N71-16008
NASA CR-116133

Technical Report 70-120
NGL-21-002-008 and GJ-231
July 1970

On P- and S-Functions and Related Classes of n-Dimensional Nonlinear Mappings

by

Jorge J. Moré
and
Werner C. Rheinboldt

UNIVERSITY OF MARYLAND
COMPUTER SCIENCE CENTER
COLLEGE PARK, MARYLAND

Technical Report 70-120
NGL-21-002-008 and GJ-231

July 1970

On P- and S-Functions and Related Classes of n-Dimensional Nonlinear Mappings

by

Jorge J. Moré
and
Werner C. Rheinboldt

This research was supported in part by Grant NGL-21-002-008 from the National Aeronautics and Space Administration and by Grant GJ-231 from the National Science Foundation.

Abstract

This paper introduces and analyzes certain classes of mappings on R^n which represent nonlinear generalizations of the P- and S-matrices of Fiedler and Pták, and of several closely related types of matrices. As in the case of the corresponding matrices, these nonlinear P- and S-functions arise frequently in applications.

Basic properties of the different functions and of their inverses and subfunctions are established, and then a number of theorems are proved about the interrelationships between the various mappings. In particular, it is shown that the well-known monotone mappings, as well as the M-functions and certain of the strictly diagonally dominant mappings recently analyzed by Rheinboldt and Moré, respectively, are special cases of the P-functions. In turn, these P-functions and also the inverse isotone mappings are subclasses of the S-functions. In a final section, a series of characterization theorems for the different functions are presented in terms of conditions on their derivatives.

On P- and S-Functions and Related Classes of n-dimensional Nonlinear Mappings[1]

by

J. Moré[2] and W. Rheinboldt[3]

1. Introduction

Consider the problem of solving an n-dimensional equation $Fx = z$, where $F:D \subset R^n \to R^n$ is a linear or nonlinear mapping of the type arising, for instance, as a discrete analog of some elliptic boundary value problem or as an equilibrium flow on a network. In the special case when the underlying problem is linear, the resulting linear mapping usually has a very special form, and--at least in the mentioned two problem areas--the following classes of matrices appear to occur rather frequently: (a) Positive definite and symmetric matrices; (b) strictly or irreducibly diagonally dominant matrices; (c) M-, or Stieltjes-matrices; (d) P- or S-matrices. For definitions see, e.g., Varga [1962] and Fiedler and Pták [1962], [1966].

In the nonlinear case the mapping F often retains certain properties of these matrices in some form, and this suggests the possibility of introducing specific classes of mappings on R^n which represent suitable nonlinear generalizations of the cited types of matrices. A natural nonlinear extension of positive-definiteness is evidently the concept of

---

1) This work was in part supported by the National Science Foundation under Grant GJ-231 and the National Aeronautics and Space Administration under Grant NGL-21-002-008.
2) Department of Computer Science, Cornell University, Ithaca, N.Y. 14850.
3) Computer Science Center, University of Maryland, College Park, Md. 20742.

a monotone mapping introduced--in a more general setting--by Kacurovskii [1960], Minty [1962], and Browder [1963]. Symmetry carries over to the assumption that F is a potential operator, which in turn leads to convex gradient mappings as the nonlinear extension of symmetric, positive (semi) definite matrices. Following an unpublished suggestion of Ortega, Rheinboldt [1969] considered a nonlinear generalization of M-matrices, the so-called M-functions, while Moré [1970] recently defined and studied a class of Ω-diagonally dominant functions on R^n which contains all matrices under (b). At the same time Moré and Rheinboldt were led to classes of mappings $F:D \subset R^n \to R^n$ which represent nonlinear extensions of the matrices in the last of the four categories and which were correspondingly named P- and S-functions. Some relevant results about these latter mappings have already been reported in the cited dissertation of Moré, and a survey --without proofs--was also included in Rheinboldt [1970]. In this article we present now a detailed account of these new classes of P- and S-functions, their properties, and their relationships with the other types of nonlinear mappings mentioned above.

Implicitly, P- and S-functions have already been used in various applications. In particular, in connection with problems in mathematical economics, Gale and Nikaido [1965] and Nikaido [1968] proved various results about differentiable functions on R^n which in our setting turn out to be results about P-functions and their relation to M-functions. Similarly, Karamardian [1968] obtained a theorem about the solvability of certain nonlinear inequalities which in our terminology shows that all continuous P-functions are S-functions. In connection with nonlinear electrical networks, Sandberg and Willson [1969] were led to special examples of P-functions, and there are probably other related results of this type in the literature.

In Section 2 we introduce the nonlinear P- and S-functions and recall at the same time the definitions of the other cited mappings. Then in Section 3 we establish a number of the basic properties of the P-functions and use them to prove several related results. This is followed in Section 4 by a discussion of the interrelationships between the various mappings, and finally Section 5 presents a series of characterization theorems in terms of differentiability conditions.

2. Background Material

In this section we collect the basic definitions of the P- and S-functions as well as of various other related classes of mappings on R^n.

Throughout the entire article $x \leq y$ denotes the component-wise partial ordering on the n-dimensional real linear space R^n of column vectors, and $x < y$ is the associated strong relation $x_i < y_i$, $i = 1,\ldots,n$. The corresponding notation is used on the space $L(R^n)$ of real $n \times n$ matrices. A rectangle Q in R^n is the Cartesian product of n intervals on the real line, each of which may be either open, closed, or semi-open; in particular, any of these intervals may be unbounded, and thus a rectangle may be all of R^n. The index set $\{1,\ldots,n\}$ will always be denoted by N, and $e^j \in R^n$, $j \in N$, is the jth unit basis vector, while $e \in R^n$ is defined by $e_i = 1$ for each $i \in N$.

The following terminology appears to have become rather standard; inverse isotonicity was introduced by Collatz [1952] under the name "operator of monotone kind".

<u>Definition 2.1</u>. (a) A mapping $F:D \subset R^n \to R^n$ is isotone (or antitone) on D if $x \leq y$, $x,y \in D$, implies that $Fx \leq Fy$ (or $Fx \geq Fy$), and <u>strictly isotone</u> (or <u>strictly antitone</u>) if, in addition, it follows from $x < y$, $x,y \in D$, that also $Fx < Fy$ (or $Fx > Fy$).
(b) The function $F:D \subset R^n \to R^n$ is inverse isotone on D if $Fx \leq Fy$, $x,y \in D$, implies that $x \leq y$.

Clearly then, an affine mapping $Fx = Ax + b$ is isotone exactly if $A \geq 0$ and inverse isotone if and only if A is nonsingular and $A^{-1} \geq 0$.

Most of the standard discretizations of Laplace's equation give rise to affine mappings $Fx = Ax + b$ which are inverse isotone. In this case, A often has non-positive off-diagonal elements as well, which means that A is an M-matrix. For the generalization of the notion of an M-matrix to nonlinear mappings, conditions about the dependence of the component functions upon the individual variables are needed. Such conditions are used in the definition of the following classes of mappings.

<u>Definition 2.2</u>. Consider a mapping $F:D \subset R^n \to R^n$ with the components f_i, $i \in N$.
(a) F is <u>diagonal</u> if for each $i \in N$, f_i depends only on the ith variable x_i.
(b) F is <u>off-diagonally</u> antitone if for any $x \in R^n$ and any $i \neq j$, $i,j \in N$, the functions

(2.1) $\psi_{ij}:\{t \in R^1 \mid x + te^j \in D\} \to R^1,\ \psi_{ij}(t) = f_i(x+te^j)$

are antitone.

(c) F is <u>diagonally</u> <u>(strictly)</u> <u>isotone</u> if for any $x \in R^n$ the functions $\psi_{11},\dots,\psi_{nn}$ defined by (2.1) are (strictly) isotone.

