Technical Report 70-123
NGL-21-002-008 and GJ-231
July 1970

A Class of Nonlinear Functions and the Convergence of Gauss-Seidel and Newton-Gauss-Seidel Iterations

by

Jorge J. Moré

UNIVERSITY OF MARYLAND
COMPUTER SCIENCE CENTER
COLLEGE PARK, MARYLAND

Technical Report 70-123 July 1970
NGL-21-002-008 and GJ-231

A Class of Nonlinear Functions and the Convergence of Gauss-Seidel and Newton-Gauss-Seidel Iterations

by

Jorge J. Moré

This research received partial support from Grant GJ-231 of the National Science Foundation and from Grant NGL-21-002-008 of the National Aeronautics and Space Administration.

ABSTRACT

Certain related classes of nonlinear functions $F: D \subset R^n \to R^n$ are introduced, and the convergence of two types of iterative methods for the solution of the corresponding equation $Fx = 0$ is studied.

The main class of these functions is a generalization of the P-matrices of Fiedler and Pták. It is shown that the strictly monotone mappings, as well as the M-functions, are special cases of these P-functions, and that the inverse isotone mappings are closely related to them. Then a generalization of the strictly and irreducibly diagonally dominant matrices is introduced, and these Ω-diagonally dominant functions are likewise shown to be closely related to P-functions.

For a Ω-diagonally dominant function F, the nonlinear generalizations of Jacobi and Gauss-Seidel iterations due to Bers are then studied, and convergence results analogous to those available for strictly and irreducibly diagonally dominant matrices are obtained. For convex and inverse isotone F on R^n, iterations of the form

$$x^{k+1} = x^k - P_k(x^k)^{-1} Fx^k, \; k = 0,1,\ldots,$$

are then considered with particular emphasis on the Newton-Gauss-Seidel methods. A general result is obtained which contains as corollaries the global convergence theorems of Baluev for Newton's method and of Greenspan and Parter for the one-step Newton-SOR method. The global convergence of the general Newton-Gauss-Seidel method also follows from this theorem.

TABLE OF CONTENTS

INTRODUCTION

In this paper we analyze the convergence of two classes of iterative schemes for finding a solution of the equation $Fx = 0$ where $F:D \subset R^n \to R^n$ is, in general, nonlinear.

We first consider the nonlinear generalizations (Bers [1953]) of the Jacobi and Gauss-Seidel methods for the solution of linear equations. Take, for example, the Gauss-Seidel method. For affine mappings $Fx = Ax - b$ where A is some $n \times n$ matrix and b is a vector in R^n, it is well-known (Varga [1962]) that the following conditions guarantee the existence of a unique solution x^* of $Fx = 0$ and that the Gauss-Seidel iterates converge to x^* for any starting vector x^0:

1. A is symmetric and positive definite.
2. A is an M-matrix.
3. A is strictly diagonally dominant.
4. A is irreducibly diagonally dominant.

The first two of these conditions have been extended to nonlinear systems, and appropriate convergence results have been given. To be specific, Schechter [1962] proved global convergence for the nonlinear Gauss-Seidel method for a certain generalization of the first condition, and Elkin [1968], using a weaker generalization, extended Schechter's results. Concerning the second condition, Rheinboldt [1969b], following an unpublished suggestion of

J. M. Ortega, investigated an extension of the M-matrix concept and proved a global convergence result for the (underrelaxed) nonlinear Jacobi and Gauss-Seidel processes. These M-functions, and the corresponding global convergence theorem, brought together a number of apparently separate results of Bers [1953], Ortega and Rheinboldt [1970a], and Porsching [1969]. In this paper we generalize the notion of strictly and irreducibly diagonally dominant matrices to nonlinear systems and show that under suitable hypotheses the (underrelaxed) nonlinear Jacobi and Gauss-Seidel schemes are globally convergent.

At first glance, the four types of matrices mentioned, and their generalizations to nonlinear mappings, seem to involve four different concepts. Yet, it can actually be shown that for linear functions, they are all special cases of the more general class of P-matrices due to Fiedler and Pták [1964]. In fact, A is a P-matrix if A is positive definite or an M-matrix; and furthermore if A is strictly or irreducibly diagonally dominant with non-negative diagonal elements, then A is again a P-matrix. It is therefore desirable to consider an extension of the definition of a P-matrix to nonlinear mappings.

In Chapter I we generalize the concept of a P-matrix to nonlinear functions and explore the basic connection between these P-functions and other well-known classes of mappings, namely, the monotone and inverse isotone mappings, as well as the M-functions. Some of the results of this chapter are new, while others were developed jointly with W. C. Rheinboldt (see Moré and Rheinboldt [1970]).

In Chapter II we present the mentioned generalization of the strictly and irreducibly diagonally dominant matrices and investigate the relationship between P-functions and these new Ω-diagonally dominant functions. In particular, it is shown that knowledge of this relationship leads to a global convergence result for Ω-diagonally dominant functions. The other results obtained in this chapter also appear to be new, even in the linear case, but they are related to the work of Duffin [1948] and Rheinboldt [1969b] if F is a so-called off-diagonally antitone function, and to the results of Walter [1967] if F is linear.

In the last chapter we turn to the study of iterations of the form

$$x^{k+1} = x^k - P_k(x^k)^{-1}Fx^k, \ k = 0,1,\ldots,$$

with special emphasis on the general Newton-Gauss-Seidel methods. Here F is required to be continuously differentiable and convex on all of R^n with $F'(x)^{-1} \geq 0$ for each x in R^n. This implies that F is inverse isotone, and it is through this fact that the results of Chapter III are related to those of the previous chapters. For linear F, results of Varga [1962] concerning "regular splittings" are extended, while in the nonlinear case, new convergence results are presented. In particular, a general result is obtained which contains as corollaries the global convergence theorems of Baluev [1952] for Newton's method and of Greenspan and Parter [1965] for the one-step Newton-SOR method. A sufficient condition for the global convergence of the general Newton-Gauss-Seidel method also follows from this theorem.

CHAPTER I

Classes of Nonlinear Functions

1.1 Preliminary Definitions and Results

We denote by R^n the real n-dimensional linear space of column vectors $x = (x_1,\ldots,x_n)^T$, topologized by any vector norm. In particular, the ℓ_∞ norm $\|x\|_\infty = \sup\{|x_i|:i = 1,\ldots,n\}$ is frequently used. Correspondingly, $L(R^n)$ denotes the linear space of all real matrices of order n topologized by any norm induced by a vector norm in R^n. For example, in the case of the ℓ_∞ norm on R^n we have

$$\|A\|_\infty = \sup\{\sum_{j=1}^{n} |a_{ij}|:i = 1,\ldots,n\}$$

where $A = (a_{ij}) \in L(R^n)$. We use the coordinate-wise partial orderings on R^n and $L(R^n)$; that is, if x,y in R^n, then $x \geqslant y$ ($x > y$) if, and only if, $x_i \geqslant y_i$ ($x_i > y_i$) for $i = 1,\ldots,n$, and similarly for $L(R^n)$. In addition, if $x \in R^n$ and $A \in L(R^n)$, then $|x| = (|x_1|,\ldots,|x_n|)^T$ and $|A| = (|a_{ij}|)$ denotes the usual absolute value induced by the coordinate-wise partial orderings on R^n and $L(R^n)$, respectively. A rectangle in R^n is the Cartesian product of n intervals, 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 line segment $[x,y]$ is the set $\{z \in R^n:z = ty + (1-t)x$ for some $t \in [0,1]\}$, and the set $\{1,\ldots,n\}$ will always be denoted by N. Finally, the vector $e \in R^n$ is defined by $e_i = 1$ for each $i \in N$.

We now recall the definitions of certain classes of matrices that will play a role in this article.

Definition 1.1.1 a) A in $L(R^n)$ is positive definite if $x^TAx > 0$

for each $x \neq 0$ in R^n.

b) A in $L(R^n)$ is an M-matrix if $a_{ij} \leq 0$ for $i \neq j$ in N, and $A^{-1} \geq 0$.

c) A in $L(R^n)$ is strictly diagonally dominant if

$$|a_{ii}| > \sum_{j \neq i} |a_{ij}|$$

for each $i \in N$, where for $n = 1$ the sum on the right is defined to be zero.

d) A in $L(R^n)$ is irreducibly diagonally dominant if

$$|a_{ii}| \geq \sum_{j \neq i} |a_{ij}|$$

for each $i \in N$, where for at least one $i \in N$ strict inequality holds, and if for every i,j in N, there is a sequence of non-zero elements of A of the form $a_{i,i_1}, a_{i_1,i_2}, \ldots, a_{i_r,j}$.

The relationship between the first and the last two definitions is given by the following result, whose proof may be found in Varga [1962].

Theorem 1.1.2 Let A in $L(R^n)$ be a strictly or irreducibly diagonally dominant matrix with nonnegative diagonal elements. If A is symmetric, then A is positive definite. If $a_{ij} \leq 0$ for $i \neq j$ in N, then A is an M-matrix.

Another important class of matrices is the following:

Definition 1.1.3 (Fiedler and Pták [1962]). A in $L(R^n)$ is a P-matrix if for each $x \neq 0$ in R^n, there is an index $k \in N$ such that $x_k \cdot y_k > 0$ where $y = Ax$.

The following result of Fiedler and Pták [1962] characterizes P-matrices in terms of well-known concepts. B is a principal submatrix of A if B = A or if B is a matrix of order k, $1 \leqslant k < n$, obtained by deleting any n-k rows and the corresponding columns of A; by a principal minor we mean the determinant of a principal submatrix of A.

Theorem 1.1.4 Assume A in $L(R^n)$. The following statements are then equivalent.

a) A is a P-matrix.

b) The real eigenvalues of each principal submatrix of A are positive.

c) All principal minors are positive.

d) If B is any principal submatrix of A and $D \geqslant 0$ is a diagonal matrix of the same order as B, then $\det(B+D) > 0$.

Fiedler and Pták [1962] proved the equivalence of a), b), and c) in the above theorem; d) is implicit in their proof. In the same paper they also established that every M-matrix is a P-matrix.

Theorem 1.1.5 Assume A in $L(R^n)$.

a) If A is positive definite or an M-matrix, then A is a P-matrix.

b) If A is a strictly or irreducibly diagonally dominant matrix with nonnegative diagonal elements, then A is a P-matrix.

Gale and Nikaido [1965] showed that if $A \in L(R^n)$ is a positive definite matrix or a strictly diagonally dominant matrix with nonnegative diagonal elements then A is a P-matrix; the observation that every irreducibly diagonally dominant matrix with nonnegative diagonal elements is a P-matrix seems to be new. The proof of the above result will be given (in a more general setting) later on; part a) in the next section, part b) in Chapter II.

1.2 Nonlinear M- and P-functions

In this section we will define nonlinear generalizations of the matrices introduced in the previous section and prove the connections between the new concepts. We begin with a well-known generalization of positive definiteness.

Definition 1.2.1 Consider $F:D \subset R^n \to R^n$.

a) F is strictly monotone on D if for each $x \neq y$ in D,

$$(x-y)^T(Fx-Fy) > 0.$$

b) F is uniformly monotone on D if there is a $c > 0$ such that

$$(x-y)^T(Fx-Fy) \geq c\|y-x\|^2$$

for each x,y in D.

If F is linear, both concepts are clearly equivalent to positive definiteness; note, however, that a continuous, strictly monotone function is not necessarily surjective, while Minty [1962] proved that this was the case for a continuous, uniformly monotone function on a Hilbert space. An elementary proof of this fact for R^n can be found in the next section.

In order to state the generalization of M-matrices, we will need the following concept due to Collatz [1952].

Definition 1.2.2 The mapping $F:D \subset R^n \to R^n$ is inverse isotone on D if $Fx \leq Fy$, x,y in D implies $x \leq y$.

It is easy to see that $Fx = Ax$ is inverse isotone on R^n if, and only if, $A^{-1} \geq 0$. The following nonlinear generalization of the M-matrix concept was proposed by J. M. Ortega in an unpublished note and then studied by Rheinboldt [1969b].

Definition 1.2.3 Let $F:D \subset R^n \to R^n$.

a) F is off-diagonally antitone if for any x in R^n, the functions $\alpha_{ij}(x,\cdot):\{t \in R^1 : x+te^j \text{ in } D\} \to R^1$,

$$\alpha_{ij}(x,t) = f_i(x+te^j), \; i \neq j, \; i,j \in N,$$

are antitone. Here e^j denotes the jth unit basis vector in R^n.

b) F is an M-function on D if F is off-diagonally antitone and inverse isotone on D.

Once again, it is easy to see that $Fx = Ax$ is an M-function on R^n if, and only if, A is an M-matrix. We now proceed to the definition of P-functions which is just a straightforward extension of the linear definition.

Definition 1.2.4 The mapping $F:D \subset R^n \to R^n$ is a P-function on D if for each $x \neq y$ in D, there is an index k in N such that

$$(x_k - y_k)[f_k x - f_k y] > 0 \tag{1.2.1}$$

where f_k is the k-th component function of F.

The concept of a P-function is new, although inequality (1.2.1) was obtained by Nikaido [1968] in some related work. He, however, never made a systematic use of this inequality.

The remainder of this section will point out the connections among the different classes of functions defined so far. We begin with the following simple observation.

Theorem 1.2.5 Let $F:D \subset R^n \to R^n$ be a strictly monotone function on D. Then F is a P-function on D.

Proof. Assume that F is not a P-function and hence that for some $x \neq y$ in D, $(x_k - y_k)(f_k x - f_k y) \leq 0$ for each $k \in N$. Adding these inequalities we obtain $(x-y)^T(Fx-Fy) \leq 0$ for $x \neq y$ in D. This contradicts the fact that F is a strictly monotone function on D.

Definition 1.2.6 Consider $F:D \subset R^n \to R^n$, and let $L = \{i_1,\ldots,i_k\}$ be a non-empty subset of N. For fixed x in R^n, define $D_G = \{(t_{i_1},\ldots,t_{i_k}):\hat{t} \in D$ where $\hat{t}_j = t_j$ if $j \in L$, and $\hat{t}_j = x_j$ if $j \notin L\}$. Then $G:D_G \subset R^k \to R^k$ is a subfunction of F belonging to L if

$$g_i(t_{i_1},\ldots,t_{i_k}) = f_i(\hat{t}),\ i = i_1,\ldots,i_k.$$

If $F:R^n \to R^n$ is linear, then a subfunction of F corresponds to a principal submatrix. If F is nonlinear, this concept of subfunction has been used implicitly by many authors, but Rheinboldt [1969b] seems to be the first one to make explicit use of this definition. We also remark that the subfunction G depends on a specific value of x in R^n, but since it will always be clear which x is being used, this x has not been made an explicit part of the notation. The next result is a direct consequence of Definition 1.2.6.

Theorem 1.2.7 Let $F:D \subset R^n \to R^n$ be a P-function. Then each subfunction of F is also a P-function.

Theorem 1.2.7 also holds for M-functions, but the proof is more involved. We will need the following intermediary result which is interesting in its own right.

Lemma 1.2.8 Let $F:D \subset R^n \to R^n$ be an off-diagonally antitone P-function on a rectangle D. Then F is an M-function on D.

Proof. We only need to show that F is inverse isotone. For this purpose, suppose that $Fy \leq Fx$ for y,x in D but that $L = \{i \in N : y_i > x_i\}$ is not empty. For ease of notation, assume that $L = \{1,\ldots,k\}$, $k \in N$, and define

$$g_i(t_1,\ldots,t_k) = f_i(t_1,\ldots,t_k,x_{k+1},\ldots,x_n)$$

for $i \in L$. By Theorem 1.2.7, $G:D_G \subset R^k \to R^k$ is a P-function, and since F is off-diagonally antitone, it follows that

$$g_i(x_1,\ldots,x_k) = f_i(x) \geq f_i(y) \geq g_i(y_1,\ldots,y_k)$$

for $1 \leq i \leq k$. Hence,

$$(y_i-x_i)[g_i(y_1,\ldots,y_k)-g_i(x_1,\ldots,x_k)] \leq 0$$

for $i = 1,\ldots,k$, which contradicts the fact that G is a P-function.

We are now in a position to prove that Theorem 1.2.7 also holds for M-functions.

Theorem 1.2.9 Assume $F:D \subset R^n \to R^n$ is an M-function on the rectangle D. Then every subfunction of F is also an M-function.

Proof. Assume that there is a subfunction $G:D_G \subset R^k \to R^k$, $1 \leq k < n$, which is not an M-function. Since G is off-diagonally antitone, Lemma 1.2.8 implies that G is not a P-function. Hence, there are $x \neq y$ in D_G, such that

(1.2.2) $(x_i-y_i)[g_i(x_1,\ldots,x_k)-g_i(y_1,\ldots,y_k)] \leq 0,\ 1 \leq i \leq k.$

Since $x \neq y$, we may assume that $L = \{i:x_i < y_i\}$ is not empty, and moreover, for ease of notation, that $L = \{1,\ldots,m\}$, $1 \leq m \leq k$. Then, if $1 \leq i \leq m$,

$$f_i(y_1,\ldots,y_k,z_{k+1},\ldots,z_n) = g_i(y_1,\ldots,y_k) \leq g_i(x_1,\ldots,x_k)$$

(1.2.3)

$$= f_i(x_1,\ldots,x_k,z_{k+1},\ldots,z_n) \leq f_i(x_1,\ldots,x_m,y_{m+1},\ldots,y_k,z_{k+1},\ldots,z_n),$$

since F is off-diagonally antitone and (1.2.2) holds, while

$$f_i(y_1,\ldots,y_k,z_{k+1},\ldots,z_n) \leq f_i(x_1,\ldots,x_m,y_{m+1},\ldots,y_k,z_{k+1},\ldots,z_n),$$

if $m < i \leq n$. But (1.2.3) implies that this last inequality holds for all $i \in N$, and since F is inverse isotone, it follows that $x_i \geq y_i$ for $i = 1,\ldots,m$. This contradicts the definition of L and concludes the proof.

The previous result extends in part a result of Rheinboldt [1969b] in which he proved that if $F:R^n \to R^n$ is a continuous and surjective M-function, then every subfunction is again a surjective M-function. To end this section, we generalize a linear result of Fiedler and Pták [1962] which points out the precise relationship between M- and P-functions.

Theorem 1.2.10 The mapping $F:D \subset R^n \to R^n$ is an M-function on the rectangle D if, and only if, F is an off-diagonally antitone P-function.

