Technical Report 68-83 December 1968

APPLICATIONS OF ALTERNATING DIRECTION METHODS TO MILDLY NONLINEAR PROBLEMS

by

Joseph Richard Caspar

UNIVERSITY OF MARYLAND
COMPUTER SCIENCE CENTER
COLLEGE PARK, MARYLAND

Technical Report 68-83 December 1968

APPLICATIONS OF ALTERNATING DIRECTION METHODS

TO MILDLY NONLINEAR PROBLEMS

by

Joseph Richard Caspar

This research was supported in part by the National Aeronautics and Space Administration under Grant NsG 398 and by the U.S. Army Research Office Durham under Grants OOR-DA-31-124-G676 and DA-HC04-67-C-0062.

Abstract

The solution of mildly nonlinear equations in R^n--especially those arising from the discretization of mildly nonlinear, self-adjoint, elliptic boundary value problems in two dimensions--is studied. Existence and uniqueness results are presented, and several iterative techniques for approximating the solution are considered. These techniques are generally two-level iterations in which an alternating direction procedure is coupled with a linearizing procedure--either of Picard or of the Newton type. Proofs of the convergence of these procedures are given.

ACKNOWLEDGMENTS

I wish to express my thanks to my thesis advisor, Professor R. B. Kellogg, for his patience and guidance. I with also to thank Professor James M. Ortega for many helpful discussions and criticisms.

I owe thanks also to the National Aeronautics and Space Administration for their financial support first as a NASA Trainee and then under Grant NsG-398 to the Computer Science Center of the University of Maryland.

Finally, I wish to thank the United States Army, Research Office, Durham, for their financial support under Grants OOR-DA-31-124-G676 and DA-HC04-67-C-0062 to the Institute for Fluid Dynamics and Applied Mathematics of the University of Maryland.

TABLE OF CONTENTS

INTRODUCTION

The alternating direction implicit--or ADI--method for approximating the solution of certain types of elliptic and parabolic partial differential equations in two space dimensions was introduced by Peaceman and Rachford [23] in 1955. For elliptic problems, the method is iterative, and, for many problems, especially those approximating so-called model problem conditions, convergence is very rapid. Variations of the Peaceman-Rachford scheme have been introduced in [6], [7], [8], [9], [11], and [14]. These variations have extensions to three or more space dimensions, but, in two dimensions, they lack some attractive convergence properties of the Peaceman-Rachford method.

In this paper, we consider the application of the Peaceman-Rachford iteration to certain types of nonlinear elliptic difference equations in two dimensions. Earlier papers in this area are [5], [12], and [13].

Chapter I consists of background material. The elliptic partial differential operators being considered yield so-called operators of positive type when discretized in the usual way. Thus, operators of positive type are defined and some properties, based on the maximum principle, are developed. The mildly nonlinear problem in R^n is defined, conditions are given which guarantee the existence of a unique solution, and a priori bounds on the solution are obtained. Finally, an analogy is drawn between the properties of the discrete and continuous operators, and a proof is given of the existence of a unique solution to a mildly nonlinear elliptic boundary value problem.

Chapter II contains background material on ADI methods. The Peaceman-Rachford method is presented for the linear problem along with the basic convergence results. Three iterations are defined for a mildly nonlinear problem (ADI, Newton-ADI, ADI-Newton) and conditions are given which guarantee that the methods are locally convergent to a solution.

Chapter III contains a closer study of the nonlinear ADI iteration introduced in Chapter II and considered by Kellogg in [15]. Convergence results analagous to those in the linear case are obtained in a Hilbert space setting.

Chapter IV contains a study of two level iterations in which a Picard iteration is coupled with an ADI iteration. Earlier results of Douglas [5] and Gunn [12], [13] are formalized and extended. Multi-step and single-step iterations are considered.

Chapter V is devoted to a closer study of the two level iterations introduced in Chapter II in which a Newton iteration is coupled with an ADI iteration. Results are given based on contraction and monotonicity principles. Finally, some miscellaneous numerical results are presented.

We summarize our results as follows. We present formal conditions which guarantee the existence of a unique solution to a mildly nonlinear elliptic boundary value problem or its discrete version. We introduce Newton-ADI iterations for approximating the solution to the discrete problem and present algorithms which are guaranteed to converge to the solution. We also formalize and generalize some previous results on Picard-ADI iterations and obtain convergence results for a one-level nonlinear ADI iteration in a Hilbert space.

CHAPTER I

MILDLY NONLINEAR PROBLEMS

1.1 Introduction. In this chapter, we discuss existence and uniqueness of solutions to mildly nonlinear problems in R^n and obtain bounds on the solutions. In particular, we discuss mildly nonlinear problems having a linear part coming from an operator of positive type. In Section 1.2, we define operators of positive type and present or extend certain known results based on the maximum principle. In Section 1.3, we define mildly nonlinear problems and present conditions under which a unique solution can be guaranteed to exist, and in Section 1.4, we obtain a priori bounds on the solutions. Finally, in Section 1.5, we consider a mildly nonlinear elliptic partial differential equation and present results analogous to those earlier in the chapter. We first present some notation and definitions.

Let $G: X \to Y$, where X and Y are Banach spaces. If G is nonlinear, $G(x)$ will denote the value of G at $x \in X$; if G is linear, we will write Gx instead.

1.1.1 Definition: Let X, Y be Banach spaces. If $F: X \to Y$ satisfies

$$\|F(x_1) - F(x_2)\| \leqq \beta\|x_1 - x_2\| \quad \text{for } x_1, x_2 \in D \subset X,$$

for some $\beta < \infty$, F is said to be *Lipschitz on D with constant* β, and we write $F \in Lip(D,\beta)$ or $F \in Lip(D)$. If F is Lipschitz on bounded sets, we write $F \in Lip_b$ (b for "bounded.")