For an affine mapping $Fx = Ax + b$, off-diagonal antitonicity is equivalent with the condition $a_{ij} \leq 0$, $i \neq j$, $i,j \in N$, while F is diagonally (strictly) isotone if and only if $a_{ii} \geq 0$ ($a_{ii} > 0$) for each $i \in N$. The diagonal mappings evidently represent a nonlinear version of the diagonal matrices.

In line with this, the nonlinear generalization of the M-matrices can now be formulated as follows:

<u>Definition 2.3.</u> A mapping $F:D \subset R^n \to R^n$ is an <u>M-function</u> on D if it is off-diagonally antitone and inverse isotone.

Clearly then, an affine mapping $Fx = As + b$ is an M-function if and only if A is an M-matrix. Several nonlinear examples of M-functions are given by Rheinboldt [1969]; they related to discrete analogs of boundary value problems of the type considered, for instance, by Bers [1953], and to equilibrium flows on networks as studied by Birkhoff and Kellogg [1966].

As mentioned in the introduction, a natural nonlinear generalization of positive definiteness is the concept of a monotone mapping introduced by Kacurovskii [1960], Minty [1962], and Browder [1963]:

Definition 2.4. A function $F:D \subset R^n \to R^n$ is monotone on D if

(2.2) $$(x-y)^T(Fx-Fy) \geq 0, \quad x,y \in D,$$

strictly monotone if, in addition, the strict inequality holds in (2.2) whenever $x \neq y$, and uniformly monotone if there is a $c > 0$ such that

$$(x-y)^T(Fx-Fy) \geq c\|x-y\|^2, \quad x,y \in D.$$

In the affine case $Fx = Ax + b$, (2.2) is equivalent with the condition $x^TAx \geq 0$, $x \in R^n$, for the quadratic form of A. This condition evidently implies that for any $x \neq 0$, and $y = Ax$, there is at least one index $k \in N$ such that $x_k \neq 0$ and $x_k y_k \geq 0$, or, in the case of strict monotonicity, that $x_k y_k > 0$.

Fiedler and Pták [1962], [1966] called a matrix $A \in L(R^n)$ a P_0-matrix or P-matrix if it satisfies the first or second of the latter two properties, respectively. The following generalization of these concepts to nonlinear mappings is then immediate:

Definition 2.5. A mapping $F:D \subset R^n \to R^n$ is a P-function (or P_0-function) if for any $x,y \in D$, $x \neq y$, there is an index $k = k(x,y) \in N$ such that $(x_k-y_k)[f_k(x)-f_k(y)] > 0$ (or $(x_k-y_k)[f_k(x)-f_k(y)] \geq 0$ and $x_k \neq y_k$).

The class of P-matrices contains not only the positive-definite-, but also the M-matrices; moreover, any strictly or irreducibly diagonally dominant matrix with non-negative elements is likewise a P-matrix. These facts are either explicitly or implicitly contained in the cited articles of Fiedler and Pták. However, simple examples show that neither the P- nor the P_0-matrices cover the matrices $A \in L(R^n)$ with nonnegative inverses $A^{-1} \geqslant 0$. On the other hand, these matrices evidently satisfy the condition

$$\text{(2.3)} \qquad Au > 0 \text{ for some } u \geqslant 0,\ u \neq 0,$$

and Ky Fan [1958] proved that any $A \in L(R^n)$ with $a_{ij} \leqslant 0$, $i \neq j$, $i,j \in N$, is an M-matrix if and only if (2.3) is satisfied, while Gale and Nikaido [1965] showed that (2.3) also holds for every P-matrix. In line with this, and following earlier work by Stiemke [1915], Fiedler and Pták [1966] called any $A \in L(R^n)$ an S-matrix if it has the property (2.3) and an S_0-matrix if $Au \geqslant 0$ for some $u \geqslant 0$, $u \neq 0$. These concepts lend themselves readily to the following nonlinear generalization:

Definition 2.6. A mapping $F:D \subset R^n \to R^n$ is an S-function (or S_0-function) on D if for any $x \in D$ there is a $y \in D$ such that $y \geqslant x$, $y \neq x$, and $Fy > Fx$ (or $Fy \geqslant Fx$).

In the subsequent sections we shall investigate the relationship between the various classes of functions introduced here, and we will also show how most of the properties of the P- and S-matrices generalize to the corresponding nonlinear mappings. For details about these matrices themselves see in all cases Fiedler and Pták [1962], [1966].

We conclude this section with the nonlinear generalization of strictly diagonally dominant matrices due to Moré [1970], and we refer to this dissertation for the more general concept of Ω-diagonal dominance.

Definition 2.7. A mapping $F:D \subset R^n \to R^n$ is strictly diagonally dominant if for any $x,y \in D$, $x \neq y$, it follows from $f_k(x) = f_k(y)$ that $|x_k-y_k| < \|x-y\|_\infty$.

Moré [1970] shows that--as in all previous cases--this nonlinear definition covers the corresponding linear concept; in other words, that $Fx = Ax + b$ is strictly diagonally dominant if and only if A is a strictly diagonally dominant matrix. The mentioned class of Ω-diagonally dominant functions includes in the linear case the irreducibly diagonal dominant as well as a somewhat more general related family of matrices.

3. Basic Properties of P_0- and P-functions and Related Results

In this section we establish a number of the basic properties of the P_0- and P-functions and use them to prove several related results. The material included here already shows that, in some form, many of the characteristics of P-matrices are indeed inherited by the P-functions.

We begin with two observations about inverse mappings which follow almost directly from the definitions; the proofs are therefore omitted.

Theorem 3.1. (a) The mapping $F:D \subset R^n \to R^n$ is inverse isotone if and only if it is injective and $F^{-1}:FD \subset R^n \to R^n$ is isotone.

(b) If $F:D \subset R^n \to R^n$ is either strictly monotone or a P-function, then it is injective and $F^{-1}:FD \subset R^n \to R^n$ is again strictly monotone or a P-function, respectively.

Note that part (b) ensures every M-function to be injective. Observe also that S-functions do not necessarily have this property. In fact, the class of S-functions is rather large; for example, any $f:(a,b) \subset R^1 \to R^1$ with $f(x) < \limsup_{z \to \beta-} f(z)$ for all $x \in (a,b)$ is already an S-function on (a,b).

A simple modification of Theorem 3.1(b) states that for any injective, monotone mapping $F:D \quad R^n \to R^n$ also F^{-1} is a monotone function on FD. The corresponding result for P_0-functions is undecided. We conjecture that for any continuous, injective P_0-function $F:Q \subset R^n \to R^n$ on an open rectangle Q also F^{-1} is a P_0-function on FQ. This result certainly holds for the linear case and in Theorem 5.12 below we will prove that it is also correct for F-differentiable P_0-functions. The following two theorems are closely related to this conjecture.

Theorem 3.2. The mapping $F:D \subset R^n \to R^n$ is injective and $F^{-1}:FD \subset R^n \to R^n$ is a P_0-function if and only if $F + \Delta$ is injective for any diagonal matrix $\Delta = \mathrm{diag}(d_1,\ldots,d_n)$, $d_i \geq 0$, $i \in N$.

Proof. Clearly, F is injective and F^{-1} is a P_0-function if and only if for each $x \neq y$ in D there is an index $k \in N$ such that $f_k(x) \neq f_k(y)$ and $(x_k-y_k)[f_k(x)-f_k(y)] \geq 0$.

Suppose now that F^{-1} is a P_0-function but that $Fx + \Delta x = Fy + \Delta y$ for some $x \neq y$ in D and some diagonal matrix $\Delta = \mathrm{diag}(d_1,\ldots,d_n)$, $d_i \geq 0$, $i = 1,\ldots,n$. Then $f_i(x)-f_i(y) = d_i(y_i-x_i)$ for each $i \in N$ and hence $(x_i-y_i)[f_i(x)-f_i(y)] = -d_i(x_i-y_i)^2 < 0$ whenever $f_i(x) \neq f_i(y)$. This contradicts the fact that F^{-1} is a P_0-function.