Proof. Lemma 1.2.8 gives us the sufficiency of the condition, so we only need to prove the necessity. In order to obtain a contradiction, assume that F is an M-function, but not a P-function. Then there are y,x in D, $y \neq x$, such that

$$(1.2.4) \qquad (y_i - x_i)(f_i y - f_i x) \leq 0, \; i \in N.$$

Since $y \neq x$, we may assume that $L = \{i \in N: y_i < x_i\}$ is not empty, and that $L = \{1,\ldots,k\}$, $k \in N$. Let $G:D_G \subset R^k \to R^k$ be the subfunction of F defined by

$$g_i(t_1,\ldots,t_k) = f_i(t_1,\ldots,t_k,x_{k+1},\ldots,x_n), \; 1 \leq i \leq k.$$

Since F is off-diagonally antitone, (1.2.4) implies that

$$(1.2.5) \qquad \begin{aligned} g_i(x_1,\ldots,x_k) &= f_i x \leq f_i y \leq f_i(y_1,\ldots,y_k,x_{k+1},\ldots,x_n) \\ &= g_i(y_1,\ldots,y_k) \end{aligned}$$

for $i = 1,\ldots,k$. Since Theorem 1.2.9 shows that G is an M-function, G is inverse isotone, and hence it follows from (1.2.5) that $x_i \leq y_i$ for $i = 1,\ldots,k$. This contradicts the construction of L, and therefore, F must be a P-function.

1.3 Properties of P-functions and Inverse Isotone Mappings

We now investigate some of the basic properties of inverse isotone mappings and P-functions. Only those results shall be proved which are relevant to the discussion in the next chapter; a more exhaustive list of the known properties of P-functions may be found in Moré and Rheinboldt [1970].

Theorem 1.3.1 Consider $F:D \subset R^n \to R^n$.

a) If F is inverse isotone on D, then F is injective on D, and $F^{-1}:F(D) \subset R^n \to R^n$ is isotone.

b) If F is a P-function on D, then F is injective on D, and F^{-1} is a P-function on $F(D)$.

Proof. Assume first that F is inverse isotone on D. If $Fx = Fy$ for x,y in D, then $x \leq y$ and $x \geq y$. Hence $y = x$ and F is injective on D. Since $Fx \leq Fy$ implies $F^{-1}(Fx) = x \leq y = F^{-1}(Fy)$, F^{-1} is isotone on $F(D)$.

If F is a P-function on D and $Fx = Fy$ for $x \neq y$ in D, then for some $k \in N$, $(x_k-y_k)(f_kx-f_ky) > 0$ which contradicts the fact that $f_kx = f_ky$. That F^{-1} is a P-function on $F(D)$ is clear from the definitions.

Part a) of the above result is well-known, while if $F:R^n \to R^n$ is linear, part b) is due to Sandberg and Willson [1969].

If A is a P-matrix, then it is clear that A has positive diagonal elements. To describe the corresponding notion we need the following definition of Ortega and Rheinboldt [1970b].

Definition 1.3.2 Consider $F:D \subset R^n \to R^n$. Then

a) For fixed x in R^n, the i-th diagonal function

$\alpha_{ii}(x,\cdot):\{t \in R^1: x+te^i \text{ in } D\} \to R^1$ is defined by

$$\alpha_{ii}(x,t) = f_i(x+te^i),$$

for each $i \in N$.

b) If each diagonal function is (strictly) isotone for each x in R^n, then F is (strictly) diagonally isotone on D.

Theorem 1.3.3 Assume $F:D \subset R^n \to R^n$ is a P-function. Then F is strictly diagonally isotone on D.

Proof. Let $x \in R^n$ and $i \in N$ be given. If $s > t$ and $x + se^i$, $x+te^i$ lie in D, then $(s-t)[f_i(x+se^i)-f_i(x+te^i)] > 0$ since F is a P-function, and hence, $f_i(x+ste^i) > f_i(x+te^i)$.

Another property of P-matrices is that SA and AS are P-matrices if A is a P-matrix and $S \geq 0$ is an invertible diagonal matrix. We now generalize this result of Fiedler and Pták [1962].

Definition 1.3.4 (Ortega and Rheinboldt [1970b]) The mapping $\phi:D \subset R^n \to R^n$ is a diagonal function on D if for each x in D and $i \in N$, $\phi_i(x) = \phi_i(x_i)$.

Theorem 1.3.5 Let $F:D \subset R^n \to R^n$ be a P-function and $\phi:D_0 \subset R^n \to R^n$ a diagonal, strictly isotone function.

a) If $\phi(D_0) \subset D$, then $F \circ \phi:D_0 \subset R^n \to R^n$ is a P-function.

b) If $F(D) \subset D_0$, then $\phi \circ F:D \subset R^n \to R^n$ is a P-function.

Proof. We only carry out the proof for a); that for b) is analogous. Let $x \neq y$ in D_0 be given. Then $\phi(x) \neq \phi(y)$ in D, and since F is a P-function on D, there is an index $k \in N$ such that $[\phi_k(x)-\phi_k(y)][f_k\phi(x)-f_k\phi(y)] > 0$. Now, ϕ_k is strictly isotone, and $\phi_k(x) - \phi_k(y) = \phi_k(x_k) - \phi_k(y_k)$, so if $\phi_k(x) - \phi_k(y) > 0$, it follows that $f_k\phi(x) - f_k\phi(y) > 0$ and $x_k - y_k > 0$. Hence, $(x_k-y_k)[f_k\phi(x)-f_k\phi(y)] > 0$. A similar argument yields the theorem when $\phi_k x - \phi_k y < 0$.

A very similar theorem is as follows:

Theorem 1.3.6 Let $F:D \subset R^n \to R^n$ be a P-function and let $\phi:D_0 \subset R^n \to R^n$ be a diagonal mapping such that $\phi(D_0) \subset D$ and $F(D) \subset D_0$. If for each k in N, ϕ_k is either strictly isotone or strictly antitone, then $G = \phi \cdot F \cdot \phi:D_0 \subset R^n \to R^n$ is also a P-function. In particular (Gale and Nikaido [1965]), if A is a P-matrix and S is a diagonal, invertible matrix, then SAS is a P-matrix.

Proof: Let $x \neq y$ in D_0 be given. Then $\phi(x) \neq \phi(y)$ in D, and since F is a P-function on D, there is an index k in N such that $[\phi_k(x)-\phi_k(y)][f_k\phi(x)-f_k\phi(y)] > 0$. Without loss of generality we assume that ϕ_k is strictly antitone. If $\phi_k(x) - \phi_k(y) > 0$, then $f_k\phi(x) - f_k\phi(y) > 0$, and since ϕ is a diagonal mapping, $x_k < y_k$, $\phi_k F\phi(x) < \phi_k F\phi(y)$. Hence, $(x_k-y_k)[\phi_k F\phi(x)-\phi_k F\phi(y)] > 0$. If $\phi_k(x) - \phi_k(y) < 0$, all the inequalities are to be reversed.

In connection with certain problems in nonlinear transistor networks, Willson [1968], and Sandberg and Willson [1969], were led to investigate the surjectivity of functions of the type $A + \phi$ where ϕ is a diagonal mapping. In particular, they obtained that $A + \phi$ is surjective if A is a P-matrix and ϕ is continuous and isotone on R^n. In order to generalize this result, we introduce the following definition.

Definition 1.3.7 $F:D \subset R^n \to R^n$ is a uniform P-function if there is a $c > 0$ such that for each $x \neq y$ in D,

$$(1.3.1) \qquad (x_k - y_k)\cdot(f_k x - f_k y) \geq c\|y-x\|^2$$

for some $k = k(x,y)$ in N.

The existence of such a $c > 0$ for the function $A + \phi$ is a consequence of the next result.

Lemma 1.3.8 If A in $L(R^n)$ is a P-matrix, then there is a $c > 0$ such that for each $x \neq 0$ in R^n and for some index $k = k(x)$ in N,

$$x_k y_k \geq c\|x\|^2$$

where $y = Ax$.

Proof. Define $g:R^n \to R^1$ by $g(x) = \max\{x_j y_j : y = Ax, j \in N\}$. Then g is continuous and positive on the unit sphere, and hence there is a $c > 0$ such that $g(x) \geq c$ for all x of unit norm. The result follows immediately.

The interesting fact is that (1.3.1) implies that F is a homeomorphism of R^n onto R^n.

Theorem 1.3.9 Let $F:R^n \to R^n$ be a continuous and uniform P-function on R^n. Then F is a homeomorphism of R^n onto R^n.

Proof. Since all the norms in R^n are equivalent, we may assume that (1.3.1) holds for the infinity norm. Equation (1.3.1) implies that $\|Fx-Fy\|_\infty \geq c\|x-y\|_\infty$ and hence that F is injective and $\|F^{-1}x-F^{-1}y\|_\infty \leq \frac{1}{c}\|x-y\|_\infty$ for all x,y in $F(D)$. Thus, only the surjectivity of F needs proof.

For $n = 1$, surjectivity follows readily from (1.3.1); therefore assume that the result is valid for some $n \geq 1$, and let $F:R^{n+1} \to R^{n+1}$ satisfy (1.3.1). Fix $t \in R^1$, and define $G(\cdot,t):R^n \to R^n$ by

$$g_i(x_1,\ldots,x_n,t) = f_i(x_1,\ldots,x_n,t)$$

for $i \in N$. For each $t \in R^1$, $G(\cdot,t)$ is clearly a uniform P-function on R^n, and therefore is surjective by the induction hypothesis. Let $b \in R^{n+1}$ be given. We can then define $H:R^1 \to R^n$ by

$$f_i(h_1(t),\ldots,h_n(t),t) = b_i = g_i(h_1(t),\ldots,h_n(t),t) \tag{1.3.2}$$

for $i \in N$, and $\psi:R^1 \to R^1$ by

$$\psi(t) = f_{n+1}(h_1(t),\ldots,h_n(t),t).$$

To prove that H is continuous, note that for $s \neq t$, (1.3.1) implies 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 therefore, that

$$\|G[H(s),t] - G[H(t),t]\|_\infty \geq c\|H(s)-H(t)\|_\infty.$$

Since (1.3.2) holds,

$$\|G[H(s),t]-G[H(s),s]\|_\infty \geq c\|H(s)-H(t)\|_\infty,$$

and the continuity of H follows from the continuity of F. Therefore, ψ is continuous, and if $s \neq t$, (1.3.1) and (1.3.2) imply that

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

Hence, $\lim_{t\to+\infty} \psi(t) = +\infty$ and $\lim_{t\to-\infty} \psi(t) = -\infty$, and since ψ is continuous, ψ is surjective. The intermediate value property for continuous function implies that there is a $t^* \in R^1$ with $\psi(t^*) = b_{n+1}$. It follows from (1.3.2) that $Fx^* = b$ where $x^* = (h_1(t^*),\ldots,h_n(t^*),t^*)^T$, and the proof is complete.

As a trivial corollary, we have the finite dimensional version of a result of Minty [1962].

<u>Corollary 1.3.10</u> Assume $F:R^n \to R^n$ is continuous and uniformly monotone on R^n. *Then* F *is a homeomorphism of* R^n *onto* R^n.

It should be noted that the last two results follow readily from the Domain Invariance Theorem and results of Rheinboldt [1969a], but the important point here is that knowledge about P-functions has allowed us to give an easy proof of an otherwise difficult result.

We finish this section with a special case of a theorem of Sandberg and Willson [1969].

Corollary 1.3.11 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 of R^n onto R^n.

Proof. The result follows directly from Lemma 1.3.8 and Theorem 1.3.9.

1.4 Differentiable P-functions and Inverse Isotone Mappings

The results of Gale and Nikaido [1965] show that if the Jacobian of a function F defined on a rectangle D is a P-matrix for each x in D then F is injective on D. The purpose of this section is to incorporate their results into the framework of P-functions, and, in general, to investigate the effect of differentiability assumptions upon the definitions.

The notions of differentiability to be used are that of the well-known Gateaux and Fréchet derivatives. Briefly: $F: D \subset R^n \to R^n$ is G-differentiable at an interior point x of D if there is an $A \in L(R^n)$ such that for any $h \in R^n$,

$$\lim_{t \to 0} \frac{1}{|t|} \|F(x+th) - Fx - tAh\| = 0,$$

and it is F-differentiable at x if

$$\lim_{h\to 0} \frac{1}{\|h\|} \|F(x+h)-Fx-Ah\| = 0.$$

In either case, there is only one such A, denoted by $F'(x)$, namely, the Jacobian matrix $(\partial_j f_i(x))$ where $\partial_j f_i(x) \equiv \frac{\partial f_i(x)}{\partial x_j}$. For a summary of the properties of G- and F-differentiable functions, see Ortega and Rheinboldt [1970b].

We begin our investigations with a well-known result--see, for example, Gale and Nikaido [1965].

Theorem 1.4.1 Let $F:D \subset R^n \to R^n$ be G-differentiable on the convex set D, and assume that $F'(x)$ is a positive definite matrix for each x in D. Then F is strictly monotone and hence, injective on D.

A similar assertion can be made if the Jacobian matrix of F is a P-matrix, but first we shall need the following result of Gale and Nikaido [1965].

Lemma 1.4.2 If A in $L(R^n)$ is a P-matrix, then there is an $h > 0$ such that $Ah > 0$.

Theorem 1.4.3 (Gale and Nikaido [1965], Nikaido [1968]) Let $F:D \subset R^n \to R^n$ be F-differentiable on the closed rectangle D, and suppose that $F'(x)$ is a P-matrix for each x in D. Then F is a P-function on D.

Proof. The proof proceeds by induction on n. For $n = 1$ the result is clear, so assume it holds for some $n - 1 \geq 1$, and let

$F:D \subset R^n \to R^n$ satisfy the conditions of the theorem.

We observe that if for any $u \neq v$ in D we have $u_i = v_i$ for some $i \in N$, then there is a $j \neq i$ in N such that

$$(1.4.1) \qquad (u_j-v_j)(f_ju-f_jv) > 0.$$

To prove this remark, assume that $i = n$, and define $G:D_G \subset R^{n-1} \to R^{n-1}$ by

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

for $i = 1,\ldots,n-1$. Since $G'(t_1,\ldots,t_{n-1})$ is a P-matrix for each $(t_1,\ldots,t_{n-1}) \in D_G$, the induction hypothesis implies that G is a P-function, and, therefore, that $(u_j-v_j)[g_j(u_1,\ldots,u_{n-1})-g_j(v_1,\ldots,v_{n-1})] > 0$ for some $j \neq n$ which is (1.4.1).

Consider now the set $D_0 = \{x \in D:Fx \leq Fz,\ x > z\}$ where $z \in D$ is, for the moment, fixed. We claim that D_0 is empty; for if this were not so, and $\{x^k\} \subset D_0$ is any decreasing sequence, then clearly $\lim_{k\to+\infty} x^k = x$ exists in D and $Fx \leq Fz$. If $x_i = z_i$ for some $i \in N$, then $x = z$ by our initial observation and

$$(1.4.2) \qquad \lim_{k\to+\infty} \frac{1}{\|x^k-z\|}\,[Fx^k-Fz-F'(z)(x^k-z)] = 0.$$

By Lemma 1.3.7, there is a $c > 0$ such that some component of $F'(z)\dfrac{x^k-z}{\|x^k-z\|}$ is greater than c, and by (1.4.2), some component of $Fx^k - Fz$ must be positive for large enough k. This contradicts the fact that $x^k \in D_0$, and therefore, we have $x \in D_0$. Zorn's Lemma then yields the existence of a minimal element in D_0; that is, a $u \in D_0$

such that $x \leqslant u$ for $x \in D_0$ implies $x = u$. This is, however, impossible: in fact, Lemma 1.4.2 shows that there is an $h < 0$ with $F'(u)h < 0$, and since

$$\lim_{t \to 0^+} \frac{F(u+th)-F(u)}{t} = F'(u)h < 0,$$

we have $F(u+t_0h) < Fu \leqslant Fz$ and $z < u + t_0h < u$ for sufficiently small $t_0 > 0$, which contradicts the minimality of u. Hence, D_0 is empty.

It now follows that F must be a P-function on D: if for some $x \neq y$ in D, $(x_k-y_k)(f_kx-f_ky) \leqslant 0$ for each $k \in N$, then $x_k \neq y_k$ for each $k \in N$ by the remark at the beginning of the proof. Define $S = \text{diag}\ (s_1,\ldots,s_n)$ by

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

and $H:S^{-1}(D) \subset R^n \to R^n$ by $H(z) = SF[Sz]$. Then $H'(z) = SF'[Sz]S$ is a P-matrix (Theorem 1.3.6) on the closed rectangle $S^{-1}(D)$; moreover, $H(S^{-1}x) \leqslant H(S^{-1}y)$ and $S^{-1}x > S^{-1}y$ which altogether contradicts what we have already proved.

The proof of this theorem uses ideas of Gale and Nikaido [1965] and Nikaido [1968]. In fact, if F satisfies the conditions of Theorem 1.4.3, Gale and Nikaido [1965] showed that F is injective on D, while Nikaido [1968] derived (1.2.1) and noted that injectivity followed from this relationship.

The converse of Theorem 1.4.3 is false as shown by the one-dimensional example $f(x) = x^3$, but certain partial converses are known (see Moré and Rheinboldt [1970]). A similar remark can be made about the next result.

Corollary 1.4.4 Assume that $F:D \subset R^n \to R^n$ is F-differentiable on the closed rectangle D, and that $F'(x)$ is an M-matrix for each x in D. Then F is an M-function on D.

Proof. Note that for each $i \neq j$ in N, and $x \in R^n$ $\alpha_{ij}(x,\cdot)$ is defined on a closed interval, and has there a non-positive derivative. Since F is a P-function, the result therefore follows from Theorem 1.2.10.

The last two results raise the question of whether or not $F'(x)^{-1} \geq 0$ for all x in a suitable set D implies that F is inverse isotone. This is not known, but a partial result using the notion of convexity is available.

Definition 1.4.5 The mapping $F:D \subset R^n \to R^n$ is convex on the convex set D if

$$F(\lambda x+(1-\lambda)y) \leq \lambda F(x)+(1-\lambda)Fy$$

for each x,y in D, and λ in $[0,1]$.

If F is differentiable, then convexity can be characterized as follows:

Lemma 1.4.6 Let $F:D \subset R^n \to R^n$ be G-differentiable on the convex set D. Then F is convex on D if, and only if,

$$Fy - Fx \geq F'(x)(y-x) \tag{1.4.3}$$

for each x,y in D.