1.1.2 Definition: Let H be a real Hilbert space. If $F: H \to H$ satisfies

$$(F(x) - F(y), x - y) \geqq \alpha \|x - y\|^2 \text{ for } x, y \in D \subset H,$$

for some $\alpha \geqq 0$, F is said to be *monotone on D with constant* α, and we write $F \in Mon(D,\alpha)$ or $F \in Mon(D)$. If $\alpha > 0$, F is said to be *uniformly monotone on D*. If F is uniformly monotone on bounded sets, we write $F \in Mon_b$. If

$$(F(x) - F(y), x - y) > 0 \quad \text{for } x, y \in D,$$

F is said to be *strictly monotone on D*.

Let X, Y be Banach spaces. Then $L(X,Y)$ denotes the set of linear operators from X to Y. Thus $L(R^n,R^n)$ is the set of n×n matrices.

For $x \in R^n$ and $A \in L(R^n,R^n)$, we have

$$x = (x_1,\cdots,x_n)^T \quad \text{and} \quad A = (a_{ij}),$$

In R^n, we use the following vector norms,

$$\|x\|_\infty = \sup_{1\leqq i\leqq n} |x_i| \quad \text{and} \quad \|x\|_p = [\sum_{i=1}^{n} |x_i|^p]^{1/p} \quad \text{for } 1 \leqq p < \infty,$$

and the corresponding matrix norms,

$$\|A\|_p = \sup_{\|x\|_p=1} \|A\|_p \quad \text{for } 1 \leqq p \leqq \infty.$$

If $A \in L(R^n,R^n)$ has eigenvalues $\lambda_1,\cdots,\lambda_n$, then $\sigma(A) = \{\lambda_1,\cdots,\lambda_n\}$ and $\rho(A) = \max_{1\leqq i\leqq n} |\lambda_i|$. To say A is positive definite (positive semi-definite) means A is symmetric and $\lambda_i > 0$ ($\lambda_i \geqq 0$) for $1 \leqq i \leqq n$. If A is symmetric, then it is well known that $\rho(A) = \|A\|_2$. If r is a scalar, $r + A$, is shorthand for $rI + A$, where I is the n×n identity matrix.

If $x, y \in R^n$, then

$$[x,y] = \{x + t(y - x) : 0 \leqq t \leqq 1\},$$

$$[x,y]^* = \{\xi : \xi_i = x_i + t_i(y_i - x_i): 0 \leqq t_i \leqq 1, 1 \leqq i \leqq n\}.$$

If $x \in R^n$, $x \geqq 0$ $(x > 0)$ means $x_i \geqq 0$ $(x_i > 0)$ for $1 \leqq i \leqq n$. If $A \in L(R^n,R^n)$, then $A \geqq 0$ $(A > 0)$ means $a_{ij} \geqq 0$ $(a_{ij} > 0)$ for $1 \leqq i,j \leqq n$. Furthermore, $|x| = (|x_1|,\cdots,|x_n|)^T$, and $|A| = (|a_{ij}|)$.

For $x \in X$, a Banach space, and $\rho > 0$, define the set

$$S(x,\rho) = \{y \in X : \|x - y\| < \rho\}.$$

1.2 Operators of Positive Type. The discretized versions of certain types of elliptic partial differential operators are often of so called positive type (see [10, P. 181].) For operators of positive type, maximum principles, similar to the differential maximum principles, are readily available. Furthermore, a bound on the inverse of an operator of positive type is often easy to obtain. We define operators of positive type in the next two definitions.

1.2.1 Definition: Let $\bar{\Omega}$ be a set with m elements, denoted $P_1,\cdots,P_m$. For $n < m$, let $\Omega = \{P_1,\cdots,P_n\}$, and let $\Omega' = \bar{\Omega} - \Omega$. Associated with each point, $P \in \Omega$, let there be a set $N(P) \subset \bar{\Omega}$, of "neighbors" of P satisfying $P \notin N(P)$. The neighborhood system $\{N(P)\}$ is said to be *irreducible* if, given $P \in \Omega$ and $Q \in \bar{\Omega}$, there are points $Q_1,\cdots,Q_k \in \Omega$ such that $Q_{i+1} \in N(Q_i)$ for $0 \leqq i \leqq k$, where $Q_0 = P$ and $Q_{k+1} = Q$. $(\Omega,\Omega', \bar{\Omega},\{N(P)\})$ is called a *mesh domain with neighborhood system* or, simply, *mesh domain*, and is called *proper* if $\{N(P)\}$ is irreducible. For $X = \Omega$, Ω', or $\bar{\Omega}$, let $\mathcal{F}(X)$ be the set of real valued functions on X, and for $u \in \mathcal{F}(X)$, let

$$\|u\|_{X,\infty} = \max_{P \in X}|u(P)|.$$

1.2.2 Definition: Let $(\Omega,\Omega', \bar{\Omega},\{N(P)\})$ be a proper mesh domain. Let the linear operator $\ell: \mathcal{F}(\bar{\Omega}) \to \mathcal{F}(\Omega)$ be defined by

(1.2.1) $$\ell u(P) = a(P,P)u(P) - \sum_{Q\epsilon N(P)} a(P,Q)u(Q),$$

where

(1.2.2) $$\begin{cases} \text{a)} \quad a(P,Q) > 0 \quad \text{for } P \in \Omega \text{ and } Q \in N(P), \\ \text{b)} \quad a(P,P) \geqq \sum_{Q\epsilon N(P)} a(P,Q) \quad \text{for } P \in \Omega \end{cases}$$

Then ℓ is of *positive type*. If equality holds in (1.2.2b), then ℓ is of *minimal positive type*. The set of operators of positive type on $\mathcal{F}(\Omega)$ will be denoted $\Pi(\Omega)$, and the set of operators of minimal positive type will be denoted $\Pi_0(\Omega)$.