Conversely, if for some $x \neq y$ in D we have $(x_i-y_i)[f_i(x)-f_i(y)] < 0$ whenever $f_i(x) \neq f_i(y)$, $i \in N$, then $Fx + \Delta x = Fx + \Delta y$ for $\Delta = \mathrm{diag}(d_1,\ldots,d_n)$ with

$$d_i = \begin{cases} -\dfrac{f_i(x)-f_i(y)}{x_i-y_i} & , \text{ if } f_i(x) \neq f_i(y) \\ 0 & , \text{ otherwise.} \end{cases}$$

This completes the proof.

Theorem 3.2 raises the question whether F itself is a P_0-function if $F + \Delta$ is injective for any diagonal matrix Δ with positive diagonal elements. In the linear case this is correct, but for the nonlinear case again only some partial answers are known. We return to this in Theorem 3.7.

<u>Theorem 3.3</u>. If $F:D \subset R^n \to R^n$ is a P_0-function (or P-function), then $F + \Phi$ is a P-function for any strictly isotone (or isotone), diagonal mapping $\Phi:D \subset R^n \to R^n$. Conversely, if for any $\varepsilon > 0$, $F + \varepsilon I$ is a P-function, where $I \in L(R^n)$ is the identity, then F is a P_0-function.

<u>Proof</u>. The first part is a direct consequence of the definitions. For the proof of the second part, suppose that F is not a P_0-function. Then there are $x \neq y$ in D such that $(x_i-y_i)[f_i(x)-f_i(y)] < 0$ whenever $x_i \neq y_i$, $i \in N$. Therefore,

$$\varepsilon = \min \{- \frac{f_i(x)-f_i(y)}{x_i-y_i} \;\Big|\; x_i \neq y_i,\ i \in N\} > 0,$$

and $(x_i-y_i)[f_i(x)+\varepsilon x_i-f_i(y)-\varepsilon y_i] \leq 0$ contradicts the fact that $F + \varepsilon I$ is a P-function.

Recall again the conjecture stated before Theorem 3.2 If $F:D \subset R^n \to R^n$ is an injective P_0-function, then by Theorem 3.3, $F + \Delta$ is a P-function and hence injective for any $\Delta = \mathrm{diag}(d_1,\ldots,d_n) \in L(R^n)$ with a strictly positive diagonal. This is not sufficient, however, to conclude from Theorem 3.2 that F^{-1} is again a P_0-function since that theorem requires $F + \Delta$ to be injective for any Δ with a nonnegative diagonal. It appears that the conjecture cannot be settled in this way.

In Theorem 3.3 we considered additive transformations of P-functions. In continuation of this, the following result concerns the composition of these functions with diagonal mappings. For simplicity we assume here that the domains of definitions are all of R^n; it should be self-evident how this generalizes to other domains.

Theorem 3.4. Let $F:R^n \to R^n$ be a P-function (or P_0-function) and $\Phi:R^n \to R^n$ a diagonal mapping.

(a) If Φ is strictly isotone, then $F\circ\Phi$ as well as $\Phi\circ F$ are again P-functions (or P_0-functions).

(b) If each component ϕ_i, $i \in N$, of Φ is either strictly isotone or strictly antitone, then $\Phi\circ F\circ\Phi$ is again a P-function (or P_0-function).

Proof. The proofs are essentially the same for all cases; we therefore consider here only (b) under the assumption that F is a P-function. If $x \neq y$, then also $\Phi x \neq \Phi y$ and hence $[\phi_k(x_k)-\phi_k(y_k)][f_k(\Phi x)-f_k(\Phi y)] > 0$ for some $k \in N$. If ϕ_k is strictly isotone and, say, $x_k > y_k$, then

$\phi_k(x_k) > \phi_k(y_k)$ and hence $f_k(\Phi x) > f_k(\Phi y)$, as well as $\phi_k(f_k(\Phi x)) > \phi_k(f_k(\Phi y))$. Thus,

$$(x_k - y_k)[\phi_k(f_k(\Phi x)) - \phi_k(f_k(\Phi y))] > 0,$$

and the same result holds in the other cases.

For P_0- and P-matrices it follows immediately from the definitions that also any principal submatrix belongs to the same class. For M-matrices the same result appears to be due to Ky Fan [1958]. The following subfunction-concept represents the nonlinear generalization of a submatrix.

Definition 3.5. Consider $F:D \subset R^n \to R^n$ and a non-empty set $M = \{i_1,\ldots,i_m\} \subset N$. For fixed constants c_j, $j \notin M$, define

$$\pi : R^m \to R^n, \ \pi(y) = \sum_{j=1}^{m} y_j e^{i_j} + \sum_{j \notin M} c_j e^j$$

where e^j are again the basis vectors of R^n. Then the subfunction G of F, corresponding to M and $\{c_j\}$, is defined on $D_G = \{y \in R^m \mid \pi(y) \subset D\}$ by $G:D_G \subset R^m \to R^m$, $g_j(y) = f_{i_j}(\pi(y))$, $j = 1,\ldots,m$.

Although a subfunction depends on the index set M as well as on the constants c_j, we chose not to burden the notation with an indication of this dependence. In all cases the set M and the c_j should be self-evident from the context.

As in the linear case, the next result follows rather directly from the definitions; no proof is therefore given.

Theorem 3.6. For any P_0- and P-function $F: D \subset R^n \to R^n$ also any subfunction belongs to the same function class.

As an application, we prove the following partial answer to the question raised after Theorem 3.2.

Theorem 3.7. Let $F: D \subset R^n \to R^n$ be such that for any $\Delta = \mathrm{diag}(d_1, \ldots, d_n) \in L(R^n)$ with $d_i > 0$, $i \in N$, the mapping $F + \Delta$ and all its subfunctions are injective. Then F is a P_0-function.

Proof. If F is not a P_0-function, there exist $x \neq y$ in D such that

$$(3.1) \qquad (x_i - y_i)[f_i(x) - f_i(y)] < 0 \text{ whenever } x_i \neq y_i,\ i \in N.$$

Let $M = \{i \in N \mid x_i \neq y_i\}$ and for ease of notation assume that $M = \{1, \ldots, m\} \subset N$. Consider the subfunction G of F with the components

$$(3.2) \qquad g_i(t_1, \ldots, t_m) = f_i(t_1, \ldots, t_m, y_{m+1}, \ldots, y_n),\ i = 1, \ldots, m.$$

Then for $\Delta = \mathrm{diag}(d_1, \ldots, d_n)$ with

$$d_i = \begin{cases} -\dfrac{g_i(x_1, \ldots, x_m) - g_i(y_1, \ldots, y_m)}{x_i - y_i}, & i = 1, \ldots, m \\ 1, & i = m+1, \ldots, n, \end{cases}$$

we have

$$g_i(x_1, \ldots, x_m) + d_i x_i = g_i(y_1, \ldots, y_m) + d_i y_i,\ i = 1, \ldots, m,$$

which contradicts the assumption that $G + \Delta$ is injective.

We turn now to the question whether Theorem 3.6 is also valid for M-functions. For continuous surjective M-functions $F:R^n \to R^n$, this was shown by Rheinboldt [1969]. In order to prove the same result for arbitrary M-functions, we need first a theorem about the relation between M- and P-functions.

Theorem 3.8. Any off-diagonally antitone P-function $F:Q \subset R^n \to R^n$ on a rectangle Q is an M-function.

Proof. Let $Fx \leq Fy$ for some $x,y \in Q$ and assume that $M = \{i \in N \mid x_i > y_i\}$ is not empty. For ease of notation let $M = \{1,\ldots,m\} \subset N$, and consider the subfunction G of F with the components (3.2). Since F is off-diagonally antitone, we obtain by definition of M that

$$g_i(y_1,\ldots,y_m) = f_i(y) \geq f_i(x) \geq g_i(x_1,\ldots,x_m),\ i = 1,\ldots,m,$$

and hence that

$$(x_i-y_i)[g_i(y_1,\ldots,y_m)-g_i(x_1,\ldots,x_m)] \leq 0,\ i = 1,\ldots,m.$$

This contradicts the fact that, by Theorem 3.6, G is a P-function.