The proof of this result can be found in Ortega and Rheinboldt [1970b]. It is now very easy to prove the following characterization of inverse isotonicity.

Theorem 1.4.7 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 on D if, and only if, $F'(x)^{-1} \geq 0$ for all x in D.

Proof. Assume first that F is inverse isotone and suppose $F'(x)h \geq 0$. By (1.4.3) it follows that $F(x+h) - Fx \geq 0$, and inverse isotonicity implies $h \geq 0$. Since $F'(x)h \geq 0$ implies $h \geq 0$, $F'(x)^{-1} \geq 0$ as desired. Conversely, if $F'(x)^{-1} \geq 0$ for each x in D and $Fy \leq Fx$ for y,x in D, (1.4.3) implies that $y \leq x$ and F is therefore inverse isotone on D.

CHAPTER II

Strictly and Ω-diagonally Dominant Functions

2.1 Definitions and Preliminary Results

In the previous chapter several classes of nonlinear functions were introduced and were shown to be natural generalizations of known types of matrices. For these classes of functions, in this and the next chapter, we study the convergence of iterative methods to the solution of $Fx = 0$ where $F:D \subset R^n \to R^n$ is known to have a zero in D.

In this chapter we consider the following two iterations:

The Gauss-Seidel iteration: Solve

$$f_i(x_1^{k+1},\ldots,x_{i-1}^{k+1},x_i,x_{i+1}^k,\ldots,x_n^k) = 0 \tag{2.1.1}$$

for x_i, and set

$$x_i^{k+1} = (1-\omega_k)x_i^k + \omega_k x_i, \quad i = 1,\ldots,n, \; k = 0,1,\ldots, \tag{2.1.2}$$

and the Jacobi iteration: Solve

$$f_i(x_1^k,\ldots,x_{i-1}^k,x_i,x_{i+1}^k,\ldots,x_n^k) = 0 \tag{2.1.3}$$

for x_i, and set

$$x_i^{k+1} = (1-\omega_k)x_i^k + \omega_k x_i, \quad i = 1,\ldots,n, \; k = 0,1,\ldots, \tag{2.1.4}$$

where $\omega_k \in (0,2)$ is a given sequence of relaxation parameters. These two iterative schemes are analyzed by generalizing the notion of strictly and irreducibly diagonally dominant matrices and examining the relationship of this generalization to P-functions.

In order to introduce our generalization of a strictly diagonally dominant matrix, we will have to look at this class of matrices from a somewhat different point of view than is usual. The following result will indicate the way.

Lemma 2.1.1 Let $v \in R^n$; then

a) $|v_k| > \sum_{j \neq k} |v_j|$ for some $k \in N$

if, and only if, for any $x \in R^n$

b) $\sum_{j=1}^{n} v_j x_j = 0$, $x \neq 0$, implies that $|x_k| < \|x\|_\infty$.

Proof. Assume that a) holds, and that $\sum_{j=1}^{n} v_j x_j = 0$, $x \neq 0$. Then $v_k x_k = -\sum_{j \neq k} v_j x_j$, and $|v_k||x_k| \leq \sum_{j \neq k} |v_j| \|x\|_\infty$, from which b) follows.

If b) holds but $|v_k| \leq \sum_{j \neq k} |v_j|$, then $\alpha |v_k| = \sum_{j \neq k} |v_j|$ for $\alpha \geq 1$. Define $x \in R^n$ by $x_k = \alpha \operatorname{sgn} v_k$, $x_j = -\operatorname{sgn} v_j$, $j \neq k$; then $\|x\|_\infty = \alpha = |x_k|$ and $\sum_{j=1}^{n} v_j x_j = 0$. This contradicts b) since $x \neq 0$. Hence, a) must hold.

If $A \in L(R^n)$, and for some $k \in N$, $v_j = a_{kj}$, $j = 1, \ldots, n$, then a) is equivalent to assuming "strict diagonal dominance on the kth row". Condition b) can be generalized to the nonlinear case.

Definition 2.1.2 a) A functional $f: D \subset R^n \to R^1$ is strictly diagonally dominant on D with respect to the kth variable if for every $x \neq y$ in D,

$$fx = fy, \text{ implies that } |x_k - y_k| < \|x-y\|_\infty.$$

b) A function $F:D \subset R^n \to R^n$ is strictly diagonally dominant on D if for each $k \in N$, the kth component function of F, f_k, is strictly diagonally dominant with respect to the kth variable.

From Lemma 2.1.1 we obtain immediately:

Theorem 2.1.3 Let $A \in L(R^n)$. Then A is a strictly diagonally dominant matrix if, and only if, the induced mapping $Fx = Ax$ is a strictly diagonally dominant function on R^n.

We next prove several results that give sufficient conditions for a function to be strictly diagonally dominant.

Theorem 2.1.4 Let $F:D \subset R^n \to R^n$ be G-differentiable on the convex set D, and assume that $F'(x)$ is a strictly diagonally dominant matrix for each x in D. Then F is a strictly diagonally dominant function on D.

Proof. Let $k \in N$ be given, and assume that $f_k x = f_k y$ for some $x \neq y$ in D. Then $\psi(t) = f_k(x+t(y-x))$ is differentiable on $[0,1]$, and $\psi(0) = \psi(1)$. By Rolle's theorem, there is a $t_0 \in (0,1)$ such that

$$\psi'(t_0) = \sum_{j=1}^{n} \partial_j f_k(x+t_0(y-x))(y_j-x_j) = 0.$$

The conclusion now follows from Lemma 2.1.1 with

$$v_j = \partial_j f_k(x+t_0(y-x)), \; j = 1,\ldots,n.$$

Later we shall see that this result admits a certain converse. On the other hand, Theorem 2.1.4 does not account for the case where $F'(x)$ is not strictly diagonally dominant at all points. The next result will point out how this theorem can be extended to cover this case.

Theorem 2.1.5 Let $f: D \subset R^n \to R^1$ be continuously differentiable on the convex set D, and assume that for some fixed $k \in N$, either

a) $$|\partial_k f(x)| > \sum_{j \neq k} |\partial_j f(x)| \text{ for each } x \in D,$$

or,

b) $$|\partial_k f(x)| \geq \sum_{j \neq k} |\partial_j f(x)| \text{ for each } x \in D, \text{ where}$$

$\partial_k f(x)$ does not change sign on D, and f is not constant on any line segment $[x,y]$ for which $x_k \neq y_k$.

Then f is strictly diagonally dominant on D with respect to the kth variable.

Proof. If a) holds, the proof is identical to the one given in Theorem 2.1.4, so assume that b) holds. If for some $x \neq y$ in D, $fx = fy$ and $|x_k - y_k| = \|x-y\|_\infty$, then $x_k \neq y_k$, and without loss of generality, we may suppose that $y_k - x_k > 0$. Since $\partial_k f(x)$ does not change sign for $x \in D$, assume that $\partial_k f(x+t(y-x)) \geq 0$ for each $t \in [0,1]$ and set $\psi(t) = f(x+t(y-x))$ for $t \in [0,1]$. Then

$$\psi'(t) = \partial_k f(x+t(y-x))(y_k-x_k) + \sum_{j\neq k} \partial_j f(x+t(y-x))(y_j-x_j),$$

and by b)

$$\psi'(t) \geq \sum_{j\neq k} |\partial_j f(x+t(y-x))| (\|y-x\|_\infty - |y_j-x_j|) \geq 0.$$

Since $0 = fy - fx = \psi(1) - \psi(0) = \int_0^1 \psi'(t)dt \geq 0,$

it follows that $\psi(t) \equiv 0$ for $t \in [0,1]$ which contradicts the fact that f is not constant on the line segment $[x,y]$.

Corollary 2.1.6 Let $F:D \subset R^n \to R^n$ be continuously differentiable on the convex set D, and assume that for every x in D, $F'(x)$ is a diagonally dominant matrix whose diagonal entries do not change sign. If for each $k \in N$, either

a) $|\partial_k f_k(x)| > \sum_{j\neq k} |\partial_j f_k(x)|$ for each x in D,

or

b) f_k is not constant on any line segment $[x,y]$ for which $x_k \neq y_k$.

Then F is strictly diagonally dominant on D.

Proof. For each $k \in N$, f_k satisfies the conditions of Theorem 2.1.5 and is therefore strictly diagonally dominant with respect to the kth variable.

The preceding result extends the class of functions which Theorem 2.1.4 identifies as strictly diagonally dominant.

Example 2.1.7 Let $F:R^2 \to R^2$ be defined by

$$F(x_1,x_2) = \begin{bmatrix} x_1 - \sin x_2 \\ x_2^3 \end{bmatrix} .$$

Direct computation shows that F satisfies the hypotheses of Corollary 2.1.6 and is therefore strictly diagonally dominant on R^2. Note, however, that $F'(x)$ is strictly diagonally dominant only if x_2 is not an even multiple of π.

Suppose now that $A \in L(R^n)$ is irreducibly diagonally dominant, but not strictly diagonally dominant; then $Fx = Ax$ is not a strictly diagonally dominant function. Since this type of matrix function arises frequently in practical situations, it is interesting to consider a corresponding extension of the diagonal dominance concept. We begin with an analog of Lemma 2.1.1.

Lemma 2.1.8 Let $v \in R^n$; then

a) $$|v_k| \geq \sum_{j \neq k} |v_j| \text{ for some } k \in N$$

if, and only if, for any $x \in R^n$,

b) $\sum_{j=1}^{m} v_j x_j = 0$, $x \neq 0$, implies that either $|x_k| < \|x\|_\infty$, or $|x_k| = \|x\|_\infty = |x_j|$ whenever $v_j \neq 0$.

Proof. Assume that a) holds and that $\sum_{j=1}^{n} v_j x_j = 0$ for some $x \neq 0$. If $|x_k| < \|x\|_\infty$ there is nothing to prove, hence suppose that $|x_k| = \|x\|_\infty$. Then

$$v_k x_k = -\sum_{j\neq k} v_j x_j,$$

and

$$\sum_{j\neq k} |v_j| \|x\|_\infty \leqslant |v_k| \|x\|_\infty = |v_k| |x_k| \leqslant \sum_{j\neq k} |v_j| |x_j| .$$

Thus,

$$\sum_{j\neq k} |v_j| (\|x\|_\infty - |x_j|) \leqslant 0,$$

which shows that $|x_j| = \|x\|_\infty$ whenever $v_j \neq 0$.

Conversely, if b) holds, and

$$|v_k| < \sum_{j\neq k} |v_j|, \tag{2.1.5}$$

then $\alpha |v_k| = \sum_{j\neq k} |v_j|$ for $\alpha > 1$. Define x in R^n by $x_k = \alpha \operatorname{sgn} v_k$, $x_j = -\operatorname{sgn} v_j$, $j \neq k$; then $\|x\|_\infty = \alpha = |x_k|$, and $\sum_{j=1}^{n} v_j x_j = 0$. By b), $|x_j| = \|x\|_\infty$ whenever $v_j \neq 0$, but since $|x_j| < \alpha = \|x\|_\infty$, we have $v_j = 0$ for all $j \neq k$. This contradicts (2.1.5).

Lemma 2.1.8 is the clue to generalizing the notion of a diagonally dominant matrix; we only need to specify the nonlinear counterpart of the condition $v_j \neq 0$ in b). To do this, we will use the well-known notion of a finite directed graph or network. For our purposes a network $\Omega = (N, \Lambda)$ consists of a set of n nodes

$N = \{1,\ldots,n\}$, and a set $\Lambda \subset N \times N$ of (directed) links which contain no loops, that is, $(i,i) \notin \Lambda$ if $i \in N$. A link from i to j is then an element (i,j) of Λ, and a (directed) path from i to j is a sequence of links of the form

$$(i,i_1),(i_1,i_2),\ldots,(i_r,j).$$

With these concepts in mind, we formulate our next definition.

Definition 2.1.9 A mapping $F:D \subset R^n \to R^n$ is diagonally dominant on D with respect to the network $\Omega = (N,\Lambda)$ if for every $x \neq y$ in D, $f_k x = f_k y$ for some k in N, implies that either a) $|x_k-y_k| < \|x-y\|_\infty$, or b) $|x_k-y_k| = \|x-y\|_\infty = |x_j-y_j|$ whenever $(k,j) \in \Lambda$.

Note that if F is diagonally dominant with respect to a network $\Omega = (N,\Lambda)$, then it is also diagonally dominant with respect to any sub-network $\Omega_0 = (N,\Lambda_0)$ in the sense that $\Lambda_0 \subset \Lambda$. The "largest" network that we will consider here is the associated network $\Omega_F = (N,\Lambda_F)$ of the mapping F. It is defined by $\Lambda_F = \{(i,j) \in N \times N$: $i \neq j$, and for some $x \in D$, $\alpha_{ij}(x,\cdot)$ is not constant$\}$ where $\alpha_{ij}(x,\cdot)$ is specified in Definition 1.2.3.

Theorem 2.1.10 Let $A \in L(R^n)$. Then A is a diagonally dominant matrix if, and only if, the induced mapping $Fx = Ax$ is diagonally dominant on R^n with respect to the associated network Ω_F.

Proof. The result follows directly from the definition and Lemma 2.1.8.

It should now be apparent how to generalize the notion of an irreducibly diagonally dominant matrix; instead, we shall generalize a weaker concept which, for a special class of off-diagonally antitone functions, seems to be due to Duffin [1948].

Definition 2.1.11 The mapping $F:D \subset R^n \to R^n$ is Ω-diagonally dominant on D if there is a network $\Omega = (N,\Lambda)$ such that

a) F is diagonally dominant with respect to the network $\Omega = (N,\Lambda)$, and

b) There is a nonempty subset $J(F)$ of N such that for each $i \in J(F)$, f_i is strictly diagonally dominant with respect to the ith variable, and for each $i \notin J(F)$, there is a path in Ω from i to some $j = j(i) \in J(F)$.

For linear $F:R^n \to R^n$ Walter [1967] has considered Definition 2.1.11 in a somewhat different form. More specifically, he defines A to satisfy condition Z_2 if A is diagonally dominant, and if there is a non-empty $J(A) \subset N$ such that for every non-empty proper subset L of N for which $L \cap J(A)$ is empty, there exists an $i \in L$ and a $j \notin L$ such that $a_{ij} \neq 0$. The equivalence of condition Z_2 and Definition 2.1.11 is a consequence of the next result.

Lemma 2.1.12 Let $\Omega = (N,\Lambda)$ be a network and J a non-empty subset of N. Then for each $i \notin J$ there is a path from i to some $j = j(i) \in J$ if, and only if, for every non-empty subset L of N such that $L \cap J$ is empty, there is an (i,j) in Λ with $i \in L$ and $j \notin L$.

Proof. Assume first that for each $i \notin J$ there is a path from i to some $j = j(i) \in J$, and let L be a non-empty subset of N such that $L \cap J$ is empty. Choose $i_0 \in L$; then $i_0 \notin J$ and hence, there is a path $(i_0,i_1),\ldots,(i_{r-1},i_r)$ to some $i_r \in J$. Let p be the first integer such that $i_p \notin L$, and note that $1 \leq p \leq r$ since $i_0 \in L$ and $i_r \notin L$. Then $(i_{p-1},i_p) \in \Lambda$ with $i_{p-1} \in L$ and $i_p \notin L$.

Conversely, let $i_0 \notin J$ be given. With $L = \{i_0\}$, our assumptions imply that there is an $i_1 \notin L$ with $(i_0,i_1) \in \Lambda$. If $i_1 \in J$, then (i_0,i_1) is the desired path; otherwise, set $L = \{i_0,i_1\}$ and note that $L \cap J$ is empty. Hence, there is an $i_2 \notin L$ with $(k_2,i_2) \in \Lambda$ for some $k_2 \in L = \{i_0,i_1\}$. Since J is a non-empty subset of N, the continuation of this process will yield an $i_p \notin L = \{i_0,i_1,\ldots,i_{p-1}\}$ with $(k_p,i_p) \in \Lambda$ for some $k_p \in L$ and such that $i_p \in J$. Our desired path is then $(i_0,k_2),(k_2,k_3),\ldots,(k_p,i_p)$ if $k_2 = i_1$; otherwise $k_2 = i_0$ and the path is (i_0,k_3), $(k_3,k_4),\ldots,(k_p,i_p)$.

Note that with $J(F) = \{1,\ldots,n\}$, Definition 2.1.11 reduces to the definition of a strictly diagonally dominant function. Therefore we will be able to develop the theory of strictly and Ω-diagonally dominant functions in parallel.

2.2 Basic Properties of Ω-diagonally Dominant Functions

In this section we will generalize some of the facts known about strictly and irreducibly diagonally dominant matrices (see, for example, Varga [1962]) to functions that satisfy Definitions 2.1.2 or 2.1.11.

Theorem 2.2.1 Let $F:D \subset R^n \to R^n$ be Ω-diagonally dominant on D. Then each subfunction of F is also Ω-diagonally dominant on D.

Proof. Let $L \subset N$ be a non-empty proper subset, and for ease of notation assume that $L = \{1,\ldots,k\}$, $1 \leq k \leq n$. Let $G:D_G \subset R^k \to R^k$ be the subfunction of F belonging to L with components

$$g_i(x_1,\ldots,x_k) = f_i(x_1,\ldots,x_k,z_{k+1},\ldots,z_n),\ i \in L.$$

In order to show that G is Ω-diagonally dominant, we must exhibit a network Ω_* and a non-empty subset $J(G)$ of L which satisfy Definition 2.1.11. Define

$$J(G) = \{i \in L:(i,j) \in \Lambda \text{ for some } j \notin L\} \cup (L \cap J(F)),$$

and set $\Omega_* = (L,\Lambda_*)$ where $\Lambda_* = \Lambda \cap (L\times L)$. We show first that $J(G)$ is not empty. For this assume that $L \cap J(F)$ is empty; otherwise there is nothing to prove. Then there exists an $i_0 \in L$ such that $i_0 \notin J(F)$ and a path

(2.2.1) $$(i_0,i_1),(i_1,i_2),\ldots,(i_{r-1},i_r)$$

connecting i_0 to some $i_r \in J(F)$. If p is the first integer such that $i_p \notin L$, then $1 \leqslant p \leqslant r$, since $i_0 \in L$ and $i_r \notin L$. Therefore, $J(G)$ is not empty, since necessarily $i_{p-1} \in J(G)$. Assume now that $i_0 \in L$ is any index for which $i_0 \notin J(G)$; then $i_0 \notin J(F)$, and thus there is a path (2.2.1) connecting i_0 to some $i_r \in J(F)$. If p is defined as above, then $1 \leqslant p \leqslant r$, and, since $i_0 \notin J(G)$, we have $p \neq 1$. Hence, $(i_0,i_1),\ldots,(i_{p-2},i_{p-1})$ is a path in L connecting i_0 to $i_{p-1} \in J(G)$. Clearly, G is diagonally dominant with respect to Ω_*, and hence, we only need to show that for $i \in J(G)$, f_i is strictly diagonally dominant with respect to the ith variable. For $i \in J(F)$, this is clear; hence assume that $i \notin J(F)$ and that for $x \neq y$ in D_G, $g_i x = g_i y$. Then

$$f_i(x_1,\ldots,x_k,z_{k+1},\ldots,z_n) = f_i(y_1,\ldots,y_k,z_{k+1},\ldots,z_n),$$

and consequently, $|x_i-y_i| < \|x-y\|_\infty$, for otherwise, $|z_j-z_j| = 0 = \|x-y\|_\infty$, since $(i,j) \in \Lambda$ for some $j \notin L$.

