thus Theorem 4.6 applies, and F is for $h \in (0, 2/\gamma)$ a continuous, surjective M-function.

Clearly, the same arguments apply also to the corresponding discretizations of mildly nonlinear elliptic problems of the form

$$\Delta u = \varphi(s, t, u, u_s, u_t), \qquad x \in \mathscr{D}^0; \qquad u = g, \qquad x \in \dot{\mathscr{D}},$$

which were considered by Bers (1953). The global convergence Theorem 3.3 then provides another proof for the convergence result given by Bers.

5. The Influence Condition for Inverse Isotonicity and Boundary Value Problems for Network Flows

In this Section we turn to influence conditions in the sense of the discussion at the beginning of Section 4; that is, to conditions concerning the influence of certain changes of the state of one node upon the efflux from all nodes of the network. Such conditions permit another (sufficient) answer to the question when an off-diagonally antitone mapping is inverse isotone.

Consider the condition (b) of Theorem 4.4 for the transpose A^T of a matrix $A \in L(R^n)$ with nonpositive off-diagonal elements. If $A^T u > 0$ for some $u > 0$, then A^T—and hence also A—is an M-matrix. In network-terminology, $A^T u > 0$ means that for each i the linear combination of all effluxes

$$\sum_{j=1}^{n} u_j f_j(x + te^i) = x^T A^T u + (A^T u)_i \, t \tag{5.1}$$

is an isotone function of any change t of the state of the i-th node. This represents an influence condition which, in this form, can also be applied to nonlinear mappings.

Theorem 5.1. *Let $F : R^n \to R^n$ be off-diagonally antitone, and suppose that there exists a diagonal M-function $H : D \subset R^n \to R^n$ such that $F(R^n) \subset D$ and that, for any $x \in R^n$, the function*

$$Q : R^1 \to R^n, \qquad q_i(t) = \sum_{j=1}^{n} h_j(f_j(x + te^i)), \qquad i = 1,..., n, \tag{5.2}$$

is isotone. Let

$$\begin{aligned} \phi &: R^1 \to R^n, \qquad \varphi_i(t), \qquad i = 1,..., n; \\ \psi &: R^1 \to R^n, \qquad \psi_i(t), \qquad i = 1,..., n, \end{aligned} \tag{5.3}$$

be isotone mappings such that $\phi + \psi$ is strictly isotone, and assume that for every node i in Ω_F there exists a node $l = l(i)$ which is strictly connected to i and for which there is strict isotonicity either of φ_l or of q_l for any $x \in R^n$. Then

$$\hat{F} : R^n \to R^n, \qquad \hat{f}_i(x) = \varphi_i(x_i) + \psi_i(f_i(x)), \qquad i = 1,..., n, \tag{5.4}$$

is an M-function.

Proof. Evidently, $\hat{F}$ is again off-diagonally antitone. In order to show that $\hat{F}$ is inverse isotone, let $\hat{F}x \leqslant \hat{F}y$ for some $x, y \in R^n$, and set

$$N^- = \{i \in N \mid y_i < x_i\}; \qquad N^+ = \{i \in N \mid y_i \geqslant x_i\}.$$

Suppose that N^- is not empty, let $i_1, i_2, ..., i_m$ be the nodes of N^-, and set

$$z^j = (x_{i_j} - y_{i_j})\, e^{i_j}, \qquad j = 1,..., m, \qquad z = \sum_{j=1}^{m} z^j.$$

Then

$$\begin{aligned} \sum_{j=1}^{n} h_j(f_j(y)) &\leqslant \sum_{j=1}^{n} h_j(f_j(y + z^1)) \leqslant \sum_{j=1}^{n} h_j(f_j(y + z^1 + z^2)) \leqslant \cdots \\ &\leqslant \sum_{j=1}^{n} h_j(f_j(y + z)) \leqslant \sum_{j\in N^-} h_j(f_j(y + z)) + \sum_{j\in N^+} h_j(f_j(y)) \qquad (5.5) \\ &\leqslant \sum_{j\in N^-} h_j(f_j(x)) + \sum_{j\in N^+} h_j(f_j(y)) \end{aligned}$$

where the sums over N^+ may be empty. Hence we have

$$\sum_{j\in N^-} h_j(f_j(y)) \leqslant \sum_{j\in N^-} h_j(f_j(x)). \tag{5.6}$$

Suppose that $f_j(y) < f_j(x)$ for some $j \in N^-$, then

$$\begin{aligned} \varphi_j(x_j) + \psi_j(f_j(x)) = \hat{f}_j(x) \leqslant \hat{f}_j(y) &= \varphi_j(y_j) + \psi_j(f_j(y)) \\ &\leqslant \varphi_j(y_j) + \psi_j(f_j(x)). \end{aligned} \tag{5.7}$$