Now we can extend Theorem 3.6 also to M-functions.

Theorem 3.9. Let $F:Q \subset R^n \to R^n$ be an M-function on the rectangle Q. Then also any subfunction of F is again an M-function.

Proof. Suppose that there is a subfunction $G:D_G \subset R^m \to R^n$ of F corresponding to an index set M and the constants c_j, $j \notin M$, which is not an M-function.

For ease of notation we assume once more that $M = \{1,\ldots,m\}$. Since D_G is again a rectangle, Theorem 3.8 implies that G is not a P-function, and hence that, for some $u \neq v$, $u,v \in D_G \subset R^m$,

$$(u_j-v_j)(g_j(u)-g_j(v)) \leq 0, \; j = 1,\ldots,m.$$

Since $u \neq v$, a possible interchange ensures that $M_0 = \{i \in M \mid u_i < v_i\}$ is not empty, and for notational simplicity let $M_0 = \{1,\ldots,k\}$, $1 \leq k \leq m$. Then, for $i = 1,\ldots,k$,

$$\begin{aligned} f_i(v_1,\ldots,v_m,c_{m+1},\ldots,c_n) &= g_i(v) \leq g_i(u) \\ &= f_i(u_1,\ldots,u_m,c_{m+1},\ldots,c_n) \\ &\leq f_i(u_1,\ldots,u_k,v_{k+1},\ldots,v_m,c_{m+1},\ldots,c_n) \end{aligned}$$

while, for $i = k+1,\ldots,n$, by the off-diagonal antitonicity,

$$f_i(v_1,\ldots,v_m,c_{m+1},\ldots,c_n) \leq f_i(u_1,\ldots,u_k,v_{k+1},\ldots,v_m,c_{m+1},\ldots,c_n).$$

But now it follows from the inverse isotonicity of F that $v_j \leq u_j$, $j = 1,\ldots,k$, in contradiction to the construction of M_0.

These theorems on subfunctions are especially helpful in proofs about P- or M-functions which use induction with respect to the dimension of the space. For example, Theorem 3.9 provides a tool for the characterization of surjectivity for continuous M-functions (see Rheinboldt [1969]). As a different illustration, we give here a simple proof of the finite-dimensional version of a well-known result of Minty [1962] which states that a continuous, uniformly monotone function on a Hilbert space is a homeomorphism. In fact, we will prove the following stronger result for the n-dimensional case.

Theorem 3.10 (Moré [1970]). Let $F:R^n \to R^n$ be a continuous and uniform P-function on R^n in the sense that there exists a $c > 0$ such that for any $x \neq y$ in R^n there is an index $k = k(x,y) \in N$ with

$$(3.3) \qquad (x_k-y_k)[f_k(x)-f_k(y)] \geq c\|x-y\|^2.$$

Then F is a homeomorphism of R^n onto itself.

Proof. Since all norms on R^n are equivalent, we may assume that (3.3) holds for the infinity norm. Then (3.3) implies that $\|Fx-Fy\|_\infty \geq c\|x-y\|_\infty$ and hence that F is injective and $\|F^{-1}x-F^{-1}y\|_\infty \leq c^{-1}\|x-y\|_\infty$ for all x,y in FR^n. Thus only the surjectivity of F needs to be proved.

For $n = 1$, surjectivity is a direct consequence of (3.3); assume, therefore, that the result is valid for some $n \geq 1$ and let $F:R^{n+1} \to R^{n+1}$ satisfy the conditions of the theorem. Then for any given $t \in R^1$ the mapping

$$G(\cdot,t):R^n \to R^n;\ g_i(x_1,\ldots,x_n,t) = f_i(x_1,\ldots,x_n,t),\ i \in N,$$

is a uniform P-function on R^n and hence surjective by induction hypothesis. This implies that for any fixed $z \in R^{n+1}$ a mapping $H:R^1 \to R^n$ is well-defined by the relations

(3.4) $\quad f_i(h_1(t),\ldots,h_n(t),t) = z_i = g_i(h_1(t),\ldots,h_n(t),t),\ i \in N;$

moreover, H is continuous. In fact, for $s \neq t$ it follows from (3.3) that

$$[h_k(s)-h_k(t)][g_k(H(s),t)-g_k(H(t),t)] \geq c\|H(s)-H(t)\|_\infty^2$$

and hence, together with (3.4), that

$$\|G(H(s),t)-G(H(s),s)\|_\infty = \|G(H(s),t)-G(H(t),t)\|_\infty$$

$$\geq c\|H(s)-H(t)\|_\infty;$$

therefore, the continuity of H is a consequence of that of F. Then also the function

$$\psi : R^1 \to R^1,\ \psi(t) = f_{n+1}(h_1(t),\ldots,h_n(t),t)$$

is continuous, and for $s \neq t$ we obtain from (3.3) and (3.4) that

$$(s-t)[\psi(s)-\psi(t)] \geq c|s-t|^2.$$

This shows that $\lim_{t\to+\infty} \psi(t) = +\infty$, and $\lim_{t\to\infty} \psi(t) = -\infty$, and, thus, that ψ is surjective. Now, there exists a $t^* \in R^1$ with $\psi(t^*) = z_{n+1}$, and this together with (3.4) shows that $Fx^* = z$ for $x^* = (h_1(t^*),\ldots,h_n(t^*),t^*)^T$. Since z was arbitrary, this completes the proof.

Clearly, this result contains as a direct corollary the cited theorem of Minty [1962].

Corollary 3.11. A continuous, uniformly monotone mapping $F:R^n \to R^n$ is a homeomorphism from R^n onto itself.

The following lemma shows that all linear P-functions satisfy the uniformity conditon (3.3).

Lemma 3.12. Let $A \in L(R^n)$ be a P-matrix. Then there exists a $c > 0$ such that for any $x \neq 0$ there is an index $k = k(x) \in N$ with

$$x_k y_k \geq c\|x\|_\infty^2,$$

where $y = Ax$.

The proof follows directly from the fact that the functional $g:R^n \to R^1$, $g(x) = \max_{j \in N} (x_j, y_j)$ is continuous and positive on the unit sphere.

This leads directly to the following result of Sandberg and Willson [1969].

Corollary 3.14. Let $A \in L(R^n)$ be a P-matrix, and $\Phi:R^n \to R^n$ a continuous, diagonal, and isotone function on R^n. Then $F = A + \Phi$ is a homeomorphism from R^n onto itself.

In fact, by Theorem 3.3, F is a P-function and from Lemma 3.12 it follows readily that F is also uniform. Thus Theorem 3.10 applies.

For a nonlinear generalization of this result, see Rheinboldt [1970].

4. Relationships between the Function Classes

In this section we will prove a number of results about the relationships between the various function classes introduced in Section 2. For clarity we have indicated the general structure of these relations in the form of a diagram (Figure 1). This diagram is meant to be only illustrative, and the number at each implication arrow gives the corresponding theorem that specifies the precise conditions under which the implication holds. Several relations were not included in order not to burden the figure. For example, by definition any isotone or strictly isotone function is diagonally isotone, or strictly diagonally isotone, respectively, and Theorem 4.9 below shows that certain P_0-functions are M-functions. Note also that the diagram contains several derived implications, such as, for example, that M-functions are strictly diagonally isotone.