Note that for an irreducibly diagonally dominant matrix A, not every principal submatrix of A is irreducibly diagonally dominant. However, each principal submatrix is, at least, nonsingular as the following result shows:

Theorem 2.2.2 Assume that $F:D \subset R^n \to R^n$ is Ω-diagonally dominant on D. Then F and all of its subfunctions are injective.

Proof. In view of the previous result, it is sufficient to prove the assertion for F itself. For this, we will use the alternate definition of Ω-diagonally dominant functions given by Lemma 2.1.12. If $Fx = Fy$ for $x \neq y$ in D, define

$$L = \{i \in N: |x_i - y_i| = \|x-y\|_\infty\},$$

and note that L is a non-empty subset of N such that $L \cap J(F)$ is empty. By Lemma 2.1.12, there is an $(i,j) \in \Lambda$ with $i \in L$ and $j \notin L$. But $f_i x = f_i y$ and $|x_i - y_i| = \|x-y\|_\infty$ implies that $|x_j - y_j| = \|x-y\|_\infty$ since $(i,j) \in \Lambda$, contradicting the fact that $j \in L$.

For our next result, recall that when $F:D \subset R^n \to R^n$ is given, then for each $x \in R^n$, the ith diagonal function

$$\alpha_{ii}(x,\cdot):\{t \in R^1 : x+te^i \in D\} \to R^1$$

is defined by $\alpha_{ii}(x,t) = f_i(x+te^i)$ for $i \in N$.

Theorem 2.2.3 Assume that $F:D \subset R^n \to R^n$ is continuous and Ω-diagonally dominant on the rectangle D. Then for any fixed $k \in N$, $\alpha_{kk}(x,\cdot)$ is, for arbitrary x in D, either strictly isotone or strictly antitone. Moreover, if for some $y \neq x$ in D, $|x_k - y_k| = \|x-y\|_\infty$ and $f_k x \neq f_k y$, then $(x_k - y_k)(f_k x - f_k y) > 0$ if $\alpha_{kk}(x,\cdot)$ is isotone, and $(x_k - y_k)(f_k x - f_k y) < 0$ if $\alpha_{kk}(x,\cdot)$ is antitone.

Proof. Let $x \in D$ and $k \in N$ be given. By Theorem 2.2.2, $\alpha_{kk}(x,\cdot)$ is injective on its domain of definition. Since the domain of $\alpha_{kk}(x,\cdot)$ is an interval I_k and $\alpha_{kk}(x,\cdot)$ is continuous, $\alpha_{kk}(x,\cdot)$

must be either strictly isotone or strictly antitone. Assume that $\alpha_{kk}(x,\cdot)$ is strictly isotone. We now claim that for any $y \in D$, $\alpha_{kk}(y,\cdot)$ is also strictly isotone. To prove this, let $D_0 = \{y \in D: \alpha_{kk}(y,\cdot)$ is strictly isotone on $I_k\}$, and note that $y \in D_0$ if and only if $\alpha_{kk}(y,s) > \alpha_{kk}(y,t)$ for at least one pair of real numbers $s > t$ with $y_k + s$ and $y_k + t$ in I_k. Of course, if I_k is only one point, then trivially $D_0 = D$; hence assume that I_k is a non-degenerate interval. We observe now that D_0 is not empty since $x \in D_0$ and show that D_0 is closed in D. For this purpose let $y \in D$ be a limit point of D_0 and $\{y^m\} \subset D_0$ such that $\lim_{m\to+\infty} y^m = y$. Since I_k is a non-degenerate interval, and $\{y^m\} \subset D$, there is a $\delta \neq 0$ such that $y_k^m + \delta$ belongs to I_k for sufficiently large m. For the sake of being definite, assume that $\delta > 0$. Then $\alpha_{kk}(y^m,\delta) > \alpha_{kk}(y^m,0)$ and hence, $\alpha_{kk}(y,\delta) \geqslant \alpha_{kk}(y,0)$ since F is continuous. But $\alpha_{kk}(y,\cdot)$ is injective and therefore, $\alpha_{kk}(y,\delta) > \alpha_{kk}(y,0)$, which in turn implies that $y \in D_0$. On the other hand, D_0 is also open in D, for if $y \in D_0$, there is a $\delta \neq 0$ (which is again assumed to be positive) such that $y_k + \delta$ belongs to I_k and consequently $\alpha_{kk}(y,\delta) > \alpha_{kk}(y,0)$. Continuity of F now implies that $\alpha_{kk}(\hat{y},\delta) > \alpha_{kk}(\hat{y},0)$ for all $\hat{y}$ in a sufficiently small relative neighborhood of y in D, and therefore $\hat{y} \in D_0$.

We have now shown that D_0 is a non-empty subset of D which is both open and closed in D. Since D is connected, this implies $D_0 = D$ as desired. To finish the proof, we must show that if $y \in D$

is such that $y \neq x$ $|y_k - x_k| = \|y-x\|_\infty$ and $f_k y \neq f_k x$ then $(x_k - y_k)(f_k x - f_k y) > 0$. In order to arrive at a contradiction, assume that $x_k > y_k$ and $f_k x < f_k y$. Define $g:[0,1] \to D$ by $g_k(t) = x_k$ and $g_i(t) = ty_i + (1-t)x_i$ for $i \neq k$. Then $|g_j(t) - y_j| < \|g(t)-y\|_\infty$ for each $t \in (0,1]$ and $j \neq k$. Since F is Ω-diagonally dominant, we must have $f_k g(t) \neq f_k y$ for each $t \in (0,1]$, and, since $f_k g(t) < f_k y$ for $t = 0$, the continuity of F implies that $f_k g(t) < f_k y$ for $t \in [0,1]$. Hence, $f_k g(1) = f_k[y + (x_k - y_k)e^k] \leqslant f_k y$, contradicting the assumptions that $\alpha_{kk}(y,\cdot)$ is strictly isotone and $x_k > y_k$.

<u>Corollary 2.2.4</u> Let $F:D \subset R^n \to R^n$ be continuous on the rectangle D. Then F is strictly diagonally dominant on D if, and only if, the following two conditions hold:

a) For each fixed $k \in N$, $\alpha_{kk}(x,\cdot)$ is, for arbitrary $x \in D$, either strictly isotone or strictly antitone.

b) If $|y_k - x_k| = \|y-x\|_\infty$ for $y \neq x$ in D and $k \in N$, then $(x_k - y_k)(f_k x - f_k y) > 0$, if $\alpha_{kk}(x,\cdot)$ is isotone, and $(x_k - y_k)(f_k x - f_k y) < 0$, if $\alpha_{kk}(x,\cdot)$ is antitone.

<u>Proof</u>. If F is strictly diagonally dominant, then it is also Ω-diagonally dominant, and the sufficiency of the two conditions follows by noting that if for some $y \neq x$ in D we have $|x_k - y_k| = \|x-y\|_\infty$, then necessarily, $f_k x = f_k y$. The necessity of the condition is clear.

Corollary 2.2.5 Let $F:D \subset R^n \to R^n$ be continuous and Ω-diagonally dominant on the rectangle D. Then F is a P-function if and only if F is diagonally isotone on D.

Proof. If F is a P-function, then F is diagonally isotone on D by Theorem 1.3.3. Conversely, if $x \neq y$ are given, we must show that there is a $k \in N$ such that $(x_k-y_k)(f_kx-f_ky) > 0$. Let

$$L = \{i \in N: f_ix=f_iy, |x_i-y_i|=\|x-y\|_\infty\},$$

and note that if L is empty, then there is necessarily a $k \in N$ such that $|x_k-y_k| = \|x-y\|_\infty$ and $f_kx \neq f_ky$. Theorem 2.2.3 then implies that $(x_k-y_k)(f_kx-f_ky) > 0$. Otherwise, L is a non-empty subset of N such that $L \cap J(F)$ is empty. Since F is Ω-diagonally dominant, Lemma 2.1.12 yields an $(i,j) \in \Lambda$ with $i \in L$ and $j \notin L$. It follows that $|x_j-y_j| = \|x-y\|_\infty$, and since $j \notin L$, we have $f_jx \neq f_jy$. Theorem 2.2.3 now shows that $(x_j-y_j)(f_jx-f_jy) > 0$.

The last result we prove in connection with Theorem 2.2.3 represents a converse of Theorems 2.1.4 and 2.1.6.

Corollary 2.2.6 Assume that $F:D \subset R^n \to R^n$ is continuous, G-differentiable, and Ω-diagonally dominant on the open set D. Then $F'(x)$ is a diagonally dominant matrix for each x in D.

Proof. Let $x \in D$ be given. Since F is continuous on the open set D, we can apply Theorem 2.2.3 to any open rectangle $D_0 \subset D$

containing x. Hence, for any given $k \in N$, $\alpha_{kk}(x,\cdot)$ is either strictly isotone or strictly antitone, and--to be definite--assume that $\alpha_{kk}(x,\cdot)$ is strictly isotone. Now let u be the vector with the components $u_k = 1$ and $u_j = -\operatorname{sgn} \partial_j f_k(x)$ for $j \neq k$, and let $\delta > 0$ be such that $x + tu \in D_0$ for $t \in [0,\delta)$. Then, for $t \in (0,\delta)$, $t = \|x+tu-x\|_\infty$, and therefore

$$t[f_k(x+tu)-f_k(x)] \geqslant 0.$$

Dividing by $t^2 > 0$ and passing to the limit as $t \to 0^+$, we obtain,

$$f_k'(x)u = \sum_{j=1}^{n} \partial_j f_k(x)u_j \geqslant 0. \tag{2.2.2}$$

But $\partial_k f_k(x) \geqslant 0$ since $\alpha_{kk}(x,\cdot)$ is strictly isotone and hence (2.2.2) is equivalent to

$$|\partial_k f_k(x)| \geqslant \sum_{j\neq k} |\partial_j f_k(x)|.$$

To conclude this section, we investigate the relationship between Ω-diagonally dominant functions and M-functions. The functions to be considered will be assumed to be defined on a set of the form

$$D = \prod_{i=1}^{n} I_i, \tag{2.2.3}$$

where each I_i is an interval of the form $(\alpha_i,+\infty)$ or $[\alpha_i,+\infty)$. In the first case, $\alpha_i = -\infty$ is permitted; otherwise, α_i is real. We set $R^1_+ = \{t \in R^1 : t \geqslant 0\}$.

Theorem 2.2.7 Let $F:D \subset R^n \to R^n$ be a continuous, off-diagonally antitone function on a set D of the form (2.2.3). The following three statements are then equivalent, and in either case F is an M-function.

a) F is a diagonally isotone, strictly diagonally dominant function on D.

b) For each x in D, $F(x+te)$ is strictly isotone as a function of $t \in R^1_+$.

c) For any $x \neq y$ in D, $(x_k-y_k)(f_kx-f_ky) > 0$ whenever $|x_k-y_k| = \|x-y\|_\infty$.

Proof: By Corollary 2.2.4, a) and c) are equivalent, and moreover, c) implies that F is a P-function. Since F is off-diagonally antitone, F is an M-function by Theorem 1.2.5, and thus, to conclude the proof, it suffices to show that b) and c) are equivalent.

If F satisfies b), and for some $x \neq y$ in D we have $|x_k-y_k| = \|x-y\|_\infty$, then $x_k \neq y_k$. If $y_k > x_k$, then b) together with the off-diagonal antitonicity of F implies that $f_kx < f_k(x+(y_k-x_k)e) \leq f_ky$. Similarly, if $x_k > y_k$ we obtain that $f_ky < f_k(y+(x_k-y_k)e) \leq f_kx$. In either case, $(x_k-y_k)(f_kx-f_ky) > 0$ and c) is satisfied. On the other hand, if F satisfies c) and $s > t \geq 0$, then $s - t = \|(x+se)-(x+te)\|_\infty$ for any $x \in D$. Hence,

$$(s-t)[f_k(x+se)-f_k(x+te)] > 0$$

for each $k \in N$ and therefore $F(x+se) > F(x+te)$.

If F is not necessarily strictly diagonally dominant, we have the following result.

Theorem 2.2.8 Let $F: D \subset R^n \to R^n$ be off-diagonally antitone on a set D of the form (2.2.3), and suppose that for any $x \in D$, $F(x+te)$ is an isotone function for t on R^1_+. Assume further that

$$J(F) = \{j \in N : f_j(x+te) \text{ is strictly isotone in } t \text{ on } R^1_+ \text{ for any } x \in D\}$$

is not empty, and define $\Omega = (N, \Lambda)$ by

$$\Lambda = \{(i,j) \in N \times N : i \neq j,\ \alpha_{ij}(x,t) = f_i(x+te^j) \text{ is strictly antitone in } t \text{ on } R^1_+ \text{ for any } x \in D\}.$$

If for any $i \notin J(F)$ there is a path in Ω from i to some $j \in J(F)$, then F is Ω-diagonally dominant, and hence, an M-function on D.

Proof. We prove first that F is diagonally dominant on D. Let $x \neq y$ in D be given, and suppose that $f_k x = f_k y$. For $|x_k - y_k| < \|x-y\|_\infty$ there is nothing to prove; thus, assume that $|x_k - y_k| = \|x-y\|_\infty$. Then $x_k \neq y_k$, and without loss of generality, we may take $x_k > y_k$. If $(k,j) \in \Lambda$ but $|x_k - y_k| = \|x-y\|_\infty > |x_j - y_j|$, then $x_k - y_k > x_j - y_j$ and $x_k - y_k \geq x_i - y_i$ for $i \neq j$. Since

$\alpha_{kj}(y,\cdot)$ is strictly antitone and $F(y+te)$ isotone,

$$f_k y \leq f_k(y+(x_k-y_k)e) < f_k x,$$

which is a contradiction. Hence, $x_k-y_k = x_j-y_j$ whenever $(k,j) \in \Lambda$ and therefore F is diagonally dominant on D.

To show that F is Ω-diagonally dominant on D, we only need to prove that for $k \in J(F)$, f_k is strictly diagonally dominant with respect to the kth variable. Let $x \neq y$ in D be given, and assume that $f_k x = f_k y$ with $|x_k-y_k| = \|x-y\|_\infty$. If we let $x_k > y_k$, it follows that

$$f_k y < f_k(y+(x_k-y_k)e) \leq f_k x,$$

since $k \in J(F)$ and F is off-diagonally antitone. A contradiction is also reached if we take $x_k < y_k$; hence, $|x_k-y_k| < \|x-y\|_\infty$ and F is therefore Ω-diagonally dominant on D.

Finally, it will follow from Theorem 1.2.5 and Corollary 2.2.5 that F is an M-function if F is shown to be diagonally isotone on D. To do this, let $x \in D$ and $k \in N$ be given, and suppose that $s > t$ with $x + te^k$ and $x + se^k$ in D. Since F is off-diagonally antitone and $F(y+\hat{t}e)$ is isotone in $\hat{t} \geq 0$ for any $y \in D$, we have

$$f_k(x+te^k) \leq f_k(x+te^k+(s-t)e) \leq f_k(x+se^k),$$

which is the desired result.

For $D = R^n$, this last result is a special case of a theorem of Rheinboldt [1969b]. His proof, however, was direct and did not use the concept of Ω-diagonal dominance.

2.3 Convergence Theorems

In this section we will see how the nonlinear generalizations of diagonal dominance allow us to extend the following classical result: If $A \in L(R^n)$ is a strictly or irreducibly diagonally dominant matrix, then for every b in R^n, $Ax = b$ has a unique solution x^*; and for any x^0 in R^n, the Jacobi and Gauss-Seidel sequences converge to x^*.

The following convergence proofs are somewhat long, but the ideas behind them are rather simple. Specifically, we will define an iteration function H for the Jacobi and Gauss-Seidel sequences, which will allow us to represent these implicit iterative methods as explicit iterative schemes $x^{k+1} = Hx^k$, $k = 0,1,\ldots$. The iteration function H will then be shown to satisfy the hypotheses of the next result which is essentially due to Diaz and Metcalf [1968].

Lemma 2.3.1 Let $H:D_0 \subset R^n \to R^n$ map the closed set D_0 into itself, and suppose that H has a fixed point x^* in D_0. If for some $m \geq 1$,

$$\|H^m x - H^m y\| < \|x-y\|, \text{ whenever } x,y \in D,\ x \neq y, \tag{2.3.1}$$

then x^* is the only fixed point of H in D_0, and for any $x^0 \in D_0$, the sequence $x^{k+1} = Hx^k$ converges to x^*.

Proof. Assume for the moment that $m = 1$. Then (2.3.1) implies that x^* is the only fixed point of H in D_0, and that $\varepsilon_k = \|x^k - x^*\|$ is a decreasing sequence of nonnegative numbers and hence convergent. Thus $\{x^k\}$ is bounded, and if $\{x^{k_i}\}$ is any convergent subsequence such that $\lim_{i\to+\infty} x^{k_i} = y^* \neq x^*$, then

$$\lim_{i\to+\infty} \varepsilon_{k_i+1} = \|Hy^*-x^*\| = \|Hy^*-Hx^*\| < \|y^*-x^*\| = \lim_{i\to+\infty} \varepsilon_{k_i},$$

which contradicts the fact that $\{\varepsilon_k\}$ is convergent. Therefore, $\lim_{i\to+\infty} x^{k_i} = x^*$, and consequently, $\lim_{k\to+\infty} x^k = x^*$. If $m > 1$, then (2.3.1) implies that H^m and H have the same number of fixed points. Moreover, the previous argument applied to H^m implies that $y^{k+1} = H^m y^k$ converges to x^* for any $y^0 \in D_0$. Setting y^0 successively equal to $x^0, \ldots, H^{m-1}x^0$, we obtain the desired result.