Here the last inequality is strict if ψ_j is strictly isotone, and then it follows from $\varphi_j(x_j) < \varphi_j(y_j)$ that $x_j < y_j$, which contradicts $j \in N^-$. If φ_j is strictly isotone, then $\varphi_j(x_j) \leqslant \varphi_j(y_j)$ implies that $x_j \leqslant y_j$, which again contradicts $j \in N^-$. Thus we must have $f_j(y) \geqslant f_j(x)$ for all $j \in N^-$ and

$$\sum_{j\in N^-} h_j(f_j(x)) \leqslant \sum_{j\in N^-} h_j(f_j(y)) \leqslant \sum_{j\in N^-} h_j(f_j(x)) \tag{5.8}$$

shows that (5.6) is an equality. But then also all inequalities in (5.5) are equalities. Moreover, it follows from (5.8) and the strict isotonicity of H that $f_j(y) = f_j(x)$, $j \in N^-$, which by (5.7) implies that $\varphi_j(x_j) \leqslant \varphi_j(y_j)$, $j \in N^-$. But then it follows from $x_j > y_j$, $j \in N^-$, that $\varphi_j(x_j) = \varphi_j(y_j)$, $j \in N^-$, and therefore that φ_j is not strictly isotone for $j \in N^-$. From Eq. (5.5) we obtain for Q taken at the point y that

$$q_i(0) = \sum_{j=1}^{n} h_j(f_j(y)) = \sum_{j=1}^{n} h_j(f_j(y + (x_i - y_i)\, e^i)) = q_i(x_i - y_i), \qquad i \in N^-,$$

which, because of $x_i > y_i$, shows that q_i, $i \in N^-$, is not strictly isotone for all points of R^n. Altogether therefore, there must be at least one node $l \in N^+$ for which there is strict isotonicity either of φ_l or of q_l at any $x \in R^n$. Then, by the connectivity assumption, there is at least one strict link $(i, i') \in \Lambda_F$ with $i \in N^+$ and $i' \in N^-$. From

$$\begin{aligned} &f_i(y) > f_i(y + (x_{i'} - y_{i'})\, e^{i'}), \\ &f_j(y) \geqslant f_j(y + (x_{i'} - y_{i'})\, e^{i'}), \qquad j \in N^+ \sim \{i'\}, \end{aligned}$$

the strict isotonicity of H, and (5.5) it follows now that

$$\begin{aligned} \sum_{j=1}^{n} h_j(f_j(y)) &= \sum_{j\in N^-} h_j(f_j(y+z)) + \sum_{j\in N^+} h_j(f_j(y)) \\ &> \sum_{j\in N^-} h_j(f_j(y+z)) + \sum_{j\in N^+} h_j(f_j(y + (x_{i'} - y_{i'})\, e^{i'})) \\ &\geqslant \sum_{j\in N^-} h_j(f_j(y+z)) + \sum_{j\in N^+} h_j(f_j(y+z)) \\ &= \sum_{j=1}^{n} h_j(f_j(y+z)) = \sum_{j=1}^{n} h_j(f_j(y)), \end{aligned}$$

which is a contradiction. Thus N^- must be empty and $\hat{F}$ is inverse isotone.

Note that for $\phi = 0$ and $\psi = I$ we obtain from this theorem a result about the mapping F itself.

As in the case of Theorem 4.6 the addition of continuity and surjectivity assumptions to Theorem 5.1 provides a result about surjective M-functions.