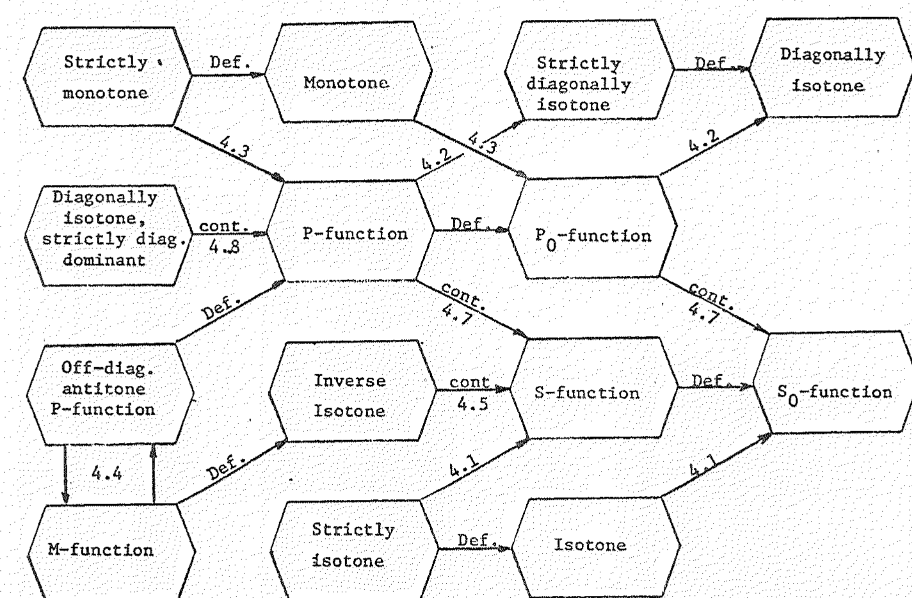

Figure 1

The following three results represent relatively straightforward conclusions from the definitions; their proofs were therefore again omitted.

Theorem 4.1. An isotone (or strictly isotone) mapping $F:D \subset R^n \to R^n$ on the open set D is an S_0- (or S-) function.

Theorem 4.2. Any P_0- (or P-) function $F:D \subset R^n \to R^n$ is diagonally isotone (or strictly diagonally isotone).

Theorem 4.3. Any monotone (or strictly monotone) function $F:D \subset R^n \to R^n$ is a P_0- (or P-) function.

In Theorem 3.8 we obtained already a relation between P- and M-functions. The following theorem extends this to the equivalence statement shown in Figure 1.

Theorem 4.4. The mapping $F:Q \subset R^n \to R^n$ on the rectangle Q is an M-function if and only if F is an off-diagonally antitone P-function.

Proof. Theorem 3.8 gives us the sufficiency of the condition; thus we only need to prove the necessity. Suppose that F is an M-function, but not a P-function. Then there exist $x,y \in D$, $x \neq y$, such that

$$(4.1) \qquad (y_i-x_i)[f_i(y)-f_i(x)] \leq 0, \; i \in N.$$

Since $y \neq x$, we may assume that $M = \{i \in N \mid y_i < x_i\}$ is not empty, and --for ease of notation--that $M = \{1,2,\ldots,m\}$, $m \in N$. Let $G:D_G \subset R^m \to R^m$ be the subfunction of F with the components

$$g_i(t_1,\ldots,t_m) = f_i(t_1,\ldots,t_m,x_{m+1},\ldots,x_n),\ i = 1,\ldots,m.$$

By the off-diagonal antitonicity, (4.1) implies that, for $i = 1,\ldots,m$,

$$\begin{aligned} g_i(x_1,\ldots,x_m) &= f_i(x) \\ &\leq f_i(y) \leq f_i(y_1,\ldots,y_m,x_{m+1},\ldots,x_n) = g_i(x_1,\ldots,x_m). \end{aligned} \tag{4.2}$$

Now, by Theorem 3.7, G is an M-function and therefore inverse isotone. But then (4.2) implies that $x_i \leq y_i$, $i = 1,\ldots,m$, against construction of M.

The next result establishes the connection between the inverse isotone and the S-functions.

Theorem 4.5. If $F:D \subset R^n \to R^n$ is a continuous, inverse isotone mapping on the open set D, then F is an S-function.

Proof. By Theorem 3.1(a), F is injective, and thus, by the Domain Invariance Theorem, FD is an open set. Hence, given $x \in D$ there is a $\delta > 0$ such that $Fx + \delta e \in FD$ and consequently $Fy = Fx + \delta e$ for some $y \in D$. Since $Fy > Fx$, the inverse isotonicity implies that $y \geq x$, $y \neq x$, as desired.

The relations in Figure 1 between P_0-, and S_0-, as well as P- and S-functions are essentially due to Karamardian [1968]. The following lemma is implicitly contained in Karamardian's proof and forms its central part.

Lemma 4.6. Let $G:S \subset R^n \to R^n$ be continuous on the (n-1)-simplex $S = S[\alpha e^1,\ldots,\alpha e^n]$ where $\alpha > 0$. Then there exists a vector $w \in S$ such that

$$(4.3) \qquad w^T G w = \alpha \min_{j \in N} g_j(w).$$

Proof. The proof is based on the following generalization of the Brouwer fixed point theorem by Kakutani [1941]: If $\Phi:C \subset R^n \to K(C)$ maps a non-empty convex, compact set C into the set $K(C)$ of all non-empty closed convex subsets of C, and if Φ is upper-semicontinuous in the sense that $\bigcup_{u \in C} (u,\Phi u) \subset C \times C$ is closed, then $u \in \Phi u$ for some $u \in C$.

For the application define

$$\Phi:S \to K(S), \ \Phi u = \{v \in S \mid v^T G u = \min_{z \in S} z^T G u\}.$$

Clearly, Φu is indeed non-empty, closed, and convex and hence a member of $K(S)$. To show that Φ is upper-semicontinuous, assume that $\{u^k\} \subset S$, $\lim_{k\to\infty} u^k = u$, $v^k \in \Phi u^k$, $k = 0,1,\ldots$, and $\lim_{k\to\infty} v^k = v$. Then we have for all $k \geq 0$ and $z \in S$, $(v^k)^T G u^k \leq z^T G u^k$ and hence, by the continuity of G and the closedness of S, $v^T G u \leq z^T G u$; that is, $v \in \Phi u$.

Thus the Kakutani theorem applies, and there exists a $w \in S$ such that

$$w^T G w = \min_{z \in S} z^T G w = \alpha \min \{ \sum_{i=1}^{n} \lambda_i g_i(w) \mid \sum_{i=1}^{n} \lambda_i = 1, \lambda_j \geq 0, j \in N\} = \alpha \min_{j \in N} g_j(w)$$

which completes the proof.

It may be noted that analogously we can prove the corresponding result with the minimum in (4.3) replaced by the maximum.

With this result we now obtain the desired relations between the P- and S-functions.

Theorem 4.7. A continuous P_0- (or P-) function $F:D \subset R^n \to R^n$ on the open set D is an S_0- (or S-) function.

Proof. Given $x \in D$, let $\alpha > 0$ be such that the n-simplex $S[x, x+\alpha e^1, \ldots, x+\alpha e^n]$ is contained in D. Then Lemma 4.6 can be applied to the mapping $G:S \to R^n$, $Gz = F(x+z)-Fx$, $z \in S = S[\alpha e^1, \ldots, \alpha e^n]$. Hence there exists a $y = x + w$, $w \in S$, such that

$$(4.4) \qquad (y-x)^T(Fy-Fx) = \alpha \min_{j \in N} [f_j(y)-f_j(x)],$$

and clearly $y \geq x$, $y \neq x$. Let $f_k(y) - f_k(x) = \min_{j \in N} [f_j(y)-f_j(x)] = \mu$. If F is a P_0- (or P-) function and $\mu \geq 0$ (or $\mu > 0$) there is nothing to prove. Assume therefore that $\mu < 0$ (or $\mu \leq 0$) and set $\gamma_j = (y_j-x_j)(f_j(y)-f_j(x))$. Then

$$(y_j-x_j)(f_k(y)-f_k(x)) \begin{cases} < \gamma_j & \text{whenever } \gamma_j \geq 0 \text{ and } y_j > x_j \\ & \qquad (\text{or } \gamma_j > 0) \\ \leq \gamma_j & \text{otherwise,} \end{cases}$$

and in both cases there are indices for which the inequality sign applies. Thus we obtain, by summation, $\alpha\mu < (y-x)^T(Fy-Fx)$ which contradicts (4.4) and hence completes the proof.