The next two results are well known maximum principles.

<u>1.2.3 Theorem</u>: *Let $(\Omega,\Omega',\bar{\Omega},\{N(P)\})$ be a proper mesh domain. Let $\ell \in \Pi_0(\Omega)$ be given by (1.2.1). Let $u \in \mathcal{F}(\bar{\Omega})$.*

i) Suppose $\ell u \leqq 0$ on Ω, then u is constant on $\bar{\Omega}$ or

(1.2.3a) $$u(P) < \max_{Q\epsilon\Omega'} u(Q) \quad \text{for } P \in \Omega.$$

ii) Suppose $\ell u \geqq 0$ on Ω, then u is constant on $\bar{\Omega}$ or

(1.2.3b) $$\min_{Q\epsilon\Omega'} u(Q) < u(P) \quad \text{for } P \in \Omega.$$

iii) Suppose $\ell u = 0$ on Ω, then u is constant on $\bar{\Omega}$ or

(1.2.3c) $$\min_{Q\epsilon\Omega'} u(Q) < u(P) < \max_{Q\epsilon\Omega'} u(Q) \quad \text{for } P \in \Omega.$$

<u>Proof</u>: *i)* Suppose u attains its maximum, M, at $P \in \Omega$. Then, since $\ell \in \Pi_0(\Omega)$,

$$a(P,P)M = \sum_{Q\epsilon N(P)} a(P,Q)M \geqq \sum_{Q\epsilon N(P)} a(P,Q)u(Q) \geqq a(P,P)u(P) = a(P,P)M.$$

Hence,

$$\sum_{Q\in N(P)} a(P,Q)[M - u(Q)] = 0.$$

But, then by (1.2.2a), $u \equiv M$ on $N(P)$. But, since $\{N(P)\}$ is irreducible, $u \equiv M$ on $\bar{\Omega}$.

ii) Apply i) to $-u$.

iii) Apply i) and ii) to u.

<u>1.2.4 Theorem</u>: *Let $(\Omega, \Omega', \bar{\Omega}, \{N(P)\})$ be a proper mesh domain. Let $\ell \in \Pi(\Omega) - \Pi_0(\Omega)$ be given by (1.2.1). Let $u \in \mathcal{F}(\bar{\Omega})$.*

i) Suppose $\ell u \leqq 0$ on Ω, then $u \equiv K \leqq 0$ or

$$u(P) < \max(0, \max_{Q\in\Omega'} u(Q)) \quad \text{for } P \in \Omega. \tag{1.2.4a}$$

ii) Suppose $\ell u \geqq 0$ on Ω, then $u \equiv K \geqq 0$ or

$$\min(0, \min_{Q\in\Omega'} u(Q))) < u(P) \quad \text{for } P \in \Omega. \tag{1.2.4b}$$

iii) Suppose $\ell u = 0$ on Ω, then $u \equiv 0$ or

$$\min(0, \min_{Q\in\Omega'} u(Q)) < u(P) < \max(0, \max_{Q\in\Omega'} u(Q)) \quad \text{for } P \in \Omega. \tag{1.2.4c}$$

<u>Proof</u>: We need only prove i). Suppose u attains a non-negative maximum, M, at $P \in \Omega$. Then

$$a(P,P)M = a(P,P)u(P) \leqq \sum_{Q\in N(P)} a(P,Q)u(Q) \leqq \sum_{Q\in N(P)} a(P,Q)M \leqq a(P,P)M.$$

Hence, as in Theorem 1.2.2, $u \equiv M \geqq 0$. Now, since $\ell \notin \Pi_0(\Omega)$,

$$a(P_0,P_0) - \sum_{Q\in N(P_0)} u(P_0,Q) > 0$$

for some $P_0 \in \Omega$. Then

$$0 \geqq \ell u(P_0) = [a(P_0,P_0) - \sum_{Q\in N(P_0)} a(P_0,Q)]M,$$

and hence, $M \leq 0$. This completes the proof.

Let $\ell \in \Pi(\Omega)$ be given by (1.2.1) where $(\Omega, \Omega', \bar{\Omega}, \{N(P)\})$ is a proper mesh domain. Define $A_\ell = (a_{ij}) \in L(R^n, R^n)$ by

$$a_{ij} = \begin{cases} a(P_i, P_i) & \text{if } i = j \\ -a(P_i, P_j) & \text{if } i \neq j \text{ and } P_j \in N(P_i) \\ 0 & \text{if } i \neq j \text{ and } P_j \notin N(P_i). \end{cases}$$

Now, since $\ell \in \Pi(\Omega)$, A_ℓ is diagonally dominant and the diagonal dominance is strict in those rows corresponding to the points P_i for which $N(P_i) \cap \Omega' \neq \phi$. Such a point P_i exists since $\Omega' \neq \phi$ and $\{N(P)\}$ is irreducible. This also shows that A_ℓ is irreducible. Thus, A_ℓ is an irreducibly diagonally dominant M-matrix (see [28, P. 85].) In particular, A_ℓ is non-singular and $A_\ell^{-1} > 0$.