The hypothesis that H has a fixed point cannot be completely removed as shown by the one-dimensional example $h(x) = \ell n(1+e^x)$; but as noted by many authors, it can be replaced by the boundedness of D_0.

Lemma 2.3.2 Let $H:D_0 \subset R^n \to R^n$ map the compact set D_0 into itself, and assume that for some $m \geq 1$, H satisfies (2.3.1). Then H has a unique fixed point x^* in D_0, and for any $x^0 \in D_0$ the sequence $x^{k+1} = Hx^k$ converges to x^*.

Proof. Assume first that $m = 1$, and define $g:D_0 \to R^1$ by $g(x) = \|Hx-x\|$. Since D_0 is compact and g is continuous on D_0, there is a $x^* \in D_0$ such that $g(x^*) \leq g(x)$ for each $x \in D_0$. If $Hx^* \neq x^*$, (2.3.1) implies that

$$\|H^2x^*-Hx^*\| < \|Hx^*-x^*\|,$$

that is, $g(Hx^*) < g(x^*)$ contradicting the definition of x^*. Hence, x^* is a fixed point of H. If $m > 1$, then the previous argument yields that H^m has a fixed point x^*. However, (2.3.1) implies that the fixed points of H and H^m are the same, and thus, $Hx^* = x^*$. The rest of the proof follows from Lemma 2.3.1.

We now prove that under suitable assumptions, the Jacobi and Gauss-Seidel sequences (2.1.1)-(2.1.4) are well-defined and are given by an iteration function which satisfies (2.3.1).

Theorem 2.3.3 Let $F:D_0 \subset R^n \to R^n$ be Ω-diagonally dominant on the rectangle D_0, and suppose that for each x in D_0 and $i \in N$, the one-dimensional equation

$$f_i(x_1,\ldots,x_{i-1},t,x_{i+1},\ldots,x_n) = 0$$

has a (necessarily unique) solution t_i^* with $(x_1,\ldots,x_{i-1},t_i^*,x_{i+1},\ldots,x_n)^T$ in D_0. Then the Jacobi and Gauss-Seidel sequences (2.1.1)-(2.1.4) with $\omega_k \equiv \omega \in (0,1]$, $k = 0,1,\ldots$, are well-defined for any $x^0 \in D_0$. Moreover, for either method there is an iteration function $H:D_0 \subset R^n \to R^n$ such that

a) The method is equivalent with $x^{k+1} = Hx^k$, $k = 0,1,\ldots$,

b) $H(D_0) \subset D_0$,

c) $\|Hx-Hy\|_\infty \leqslant \|x-y\|_\infty$ for every x,y in D_0, and

d) $\|H^{n-\ell+1}x - H^{n-\ell+1}y\|_\infty < \|x-y\|_\infty$ for every $x \neq y$ in D_0 where ℓ denotes the number of elements in $J(F)$.

Proof. We will first present the proof for the Gauss-Seidel method.

Let $x \in D_0$ be given, and define the iteration function $H:D_0 \subset R^n \to R^n$ of the Gauss-Seidel method as follows: By assumption there is t_1^* -- which by Theorem 2.2.2 is unique -- such that

$$f_1(t_1^*,x_2,\ldots,x_n) = 0$$

and $(t_1^*,x_2,\ldots,x_n)^T \in D_0$. Set $h_1(x) = (1-\omega)x_1 + \omega t_1^*$, and note that since D_0 is convex, $(h_1(x),x_2,\ldots,x_n)^T \in D_0$. Assume that $h_j(x)$ for $j = 1,\ldots,i-1$ have been defined such that $(h_1(x),\ldots,h_{i-1}(x), x_i,\ldots,x_n)^T \in D_0$. Once again, there is a unique t_i^* such that

$$f_i(h_1(x),\ldots,h_{i-1}(x),\ t_i^*,x_{i+1},\ldots,x_n) = 0,$$

and we set $h_i(x) = (1-\omega)x_i + \omega t_i^*$. In this way, we have defined $H:D_0 \subset R^n \to R^n$ such that $H(D_0) \subset D_0$, and that for any $x^0 \in D_0$, the Gauss-Seidel method is well-defined and equivalent with $x^{k+1} = Hx^k$.

To show that H satisfies c) and d), we first prove that for every $x \neq y$ in D_0 and $i \in N$, either

$$|h_i(x)-h_i(y)| < \|x-y\|_\infty, \tag{2.3.2}$$

or

$$|h_i(x)-h_i(y)| = \|x-y\|_\infty, \text{ and,}$$

(2.3.3) $$|h_i(x)-h_i(y)| = |h_j(x)-h_j(y)| \text{ if } (i,j) \in \Lambda,\ i > j,$$

$$|h_i(x)-h_i(y)| = |x_j-y_j| \text{ if } (i,j) \in \Lambda,\ i < j.$$

The proof is by induction. Set $\hat{x} = (t_1{*}(x), x_2, \ldots, x_n)^T$ where $f_1(\hat{x}) = 0$, and similarly for $\hat{y}$. Since

(2.3.4) $$|h_1(x)-h_1(y)| \leq (1-\omega)|x_1-y_1| + \omega|\hat{x}_1-\hat{y}_1|,$$

$\hat{x}_1 = \hat{y}_1$ implies that $|h_1(x)-h_1(y)| < \|x-y\|_\infty$ and (2.3.2) applies. If $\hat{x}_1 \neq \hat{y}_1$, then $\hat{x} \neq \hat{y}$, and, since $f_1(\hat{x}) = f_1(\hat{y})$ and F is Ω-diagonally dominant, either $|\hat{x}_1-\hat{y}_1| < \|\hat{x}-\hat{y}\|_\infty \leq \|x-y\|_\infty$ and (2.3.4) implies that (2.3.2) occurs; or $|\hat{x}_1-\hat{y}_1| = \|\hat{x}-\hat{y}\|_\infty = |x_j-y_j|$ for $j > 1$, and the second part of (2.3.3) holds. Since $\|\hat{x}-\hat{y}\|_\infty = |x_j-y_j| \leq \|x-y\|_\infty$, (2.3.4) yields that $|h_1(x)-h_1(y)| \leq \|x-y\|_\infty$ as desired.

Assume now that (2.3.2) and (2.3.3) hold for $i = 1,\ldots,k-1$, and set $\hat{x} = (h_1(x),\ldots,h_{k-1}(x), t_k^*(x), x_{k+1},\ldots,x_n)^T$ where $f_k(\hat{x}) = 0$, and similarly for $\hat{y}$. If $\hat{x}_k = \hat{y}_k$, the result follows from

(2.3.5) $$|h_k(x)-h_k(y)| \leq (1-\omega)|x_k-y_k| + \omega|\hat{x}_k-\hat{y}_k|,$$

and if $\hat{x}_k \neq \hat{y}_k$, then $\hat{x} \neq \hat{y}$. Since $f_k(\hat{x}) = f_k(\hat{y})$ and F is Ω-diagonally dominant, either $|\hat{x}_k-\hat{y}_k| < \|\hat{x}-\hat{y}\|_\infty \leq \|x-y\|_\infty$ and (2.3.5) implies that (2.3.2) holds, or $|\hat{x}_k-\hat{y}_k| = \|\hat{x}-\hat{y}\|_\infty = |\hat{x}_j-\hat{y}_j|$ for any $(k,j) \in \Lambda$. If $k < j$ the third part of (2.3.3) takes place, while if $k > j$, the second part holds. In either case, $|\hat{x}_k-\hat{y}_k| \leq \|x-y\|_\infty$, and

(2.3.5) yields $|h_k(x)-h_k(y)| \leq \|x-y\|_\infty$.

Note that (2.3.2) and (2.3.3) together imply that $\|H(x)-H(y)\|_\infty \leq \|x-y\|_\infty$, and thus, we only need to verify that d) holds. For the proof we will use the notation $h_i^k(x) \equiv h_i(H^{k-1}(x))$ where $k \geq 1$ and $i \in N$; it will also be important to note that (2.3.2) applies whenever $i \in J(F)$. Assume for the moment that $\ell = 1$, and let $x \neq y$ in D_0 be given. If $H^{n-1}(x) = H^{n-1}(y)$, then $\|H^n(x)-H^n(y)\|_\infty < \|x-y\|_\infty$; otherwise $H^{n-1}(x) \neq H^{n-1}(y)$. Let $i \in N$ be given; we prove that $|h_i^n(x)-h_i^n(y)| < \|x-y\|_\infty$. If $|h_i^n(x)-h_i^n(y)| < \|H^{n-1}(x)-H^{n-1}(y)\|_\infty$ there is nothing to prove; otherwise $i \notin J(F)$ and there is a path

(2.3.6) $$(i,i_1),(i_1,i_2),\ldots,(i_{r-1},i_r),$$

with $i_r = j \in J(F)$ and $r \leq n-1$. Hence,

$$|h_i^n(x)-h_i^n(y)| = |h_{i_1}^n(x)-h_{i_1}^n(y)| \text{ if } i > i_1,$$

or

$$|h_i^n(x)-h_i^n(y)| = |h_{i_1}^{n-1}(x)-h_{i_1}^{n-1}(y)| \text{ if } i < i_1.$$

Repeat the procedure until $|h_i^n(x)-h_i^n(y)| < \|x-y\|_\infty$, or $|h_i^n(x)-h_i^n(y)| = |h_{i_r}^k(x)-h_{i_r}^k(y)|$ for some k with $1 \leq n-r \leq k \leq n$. Since $i_r = j \in J(F)$,

$$|h_j^k(x)-h_j^k(y)| < \|H^{k-1}(x)-H^{k-1}(y)\|_\infty \leq \|x-y\|_\infty.$$

Hence, $|h_i^n(x)-h_i^n(y)| < \|x-y\|_\infty$ for each $i \in N$, and thus,

$\|H^n(x)-H^n(y)\|_\infty < \|x-y\|_\infty$.

If $1 < \ell \leq n$, and $H^{n-\ell}(x) = H^{n-\ell}(y)$, there is nothing to prove; otherwise, $H^{n-\ell}(x) \neq H^{n-\ell}(y)$, and the proof proceeds as before by noting that each $i \notin J(F)$ can be joined to an $i_r \in J(F)$ by a path (2.3.6) with $r \leq n-\ell$.

The proof for the Jacobi method is very similar, but now full use is made of the assumption that D_0 is a rectangle. The distinction occurs in the definition of the iteration function H for the Jacobi method. The first component function of H is defined by $h_1(x) = (1-\omega)x_1 + \omega t_1^*$ where

$$f_1(t_1^*,x_2,\ldots,x_n) = 0$$

and $(t_1^*,x_2,\ldots,x_n)^T \in D_0$. Assume that $h_j(x)$ for $j = 1,\ldots,i-1$ have been defined such that $(x_1,\ldots,x_{j-1},h_j(x),x_{j+1},\ldots,x_n)^T \in D_0$ for $j = 1,\ldots,i-1$. If we set $h_i(x) = (1-\omega)x_i + \omega t_i^*$ where

$$f_i(x_1,\ldots,x_{i-1},t_i^*,x_{i+1},\ldots,x_n) = 0$$

and $(x_1,\ldots,x_{i-1},t_i^*,x_{i+1},\ldots,x_n)^T \in D_0$, then $(x_1,\ldots,x_{i-1},h_i(x),x_{i+1},\ldots,x_n)^T \in D_0$ since D_0 is convex, and, since D_0 is a rectangle, $(h_1(x),\ldots,h_i(x),x_{i+1},\ldots,x_n)^T \in D_0$. In this way the iteration function for the Jacobi method is defined, and it satisfies b). The rest of the proof proceeds along steps similar to those for the Gauss-Seidel sequence. This completes the proof.

If $F: R^n \to R^n$ is linear and irreducibly diagonally dominant, Mařík and Pták [1960] have proved a result related to d) of Theorem 2.3.3 in the sense that their theorem can be used to prove that for the iteration matrix H of the Jacobi method, $\|H^{n-\ell+1}\|_\infty < 1$. A consequence of the next example is that $n - \ell$ is the largest number k such that $\|H^k\|_\infty = 1$.

Example 2.3.4 Consider the matrix of order $n > 1$:

$$A = \begin{bmatrix} 1 & -1 & & & & & & \\ 0 & 1 & -1 & & & & \bigcirc & \\ & & \ddots & \ddots & & & & \\ & & & 1 & -1 & & & \\ & \bigcirc & & & 1 & -\alpha & & \\ & & & & & \ddots & \ddots & \\ & & & & & & 1 & -\alpha \\ -\alpha & 0 & \cdots & & & \cdots & 0 & 1 \end{bmatrix}, \quad |\alpha| < 1,$$

where there are ℓ alphas. Then $J(A) = \ell$, and the iteration matrix for the Jacobi method is given by

$$H = \begin{bmatrix} 0 & 1 & & & & & & \\ & 0 & 1 & & & & \bigcirc & \\ & & \ddots & \ddots & & & & \\ & & & 0 & 1 & & & \\ & \bigcirc & & & 0 & \alpha & & \\ & & & & & \ddots & \ddots & \\ & & & & & & 0 & \alpha \\ \alpha & 0 & \cdots & & & \cdots & & 0 \end{bmatrix}.$$

By direct computation, $\|H^k\|_\infty = 1$ for $0 \leq k \leq n-\ell$, and $\|H^{n-\ell+1}\|_\infty = |\alpha| < 1$. Note that if $\alpha = 0$, A is not irreducibly diagonally dominant but Ω-diagonally dominant with respect to the associated network Ω_A.

With the help of the previous theorem, we can now prove our first convergence result.

Corollary 2.3.5 Assume that $F:D_0 \subset R^n \to R^n$ satisfies the hypotheses of Theorem 2.3.3 on the closed rectangle D_0. If either D_0 is bounded or $Fx = 0$ has a solution in D_0, then $Fx = 0$ has a unique solution x^* in D_0; and for any $x^0 \in D_0$, the Jacobi and Gauss-Seidel sequences (2.1.1)-(2.1.4) with $\omega_k \equiv \omega \in (0,1]$, $k = 0,1,\ldots,$ are well-defined and converge to x^*.

Proof. We only carry out the proof for the Jacobi method; the proof for the Gauss-Seidel method is similar.

Suppose first that D_0 is bounded. By Theorem 2.3.3, the Jacobi method has a well-defined iteration function $H:D_0 \subset R^n \to R^n$ which satisfies (2.3.1) with $m = n - \ell + 1$. Since D_0 is compact, Lemma 2.3.2 implies that H has a fixed point x^* in D_0. But

$$f_i(x_1,\ldots,x_{i-1},(1 - \frac{1}{\omega})x_i + \frac{1}{\omega} h_i(x), x_{i+1},\ldots,x_n) = 0$$

for each $i \in N$, and therefore, $Fx^* = 0$. The uniqueness of x^* follows from Theorem 2.2.2 while the convergence of the Jacobi iterates to x^* is a consequence of Lemma 2.3.2 and Theorem 2.3.3.

If $Fx = 0$ is assumed to have a solution in D_0, then this solution is unique by Theorem 2.2.2, and the result now follows from Lemma 2.3.1 and Theorem 2.3.3.

An important case of Corollary 2.3.5 occurs if $F:R^n \to R^n$ is Ω-diagonally dominant on all of R^n. In this case the assumption that the one-dimensional equation

$$f_i(x_1,\ldots,x_{i-1},t,x_{i+1},\ldots,x_n) = 0$$

has a solution t_i^* for each $i \in N$ and $x \in R^n$ is essentially "diagonal surjectivity". We make this precise.

<u>Definition 2.3.6</u> The mapping $F:R^n \to R^n$ is diagonally surjective if for each $x \in R^n$ and $k \in N$, $\alpha_{kk}(x,\cdot):R^1 \to R^1$ is surjective. Here, as usual, $\alpha_{kk}(x,t) = f_k(x+te^k)$.

In connection with Corollary 2.3.5 note that the next example shows that if $F:R^n \to R^n$ is a diagonally surjective, Ω-diagonally dominant function on R^n, then $Fx = 0$ does not necessarily have a solution.

<u>Example 2.3.7</u> Define $F:R^2 \to R^2$ by

$$F(x_1,x_2) = \begin{bmatrix} x_1 - x_2 + g(x_1) \\ \\ x_2 - x_1 + g(x_2) \end{bmatrix} ,$$

where $g(t) = \arctan t - \pi/2$. By Theorem 2.2.7, F is a strictly diagonally dominant M-function. However, $Fx = 0$ does not have a solution, for otherwise, there would be an $x = (x_1, x_2)^T$ such that $F(t,t) \leq 0 = F(x_1, x_2)$ and, since F is inverse isotone, $t \leq x_1$, $t \leq x_2$ for every $t \in R^1$. This is clearly impossible.

In contrast with this example and Corollary 2.3.5, the next result shows, in particular, that if $F: R^n \to R^n$ is Ω-diagonally dominant on all of R^n and $Fx = 0$ has a solution x^*, then the Jacobi and Gauss-Seidel iterates are well-defined and converge to x^* provided F is continuous on R^n.

Theorem 2.3.8 Let $F: D \subset R^n \to R^n$ be continuous on the set D, and assume that $Fx = 0$ has a solution x^* in D, and that for some $r \geq 0$, $D_0 = \{x \in R^n : \|x - x^*\|_\infty \leq r\} \subset D$. If F is Ω-diagonally dominant on D_0, then x^* is unique in D_0, and for any x^0 in D_0 the Jacobi and Gauss-Seidel sequences (2.1.1)-(2.1.4) with $\omega_k \equiv \omega \in (0,1]$, $k = 0,1,\ldots$, are well-defined and converge to x^*.