THEOREM 5.2. *Let $F : R^n \to R^n$ be continuous and off-diagonally antitone. Suppose that there is a continuous, diagonal, and surjective M-function $H : R^n \to R^n$ such that for any $x \in R^n$ the function Q of* (5.2) *is isotone. Let the functions ϕ and ψ of* (5.3) *be continuous, isotone mappings such that each component of $\phi + \psi$ is strictly isotone and surjective. Assume further that for every node i of Ω_F there exists a node $l = l(i)$ which is surjectively connected to i and at which there is strict isotonicity and surjectivity of φ_l, or of q_l for any $x \in R^n$. Then the mapping $\hat{F}$ of* (5.4) *is a continuous, surjective M-function.*

Proof. By Theorem 5.1, $\hat{F}$ is certainly an M-function, and hence, by Theorem 3.7 it suffices to show that $\hat{F}$ is order-coercive.

Let $\{x^k\} \subset R^n$ be such that $x^k \leqslant x^{k+1}$, $k = 0, 1,\ldots$, and $\lim_{k\to\infty} x^k = +\infty$, and suppose that $\hat{F}x^k \leqslant a \in R^n$ for all $k \geqslant 0$. Set

$$N_\infty = \{i \in N \mid \lim_{k\to\infty} x_i^k = +\infty\},$$

$$N_0 = \{i \in N \mid x_i^k \leqslant \alpha_i < +\infty, k = 0, 1,\ldots\}$$

where N_0 may be empty. Then $f_j(x^k) \leqslant b_j < +\infty$ for $j \in N_\infty$. In fact, $\lim_{k\to\infty} f_j(x^k) = +\infty$ for some $j \in N_\infty$ and the surjectivity of all components of $\phi + \psi$ imply, otherwise, that

$$\lim_{k\to\infty} \hat{f}_j(x^k) = \lim_{k\to\infty}[\varphi_j(x^k) + \psi_j(f_j(x^k))] = +\infty,$$

which contradicts $\hat{f}_j(x^k) \leqslant a_j < +\infty$, $k \geqslant 0$.

Now set

$$z_i^k = x_i^k - x_i^0 \ (\geqslant 0), \quad i \in N_\infty, \qquad z^k = \sum_{i\in N_\infty} z_i^k e^i, \quad k = 0, 1,\ldots$$

and

$$y = \sum_{i\in N_\infty} x_i^0 e^i + \sum_{i\in N_0} \alpha_i e^i$$

where the sum over N_0 may, of course, be vacuous. Then, as in (5.5), we find that, for any $i \in N_\infty$,

$$\begin{aligned}
\sum_{j=1}^n h_j(f_j(y)) &\leqslant \sum_{j=1}^n h_j(f_j(y + z_i^k e^i)) \leqslant \sum_{j=1}^n h_j(f_j(y + z^k)) \\
&\leqslant \sum_{j\in N_\infty} h_j(f_j(y + z^k)) + \sum_{j\in N_0} h_j(f_j(y)) \\
&\leqslant \sum_{j\in N_\infty} h_j(f_j(x^k)) + \sum_{j\in N_0} h_j(f_j(y)) \\
&\leqslant \sum_{j\in N_\infty} h_j(b_j) + \sum_{j\in N_0} h_j(f_j(y)) = \beta < +\infty.
\end{aligned} \tag{5.9}$$

This shows that, for Q taken at the point y, we have

$$q_i(0) = \sum_{j=1}^{n} h_j(f_j(y)) \leqslant q_i(z_i^k) \leqslant \beta < +\infty, \qquad i \in N_\infty, \qquad k \geqslant 0,$$

and hence that the components q_i, $i \in N_\infty$, of Q are not surjective for all points of R^n. Moreover, from (5.9) it follows that

$$\sum_{j \in N_\infty} h_j(f_j(y)) \leqslant \sum_{j \in N_\infty} h_j(f_j(x^k)) \leqslant \sum_{j \in N_\infty} h_j(b_j), \qquad k \geqslant 0,$$

which, together with $h_j(f_j(x^k)) \leqslant h_j(b_j)$, $k \geqslant 0$, and the surjectivity of all h_j, implies that $f_j(x^k) \geqslant c_j > -\infty$ for $j \in N_\infty$ and all $k \geqslant 0$. But then