For $v \in \mathcal{F}(\Omega')$ or $\mathcal{F}(\bar{\Omega})$, let $b_v \in R^n$ be defined by

$$(b_v)_i = \sum_{Q \in N(P_i) \cap \Omega'} a(P_i, Q) v(Q)$$

for $1 \leq i \leq n$. Let $v \in \mathcal{F}(\Omega')$, and suppose $u \in \mathcal{F}(\bar{\Omega})$ satisfies

(1.2.5) $$u(P) = v(P) \quad \text{for } P \in \Omega'.$$

Let $f \in \mathcal{F}(\Omega)$ and define $x, \phi \in R^n$ by $x_i = u(P_i)$ and $\phi_i = f(P_i)$. Then

(1.2.6) $$A_\ell x - b_v = -\phi$$

if and only if

(1.2.7) $$\ell u(P) = -f(P) \quad \text{for } P \in \Omega.$$

Thus, to find the solution, u, to (1.2.7) subject to the boundary condition (1.2.5), it is sufficient to find the solution, x, to (1.2.6).

Since A_ℓ is non-singular, both u and x exist and are unique.

1.2.5 Example: Consider the boundary value problem

(1.2.8) $$\begin{cases} -u_{ss} - u_{tt} + \gamma(s,t)u = -f(s,t) & ; (s,t) \in D = (0,1)\times(0,1) \\ u(s,t) = v(s,t) & ; (s,t) \in \partial D \end{cases},$$

where

(1.2.9) $$\gamma(s,t) \geqq 0.$$

Let $h = {}^1/_{N+1}$ for some positive integer, N, and define

$$\Omega = \{(ih,jh) : 1 \leqq i,j \leqq N\},$$
$$\bar{\Omega} = \{(ih,jh) : 0 \leqq i,j \leqq N+1\} - \{(0,0),(0,1),(1,0),(1,1)\},$$

and $\Omega' = \bar{\Omega} - \Omega$. The usual 5-point difference approximation to (1.2.8) takes the form

(1.2.10) $$\begin{cases} \ell u(s,t) \equiv -\Delta_h u(s,t) + \gamma(s,t)u(s,t) = -f(s,t) & ; (s,t) \in \Omega \\ u(s,t) = v(s,t) & ; (s,t) \in \Omega' \end{cases},$$

where

(1.2.11) $$-h^2\Delta_h u(s,t) = 4u(s,t) - u(s+h,t) - u(s-h,t) - u(s,t+h) - u(s,t-h)$$

for $(s,t) \in \Omega$. Here $N(s,t) = \{(s+h,t),(s-h,t),(s,t+h),(s,t-h)\}$ for $(s,t) \in \Omega$. Then $(\Omega,\Omega',\bar{\Omega},\{N(P)\})$ is a proper mesh domain, $\ell \in \Pi(\Omega)$, and $-\Delta_h \in \Pi_0(\Omega)$.

Let ℓ be defined by (1.2.10) where (1.2.9) holds. In [1], Bers proved $\|A_\ell^{-1}\|_\infty \leqq \max_{\bar{\Omega}} \Phi / \min_{\Omega}(-\Delta_h\Phi)$ where Φ is any function in $\mathcal{F}(\bar{\Omega})$ which satisfies $\Phi \geqq 0$ on $\bar{\Omega}$ and $-\Delta_h\Phi > 0$ on Ω. A bound independent of h is obtained by noting that $\Delta_h\Phi = \Delta\Phi$ when Φ is a quadratic polynomial. We now extend this result to general operators of positive type.

1.2.6 Theorem: *Let $(\Omega, \Omega', \bar{\Omega}, \{N(P)\})$ be a proper mesh domain. Let $\ell \varepsilon \Pi(\Omega)$ and $u \varepsilon \mathfrak{J}(\bar{\Omega})$ and suppose $\ell u > 0$ on Ω. Then*

(1.2.12) i) $$\|A_\ell^{-1}\|_\infty = \frac{\max\limits_{\bar{\Omega}} u - \min(0, \min\limits_{\bar{\Omega}} u)}{\min\limits_{\Omega} \ell u}$$

ii) If $\ell \varepsilon \Pi_0(\Omega)$, then

(1.2.13) $$\|A_\ell^{-1}\|_\infty \leq \frac{\max\limits_{\bar{\Omega}} u - \min\limits_{\bar{\Omega}} u}{\min\limits_{\Omega} \ell u}.$$

iii) If $\ell_1 \varepsilon \Pi(\Omega)$ satisfies

$$\ell_1 w(P) = \ell w(P) + \gamma(P) w(P)$$

for $w \varepsilon \mathfrak{J}(\bar{\Omega})$ and $P \varepsilon \Omega$, where $\gamma(P) \geqq 0$, then

(1.2.14) $$\|A_{\ell_1}^{-1}\|_p \leq \|A_\ell^{-1}\|_p \quad \text{for } 1 \leqq p \leqq \infty$$

Proof: i) Let $w \varepsilon \mathfrak{J}(\bar{\Omega})$ satisfy

$$\begin{cases} \ell w(P) = \ell u(P) & ; \quad P \varepsilon \Omega \\ w(P) = 0 & ; \quad P \varepsilon \Omega' \end{cases}$$

Let $y, z \varepsilon R^n$ satisfy $y_i = w(P_i)$ and $z_i = \ell u(P_i)$. Then $A_\ell y = z$. Let $A_\ell^{-1} = (b_{ij}) \geqq 0$. Then for $1 \leqq i \leqq n$,

$$\max_{1 \leqq k \leqq n} y_k \geqq y_i = \sum_{j=1}^n b_{ij} z_j \geqq \min_{1 \leqq k \leqq n} z_k \sum_{j=1}^n b_{ij}$$

But, $\|A_\ell^{-1}\|_\infty = \max\limits_{1 \leqq i \leqq n} \sum\limits_{j=1}^n b_{ij}$. Hence, since $z > 0$,

$$\|A_\ell^{-1}\|_\infty \leq \frac{\max\limits_{1 \leqq k \leqq n} y_k}{\min\limits_{1 \leqq k \leqq n} z_k} = \frac{\max\limits_{\Omega} w}{\min\limits_{\Omega} \ell u}$$