The last as yet unproved implication in Figure 1 concerns the relation between P-functions and the strictly diagonally dominant mappings introduced in Definition 2.7.

Theorem 4.8 (Moré [1970]). A continuous, diagonally isotone and strictly diagonally dominant mapping $F:Q \subset R^n \to R^n$ on a rectangle Q is a P-function.

Proof. Consider any $x,y \in Q$, $x \neq y$, and let $k \in N$ be such that $|x_k-y_k| = \|x-y\|_\infty$. Then $f_k(x) \neq f_k(y)$ and $x_k \neq y_k$. Suppose now that $(x_k-y_k)[f_k(x)-f_k(y)] < 0$, and, for the sake of definiteness, let $x_k > y_k$; that is, $f_k(x) < f_k(y)$. Define $\Psi:[0,1] \to Q$ by $\psi_k(t) = x_k$ and $\psi_i(t) = ty_i + (1-t)x_i$ for $i \neq k$, $i \in N$. Then $|\psi_k(t)-y_k| = \|\Psi(t)-y\|_\infty$ for $t \in [0,1]$ and, by the strict diagonal dominance, we have $f_k(\Psi(t)) \neq f_k(y)$ for $t \in [0,1]$. Since $f_k(\Psi(0)) < f_k(y)$, it follows by continuity that $f_k(\Psi(t)) < f_k(y)$ for $t \in [0,1]$, and that, in particular,

$$f_k(\Psi(1)) = f_k(y+(x_k-y_k)e^k) < f_k(y)$$

contradicting the diagonal isotonicity of F.

A similar result holds for the mentioned Ω-diagonally dominant functions; we refer to Moré [1970] for the proof.

We conclude this section with a variation of the sufficiency part of Theorem 4.4.

Theorem 4.9. Let $F:Q \subset R^n \to R^n$ be a continuous, off-diagonally antitone, injective P_0-function on the open rectangle Q. Then F is an M-function.

Proof. We first show that $Fy > Fx$ for $x,y \in Q$ implies that $y \geqslant x$. For this observe that for sufficiently small $\varepsilon > 0$, $Fy + \varepsilon y > Fx + \varepsilon x$. By Theorem 3.4, $F + \varepsilon I$ is a P-function and hence by Theorem 4.4 an M-function. The inverse isotonicity then implies that indeed $y \geqslant x$.

In order to prove the inverse isotonicity of F, assume now that $Fy \geqslant Fx$, $x,y \in Q$. By the Domain Invariance Theorem, F is a homeomorphism of Q onto FQ and consequently $u(t) = F^{-1}(Fy+te)$ is a well-defined continuous function for $t \in [0,\varepsilon)$ and some $\varepsilon > 0$. Then $Fu(t) > Fy$ for $0 < t < \varepsilon$ and, by the first part of the proof, $u(t) \geqslant x$ for $t \in (0,\varepsilon)$. Since u is continuous, therefore also $x \leqslant u(0) = y$, and F is inverse isotone.

5. Differentiable P- and S-functions

In this section we present a number of characterization theorems for differentiable P- and S-functions in terms of properties of their derivatives. The first theorem is rather typical for many of the results which follow.

Theorem 5.1. Let $F:D \subset R^n \to R^n$ be G-differentiable on the open set D. If $F'(x)$ is an S-matrix for each x in D, then F is an S-function.

Proof. Given any $x \in D$, there exists a vector $\hat{h} \geqslant 0$ $\hat{h} \neq 0$ such that $F'(x)\hat{h} > 0$. Clearly, for small $\varepsilon > 0$ we then have $h = \hat{h} + \varepsilon e > 0$ as well as $F'(x)h > 0$. Now

$$\lim_{t\to 0+} \frac{1}{t} [F(x+th)-Fx] = F'(x)h > 0$$

shows that $x + th \in D$, $x + th > x$, and $F(x+th) > Fx$ for sufficiently small $t > 0$. Thus F is an S-function.

Corresponding results for P- and M-functions are harder to prove. The following theorem is essentially due to Gale and Nikaido [1965] and Nikaido [1968], but our proof appears to be more direct.

Theorem 5.2. Let $F:Q \subset R^n \to R^n$ be F-differentiable on the rectangle Q. If $F'(x)$ is a P-matrix (or M-matrix) for each x in Q, then F is a P-function (or M-function) on Q.

Proof. Evidently the result for M-functions follows directly from that for P-functions as a consequence of Theorem 4.4. Moreover, the P-function case will have been proved once we have shown that F is a P-function on any closed rectangle $\hat{Q} \subset Q$. For that proof we proceed by induction on n. For $n = 1$ the result is trivial; hence assume that it holds for some $n - 1 \geq 1$ and let $F:Q \subset R^n \to R^n$ satisfy the condition of the theorem and consider any closed rectangle $\hat{Q} \subset Q$.

Let $x \neq y$, $x,y \in \hat{Q}$, be given and suppose that $x_i = y_i$ for some $i \in N$. For simplicity let $i = n$ and consider the subfunction $G:Q_G \subset R^{n-1} \to R^{n-1}$ of F with the components

$$g_i(t_1,\ldots,t_{n-1}) = f_i(t_1,\ldots,t_{n-1},y_n),\ i = 1,\ldots,n-1.$$

Since $G'(t_1,\ldots,t_{n-1})$ is again a P-matrix for each $(t_1,\ldots,t_{n-1})$ in the closed rectangle $\hat{Q}_G \subset Q_G \subset R^{n-1}$, the induction hypothesis implies that G

is a P-function and therefore that, for some $j \in N$, $j \neq n$,

$$(5.1) \qquad (x_j - y_j)[f_j(x) - f_j(y)] > 0.$$

For any fixed $y \in \hat{Q}$ we now show that the set $\hat{Q}_y = \{x \in \hat{Q} \mid Fx \leqslant Fy, x > y\}$ is empty. Suppose that this were not so, and let $\{x^k\} \subset \hat{Q}_y$ be any convergent sequence. Then clearly $\lim_{k\to\infty} x^k = x \geqslant y$, and by the closedness of $\hat{Q}$, $x \in \hat{Q}$ and $Fx \leqslant Fy$. We now have three cases; namely, (a) $x > y$, (b) $x = y$, and (c) $x \neq y$, but $x_i = y_i$ for some $i \in N$. By the first part of the proof, (c) implies that (5.1) holds for some $j \in N$, $j \neq i$, which contradicts the construction of x and y. If (b) holds, then

$$\lim_{k\to\infty} \frac{1}{\|x^k - y\|} [Fx^k - Fy - F'(y)(x^k - y)] = 0,$$

and, since $F'(y)$ is a P-matrix, there exists by Lemma 3.12 a constant $c > 0$ such that some component of $F'(y)(x^k - y)$ exceeds $c\|x^k - y\| > 0$. Hence also some component of $Fx^k - Fy$ is positive for large k, which contradicts $x^k \in \hat{Q}_y$. Thus only case (a) remains and we have $x \in \hat{Q}_y$, that is $\hat{Q}_y$ is closed. In particular, therefore any linearly-ordered subset of $\hat{Q}_y$ has a lower bound in $\hat{Q}_y$ and by Zorn's Lemma there is a minimal element $u \in \hat{Q}_y$, that is $x \in \hat{Q}_y$, $x \leqslant u$. Since $F'(u)$ is by Theorem 5.1 an S-matrix, there is a vector $h < 0$ such that $F'(u)h < 0$. Now

$$\lim_{t\to 0+} \frac{1}{t} [F(u+th) - Fu] = F'(u)h < 0$$

implies that $F(u+th) < Fu \leqslant Fy$ and $y < u + th < u$ for sufficiently small $t > 0$ which contradicts the minimality of u. Hence, also (a) does not apply and $\hat{Q}_y$ is empty.