$$\begin{aligned} -\infty < \varphi_j(x_j^0) \leqslant \varphi_j(x_j^k) &= \hat{f}_j(x^k) - \psi_j(f_j(x^k)) \\ &\leqslant \hat{f}_j(x^k) - \psi_j(h_j(c_j)) \leqslant a_j - \psi_j(h_j(c_j)) < +\infty \end{aligned}$$

shows that φ_j cannot be surjective for $j \in N_\infty$. Hence, there must be one node $l \in N_0$ at which there is strict isotonicity and surjectivity either of φ_l, or of q_l for any $x \in R^n$. Moreover, there is at least one surjective link $(i', i) \in \Lambda_F$ with $i' \in N_0$ and $i \in N_\infty$. Then

$$\lim_{k \to \infty} f_{i'}(y + z_i^k e^i) = -\infty$$

and hence, because of the surjectivity of all h_j,

$$\gamma_k = \sum_{j \in N_0} h_j(f_j(y)) - \sum_{j \in N_0} h_j(f_j(y + z_i^k e^i)) \geqslant 0, \qquad k \geqslant 0$$

$$\lim_{k \to \infty} \gamma_k = +\infty.$$

Let $k_0 \geqslant 0$ be such that

$$\gamma_k > \sum_{j \in N_\infty} h_j(b_j) - \sum_{j \in N_\infty} h_j(f_j(y)), \qquad k \geqslant k_0,$$

then, together with (5.9), we get

$$\begin{aligned} \sum_{j=1}^{n} h_j(f_j(y)) &\leqslant \sum_{j=1}^{n} h_j(f_j(y + z^k)) \\ &= \sum_{j \in N_\infty} h_j(f_j(y + z^k)) + \sum_{j \in N_0} h_j(f_j(y + z^k)) \\ &\leqslant \sum_{j \in N_\infty} h_j(f_j(x^k)) + \sum_{j \in N_0} h_j(f_j(y + z_i^k e^i)) \\ &\leqslant \sum_{j \in N_\infty} h_j(b_j) + \sum_{j \in N_0} h_j(f_j(y)) - \gamma_k < \sum_{j=1}^{n} h_j(f_j(y)), \qquad k \geqslant k_0 \end{aligned}$$

which is a contradiction. Therefore, $\hat{F}x^k \leqslant a$, $k = 0, 1,\ldots$, is impossible and we must have $\lim_{k\to\infty} \hat{F}x^k = +\infty$.

Analogously it follows that $\{x^k\} \subset R^n$ with $x^k \geqslant x^{k+1}$, $k = 0, 1,\ldots$, and $\lim_{k\to\infty} x^k = -\infty$ implies that $\lim_{k\to\infty} \hat{F}x^k = -\infty$. Therefore, $\hat{F}$ is order-coercive and hence a surjective M-function.

For a mapping $F: R^n \to R^n$ representing a flow on a given network, the problem of finding the state vector, for which the efflux from each node equals a prescribed value, is frequently modified by the assumption that the flow satisfies certain additional conditions at a specified set of "boundary nodes". Following Birkhoff and Kellogg (1966), we consider here the case when the state x_i at a boundary node is a given function of the efflux from that node. This appears to cover all practically important boundary conditions.

Since the set of boundary nodes will always be fixed, it is no restriction to assume that these nodes are numbered $m+1, m+2,\ldots, n$. As an abbreviation, we denote by $\mathscr{A}_k$ the class of all continuous, antitone functions

$$S: R^1 \to R^k, \qquad \sigma_i(t), \qquad i = 1,\ldots, k. \tag{5.10}$$

DEFINITION 5.3. Let $F: R^n \to R^n$ and $N_b = \{m+1,\ldots, n\}$, $m < n$, be a set of boundary nodes in Ω_F.

(a) For any $z \in R^m$ and $S \in \mathscr{A}_{n-m}$ denote the system of equations

$$\begin{aligned} f_i(x) &= z_i, & i &= 1,\ldots, m \\ x_i &= \sigma_{i-m}(f_i(x)), & i &= m+1,\ldots, n \end{aligned} \tag{5.11}$$

by $\{z, S\}$. The class $B(F, N_b)$ of all these systems $\{z, S\}$ is the *boundary value problem* for F on N_b.