Now, let $v \varepsilon \mathfrak{J}(\bar{\Omega})$ satisfy

$$\begin{cases} \ell v(P) = 0 & ; \quad P \varepsilon \Omega \\ v(P) = u(P) & ; \quad P \varepsilon \Omega' \end{cases}$$

Then, by Theorem 1.2.4,

$$\min(0, \min_{\bar{\Omega}} u) = \min(0, \min_{\Omega'} u) \leqq v(P)$$

Now, $w = u - v$. Hence

$$\max_{\Omega} w = \max_{\Omega} u - \min_{\Omega} v \leqq \max_{\bar{\Omega}} u - \min(0, \min_{\bar{\Omega}} u) \quad ,$$

which establishes (1.2.12).

ii) The proof is the same except, since $\ell \varepsilon \Pi_0(\Omega)$, we can use Theorem 1.2.3 instead of Theorem 1.2.4.

iii) Let D be the non-negative diagonal matrix with diagonal entries, $d_{ii} = \gamma(P_i)$. Then $A_{\ell_1} = A_\ell + D \geqq A_\ell$. Now, since $A_\ell^{-1} \geqq 0$ and $A_{\ell_1}^{-1} \geqq 0$, we have

$$0 \leqq A_{\ell_1}^{-1} \leqq A_\ell^{-1} \quad , \tag{1.2.15}$$

from which (1.2.14) follows.

Theorem 1.2.6 can be used with the following theorem.

<u>1.2.7 Theorem</u>: *Let $(\Omega, \Omega', \bar{\Omega}, \{N(P)\})$ be a proper mesh domain. Let $\ell \varepsilon \Pi(\Omega)$ and suppose $u \varepsilon \mathcal{F}(\bar{\Omega})$ satisfies*

$$\begin{cases} \ell u(P) = -f(P) & ; \quad P \varepsilon \Omega \\ u(P) = v(P) & ; \quad P \varepsilon \Omega' \end{cases}$$

Then

$$\|u\|_{\Omega,\infty} \leqq \|A_\ell^{-1}\|_\infty \|f\|_{\Omega,\infty} + \|v\|_{\Omega',\infty} \tag{1.2.16}$$

<u>Proof</u>: Let $u_1, u_2 \varepsilon \mathcal{F}(\bar{\Omega})$ satisfy

$$\begin{cases} \ell u_1(P) = -f(P) & ; \quad P \in \Omega \\ u_1(P) = 0 & ; \quad P \in \Omega' \end{cases} \quad \text{and} \quad \begin{cases} \ell u_2(P) = 0 & ; \quad P \in \Omega \\ u_2(P) = v(P) & ; \quad P \in \Omega' \end{cases}$$

respectively. Then $u = u_1 + u_2$. But $||u_1||_{\Omega,\infty} \leqq ||A_\ell^{-1}||_\infty \, ||f||_{\Omega,\infty}$, and, by Theorem 1.2.4, $||u_2||_{\Omega,\infty} \leqq ||v||_{\Omega',\infty}$. The result follows from the triangle inequality.

Consider the uniformly elliptic boundary value problem

$$(1.2.17) \quad \begin{cases} Lu(s,t) = -f(s,t) \quad ; \quad (s,t) \in D = (0,1)\times(0,1) \\ u(s,t) = v(s,t) \quad ; \quad (s,t) \in \partial D \\ \\ Lu(s,t) \equiv -(a(s,t)u_s)_s - (b(s,t)u_t)_t \\ a,b \in C^1(D),\ v \in C(\partial D),\ a(s,t) \geqq a_0 > 0,\ b(s,t) \geqq b_0 > 0. \end{cases}$$

Let $(\Omega,\Omega',\bar{\Omega},\{N(P)\})$ be as in Example 1.2.5. Then, approximating (1.2.17) by central differences, we obtain the discrete boundary value problem,

$$(1.2.18) \quad \begin{cases} \ell u(s,t) = -f(s,t) \quad ; \quad (s,t) \in \Omega \\ u(s,t) = v(s,t) \quad ; \quad (s,t) \in \Omega' \end{cases} ,$$

where, for $(s,t) \in \Omega$,

$$(1.2.19) \quad \begin{aligned} h^2 \ell u(s,t) = {} & [a(s+\tfrac{h}{2},t) + a(s-\tfrac{h}{2},t) + b(s,t+\tfrac{h}{2}) + b(s,t-\tfrac{h}{2})]u(s,t) \\ & - a(s+\tfrac{h}{2},t)u(s+h,t) - a(s-\tfrac{h}{2},t)u(s-h,t) \\ & - b(s,t+\tfrac{h}{2})u(s,t+h) - b(s,t-\tfrac{h}{2})u(s,t-h) \end{aligned}$$

Let

$$(1.2.20) \quad w_\alpha(s,t) = -e^{\alpha s}$$

Then

$$(1.2.21) \quad Lw_\alpha(s,t) = \alpha e^{\alpha s}[\alpha a(s,t) + a_s(s,t)]$$