Proof. The result will follow from Theorem 2.3.4 if we prove that for each x in D_0 and $i \in N$, the equation

$$f_i(x_1, \ldots, x_{i-1}, t, x_{i+1}, \ldots, x_n) = 0$$

has a unique solution t_i^* with $(x_1, \ldots, x_{i-1}, t_i^*, x_{i+1}, \ldots, x_n)^T \in D_0$. To show this, let $x \in D_0$ and $i \in N$ be given, and define $\psi: R^1 \to R^1$ by

$$\psi(t) = f_i(x_1,\ldots,x_{i-1},t,x_{i+1},\ldots,x_n).$$

Clearly, ψ is defined for $|t-x_i^*| \leq r$, and by Theorem 2.2.3, ψ is either strictly isotone or strictly antitone. In either case, $\psi(t) = 0$ has at most one solution. We now proceed to show that such a solution exists if ψ is strictly isotone; for the strictly antitone case the proof is analogous. Since $x \in D_0$, $\|x-x^*\|_\infty \equiv \rho \leq r$, and if $\rho = 0$ there is nothing to prove; hence assume that $\rho > 0$ and let $t_i^+ = x_i^* + \rho$. If $v = (x_1,\ldots,x_{i-1},t_i^+,x_{i+1},\ldots,x_n)^T$, then $|t_i^+-x_i^*| = \|v-x^*\|_\infty$, and by Theorem 2.2.3,

$$(t_i^+-x_i^*)[f_iv-f_ix^*] = \rho\psi(t_i^+) \geq 0,$$

or $\psi(t_i^+) \geq 0$. Similarly, if $t_i^- = x_i^* - \rho$, then $\psi(t_i^-) \leq 0$. The continuity of ψ yields a $t_i^* \in [t_i^-,t_i^+]$ with $\psi(t_i^*) = 0$, and since $|t_i^*-x_i^*| \leq r$, it follows that $(x_1,\ldots,x_{i-1},t_i^*,x_{i+1},\ldots,x_n)^T \in D_0$. This completes the proof.

If $F:R^n \to R^n$ is linear and Ω-diagonally dominant on R^n with respect to the associated network Ω_F, then F is necessarily surjective; and in the particular case where $\omega = 1$, the previous theorem is due to Walter [1967]. If F is not defined on all of R^n, then, in general, it is very difficult to find a set D_0 which satisfies the hypotheses of the last two results; however, if F is off-diagonally antitone and for some u,v, $Fu \leq 0 \leq Fv$, then D_0 can be taken to be the set $\langle u,v\rangle = \{z \in R^n : u \leq z \leq v\}$.

Theorem 2.3.9 Let $F:D \subset R^n \to R^n$ be continuous, off-diagonally antitone, and Ω-diagonally dominant on the set D. If there are u,v in D such that $\langle u,v\rangle \subset D$ with $Fu \leq 0 \leq Fv$, then $Fx = 0$ has a solution x^* in $\langle u,v\rangle$ which is unique in D; and for any x^0 in $\langle u,v\rangle$, the Jacobi and Gauss-Seidel sequences (2.1.1)-(2.1.4) with $\omega_k \equiv \omega \in (0,1]$ for $k = 0,1,\ldots,$ are well-defined and converge to x^*.

Proof. By Corollary 2.3.5 we only need to verify that for each $x \in \langle u,v\rangle$, $\psi(t) = f_i(x_1,\ldots,x_{i-1},t,x_{i+1},\ldots,x_n) = 0$ has a unique solution $t_i^* \in [u_i,v_i]$. Note that ψ is defined on $[u_i,v_i]$. Moreover, since F is off-diagonally antitone,

$$f_i(v_1,\ldots,v_{i-1},t,v_{i+1},\ldots,v_n) \leq \psi(t) \leq f_i(u_1,\ldots,u_{i-1},t,u_{i+1},\ldots,u_n),$$

and hence, $\psi(u_i) \leq 0 \leq \psi(v_i)$. The continuity of ψ on $[u_i,v_i]$ then implies that there is a $t_i^* \in [u_i,v_i]$ with $\psi(t_i^*) = 0$.

Under the hypotheses of the previous theorem, F is necessarily diagonally isotone and hence Theorem 1.2.10 and Corollary 2.2.5 imply that F is an M-function. In fact, Theorem 2.3.9 can be shown to hold if F is a continuous M-function on the set D. In this form this last theorem is implicit in the work of Rheinboldt [1969b].

Theorem 2.3.9 can be used, for example, to obtain results about non-negative solutions of two-point boundary value problems. Consider

(2.3.7) $u''(t) = g(t,u(t),u'(t))$ for $a < t < b$; $u(a) = \alpha$, $u(b) = \beta$,

where g is continuously differentiable on

(2.3.8) $S = \{(t,u,u') \in R^3 : a \leq t \leq b,\ 0 \leq u < +\infty,\ -\infty < u' < +\infty\}$,

and

(2.3.9) $g_u(t,u,u') \geq 0$, $|g_{u'}(t,u,u')| \leq M < +\infty$, for all $(t,u,u') \in S$.

Then, as shown, for instance, by Bailey, Shampine, and Waltman [1968] it is known that (2.3.7)-(2.3.9) has a unique, twice-continuously differentiable, non-negative solution provided that $\alpha, \beta \geq 0$, and $g(t,0,0) \leq 0$ for $t \in [a,b]$. To obtain a numerical solution of this problem, we introduce the partition $s_j = a + jh$, $j = 0,\ldots,n+1$, $h = \frac{b-a}{n+1}$, of $[a,b]$, and use the standard finite difference approximation of (2.3.7)

(2.3.10) $$Ax + \phi(x) - c = 0$$

where

(2.3.11) $$A = \begin{bmatrix} 2 & -1 & & & \bigcirc \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ \bigcirc & & & -1 & 2 \end{bmatrix}, \quad c = (\alpha,0,\ldots,0,\beta)$$

and $\phi : R^n \to R^n$ is defined by

(2.3.12) $\phi_i(x) = h^2 g(s_i, x_i, \frac{x_{i+1} - x_{i-1}}{2h})$, $i = 1,\ldots,n$.

We want to find a solution of (2.3.10) in $R^n_+ = \{x \in R^n : x \geq 0\}$.

Theorem 2.3.10 Consider the mapping $Fx = Ax + \phi(x) - c$ defined by (2.3.10)-(2.3.12) where g is continuously differentiable on the set S of (2.3.8) and satisfies (2.3.9). If $\alpha, \beta \geq 0$, $g(t,0,0) = 0$ for $t \in [a,b]$, and $h = \frac{b-a}{n} \in (0, \frac{2}{M})$, then the equation (2.3.10) has a unique nonnegative solution $x^* \in R^n_+$, and for any $x^0 \in R^n_+$, the Jacobi and Gauss-Seidel iterates (2.1.1)-(2.1.4) with $\omega_k \equiv \omega \in (0,1]$, $k = 0,1,\ldots$, are well-defined and converge to x^*. Moreover, $x^* \in \langle 0, t_0 e \rangle$ where t_0 satisfies

$$t_0 + h^2 g(s_1, t_0, \frac{t_0 - \alpha}{2h}) \geq \alpha, \quad t_0 + h^2 g(s_n, t_0, \frac{\beta - t_0}{2h}) \geq \beta.$$

Proof. Clearly F is off-diagonally antitone, and Theorem 2.2.8 implies that F is Ω-diagonally dominant on R^n_+ with respect to the associated network Ω_F. Moreover, $F(0) = \phi(0) - c \leq 0$, and $F(te) \geq 0$ for $t \geq t_0$. Hence, the result follows directly from Theorem 2.3.9.

The use of the approximation (2.3.10) is of course standard, but it is usually assumed that (2.3.9) holds for all $(t,u,u') \in [a,b] \times R^1 \times R^1$. Discrete analogues of mildly nonlinear elliptic boundary value problems of the form $\Delta u = g(t,u)$ with $g(t,0) \equiv 0$ and $g(t,u) \geq 0$ for $u \geq 0$, have been considered, for example, by Greenspan and Parter [1965], and these authors obtain an existence result similar to ours. However, they do not treat either the (nonlinear) Jacobi or Gauss-Seidel method.

In the case where (2.3.9) holds for all $(t,u,u') \in [a,b] \times R^1 \times R^1$ and hence when F is defined on all of R^n, the techniques used in this paper allow us to prove also some results about overrelaxation. For this we introduce the following concept.

Definition 2.3.11 The mapping $F:R^n \to R^n$ is uniformly diagonally dominant if there is a $q \in [0,1)$ such that for every $x \neq y$ and $k \in N$, $f_k x = f_k y$ implies that $|x_k - y_k| \leq q\|x-y\|_\infty$.

Clearly, every uniformly diagonally dominant function is strictly diagonally dominant on R^n, but the following example shows that the converse does not hold.

Example 2.3.12 Let $F:R^2 \to R^2$ be the function of Example 2.1.7 and assume that there is a $q \in [0,1)$ such that Definition 2.3.10 is satisfied. Since, for each $k \geq 1$, $f_1(\sin \frac{1}{k}, \frac{1}{k}) = f_1(0,0)$, we have then $|\sin \frac{1}{k}| \leq q|\frac{1}{k}|$ and for $k \to +\infty$ it follows that $q \geq 1$ which is a contradiction. Hence, F is not uniformly diagonally dominant.

Uniformly diagonally dominant functions usually appear in the literature under the disguise of a uniformity condition on the Jacobian. The next result illustrates this point.

Theorem 2.3.13 Let $F:R^n \to R^n$ be G-differentiable on R^n, and assume that for every x in R^n, $F'(x)$ has non-zero diagonal entries. If for each $x \in R^n$ and $k \in N$, there is a constant $q \in [0,1)$ such that

$$\sum_{j \neq k} |\partial_j f_k(x)| \leq q|\partial_k f_k(x)|,$$

then F is uniformly diagonally dominant.

Proof. Assume that $f_k(x) = f_k(y)$ for $x \neq y$ and some index $k \in N$. Then $\psi(t) = f_k(x+t(y-x))$ is differentiable on $[0,1]$, $\psi(0) = \psi(1)$, and by Rolle's theorem, there is a $t_0 \in (0,1)$ such that

$$\psi'(t_0) = \sum_{j=1}^{n} \partial_j f_k(x+t_0(y-x))(y_j-x_j) = 0.$$

The result now follows from this equation.

If, in addition to the hypotheses of the previous theorem, we assume that there is an $m > 0$ such that for each $x \in R^n$ and $k \in N$, $|\partial_k f_k(x)| \geq m$, then F is diagonally surjective. The next result shows, in particular, that F must then be surjective.

Theorem 2.3.14 Let $F:R^n \to R^n$ be diagonally surjective and uniformly diagonally dominant. Then F is surjective and for any $x^0 \in R^n$ the Jacobi and Gauss-Seidel iterates (2.1.1)-(2.1.4) with $\omega_k \in [\underline{\omega},\bar{\omega}]$, $k = 0,1,\ldots,$ and

$$0 < \underline{\omega} \leq \bar{\omega} < \frac{2}{1+q} ,$$

are well-defined and converge to the unique solution x^* of $Fx = 0$. Moreover, there is a $q \in [0,1)$ such that

(2.3.13) $$\|x^{k+1}-x^*\|_\infty \leq (|1-\omega_k|+\omega_k q)\|x^k-x^*\|_\infty \leq \beta\|x^k-x^*\|_\infty$$

for each $k \geq 0$ where $\beta = \max\{1-\underline{\omega}(1-q), \bar{\omega}(1+q)-1\} < 1$.

Proof. We will only carry out the proof for the Jacobi method; for the Gauss-Seidel method it proceeds in a similar fashion.

To prove the surjectivity of F and hence the existence of x^*, let $b \in R^n$ be given, and consider the iterates $\{y^k\}$ of the Jacobi method for $\hat{F}(x) = F(x) - b$ with $\omega_k \equiv 1$ and arbitrary $y^0 \in R^n$. The diagonal surjectivity of F together with Theorem 2.2.2 implies that the iterates $\{y^k\}$ are well-defined. Now let $k \geq 1$, and note that, since

$$f_i(y_1^{k-1},\ldots,y_{i-1}^{k-1},y_i^k,y_{i+1}^{k-1},\ldots,y_n^{k-1}) = f_i(y_1^k,\ldots,y_{i-1}^k,y_i^{k+1},y_{i+1}^k,\ldots,y_n^k)$$

for each $i \in N$, the uniform diagonal dominance of F implies that $|y_i^{k+1}-y_i^k| \leq q\|y^k-y^{k-1}\|_\infty$ for some $q \in [0,1)$. Hence, $|y^{k+1}-y^k|_\infty \leq q\|y^k-y^{k-1}\|_\infty$, and by the triangle inequality, it follows that $\{y^k\}$ is a Cauchy sequence. Consequently, $\{y^k\}$ converges to some x^* where $\hat{F}x^* = Fx^* - b = 0$.

The uniqueness of x^* follows from Theorem 2.2.2, and thus, to complete the proof it suffices to prove (2.3.13). For this, note that $x_i^{k+1} = (1-\omega)x_i^k + \omega t_i^*(x^k)$ where

$$f_i(x_1^k,\ldots,x_{i-1}^k,t_i^*,(x^k),x_{i+1}^k,\ldots,x_n^k) = 0 = f_i(x^*)$$

for each $i \in N$. Hence, $|t_i^*(x^k)-x_i^*| \leqslant q\|x^k-x^*\|_\infty$, and therefore,

$$|x_i^{k+1}-x_i^*| \leqslant |1-\omega_k||x_i^k-x_i^*| + \omega_k|t_i^*(x^k)-x_i^*| \leqslant \{|1-\omega_k|+\omega_k q\}\|x^k-x^*\|_\infty$$

for each $i \in N$, which in turn implies (2.3.13).

To apply this theorem to functions arising as discrete analogues of (2.3.7), we will need the following result of Ostrowski [1956].

Lemma 2.3.15 Let $H \in L(R^n)$ be nonnegative. Then for each $\varepsilon > 0$ there exists a diagonal and invertible matrix $D \geqslant 0$ such that

$$\|D^{-1}HD\|_\infty \leqslant \rho(H) + \varepsilon \text{ (= spectral radius of } H + \varepsilon).$$

Consider now the two-point boundary value problem

(2.3.14) $u''(t) = g(t,u(t))$ for $a < t < b$; $u(a) = \alpha$, $u(b) = \beta$

where g is continuously differentiable and

(2.3.15) $g_u(t,u) \geqslant 0$ for all $(t,u) \in [a,b] \times R^1$.

The discrete analog corresponding to (2.3.10)-(2.3.12) now assumes the simpler form

(2.3.16) $$Fx = Ax + \phi(x) - c$$

where $A \in L(R^n)$ and $c \in R^n$ are again given by (2.3.11) and $\phi: R^n \to R^n$ has components

$$\phi_i(x) = h^2 g(s_i, x_i)$$

where $s_j = a + jh$, $j = 0,\ldots,n+1$, $h = \frac{b-a}{n+1}$.

Theorem 2.3.16 Fix $h > 0$. Under the stated assumptions concerning the mapping $Fx = Ax + \phi(x) - c$, $Fx = 0$ has a unique solution x^* and for any $x^0 \in R^n$ the Jacobi and Gauss-Seidel iterates (2.1.1)-(2.1.4) with $\omega_k \in [\underline{\omega}, \bar{\omega}]$, $k = 0,1,\ldots,$ and

$$0 < \underline{\omega} \leq \bar{\omega} < \frac{2}{1+\cos(\pi/n+1)}$$

are well-defined and converge to x^*.

Proof. Let H be the Jacobi iteration matrix for the matrix A; that is, H has the elements $h_{ij} = -\frac{a_{ij}}{a_{ii}}$, $i \neq j$, $h_{ii} = 0$, $i,j = 1,\ldots,n$. Then $H \geq 0$, and it is well-known (see Varga [1962]) that $\rho(H) = \cos(\pi/n+1)$. If $\varepsilon > 0$ is chosen so that $\cos(\pi/n+1) + \varepsilon < 1$ and $\bar{\omega} < 2/(1+\varepsilon+\cos(\pi/n+1))$, then Lemma 2.3.15 yields a diagonal, invertible matrix $D \geq 0$ such that $\|D^{-1}HD\|_\infty \leq \cos(\pi/n+1) + \varepsilon < 1$. Hence, AD is uniformly diagonally dominant with $q = \cos(\pi/n+1) + \varepsilon$. Define $\hat{F}: R^n \to R^n$ by $\hat{F} = F \cdot D$ and note that Theorem 2.3.13 ensures $\hat{F}$

to be uniformly diagonally dominant with $q = \cos(\pi/n+1) + \varepsilon$. Since, in addition, F is diagonally surjective, Theorem 2.3.14 guarantees that $Fx = 0$ has a solution y^* and that the Jacobi or Gauss-Seidel iterates $\{y^k\}$ for $\hat{F}$ converge to y^* for any set of relaxation parameters $\omega_k \in [\underline{\omega},\bar{\omega}]$, $k = 0,1,\ldots$. But D is a diagonal matrix, and $\hat{F} = F \cdot D$. Thus, $Fx^* = 0$ where $x^* = D^{-1}y^*$; and if $\{x^k\}$ are the Jacobi or Gauss-Seidel iterates for F, then $x^k = D^{-1}y^k$ for each $k \geq 0$. Consequently, $\{x^k\}$ also converges to x^* as long as $\omega_k \in [\underline{\omega},\bar{\omega}]$ for $k = 0,1,\ldots$.

The last two convergence results for discrete analogues of two-point boundary value problems were only meant to illustrate the convergence theorems and thus were not stated in their most general form. In particular, Theorem 2.3.16 could have been stated for a function of the form $Ax + \phi(x) - c = 0$ where $A \in L(R^n)$ is any matrix with nonnegative diagonal elements and such that for some diagonal and invertible $D \geq 0$, AD is strictly diagonally dominant, $\phi: R^n \to R^n$ is a continuous, diagonal, and isotone function, and c is a vector in R^n. In this case, $\cos(\pi/n+1)$ should be replaced by the spectral radius of the absolute value of the Jacobi iteration matrix for A.