(b) The boundary value problem $B(F, N_b)$ is *solvable* (or *uniquely solvable*) if for any $\{z, S\} \in B(F, N_b)$ there exists a vector $x = \mathrm{sol}(z, S) \in R^n$ (or a unique vector $x = \mathrm{sol}(z, S) \in R^n$) which solves the system (5.11).

(c) The boundary value problem $B(F, N_b)$ is *inverse isotone* if for any $\{z', S'\}, \{z'', S''\} \in B(F, N_b)$ it follows from $z' \leqslant z''$ and $S'(t) \leqslant S''(t)$, $t \in R^1$, that $x \leqslant y$ for any $x = \mathrm{sol}(z', S')$ and $y = \mathrm{sol}(z'', S'')$.

The connection between the inverse isotonicity of boundary value problems and that of mappings on R^n is given by the following result:

THEOREM 5.4. *For given $F: R^n \to R^n$ and $N_b = \{m+1,\ldots, n\}$, $m < n$, the boundary value problem $B(F, N_b)$ is inverse isotone if and only if the mapping*

$$F^S: R^n \to R^n, \qquad f_i^S(x) = \begin{cases} f_i(x), & i = 1,\ldots, m \\ x_i - \sigma_{i-m}(f_i(x)), & i = m+1,\ldots, n \end{cases} \tag{5.12}$$

is inverse isotone for any $S \in \mathscr{A}_{n-m}$. *Moreover,* $B(F, N_b)$ *is solvable* (*or uniquely solvable*) *if and only if* F^S *is surjective* (*or bijective*) *for any* $S \in \mathscr{A}_{n-m}$.

Proof. Let $B(F, N_b)$ be inverse isotone and $u = F^S y \geqslant F^S x = v$ for some $S \in \mathscr{A}_{n-m}$. Then $\{z', S'\}$, $\{z'', S''\} \in B(F, N_b)$ for $z' = (v_1, ..., v_m)^T$, $z'' = (u_1, ..., u_m)^T$ and

$$S', S'' : R^1 \to R^{n-m}, \qquad \sigma_i'(t) = \sigma_i(t) + v_{m+i},$$
$$\sigma_i''(t) = \sigma_i(t) + u_{m+i}, \qquad i = 1, ..., n - m.$$

Since evidently $z'' \geqslant z'$ and $S''(t) \geqslant S'(t)$, $t \in R^1$, as well as $x = \mathrm{sol}(z', S')$ and $y = \mathrm{sol}(z'', S'')$, it follows that $y \geqslant x$ and therefore that F^S is inverse isotone.

Conversely, suppose that F^S is inverse isotone for any $S \in \mathscr{A}_{n-m}$. If $\{z', S'\}, \{z'', S''\} \in B(F, N_b)$ satisfy $z'' \geqslant z'$ and $S''(t) \geqslant S'(t)$, $t \in R^1$, then for any $x = \mathrm{sol}(z', S')$ and $y = \mathrm{sol}(z'', S'')$ it follows that

$$f_i(y) = z_i'' \geqslant z_i' = f_i(x), \qquad i = 1, ..., m$$
$$y_i - \sigma_{i-m}'(f_i(y)) \geqslant y_i - \sigma_{i-m}''(f_i(y)) = 0 = x_i - \sigma_{i-m}'(f_i(x)),$$
$$i = m + 1, ..., n,$$

and hence that $F^{S'} y \geqslant F^{S'} x$. This implies that $y \geqslant x$ and therefore that $B(F, N_b)$ is inverse isotone.

The proof of the second statement is a direct consequence of the fact that, for any $S \in \mathscr{A}_{n-m}$ and $z \in R^n$, a vector $x \in R^n$ is a solution of $F^S x = z$ if and only if $x = \mathrm{sol}(z', S')$ with $z' = (z_1, ..., z_m)^T$ and

$$S' : R^1 \to R^{n-m}, \qquad \sigma_i'(t) = \sigma_i(t) + z_{i+m}, \qquad i = 1, ..., n - m.$$

As a consequence of this theorem, all our results about inverse isotone mappings can be carried over to boundary value problems. For instance, it follows from Lemma 2.3 that a solvable and inverse isotone boundary value problem is necessarily uniquely solvable. Note that when $F : R^n \to R^n$ is off-diagonally antitone, then for any $S \in \mathscr{A}_{n-m}$ also the mapping F^S of (5.12) is off-diagonally antitone. Hence in this case, $B(F, N_b)$ is inverse isotone if and only if F^S is an M-function for any $S \in \mathscr{A}_{n-m}$. This permits, for example, the following important extension of Theorem 3.3 to boundary value problems.