Since $a(s,t) \geqq a_0 > 0$, we can pick α such that

(1.2.22) $$Lw_\alpha \geqq m > 0 \quad \text{on } D$$

Now, $$h^2 \ell w_\alpha(s,t) = a(s+\tfrac{h}{2},t)(e^{\alpha(s+h)} - e^{\alpha s}) - a(s-\tfrac{h}{2},t)(e^{\alpha s} - e^{\alpha(s-h)})$$

$$= e^{\alpha s}[2a(s+\tfrac{h}{2},t)e^{\alpha\frac{h}{2}}\sinh \alpha\tfrac{h}{2} - 2a(s-\tfrac{h}{2},t)e^{-\alpha\frac{h}{2}}\sinh \alpha\tfrac{h}{2}]$$

$$= \frac{\sinh \alpha\frac{h}{2}}{\alpha\frac{h}{2}} \alpha h e^{\alpha s}[\xi(s,t,\tfrac{h}{2}) - \xi(s,t,-\tfrac{h}{2})],$$

where $\xi(s,t,\theta) = a(s+\theta,t)e^{\alpha\theta}$. Now

$\frac{\partial}{\partial\theta}\xi(s,t,\theta) = e^{\alpha\theta}[\alpha a(s+\theta,t) + a_s(s+\theta,t)]$. Hence, from the Mean Value Theorem, for some $\theta = \theta(s,t) \varepsilon [-\frac{h}{2},\frac{h}{2}]$,

$$\ell w_\alpha(s,t) = \frac{\sinh \alpha\frac{h}{2}}{\alpha\frac{h}{2}} \alpha e^{\alpha s} e^{\alpha\theta}[\alpha a(s+\theta,t) + a_s(s+\theta,t)]$$

$$= \frac{\sinh \alpha\frac{h}{2}}{\alpha\frac{h}{2}} Lw_\alpha(s+\theta,t)$$

Now, $\frac{1}{s}\sinh(s) \geqq 1$ for all $s \neq 0$. Hence there is a $K = K(\alpha,h) \geqq 1$ such that

(1.2.23) $$\ell w_\alpha(s,t) = KLw_\alpha(s+\theta,t)$$

Now, if $(s,t) \varepsilon \Omega$, then $(s+\theta,t) \varepsilon \bar{D}$. Thus, if (1.2.22) holds, then

(1.2.24) $$\min_\Omega \ell w_\alpha = \min_D Lw_\alpha > 0$$

Thus, by Theorem 1.2.6, if (1.2.22) holds, then

(1.2.25) $$\|A_\ell^{-1}\|_\infty \leqq \frac{\max_{\bar{D}} w_\alpha - \min_{\bar{D}} w_\alpha}{\min_D Lw_\alpha}$$

We note that (1.2.25) gives a bound independent of the mesh size, h.

Let $a(s,t) = b(s,t) = 1$. Then $-L = \Delta$, the Laplacian, and $\ell = -\Delta_h$, which was given in (1.2.11). Now, $-\Delta e^{\alpha s} = \alpha^2 e^{\alpha s} > 0$ whenever $\alpha \neq 0$. Hence

$$\|A_{-\Delta_h}^{-1}\|_\infty \leqq \min\left(\min_{\alpha>0} \frac{e^\alpha - 1}{\alpha^2}, \min_{\alpha<0} \frac{1 - e^\alpha}{\alpha^2 e^\alpha}\right) \approx \frac{e^{1.6} - 1}{(1.6)^2} < 1.545$$

In this case, a sharper bound can be obtained by employing a different test function. Let

(1.2.26) $$v_\alpha(s,t) = \alpha s - s^2$$

Then , if L is given by (1.2.17),

(1.2.27) $$Lv_\alpha(s,t) = 2a(s,t) - (\alpha - 2s)a_s(s,t).$$

Thus, if $a_s \not\equiv 0$ on $\bar{D}$, we can pick α in order to insure

(1.2.28) $$Lv_\alpha \geqq m > 0 \quad \text{on } D$$

Now, $h^2 \ell v_\alpha(s,t) = a(s+\frac{h}{2},t)[v_\alpha(s,t) - v_\alpha(s+h,t)] - a(s-\frac{h}{2},t)[v_\alpha(s-h,t) - v_\alpha(s,t)]$

$$= a(s+\tfrac{h}{2},t)(2sh + h^2 - \alpha h) - a(s-\tfrac{h}{2},t)(2sh - h^2 - \alpha h)$$

$$= h[\eta(s,t,\tfrac{h}{2}) - \eta(s,t,-\tfrac{h}{2})],$$

where

$$\eta(s,t,\theta) = a(s+\theta,t)(2s - \alpha + 2\theta)$$

Now, $\frac{\partial}{\partial\theta}\eta(s,t,\theta) = a_s(s+\theta,t)(2s - \alpha + 2\theta) + 2a(s+\theta,t)$. Hence, for some $\theta = \theta(s,t) \in [-\frac{h}{2},\frac{h}{2}]$, from the Mean Value Theorem,

$$\ell v_\alpha(s,t) = 2a(s+\theta,t) - a_s(s+\theta,t)(\alpha - 2(s + \theta))$$

Hence

(1.2.29) $$\ell v_\alpha(s,t) = Lv_\alpha(s+\theta,t)$$

Now, if $(s,t) \in \Omega$, then $(s+\theta,t) \in \bar{D}$. Thus, if (1.2.28) holds, then

(1.2.30) $$\min_\Omega \ell v_\alpha \geqq \min_D Lv_\alpha \geqq m > 0$$

Hence, by Theorem 1.2.6, if (1.2.28) holds, then

$$\left\|A_{\ell}^{-1}\right\|_{\infty} \leq \frac{\max\limits_{\bar{D}} v_{\alpha} - \min\limits_{\bar{D}} v_{\alpha}}{\min\limits_{D} Lv_{\alpha}} \tag{1.2.31}$$

Here again, (1.2.31) gives a bound independent of the mesh size, h. Now

$$\max_{\bar{D}} v_{\alpha} - \min_{\bar{D}} v_{\alpha} = \begin{cases} 1 - \alpha \ ; & \alpha \leq 0 \\ \frac{\alpha^2}{4} + 1 - \alpha \ ; & 0 \leq \alpha \leq 1 \\ \frac{\alpha^2}{4} \ ; & 1 \leq \alpha \leq 2 \\ \alpha - 1 \ ; & 2 \leq \alpha \end{cases} \tag{1.2.32}$$

and $-\Delta_h v_{\alpha} = 2 > 0$. So minimizing (1.2.31) with respect to α, we find at $\alpha = 1$,

$$\left\|A_{-\Delta_h}^{-1}\right\|_{\infty} \leq \frac{1}{8} \tag{1.2.33}$$

This gives a considerably better estimate than when w_{α} is used as a test function.