CHAPTER III

Global Convergence of Newton-Gauss-Seidel Methods

3.1 Introduction and Preliminaries

In the previous chapter, the convergence of the nonlinear Jacobi- and Gauss-Seidel iterations was considered and in particular global convergence was proved provided that $F:R^n \to R^n$ is continuous and Ω-diagonally dominant on R^n and $Fx = 0$ has a solution x^*. If $F:R^n \to R^n$ is continuously differentiable, then there are many iterative methods which make use of the derivative of F to find the solution x^*. The best known method of this kind is Newton's method:

$$x^{k+1} = x^k - F'(x^k)^{-1}Fx^k, \quad k = 0,1,\ldots \; .$$

The use of Newton's method, however, implies that at the k-th stage the linear system

$$F'(x^k)z = F'(x^k)x^k - Fx^k \tag{3.1.1}$$

has to be solved for z in order to obtain x^{k+1}, and this fact leads to iterative methods of the form

$$x^{k+1} = x^k - P_k(x^k)^{-1}Fx^k, \quad k = 0,1,\ldots, \tag{3.1.2}$$

where $P_k(x)$ is "easily" invertible. In particular, the use of a relaxation parameter ω_k and $m_k \geq 1$ steps of the SOR method to solve for z in (3.1.1) leads to the Newton-Gauss-Seidel methods:

$$(3.1.3)\qquad P_k(x)^{-1} = \omega_k[I+\ldots+H_k(x)^{m_k-1}][D(x)-\omega_k L(x)]^{-1}$$

where

$$(3.1.4)\qquad H_k(x) = [D(x)-\omega_k L(x)]^{-1}[(1-\omega_k)D(x)+\omega_k U(x)]$$

and

$$(3.1.5)\qquad F'(x) = D(x) - L(x) - U(x)$$

is a splitting of the Jacobian matrix $F'(x)$ into diagonal, strictly lower and strictly upper triangular parts. For a full discussion of these and related methods, see Ortega and Rheinboldt [1970b].

In this chapter we will be interested in global convergence theorems for iterative methods of the form (3.1.2). One of the first such theorems was given by Baluev [1952] for Newton's method, while Greenspan and Parter [1965] gave a global convergence theorem for the special case of (3.1.2)-(3.1.5) in which $m_k \equiv \omega_k \equiv 1$ and F arises as a discrete analogue of certain nonlinear boundary value problems. Ortega and Rheinboldt [1970a] presented a more abstract formulation (see Corollary 3.2.4 for a precise statement) but not sufficiently general to permit either $\omega_k \not\equiv 1$ or $m_k \not\equiv 1$. In Section 3.2, we give a still more general result which contains the others, but allows an arbitrary sequence $m_k \geq 1$ and ω_k in $[\omega,1]$, $\omega > 0$, under the basic assumption that F is a convex mapping. In the remainder of this section, we collect various definitions and lemmas dealing specifically with the case when F is linear.

Definition 3.1.1 Let G, P, Q be mappings from R^n to $L(R^n)$. The ordered pair (P,Q) is a splitting of G if P(x) is invertible for each x in R^n and $Gx = Px - Qx$ for all x in R^n. Moreover, if $P(x)^{-1} \geq 0$ for all x in R^n, then the splitting is

a) regular, if $Q(x) \geq 0$ for all x in R^n,

b) R-weak regular, if $Q(x)P(x)^{-1} \geq 0$ for all x in R^n,

c) L-weak regular, if $P(x)^{-1}Q(x) \geq 0$ for all x in R^n,

and

d) weak regular, if both R- and L-weak regular.

In the important special case in which G is constant, that is, G(x) is a fixed matrix for all x in R^n, we will assume that P and Q are also constant. In this case regular splittings were first considered by Varga [1962] while weak regular splittings were introduced by Ortega and Rheinboldt [1967] who also gave examples to show that weak regular splittings need not be regular. Similar examples show that R- or L-weak regular splittings need not be weak regular.

If we are trying to find the solution of $Ax = b$, where A is invertible, then a splitting (P,Q) of A induces the iteration $x^{k+1} = x^k - P^{-1}(Ax^k - b)$ which, as is well-known, converges to $A^{-1}b$ for all $x^0 \in R^n$ if, and only if, $\rho(QP^{-1}) < 1$ where ρ denotes the spectral radius. Necessary and sufficient conditions for $\rho(QP^{-1}) < 1$ have been obtained by several authors. In particular, Varga [1962] showed that if (P,Q) is a regular splitting of A and $A^{-1} \geq 0$, then

$\rho(QP^{-1}) < 1$; later, Ortega and Rheinboldt [1967] showed that this remains true for weak regular splittings and that the converse also holds. Their proof holds verbatim for L-weak regular splittings while a trivial modification gives the following result.

<u>Lemma 3.1.2</u> Let (P,Q) be an R-weak regular splitting of A in $L(R^n)$. Then $A^{-1} \geqslant 0$ if, and only if, $\rho(QP^{-1}) < 1$.

<u>Proof</u>. Assume $A^{-1} \geqslant 0$, and let $H = QP^{-1}$. Then $H \geqslant 0$, and since

$$(I-H)(I+\ldots+H^m) = I - H^{m+1}, \quad P^{-1} = A^{-1}(I-H),$$

we have, since $A^{-1} \geqslant 0$,

$$0 \leqslant P^{-1}[I+\ldots+H^m] = A^{-1}[I-H^m] \leqslant A^{-1}, \quad m = 1,2,\ldots .$$

But P^{-1} contains at least one non-zero element in each row, and therefore, $I + \ldots + H^m$ is bounded for all m. Since $H \geqslant 0$, the series converges, and consequently, $\rho(H) < 1$. Conversely, if $\rho(H) < 1$, then $(I-H)^{-1} \geqslant 0$, and $A^{-1} = P^{-1}(I-H)^{-1} \geqslant 0$.

The next two lemmas are also of interest in connection with Lemma 3.1.2. The first lemma is essentially due to Price [1968].

<u>Lemma 3.1.3</u> Let $A \in L(R^n)$. Then $A^{-1} \geqslant 0$ if, and only if, there is an R-weak regular splitting (P,Q) of A such that $\rho(QP^{-1}) < 1$.

Proof. If $A^{-1} \geqslant 0$, then $P = A$, $Q = 0$, is a splitting of A with the desired properties.

Conversely, if (P,Q) is an R-weak regular splitting of A with $\rho(QP^{-1}) < 1$, then Lemma 3.1.2 shows that $A^{-1} \geqslant 0$.

The next result is closely connected to work of Bramble and Hubbard [1964].

Lemma 3.1.4 Let $A \in L(R^n)$. Then $A^{-1} \geqslant 0$ and has an R-weak regular splitting if, and only if, there is an $S \geqslant 0$ such that AS is an M-matrix.

Proof. Assume that $A^{-1} \geqslant 0$, and that (P,Q) is an R-weak regular splitting of A. Then $AP^{-1} = I - QP^{-1}$, and if we let $S = P^{-1}$, we only need to show that $AS = I - QP^{-1}$ is an M-matrix. Let $H = QP^{-1} \geqslant 0$. Then $I - H$ has non-positive off-diagonal elements, and by Lemma 3.1.2, $\rho(H) < 1$. Hence, $(I-H)^{-1} \geqslant 0$, and therefore $AS = I - H$ is an M-matrix.

Conversely, if for some $S \geqslant 0$, AS is an M-matrix, then S is nonsingular and $A^{-1} \geqslant 0$. Let $D = \text{diag}\ (AS)$ and $B = D - AS$. Then (D,B) is a regular splitting of the M-matrix AS, and by Lemma 3.1.2, $\rho(BD^{-1}) < 1$. Hence, $A^{-1} \geqslant 0$ and, clearly, (DS^{-1}, BS^{-1}) is an R-weak regular splitting of A.

We now return to the connection between the splittings of A and the convergence of the iteration

(3.1.6) $$x^{k+1} = x^k - P_k^{-1}(Ax^k - b), \quad k = 0,1,\ldots .$$

Theorem 3.1.5 Let $A \in L(R^n)$ be invertible, and (P_k, Q_k) a sequence of R-weak regular splittings of A. Then each of the following statements implies the next statement.

a) The iterates (3.1.6) converge to $x^* = A^{-1}b$ for any x^0 in R^n.

b) $A^{-1} \geq 0$.

c) The iterates (3.1.6) converge for any x^0 in R^n.

Proof. Without loss of generality we may assume that $b = 0$. By (3.1.6),

(3.1.7) $$Ax^{k+1} = (I - AP_k^{-1})Ax^k = \left(\prod_{j=0}^{k} R_j\right)Ax^0$$

where $R_j = I - AP_j^{-1} = Q_jP_j^{-1} \geq 0$.

Now assume that a) holds; then it suffices to show that $Ax^0 \geq 0$ implies $x^0 \geq 0$ for any x^0. But (3.1.7) shows that $Ax^k \geq 0$ for all k, and by (3.1.6) that $x^{k+1} \leq x^k \leq x^0$ for all k. Since a) guarantees that $\lim x^k = 0$, it follows that $x^0 \geq 0$.

Next assume that b) holds. We first prove convergence for all x^0 such that $Ax^0 \geq 0$. In this case, as before, $Ax^k \geq 0$, and $x^{k+1} \leq x^k$ for all k. Since $A^{-1} \geq 0$, $x^k \geq x^{k+1} \geq 0$ for all k, so $\{x^k\}$ converges. Now note that since A is invertible, the iterates (3.1.6) converge for arbitrary x^0 if, and only if, $\left(\prod_{j=0}^{k} R_j\right)x$ converges for each $x \in R^n$. But we have already shown convergence if $Ax^0 = x \geq 0$; therefore

(3.1.6) converges for each basis vector e^i and hence for all x in R^n. This completes the proof.

Note that (3.1.6) implies that $P_k(x^{k+1}-x^k) = Ax^k - b$, and therefore c) implies a) if $\{P_k\}$ is bounded. In particular, this occurs if we have only one splitting. In general, however, none of the implications of Theorem 3.1.5 can be reversed without additional hypotheses as the following examples show.

Example 3.1.6 i) Consider the one-dimensional example in which $A = 1$, $b = 0$ and $P_k = (k+2)^2$ for $k = 0,1,\ldots$. Then $\lim x^k = \frac{1}{2} x^0 \neq 0$ if $x^0 \neq 0$. Hence, b) does not imply a) in general. This example also illustrates that if $A^{-1} \geqslant 0$, and $x^0 \neq x^*$, then the iterates (3.1.6) need not converge to the solution of $Ax = b$, but as Theorem 3.1.5 shows, these iterates must converge.

ii) Consider now the one-dimensional example where $A = -1$, $b = 0$, and $P_k = \frac{1}{(k+1)(k+3)}$ for $k = 0,1,\ldots$. Then $\lim x^k = 2x^0 \neq x^0$ if $x^0 \neq 0$. Hence, c) does not always implies b).

The next lemma is essentially a rewording of the conclusion a) implies b) in Theorem 3.1.5, but now A is not assumed to be invertible.

Lemma 3.1.7 Let (P_k, Q_k) be a sequence of R-weak regular splittings of A in $L(R^n)$, and set $R_k = Q_k P_k^{-1}$. If $\lim_{k\to+\infty} \prod_{j=0}^{k} R_j = 0$, then A^{-1} exists, and $A^{-1} \geqslant 0$.

Proof. It suffices to show that $A^T x \geq 0$ implies $x \geq 0$. Since

$$x^T A P_j^{-1} = x^T(I-R_j),$$

$A^T x \geq 0$ implies that $x^T R_j \leq x^T$ for $j = 1,2,\ldots$. But $R_j \geq 0$, and

$$x^T \geq x^T \prod_{j=0}^{k} R_j \to 0.$$

Hence, $x \geq 0$.

The next result will play an important role in the next section; it also contains a partial converse of the previous lemma.

Lemma 3.1.8 Assume (P_k, Q_k) is an R-weak regular splitting of $A_k \in L(R^n)$, and set $R_k = Q_k P_k^{-1}$ for $k = 0,1,\ldots$. If there is a nonsingular $C \geq 0$ in $L(R^n)$ such that $CA_k \geq I$ for all $k \geq 0$, then

a) $A_k^{-1} \geq 0$ for $k = 0,1,\ldots,$

b) $\prod_{j=0}^{k} R_j$ is convergent (and hence bounded) as $k \to +\infty$, and

c) $S_m = \sum_{k=1}^{m} P_k^{-1} \prod_{j=0}^{k-1} R_j$ converges as $m \to +\infty$.

Proof. For fixed k, we first show that $A_k^{-1} \geq 0$. Since $CA_k \geq I$, and $I - R_k = A_k P_k^{-1}$ for all $k \geq 0$, it follows that

$$0 \leq P_k^{-1} \leq C[I-R_k]. \tag{3.1.9}$$

Now, since P_k^{-1} is nonsingular, (3.1.9) implies that $[I-R_k^T]u > 0$ where $u = C^T e > 0$, and by a well-known theorem of Fan [1958], we conclude that $\rho(R_k^T) = \rho(R_k) < 1$, and Lemma 3.1.2 then implies that $A_k^{-1} \geq 0$.

To prove b) and c), let $T_k = \prod_{j=0}^{k} R_j$, and note that (3.1.9) implies that

$$(3.1.10) \quad S_m \leq C \sum_{k=1}^{m} (T_{k-1}-T_k) = CT_0 - CT_m \leq CT_0,$$

for all $m \geq 1$. Since $S_m \leq S_{m+1}$ for all $m \geq 1$, it follows that S_m is convergent and therefore, that c) holds. From (3.1.9) we obtain that $0 \leq CT_m \leq CT_{m-1}$. Hence, CT_m converges as $m \to +\infty$, and since C is nonsingular, so does T_m.

It is very important to note that in parts b) and c) of Lemma 3.1.8 the condition $CA_k \geq I$ cannot be replaced by $0 \leq A_k^{-1} \leq C$ for all $k \geq 0$, as the following example shows.

<u>Example 3.1.9</u> Let

$$A_k = \begin{bmatrix} 1 & -\alpha\delta_k \\ -\alpha\delta_{k+1} & 1 \end{bmatrix}$$

where $\delta_k = 1$ if k is even, and $\delta_k = 0$ if k is odd. If $\alpha \geq 0$, then

$$0 \leq A_k^{-1} \leq \begin{bmatrix} 1 & \alpha \\ \alpha & 1 \end{bmatrix}$$

for each $k \geq 0$, and if $P_k = I$ and $Q_k = P_k - A_k$, then (P_k, Q_k) is a regular splitting of A_k for each $k \geq 0$. However,

$$\prod_{j=0}^{k} R_j = \alpha^{k+1} \begin{bmatrix} 0 & \delta_k \\ 0 & \delta_{k+1} \end{bmatrix},$$

and therefore, the sequence $\prod_{j=0}^{k} R_j$ diverges whenever $\alpha > 1$. It is also clear that S_m diverges as $m \to +\infty$ if $\alpha > 1$.

3.2 Global Convergence Theorems

We begin with our main convergence result.

Theorem 3.2.1 Let $F:R^n \to R^n$ be a continuously differentiable and convex function. Suppose that (P_k, Q_k) is a sequence of R-weak regular splittings of F' with $\{P_k\}$ uniformly bounded on compact sets and such that for each $\{y^k\}$ in R^n,

$$(3.2.1) \qquad \sum_{k=1}^{\infty} P_k(y^k)^{-1} \prod_{j=0}^{k-1} R_j(y^j)$$

converges, where $R_j(x) \equiv Q_j(x)P_j(x)^{-1}$. Assume furthermore that either

a) F is surjective, and

$$(3.2.2) \qquad \prod_{j=0}^{k} R_j(y^j)$$

is bounded for each $\{y^k\}$ in R^n, or

b) $Fx = 0$ has a solution, and (3.2.2) converges to zero for each $\{y^k\}$ in R^n.

Then $Fx = 0$ has one and only one solution x^*, and for any $x^0 \in R^n$, the sequence

$$(3.2.3) \qquad x^{k+1} = x^k - P_k(x^k)^{-1}Fx^k, \quad k = 0,1,\ldots,$$

converges to x^*.

Proof. We first show that F is inverse isotone on R^n. To do this, note that $\{P_k(x)\}$ is bounded for each $x \in R^n$, and since (3.2.1) converges when $y^k \equiv x$ for all $k \geq 0$, we have

$$\lim_{k\to+\infty} \prod_{j=0}^{k} R_j(x) = 0$$

for each $x \in R^n$. Furthermore, $(P_k(x),Q_k(x))$ is a sequence of R-weak regular splittings of $F'(x)$, and thus, Lemma 3.1.7 implies that $F'(x)^{-1} \geq 0$ for each $x \in R^n$, and hence, F is inverse isotone on R^n by Theorem 1.4.7. But by either a) or b), $Fx = 0$ has a solution x^*, and since F is injective by Theorem 1.3.1, $Fx = 0$ has only the solution x^*.

To prove the convergence of (3.2.3) to x^*, we begin by showing that this sequence is bounded below. Since F is convex on R^n, Lemma 1.4.6 yields

$$Fx^{k+1} \geq F'(x^k)(x^{k+1}-x^k) + Fx^k = [I-F'(x^k)P_k(x^k)^{-1}]Fx^k,$$

and since $R_k(x) = I - F'(x)P_k(x)^{-1} \geq 0$, it follows that

$$(3.2.4) \qquad Fx^{k+1} \geq R_k(x^k)Fx^k \geq [\prod_{j=0}^{k} R_j(x^j)]Fx^0.$$

Assume now that F is surjective, and that (3.2.2) is bounded for the sequence defined by (3.2.3). Then there is a $v \in R^n$ such that

$Fx^{k+1} \geqslant v$ for all $k \geqslant 0$, and by the surjectivity of F, there is a $u \in R^n$ such that $Fx^{k+1} \geqslant Fu$. The inverse isotonicity of F now yields that $x^{k+1} \geqslant u$, and $\{x^k\}$ is therefore bounded below. Now, instead of a), assume that b) holds. By the classical inverse function theorem, F is a local homeomorphism, and thus there are two open balls B_1 and B_2, centered at x^* and zero respectively such that F is a homeomorphism from B_1 onto B_2. Choose $v > 0$ such that $v \in B_2$. Then by (3.2.4) and the convergence of (3.2.2) to zero, there is a $k_0 \geqslant 0$ such that if $k \geqslant k_0$,

$$Fx^{k+1} \geqslant [\prod_{j=0}^{k} R_j(x^j)]Fx^0 \geqslant -v.$$

But $-v \in B_2$, and hence, there is a $u \in B_1$ such that $Fu = -v$. Consequently, $Fx^{k+1} \geqslant Fu$ for $k \geqslant k_0$, and the inverse isotonicity of F again implies that $\{x^k\}$ is bounded below.