Suppose now that $x \neq y$ in $\hat{Q}$ are such that $(x_i-y_i)[f_i(x)-f_i(y)] \leq 0$ for $i \in N$. Then $x_i \neq y_i$ for $i \in N$; otherwise the first part of the proof would imply that (5.1) holds for some $j \in N$. Define $\Delta = \mathrm{diag}(d_1,\ldots,d_n)$ by

$$d_i = \begin{cases} +1, & \text{if } x_i > y_i \\ -1, & \text{if } x_i < y_i, \end{cases}$$

and $H:\Delta^{-1}\hat{Q} \subset R^n \to R^n$ by $Hx = \Delta(F(\Delta x))$. Then, for any $x \in \Delta^{-1}\hat{Q}$, $H'(x) = \Delta F'(\Delta x)\Delta$ is by Theorem 3.4(b) again a P-matrix. Moreover, by construction of H, we have $\Delta^{-1}x > \Delta^{-1}y$ as well as $H(\Delta^{-1}x) \leq H(\Delta^{-1}y)$, which is a contradiction to the result of the second part of the proof applied to H.

<u>Corollary 5.3</u>. Let $F:Q \subset R^n \to R^n$ be F-differentiable on the rectangle Q. If $F'(x)$ is a P_0-matrix for each x in Q, then F is a P_0-function on Q.

<u>Proof</u>. For any $\varepsilon > 0$ consider the mapping $F_\varepsilon:Q \subset R^n \to R^n$, $F_\varepsilon x = Fx + \varepsilon x$. By the first part of Theorem 3.4, $F'_\varepsilon(x) = F'(x) + \varepsilon I$ is a P-matrix for each $x \in Q$, and hence, by Theorem 5.2, F_ε is a P-function for any $\varepsilon > 0$. Now the second part of Theorm 3.3 implies that F itself is a P_0-function.

We note here that it is not known whether the last two results also hold for more general domains than rectangles.

The next theorem is included only for the sake of completeness. The proof for the monotone case is well-known while that for strictly diagonally dominant functions is contained in Moré [1970].

Theorem 5.4. Let $F:D \subset R^n \to R^n$ be G-differentiable on the convex set D. If $F'(x)$ is a positive definite (or a strictly diagonally dominant) matrix for each x in D, then F is a strictly monotone (or a strictly diagonally dominant) function on D.

Theorems 5.1 through 5.4 raise the natural question whether F is inverse isotone if $F'(x)^{-1} \geq 0$ for each x in a suitable domain. The answer to this is not known in general, but the next theorem provides a partial answer. For another partial result, see Rheinboldt [1970].

Theorem 5.5. Let $F:D \subset R^n \to R^n$ be convex and G-differentiable on the open, convex set D. Then F is inverse isotone if and only if $F'(x)$ is nonsingular and $F'(x)^{-1} \geq 0$ for each $x \in D$.

Proof. Assume that F is inverse isotone and that $F'(x)h \geq 0$ for some $x \in D$ and $h \in R^n$. Then, by the convexity of F,

$$F(x+h) - Fx \geq F'(x)h \geq 0$$

(see, for example, Ortega and Rheinboldt [1970], p. 448), which by the inverse isotonicity implies that $h \geq 0$. Since $h \in R^n$ was arbitrary, it follows that $F'(x)$ is nonsingular and that $F'(x)^{-1} \geq 0$. Conversely, let $F'(x)^{-1} \geq 0$ for each $x \in D$. Then, if $Fx \geq Fy$ for some $y,x \in D$, we obtain analogously from $F'(y)(y-x) \geq Fy - Fx \geq 0$ that $y - x \geq 0$ and hence that F is inverse isotone.

The converses of Theorems 5.3 and 5.2 are certainly false as the simple one-dimensional example $f(x) = x^3$ shows. In fact, in both cases, even the non-singularity of the derivative does not guarantee that $F'(x)$ is an S-matrix or P-matrix, respectively. For example,

$$F:R^2 \to R^2,\ F(x_1,x_2) = \begin{pmatrix} x_1^3 - x_2 \\ \\ x_1 + x_2^3 \end{pmatrix},$$

is an F-differentiable P-function on R^2, and hence, by Theorem 4.7, also an S-function; but $F'(0)$ is neither a P- nor an S-matrix.

In contrast to this example the following result holds in the inverse isotone case.

<u>Theorem 5.6.</u> Let $F:D \subset R^n \to R^n$ be inverse isotone and G-differentiable on the open set D. Then $F'(x)^{-1} \geqslant 0$ for any point x in D at which $F'(x)$ is nonsingular.

<u>Proof.</u> Suppose that $F'(x)$ is nonsingular at the point $x \in D$. If $F'(x)h > 0$ for some $h \in R^n$, then

$$\lim_{t \to 0+} \frac{1}{t}[F(x+th)-Fx] = F'(x)h > 0$$

shows that $F(x+th) > Fx$ for sufficiently small $t > 0$, and hence, by the inverse isotonicity, that $h \geqslant 0$. Now let $F'(x)h \geqslant 0$ for some $h \in R^n$ and set $h^k = h + \frac{1}{k} F'(x)^{-1}e$, $k = 0,1,\ldots$. Then $F'(x)h^k > 0$, and it follows from the first part that $h^k \geqslant 0$, $k = 0,1,\ldots$. Therefore we find that also $h = \lim_{k\to\infty} h^k \geqslant 0$ and hence that $F'(x)^{-1} \geqslant 0$.

Corollary 5.7. Let $F:D \subset R^n \to R^n$ be a G-differentiable M-function on the open set D. Then $F'(x)$ is an M-matrix whenever it is nonsingular.

Proof. Because of Theorem 5.6 we need to show only that the off-diagonal elements of $F'(x)$ are non-positive. This follows immediately from the off-diagonal antitonicity together with

$$\partial_j f_i(x) = \lim_{t \to 0+} \frac{1}{t} [f_i(x+te^j)-f_i(x)] \leq 0, \; i \neq j, \; i,j \in N.$$

For P_0-functions we obtain an even simpler result:

Theorem 5.8. Let $F:D \subset R^n \to R^n$ be a G-differentiable P_0-function on the open set D. Then $F'(x)$ is a P_0-matrix for each $x \in D$.

Proof. Let $x \in D$ and $h \in R^n$, $h \neq 0$, be given and let $\{t_k\}$ be any decreasing, positive sequence with $\lim_{k\to\infty} t_k = 0$. Then $x + t_k h \in D$ for $k \geq k_0$, and, since F is a P_0-function, there is an index $j \in N$ and a subsequence $\{t_{j_k}\}$ such that $h_j \neq 0$ and

$$t_{j_k} h_j [f_j(x+t_{j_k} h)-f_j(x)] \geq 0, \; k = 0,1,\ldots .$$

Hence

$$h_j \sum_{i=1}^{n} \partial_j f_i(x) h_i = \lim_{k\to\infty} h_j \frac{1}{t_{j_k}} [f_j(x+t_{j_k} h)-f_j(x)] \geq 0$$

and $F'(x)$ is a P_0-matrix.

Gale and Nikaido [1968] showed that if $F'(x)$ is a nonsingular P_0-matrix for each x in an open rectangle Q then F is injective. We now derive a stronger result which, among other things, settles, in the case of differentiable functions, the conjecture stated before Theorem 3.2. For the details about degree theory used in the following proof, we refer, for instance, to Ortega and Rheinboldt [1970].