<u>1.2.8 Example</u>: Let $h = {}^{1}/_{N+1}$ for some positive integer, N, and set $\Omega = \{h, 2h, \cdots, Nh\}$, $\Omega' = \{0,1\}$, $\bar{\Omega} = \Omega \cup \Omega'$, and $N(ih) = \{(i-1)h, (i+1)h\}$. Then $(\Omega, \Omega', \bar{\Omega}, \{N(P)\})$ is a proper mesh domain. We may approximate the problem

$$\begin{cases} Lu(s) \equiv -(a(s)u'(s))' = -f(s) \ ; \quad s \in D = (0,1) \\ u(0) = v_0, \quad u(1) = v_1 \\ a \in C^1(0,1), \quad a(s) \geq a_0 > 0 \end{cases} \tag{1.2.34}$$

by the discrete problem

$$\begin{cases} \ell u(s) = -f(s) \ ; \quad s \in \Omega \\ u(0) = v_0, \ u(1) = v_1 \end{cases} \tag{1.2.35}$$

where

(1.2.36) $h^2 u(s) = [a(s+\frac{h}{2}) + a(s-\frac{h}{2})]u(s) - a(s+\frac{h}{2})u(s+h) - a(s-\frac{h}{2})u(s-h)$.

Now, since w_α and v_α, given in (1.2.20) and (1.2.26) respectively are independent of t, we see that (1.2.25) or (1.2.31) gives a bound on $\|A_\ell^{-1}\|_\infty$ provided (1.2.22) or (1.2.28) respectively is satisfied. In particular, if $a \equiv 1$, so that $L = -\frac{d^2}{ds^2}$ and $\ell \equiv -\delta_h^2$, we have

(1.2.37) $$\|A^{-1}_{-\delta_h^2}\|_\infty \leqq \frac{1}{8}$$

This is the best possible bound independent of h since $\lim_{h\to 0}\|A^{-1}_{-\delta_h^2}\|_\infty = \frac{1}{8}$. We see this as follows. For any $\ell \in \Pi(\Omega)$, $A_\ell^{-1} > 0$, and so $\|A_\ell^{-1}\|_\infty = \|A_\ell^{-1}(1,1,\cdots,1)^T\|_\infty$. I.e.

(1.2.38a) $$\|A_\ell^{-1}\|_\infty = \|u\|_{\Omega,\infty}$$,

where

(1.2.38b) $$\begin{cases} \ell u(P) = 1 & ; \quad P \in \Omega \\ u(P) = 0 & ; \quad P \in \Omega' \end{cases}$$

So, let $u_h \in \mathcal{J}(\Omega)$ be the solution of

$$\begin{cases} -\delta_h^2 u_h = 1 & ; \quad P \in \Omega \; (= \Omega_h) \\ u_h = 0 & ; \quad P \in \Omega' \end{cases}$$

and $u \in C^2(0,1)$ the solution of

$$-u''(s) = 1, \; s \in (0,1) \quad ; \quad u(0) = u(1) = 0.$$

It is well known that $\sup_{P\in\Omega_h} |u_h(P) - u(P)| \to 0$ $(h \to 0.)$ But $u(s) = \frac{1}{2}(s - s^2)$ and $\sup_{0<s<1} |u(s)| = \frac{1}{8}$. Hence

$$\|A^{-1}_{-\delta_h^2}\|_\infty = \|u_h\|_{\Omega_h,\infty} \to \frac{1}{8}\ (h \to 0.)$$

1.2.9 Example: Let L be given by (1.2.17) (or (1.2.34)) where $a_s \geqq d$. Then we can insure $\min_D Lw_\alpha > 0$ by picking $\alpha > \max(0,-d/a_0)$. We find from (1.2.25)

(1.2.39) $$\|A_\ell^{-1}\|_\infty \leqq \frac{e^\alpha - 1}{\alpha^2 a_0 + \alpha d} \quad \text{when } \alpha > \max(0,-\frac{d}{a_0})$$

Bounds on $\|A_\ell^{-1}\|_2$ will also be useful. We note that with ℓ given by (1.2.19), A_ℓ has the form

(1.2.40) $$A_\ell = A_{\ell_H} + A_{\ell_V}$$

where

(1.2.41a) $$h^2\ell_H u(s,t) = [a(s+\tfrac{h}{2},t) + a(s-\tfrac{h}{2},t)]u(s,t) - a(s+\tfrac{h}{2},t)u(s+h,t) - a(s-\tfrac{h}{2},t)u(s-h,t)$$

and

(1.2.41b) $$h^2\ell_V u(s,t) = [b(s,t+\tfrac{h}{2}) + b(s,t-\tfrac{h}{2})]u(s,t) - b(s,t+\tfrac{h}{2})u(s,t+h) - b(s,t-\tfrac{h}{2})u(s,t-h)$$

For positive definite $B \in L(R^n,R^n)$, let $\mu(B)$ be the smallest eigenvalue of B. Then $\|B^{-1}\|_2 = \frac{1}{\mu(B)}$. Now, A_{ℓ_H} and A_{ℓ_V} are positive definite, and $\mu(A_{\ell_H}) + \mu(A_{\ell_V}) \leqq \mu(A_\ell)$. Thus

$$\|A_\ell^{-1}\|_2 \leqq \frac{1}{\mu(A_{\ell_H}) + \mu(A_{\ell_V})}$$

Then, by [28, P. 219, Pbm. 6],

(1.2.42) $$\|A_\ell^{-1}\|_2 \leqq \frac{1}{2(a_0 + b_0)(1 - \cos{}^{\pi}/_{N+1})}$$

This can be extended to more general regions and discretizations in an obvious way. Other and generally sharper estimates for determining

$\mu(A_{\ell_H})$ and $\mu(A_{\ell_V})$ are obtained by other methods. See, e.g., [27].