THEOREM 5.5. *Let* $F : R^n \to R^n$ *be continuous and off-diagonally antitone, and suppose that for some set* N_b *of boundary nodes the boundary value problem* $B(F, N_b)$ *is solvable and inverse isotone. Then both the Gauss–Seidel process* (3.1) *and the Jacobi process* (3.2), *applied to any system* $\{z, S\} \in B(F, N_b)$, *converge to the unique solution* $x = \mathrm{sol}(z, S)$ *for any starting vector in* R^n *and any sequence* $\{\omega_k\} \subset [\epsilon, 1]$, $\epsilon > 0$, *of relaxation factors.*

The proof is immediate since F^S is a continuous, surjective M-function and the system $\{z, S\}$ is equivalent with the system $F^S x = z$ with

$$z = (z_1, ..., z_m, 0, ..., 0)^T.$$

For the practical application of this global convergence theorem, we need to know when a boundary value problem is solvable and inverse isotone. Certainly, Theorems 4.6 and 5.2, applied to the maps F^S, will provide answers to this question. However, more useful is the following corollary of Theorems 5.1 and 5.2 which uses only assumptions about the mapping F itself and not about S or F^S. For the sake of simplicity, this corollary has only be phrased for $H = I$. It should be evident that other test-functions satisfying the conditions of Theorems 5.1 or 5.2 can also be used.

THEOREM 5.6. *Let* $F : R^n \to R^n$ *be continuous and off-diagonally antitone, and* $N_b = \{m + 1, ..., n\}$, $m < n$, *some set of boundary nodes. Suppose further that*

$$Q : R^1 \to R^n, \qquad q_i(t) = \sum_{j=1}^{n} f_j(x + te^i), \qquad i = 1, ..., n, \tag{5.13}$$

is isotone. Finally, assume that for every node i *of* Ω_F *there exists a boundary node* $l = l(i) \in N_b$ *which is strictly connected to* i. *Then the boundary value problem* $B(F, N_b)$ *is inverse isotone. If for each* $i \in N$ *there is even a surjective path from some* $l = l(i) \in N_b$ *to* i, *then* $B(F, N_b)$ *is also solvable, and hence uniquely solvable.*

Proof. By Theorem 5.4 it suffices to show that for any $S \in \mathscr{A}_{n-m}$ the mapping F^S of (5.12) is an M-function and that F^S is surjective if the paths from $l(i)$ to i are surjective. Let $S \in \mathscr{A}_{n-m}$ be given, then the functions $\phi, \psi : R^1 \to R^n$ with the components

$$\varphi_i(t) = \begin{cases} 0, & i = 1, ..., m \\ t, & i = m + 1, ..., n \end{cases}; \qquad \psi_i(t) = \begin{cases} t, & i = 1, ..., m \\ -\sigma_{i-m}(t), & i = m + 1, ..., n \end{cases}$$

are clearly isotone and continuous, and each component of $\phi + \psi$ is strictly isotone and surjective. Moreover, for $i \in N_b$, φ_i is strictly isotone and surjective. Thus all conditions of Theorem 5.1 are satisfied with $H = I$, and

$$\hat{F} : R^n \to R^n, \qquad \hat{f}_i(x) = \varphi_i(x_i) + \psi_i(f_i(x)) = f_i{}^S(x), \qquad i = 1, ..., n$$

is an M-function, which, by Theorem 5.2, is surjective if the path from $l(i)$ to i is always surjective.

Note that under the conditions of this theorem the function F itself need not be an M-function.