Next, we show that $\{x^k\}$ is also bounded above, and that it has only one limit point. By (3.2.3) and (3.2.4)

$$x^{k+1} - x^k \leqslant P_k(x^k)^{-1}\prod_{j=0}^{k-1} R_j(x^j)\,|Fx^0|,$$

and therefore,

$$(3.2.5) \qquad x^{m+p} - x^m \leqslant \sum_{k=m}^{\infty} (P_k(x^k)^{-1}\prod_{j=0}^{k-1} R_j(x^j))\,|Fx^0|,$$

for any $m,p \geqslant 1$. The series on the right is convergent by assumption, and thus, for fixed $m \geqslant 1$, (3.2.5) shows that $\{x^k\}$ is bounded above. Hence, $\{x^k\}$ is bounded and has limit points. If now u is any limit point of $\{x^k\}$, then there is a subsequence of $\{x^k\}$ converging to u, and (3.2.5) implies that

$$u - x^m \leqslant \sum_{k=m}^{\infty} (P_k(x^k)^{-1} \prod_{j=0}^{k-1} R_j(x^j)) |Fx^0|.$$

If v is any other limit point of $\{x^k\}$, the above inequality shows that $u - v \leqslant 0$. By reversing the roles of u and v, we can similarly obtain $v - u \leqslant 0$, and hence, $u = v$. Since $\{x^k\}$ has only one limit point, $\{x^k\}$ converges to, say, $\hat{x}$. But $\{P_k(x^k)\}$ is bounded, and it follows from $-Fx^k = P_k(x^k)(x^{k+1}-x^k)$ that $F\hat{x} = 0$. The injectivity of F now implies that $\hat{x} = x^*$ which is the desired result.

The proof of the previous theorem uses ideas found in the papers by Greenspan and Parter [1965] and Ortega and Rheinboldt [1970a]. Note also that the proof shows that assumptions a) and b) can be replaced by any assumptions which guarantee that $Fx = 0$ has a solution x^*, and that the sequence defined by (3.2.3) is bounded below. Finally, note that if b) holds, then the use of Lemma 3.1.7 yields that $F'(x)^{-1} \geqslant 0$ for all x in R^n, and hence that F is inverse isotone without using the uniform boundedness of $\{P_k\}$. Therefore, if b) holds, (3.2.3) converges even if $\{P_k\}$ is not uniformly bounded on compact sets, but not necessarily to x^* as shown by i) of Example 3.1.6.

We now derive several important corollaries of Theorem 3.2.1, which illustrate how the different hypotheses of the theorem are satisfied. We begin with a result of Baluev [1952].

Corollary 3.2.2 Let $F:R^n \to R^n$ be continuously differentiable and convex on R^n, and suppose that $F'(x)^{-1} \geq 0$ for all x in R^n. If $Fx = 0$ has a solution x^*, then this solution is unique, and for any $x^0 \in R^n$, the sequence

$$x^{k+1} = x^k - F'(x^k)^{-1}Fx^k, \quad k = 0,1,\ldots \tag{3.2.6}$$

converges to x^*.

Proof. Define $P_k(x) = F'(x)$ and $Q_k \equiv 0$, so that $R_k(x) \equiv 0$. Then, trivially, Theorem 3.2.1, using assumption b), applies.

Note that, as is well-known, the iterates (3.2.6) exhibit monotone convergence under the conditions of Corollary 3.2.2 at least for $k \geq 1$, although this is not a direct consequence of Theorem 3.2.1. Also note that the hypothesis that $Fx = 0$ has a solution is not implied by the remaining assumptions as the one-dimensional example $f(x) = e^x$ shows.

In the remaining results, the existence of a solution to $Fx = 0$ will be a consequence of the following theorem of Hadamard [1906].

Lemma 3.2.3 Let $F:R^n \to R^n$ be continuously differentiable, and assume there is a constant M such that $\|F'(x)^{-1}\| \leq M$ for all x in R^n. Then F is a homeomorphism of R^n onto R^n.

For a more accessible proof of the above result, see Ortega and Rheinboldt [1970b].

The next theorem illustrates how Hadamard's result can be used in conjunction with Theorem 3.2.1.

Corollary 3.2.4 (Ortega and Rheinboldt [1970a]) Let $F:R^n \to R^n$ be continuously differentiable and convex, and assume there is a regular splitting (P,Q) of F' such that

$$0 \leqslant P(x)^{-1} \leqslant C_0$$

$$0 \leqslant Q(x) \leqslant C_1$$

for all x in R^n, where $\rho(C) < 1$ with $C = C_0 C_1$. Then for any x^0 in R^n, the iterates

$$x^{k+1} = x^k - P(x^k)^{-1} F x^k, \quad k = 0,1,\ldots,$$

converge to the unique solution x^* of $Fx = 0$.

Proof. By assumption, (P,Q) is an R-weak regular splitting of F', and since

$$F'(x) \leqslant P(x) = F'(x) + Q(x) \leqslant F'(x) + C_1,$$

the continuity of F' yields that P is bounded on compact sets. In order to show that (3.2.1) converges for each $\{y^k\}$ in R^n, note that $0 \leqslant R(x) = Q(x)P(x)^{-1} \leqslant C_1 C_0 \equiv \hat{C}$ where $\rho(\hat{C}) = \rho(C) < 1$. Then,

$$0 \leqslant \prod_{j=0}^{k} R(y^j) \leqslant \hat{C}^{k+1}$$

which shows that (3.2.2) converges to zero. Moreover,

$$\sum_{k=1}^{m} P(y^k)^{-1} \prod_{j=0}^{k-1} R(y^j) \leq C_0 \sum_{k=1}^{m} \hat{C}^k \leq C_0[I-\hat{C}]^{-1},$$

for any $m \geq 1$, and since (3.2.1) is a series of nonnegative terms, it converges. We conclude the proof by showing that F is surjective. This follows from Lemma 3.2.3 since $F'(x) = [I-R(x)]P(x)$, and thus

$$0 \leq F'(x)^{-1} = P(x)^{-1}[I-R(x)]^{-1} \leq C_0[I-\hat{C}]^{-1}.$$

The previous result could have been proven under either assumption a) or b) of Theorem 3.2.1 since F was surjective and (3.2.2) converged to zero. We now present a case in which (3.2.2) does not necessarily converge to zero, but it is bounded.

Theorem 3.2.5 Let $F:R^n \to R^n$ be a continuously differentiable and convex mapping, and suppose that (P_k,Q_k) is a sequence of R-weak regular splittings of F' with $\{P_k\}$ uniformly bounded on compact sets. Assume there is an A in $L(R^n)$ such that

a) $A^{-1} \geq 0$,

and

b) $F'(x) \geq A$,

for all x in R^n. Then $Fx = 0$ has one and only one solution x^*, and for any x^0 in R^n, the sequence

$$x^{k+1} = x^k - P_k(x^k)^{-1}Fx^k, \quad k = 0,1,\ldots,$$

converges to x^*.

Proof. Since for each fixed $x \in R^n$, $(P_k(x), Q_k(x))$ is an R-weak regular splitting of $F'(x)$, and $CF'(x) \geq I$ where $C = A^{-1} \geq 0$, Lemma 3.1.8 applies, and shows that $F'(x)^{-1} \geq 0$ for all $x \in R^n$. Therefore, by a) and b), we also obtain that $F'(x)^{-1} \leq A^{-1}$. Hence, $F'(x)^{-1}$ is uniformly bounded on R^n, and Lemma 3.2.3 implies that F is surjective. The result will follow from Theorem 3.2.1 if we show that for any sequence $\{y^k\}$ in R^n, the series (3.2.1) converges, and the sequence (3.2.2) is bounded. To do this, note that for any $\{y^k\}$ in R^n, $(P_k(y^k), Q_k(y^k))$ is an R-weak regular splitting of $A_k = F'(y^k)$, and that Lemma 3.1.8 applies.

Note that the above theorem would still hold if instead of a) and b), we would have assumed the existence of a nonsingular $C \geq 0$ such that $CF'(x) \geq I$ for all x in R^n. A similar remark can be made in the following results when conditions of the form a) and b) appear, but note that the next example shows that a) and b) of Theorem 3.2.5 cannot be replaced by the assumption that

$$0 \leq F'(x)^{-1} \leq B$$

for some B in $L(R^n)$.

Example 3.2.6 Define $F:R^2 \to R^2$ by

$$Fx \equiv F(x_1,x_2) = \begin{bmatrix} x_1 + g(x_1-x_2) \\ x_2 + g(x_2-x_1) \end{bmatrix}$$

where $g:R^1 \to R^1$ is a continuously differentiable, isotone and convex mapping. Then F is continuously differentiable, convex, and

$$0 \leq F'(x_1,x_2)^{-1} \leq B = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} .$$

Hence, by Lemma 3.2.3, F is a homeomorphism from R^n onto R^n, and consequently $F(x_1,x_2) = 0$ has a unique solution $x^* = (x_1^*, x_2^*)^T$. Consider now the (one-step) Newton-Jacobi method, that is, $P(x) = \text{diag } F'(x)$; we will show that it is possible to choose g in such a way that the Newton-Jacobi method does not converge to x^*.

Choose $g:R^1 \to R^1$ continuously differentiable, isotone, convex and such that

$$g(-1) = -1, \; g'(-1) = 0, \; g(1) = g'(1).$$

For example, if $g(s) = \frac{2}{2-\alpha} \left(\frac{s+1}{2} \right)^{\alpha} - 1$, $1 < \alpha < 2$, for $s \geq -1$ and $g(s) \equiv -1$ otherwise, then g satisfies the above conditions.

If now $x^0 = (0,1)^T$ it is easy to verify that $x^1 = (1,0)^T$ and $x^{2k} = (0,1)^T$ for $k \geq 0$, and thus, the Newton-Jacobi sequence will not converge to x^*.

Theorem 3.2.5 can in turn be used to give a global convergence result for the Newton-Gauss-Seidel method (3.1.2)-(3.1.5). We shall need the following lemma whose proof is due to Ortega and Rheinboldt [1967] for weak regular splittings.

Lemma 3.2.7 Let (B,C) be an R-weak regular splitting of A in $L(R^n)$, and assume that $A^{-1} \geq 0$. Set $H = CB^{-1} \geq 0$, and

$$R = B^{-1}[I+\ldots+H^m]$$

for some $m \geq 1$. Then R is invertible, and $(R^{-1}, R^{-1}-A)$ is an R-weak regular splitting of A.

Proof. Lemma 3.1.2 yields that $\rho(H) < 1$, and hence, $I - H$ and $I - H^{m+1}$ are invertible. Therefore, since $[I+\ldots+H^m][I-H] = I - H^{m+1}$, R is invertible. Next, $(R^{-1}-A)R = I - AR$, so $(R^{-1}, R^{-1}-A)$ is an R-weak regular splitting of A if $AR \leq I$. Now

$$AR = [I-H]BR = I - H^{m+1} \leq I,$$

and the proof is complete.

Corollary 3.2.8 Let $F:R^n \to R^n$ be continuously differentiable and convex on R^n, and assume that $F'(x)$ is an M-matrix for each x in R^n. Assume further that there is an M-matrix A in $L(R^n)$ such that $F'(x) \geq A$ for all x in R^n. Then, for any x^0 in R^n, any sequence $\{m_k\}$ of positive integers, and any sequence $\{\omega_k\}$ in $[\omega, 1]$, $\omega > 0$, the Newton-Gauss-Seidel iterates (3.1.2)-(3.1.5) are well-defined and converge to the unique solution of $Fx = 0$.

Proof. Set

$$B_k(x) = \frac{1}{\omega_k}[D(x)-\omega_k L(x)], \quad C_k(x) = \frac{1}{\omega_k}[(1-\omega_k)D(x)+\omega_k U(x)],$$

and note that for each $x \in R^n$, and $k \geq 0$, $(B_k(x), C_k(x))$ is a weak regular splitting of $F'(x)$. Let $\hat{H}_k(x) = C_k(x)B_k(x)^{-1}$ and $R_k(x) = B_k(x)^{-1}[I+\ldots+\hat{H}_k(x)^{m_k-1}]$. Since Lemma 3.2.6 applies, $R_k(x)$ is invertible, and if $\hat{P}_k(x) = R_k(x)^{-1}$, then $(\hat{P}_k(x), \hat{P}_k(x)-F'(x))$ is an R-weak regular splitting of $F'(x)$ for each $x \in R^n$ and $k \geq 0$. Since $\hat{P}_k(x)^{-1} = P_k(x)^{-1}$ where P_k^{-1} is defined by (3.1.3), the Newton-Gauss-Seidel iterates are well-defined. To conclude the proof, we need to show that $\{P_k\}$ is uniformly bounded on compact sets; the rest follows from Theorem 3.2.5.

Since $F'(x) = B_k(x)[I-H_k(x)]$, (3.1.3) implies that $P_k(x) = F'(x)[I-H_k(x)^{m_k}]^{-1}$, and therefore,

$$(3.2.7) \qquad |P_k(x)| \leq |F'(x)|[I-H_k(x)]^{-1} \leq |F'(x)||F'(x)^{-1}||B_k(x)|.$$

Moreover, $|B_k(x)| \leq \frac{1}{\omega}[D(x)+L(x)]$, and by continuity, all three factors on the far right of (3.2.7) are bounded on compact sets. Thus, $\{P_k\}$ is uniformly bounded on compact sets.

In a completely analogous manner, we can state a global convergence result for the general Newton-Jacobi method which is obtained by applying m_k steps of the Jacobi method to solve for z

in (3.1.1). This can also be done for the general Newton-Peaceman-Rachford iteration (see Ortega and Rheinboldt [1970b]), but in this case the acceleration parameters $\{r_k\}$ must be chosen sufficiently large so as to have R-weak regular splittings; otherwise divergence may occur as shown by Caspar [1968]. Finally, it is important to note that Corollary 3.2.7 contains, as a special case, the following result:

Corollary 3.2.9 Let A in $L(R^n)$ be an M-matrix, and $\Phi: R^n \to R^n$ a diagonal, isotone, convex and continuously differentiable mapping. Then for any b in R^n, $Ax + \Phi(x) = b$ has a unique solution x^*, and for any x^0 in R^n, any sequence $\{m_k\}$ of positive integers, and any sequence $\{\omega_k\}$ in $[\omega, 1]$, $\omega > 0$, the Newton-Gauss-Seidel iterates (3.1.2)-(3.1.5) with $Fx = Ax + \Phi(x) - b$ are well-defined and converge to x^*.

Proof. If $Fx = Ax + \Phi(x) - b$, then F is continuously differentiable and convex on R^n. Since Φ is diagonal and isotone, $\Phi'(x)$ is a diagonal, nonnegative matrix for each $x \in R^n$, and hence, $F'(x) = A + \Phi'(x) \geq A$ for all $x \in R^n$. That $A + \Phi'(x)$ is an M-matrix for each $x \in R^n$ follows from many considerations (see, e.g., Varga [1962]); in particular, note that $A + \Phi'(x)$ has a weak regular splitting (P,Q) where P is the diagonal part of A, and that $A^{-1}(A+\Phi'(x)) = I + A^{-1}\Phi'(x) \geq I$, so that Lemma 3.1.8 yields the desired result.

If $m_k \equiv \omega_k \equiv 1$, then the previous result was proved by Greenspan and Parter [1965].

REFERENCES

Bailey, P., Shampine, L., and Waltman, P. [1968], Nonlinear Two-Point Boundary Value Problems, Academic Press, Inc., New York, N.Y.

Baluev, A. [1952], On the method of Chaplygin, Dokl. Akad. Nauk SSSR 83, 781-784.

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

Bramble, J., and Hubbard, B. [1964], On a finite difference analogue of an elliptic boundary problem which is neither diagonally dominant nor of non-negative type, J. Math. and Phys. 43, 117-132.

Caspar, J. [1968], Applications of alternating direction methods to mildly nonlinear problems, Ph.D. Dissertation, University of Maryland.

Collatz, L. [1952], Aufgaben monotoner Art, Arch. Math. 3, 366-376.

Diaz, J., and Metcalf, F. [1968], On the set of subsequential limit points of successive approximations, Trans. Amer. Math. Soc. 135, 459-485.

Duffin, R. [1948], Nonlinear networks IIb, Bull. A.M.S. 54, 119-127.

Elkin, R. [1968], Convergence theorems for Gauss-Seidel and other minimization algorithms, Ph.D. Dissertation, University of Maryland.

Fan, Ky [1958], Topological proofs for certain theorems on matrices with nonnegative elements, Monatsh. Math. 62, 219-232.

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, Numer. Math. 9, 163-172.

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

Greenspan, D., and Parter, S. [1965], Mildly nonlinear elliptic partial differential equations and their numerical solution II, Numer. Math. 7, 129-147.

Hadamard, J. [1906], Sur les transformations ponctuelles, Bull. Soc. Math. France 34, 71-84.

Mařík, J., and Pták, V. [1960], Norms, spectra, and combinatorial properties of matrices, Czech. Math. J. 10, 181-196.

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

Moré, J., and Rheinboldt, W. [1970], On P- and S-functions and related classes of n-dimensional, nonlinear mappings, Computer Science Center, University of Maryland, Technical Report 70-120.

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

Ortega, J., and Rheinboldt, W. [1967], Monotone iterations for nonlinear equations with applications to Gauss-Seidel methods, SIAM J. Num. Anal. 4, 171-190.

Ortega, J., and Rheinboldt, W. [1970a], Local and global convergence of generalized linear iterations, Studies in Numerical Analysis 2, 122-143, SIAM Publications, Philadelphia, Pa.

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

Ostrowski, A. [1956], Determinanten mit überwiegender Hauptdiagonale und die absolute Konvergenz von linearen Iterationsprozessen, Commentari. Math. Helv. 30, 175-210.

Porsching, T. [1969], Jacobi and Gauss-Seidel methods for nonlinear network problems, SIAM J. Num. Anal. 6, 437-448.

Price, H. [1968], Monotone and oscillation matrices applied to finite difference approximations, Math. Comp. 22, 489-516.

Rheinboldt, W. [1969a], Local mapping relations and global implicit function theorems, Trans. Amer. Math. Soc. 138, 183-198.

Rheinboldt, W. [1969b], On M-functions and their application to nonlinear Gauss-Seidel iterations and network flows, J. Math. Anal. Appl. 32, 1, in press; Ges. f. Math. u. Datenverarb., Birlinghoven/Germany, Tech. Rep. BMwF-GMD-23.

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

Schechter, S. [1962], Iteration methods for nonlinear problems, Trans. Am. Math. Soc. 104, 179-189.

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

Walter, W. [1967], Bemerkungen zu Iterationsverfahren bei linearen Gleichungssystemen, Numer. Math. 10, 80-85.

Willson, A., Jr. [1968], On the solutions of equations for nonlinear resistive networks, Bell System Tech. J. 48, 1755-1773.