Theorem 5.9. Let $F:D \subset R^n \to R^n$ be continuous on the open set D. Assume that $F + \varepsilon I$ is injective for any $\varepsilon > 0$, and that for each $x \in D$ there is an open neighborhood $U(x) \subset D$ of x such that $Fy = Fx$ for any $y \in U(x)$ implies that $y = x$. Then F is injective on D.

Proof. Suppose that $Fx^1 = Fx^2 = y$ for some $x^1, x^2 \in D$, $x^1 \neq x^2$. By assumption we can choose open neighborhoods U_1 and U_2 of x^1 and x^2 respectively such that $\bar{U}_1 \subset D$, $\bar{U}_2 \subset D$, $\bar{U}_1 \cap \bar{U}_2 = \emptyset$, and that $Fx = y$ for any $x \in \bar{U}_1 \cup \bar{U}_2$ implies that either $x = x^1$ or $x = x^2$. Then deg (F,C,y) is well-defined for $C = \bar{U}_1$, $C = \bar{U}_2$, as well as $C = \bar{U}_1 \cup \bar{U}_2$.

For any one of these three sets and a fixed i (=1,2) consider now the homotopy

$$H:D \times [0,1] \to R^n,\ H(x,t) = (1-t)Fx + t[x-x^i+y].$$

Then $H(x,t) \neq y$ for $x \in \partial C$ and $t \in [0,1]$. In fact, $H(x,0) = Fx \neq y$ and $H(x,1) = x - x^i + y \neq y$, while $H(x,t) = y$ for some $t \in (0,1)$ leads to the contradiction $Fx + \frac{t}{1-t}x = y + \frac{t}{1-t}x^i$ with the injectivity of $F + \varepsilon I$. Hence, by the homotopy invariance theorem of degree theory, it follows that

$$\deg (F,C,y) = \deg (x-x^i+y,C,y).$$

This means that $\deg(F,\bar{U}_i,y) = 1$ as well as $\deg(F,\bar{U}_1 \cup \bar{U}_2,y) = 1$. On the other hand, our assumptions about $\bar{U}_1,\bar{U}_2$ imply that $\deg(F,\bar{U}_1 \cup \bar{U}_2,y) = \deg(F,\bar{U}_1,y) + \deg(F,\bar{U}_2,y)$ which is a contradiction.

Corollary 5.10. Let $F:D \subset R^n \to R^n$ be a continuous P_0-function on the open set D. Then F is injective on D if and only if it is locally injective.

Proof. By Theorem 3.4, $F + \varepsilon I$ is injective for any $\varepsilon > 0$. Hence, if F is locally injective, then Theorem 5.9 proves the injectivity. The converse is trivial.

Corollary 5.11 (Gale and Nikaido [1965]). Let $F:Q \subset R^n \to R^n$ be F-differentiable on the open rectangle Q. If $F'(x)$ is a nonsingular P_0-matrix for each x in Q, then F is injective in Q.

Proof. By Corollary 5.3, F is a P_0-function and hence, by Theorem 3.3, $F + \varepsilon I$ is injective for any $\varepsilon > 0$. For any $x \in Q$ and $\varepsilon \in (0,\|F'(x)^{-1}\|^{-1})$ choose $\delta > 0$ such that $U(x) = \{y \in R^n \mid \|y-x\| < \delta\} \subset D$ and

$$\|Fy-Fx-F'(x)(y-x)\| \leq \varepsilon\|y-x\|, \quad y \in U(x).$$

Then $Fy = Fx$ for any $y \in U(x)$ would lead to the contradiction $\|F'(x)(y-x)\| \leq \varepsilon\|y-x\| < \|F'(x)^{-1}\|^{-1}\|y-x\|$, and hence $U(x)$ is a neighborhood of the type required in Theorem 5.9. The result therefore follows from that theorem.

Note that in this case F need not be a P-function as is shown by the example

$$F(x_1,x_2) = \begin{pmatrix} x_2 \\ \\ x_2 - x_1 \end{pmatrix} .$$

As our final result we prove that indeed the conjecture stated before Theorem 3.2 is valid for F-differentiable functions.

Corollary 5.12. Let $F:Q \subset R^n \to R^n$ be an F-differentiable, injective P_0-function on the open rectangle Q. Then $F^{-1}:FQ \subset R^n \to R^n$ is again a P_0-function.

Proof. By Theorem 3.2 it suffices to show that $F + \Delta$ is injective for every $\Delta = \mathrm{diag}(d_1,\ldots,d_n)$, $d_i \geq 0$, $i \in N$. Since $F + \Delta$ is a P_0-function, Theorem 5.8 ensures that $F'(x) + \Delta$ is a P_0-matrix for each $x \in Q$, and by the basic characterization theorem for P_0-matrices of Fiedler and Pták [1966], $F'(x) + \Delta$ is again nonsingular. But then Corollary 5.11 implies that $F + \Delta$ is indeed injective, and the proof is complete.

6. References

Bers, L. [1953]. On mildly nonlinear partial difference equations of elliptic type. J. Res. NBS, B 51, 229-236.

Birkhoff, G., and Kellogg, R. [1966]. Solution of equilibrium equations in thermal networks, in Proc. Symp. on Generalized Networks, pp. 443-452, Brooklyn Polytechnic Press, Brooklyn, N.Y.

Browder, F. [1963]. The solvability of nonlinear functional equations. Duke Math. J. 33, 557-567.

Collatz, L. [1952]. Aufgaben monotoner Art. Arch. Math. 3, 365-376.

Fan, Ky [1958]. Topological proofs for certain theorems on matrices with nonnegative elements. Monatsh. Math. 62, 219-237.

Fiedler, M., and Pták, V. [1962]. On matrices with non-positive off-diagonal elements and positive principal minors. Czech. Math. J. 12, 382-400.

Fiedler, M., and Pták, V. [1966]. Some generalizations of positive definiteness and monotonicity. Num. Math. 9, 163-172.

Gale, D., and Nikaido, H. [1965]. The Jacobian matrix and global univalence of mappings. Mathem. Annalen 159, 81-93.

Kacurovskii, R. [1960]. On monotone operators and convex functions (Russian). Uspehi Mat. Nauk 15, 213-215.

Kakutani, S. [1941]. A generalization of Brouwer's Fixed Point Theorem. Duke Math. J. 8, 457-459.

Karamardian, S. [1968]. Existence of solutions of certain systems of nonlinear inequalities. Num. Math. 12, 327-334.

Minty, G. [1962]. Monotone (nonlinear) operators in Hilbert space. Duke Math. J. 29, 341-346.

Moré, J. [1970]. A Class of Nonlinear Functions and the Convergence of Gauss-Seidel and Newton-Gauss-Seidel Iterations. Ph.D. Dissertation, University of Maryland.

Nikaido, H. [1968]. Convex Structures and Economic Theory, Academic Press, Inc., New York, N.Y.

Ortega, J., and Rheinboldt, W. [1970], Iterative Solution of Nonlinear Equations in Several Variables. Academic Press, Inc., New York, N.Y.

Rheinboldt, W. [1969]. On M-functions and their application to nonlinear Gauss-Seidel iterations and network flows. Ges. f. Math. u. Datenverarbeitung, Birlinghoven/Germany, Tech. Rep. BMwF-GMD-23; J. Math. Anal. Appl. 32, in press.

Rheinboldt, W. [1970]. On classes of n-dimensional nonlinear mappings generalizing several types of matrices, in Proc. Symp. on Num. Sol. of Part. Diff. Eqn. II, Academic Press, Inc., New York, N.Y., in press.

Sandberg, I., and Willson, A., Jr. [1969]. Some theorems on properties of DC equations of nonlinear networks. The Bell Syst. Tech. J. 48, 1-34.

Stiemke, E. [1915]. Über positive Lösungen homogener linearer Gleichungen, Mathem. Annalen 76, 340-342.

Varga, R. [1962]. Matrix Iterative Analysis, Prentice-Hall, Englewood Cliffs, N.J.