These theorems about boundary value problems extend the results of Birkhoff and Kellogg (1966) and Porsching (1968) to a considerably wider class of mappings. The network functions $F : R^n \to R^n$ considered by these authors are defined as follows: For each link $(i, j) \in \Lambda$ of some connected network $\Omega = (N, \Lambda)$ let a "conductance" function $\varphi_{ij} : R^1 \times R^1 \to R^1$ be given which is strictly isotone in its first variable and strictly antitone in its second. For ease of notation set $\varphi_{ij} \equiv 0$ if $(i, j) \notin \Lambda$. Then

$$F : R^n \to R^n, \qquad f_i(x) = \sum_{j=1}^{n} \varphi_{ij}(x_i, x_j), \qquad i = 1,..., n$$

is the desired network function. Clearly, F is off-diagonally antitone and the associated network Ω_F of F is (in the graph-theoretical sense) isomorphic with Ω. Moreover, every link in Ω_F is strict. Birkhoff and Kellogg (1966) assume now that for any i, j the function

$$\varphi_{ij}(s, t) + \varphi_{ji}(t, s) \tag{5.14}$$

is isotone in s and t. Since, in this case,

$$\sum_{j=1}^{n} f_j(x) = \sum_{i<j} [\varphi_{ij}(x_i, x_j) + \varphi_{ji}(x_j, x_i)],$$

it is evident that then the mapping Q of (5.13) is isotone. Hence, because of the assumed connectedness of Ω, Theorem 5.6 shows that the boundary value problem $B(F, N_b)$ is inverse isotone for any set N_b of boundary nodes. If for each $(i, j) \in \Lambda$ the function φ_{ij} is supposed to be continuous and surjective separately in its first and its second variable, then Ω_F is surjectively connected and, by Theorem 5.6, $B(F, N_b)$ is also solvable.

References

L. Bers (1953), On mildly nonlinear partial difference equations of elliptic type, *J. Res. Nat. Bur. Stand.* **51** (1953), 229–236.

G. Birkhoff and B. Kellogg (1966), Solution of equilibrium equations in thermal networks, Proc. Symp. on Generalized Networks, Brooklyn Polytechnic Press, 1966.

L. Collatz (1952), Aufgaben monotoner Art, *Arch. Math.* **3** (1953), 366–376.

L. Collatz (1964), "Funktionalanalysis und Numerische Mathematik," Springer–Verlag, Berlin/New York, 1964.

R. Duffin (1948), Nonlinear Networks IIb, *Bull. Amer. Math. Soc.* **54** (1948), 119–127.

Ky Fan (1958), Topological proofs for certain theorems on matrices with non-negative elements, *Monatsh. Math.* **62** (1958), 219–232.

J. ORTEGA (1969), Notes on monotone convergence, unpublished personal note.
J. ORTEGA AND W. RHEINBOLDT (1967), On a class of approximate iterative processes, *Arch. Rational Mech. Anal.* **23** (1967), 352–365.
J. ORTEGA AND W. RHEINBOLDT (1968), Local and global convergence of generalized linear iterations, Proc. 1968 SIAM Symp. on Num. Sol. of Nonlinear Problems, SIAM Publ., Philadelphia, Pa., 1970.
T. PORSCHING (1968), Jacobi and Gauss–Seidel methods for nonlinear network problems, Bettis Atomic Power Lab., Pittsburgh, Pa., Techn. Report WAPD-T-2193, September 1968, and *SIAM J. Numer. Anal.*, **6** (1969), 437-449.
S. SCHECHTER (1962), Iteration methods for nonlinear problems, *Trans. Amer. Math. Soc.* **104** (1962), 179–189.
J. SCHRÖDER (1961), Lineare Operatoren mit positiver Inversen, *Arch. Rational Mech. Anal.* **8** (1961), 408–434.
J. SCHRÖDER (1962), Invers-Monotone Operatoren, *Arch. Rational Mech. Anal.* **10** (1962), 323–337.
J. SCHRÖDER (1966), Operator Ungleichungen und ihre numerische Anwendung bei Randwertaufgaben, *Numer. Math.* **9** (1966), 149–162.
T. STIELTJES (1886), Sur les racines de l'équation $X_n = 0$, *Acta Math.* **9** (1886), 385–400.
R. VARGA (1962), "Matrix Iterative Analysis," Prentice–Hall, Englewood Cliffs, N. J., 1962.