1.3 Existence and uniqueness of solutions of the discrete mildly nonlinear problem. Consider (1.2.17). If f depends on u, we obtain the nonlinear problem

$$(1.3.1)\qquad \begin{cases} Lu(s,t) = -f(s,t,u) & ; \quad (s,t) \in D \\ u(s,t) = v(s,t) & ; \quad (s,t) \in \partial D \end{cases}$$

If L is discretized as in (1.2.19), we obtain the discrete nonlinear problem

$$(1.3.2)\qquad \begin{cases} \ell u(P) = -f(P,u(P)) & ; \quad P \in \Omega \\ u(P) = v(P) & ; \quad P \in \Omega' \end{cases} ,$$

which is equivalent to the problem

$$(1.3.3)\qquad \begin{cases} \text{a)} & F(x) \equiv A_\ell x - b_V + \psi(x) = 0 \\ \text{b)} & \psi(x) \equiv (f(P_1,x_1),\cdots,f(P_n,x_n))^T \end{cases} ,$$

which motivates the following definition.

1.3.1 Definition: Let ϕ: $R^n \to R^n$ be of the form

$$(1.3.4)\qquad \phi(x) = (\phi_1(x_1),\cdots,\phi_n(x_n))^T$$

Then ϕ is said to be *diagonally nonlinear*, and we write $\phi \in D(R^n)$.

Let F: $R^n \to R^n$ be of the form

$$(1.3.5)\qquad F(x) = Ax + \phi(x) ,$$

where $A \in L(R^n,R^n)$ and $\phi \in D(R^n)$. Then F is said to be *mildly nonlinear*, and we write $F \in M(R^n)$.

1.3.2 Definition: Let F: X $\to$ Y where X and Y are real Banach spaces. If for some $x \in X$ and some $L \in L(X,Y)$,

$$\lim_{\|h\|\to 0} \frac{\|F(x+h) - F(x) - Lh\|}{\|h\|} = 0 \quad ,$$

then F is said to be *(Frechet)-differentiable at x*, and the derivative is denoted $F'(x) = L$.

A complete discussion of Frechet-differentiation can be found in Vainberg [29]. In this paper, the term *differentiable* will mean *Frechet-differentiable*.

If $F: R^n \to R^n$, it is not sufficient for F to be differentiable that each of the partial derivatives, $\partial f_i/\partial x_j$, exist. However, if $\phi \in D(R^n)$, and ϕ_i is differentiable on R for $1 \leqq i \leqq n$, then it is easy to verify that ϕ is differentiable on R^n and that $\phi'(x)$ is the diagonal matrix with diagonal entries $(\phi'(x))_{ii} = \phi_i'(x_i)$. If $F: R^n \to R^n$ is given by (1.3.5), then F is differentiable on R^n and

$$F'(x) = A + \phi'(x) \quad .$$

Let $x,y \in R^n$. Then, by the Mean Value Theorem applied component-wise,

$$\phi(x) - \phi(y) = \phi'(\xi)(x - y) \quad \text{for some } \xi \in [x,y]^* \quad .$$

Hence

$$F(x) - F(y) = F'(\xi)(x - y) \quad \text{for some } \xi \in [x,y]^* \quad . \tag{1.3.6}$$

For continuous $G: R^n \to L(R^n,R^n)$ and continuous $z: R \to R^n$, define

$$\int_0^1 G(z(t))dt \equiv \left(\int_0^1 G_{ij}(z(t))dt\right)$$

Then it is immediately verified that

$$F(x) - F(y) = \left[\int_0^1 F'(y + t(x-y))dt\right](x - y) \tag{1.3.7}$$

When $F \in M(R^n)$ is not necessarily differentiable, there is a natural way to define a "divided difference" of F.

<u>1.3.3 Definition</u>: Let $\phi \in D(R^n)$. Let $x,y \in R^n$. Then $\phi^D(x,y)$ is the diagonal matrix with diagonal entries

$$(1.3.8) \quad \phi^D_{ii}(x,y) = \begin{cases} \dfrac{\phi_i(x_i) - \phi_i(y_i)}{x_i - y_i} & \text{, if } x_i \neq y_i \\ \left.\begin{cases} \liminf\limits_{t \to 0} \dfrac{\phi_i(y_i+t) - \phi_i(y_i)}{t} & \text{, if finite} \\ \phi_i(y_i+1) - \phi_i(y_i) & \text{, otherwise} \end{cases}\right\} & \text{, if } x_i = y_i \end{cases}.$$

If $F(x) = Ax + \phi(x)$ where $A \in L(R^n,R^n)$, then define $F^D(x,y)$ by

$$(1.3.9) \qquad F^D(x,y) = A + \Phi^D(x,y)$$

Let $F \in M(R^n)$, the we see immediately that

$$(1.3.10) \qquad F(x) - F(y) = F^D(x,y)(x - y)$$

for $x,y \in R^n$.

We now consider conditions under which the equation

$$F(x) = 0$$

has a unique solution.

<u>1.3.4 Definiton</u>: Let X, Y be Banach spaces. $F: X \to Y$ is said to be *norm coercive* if $\|F(x)\| \to \infty$ when $\|x\| \to \infty$.

We now state the Domain Invariance Theorem (see [26, P. 98] or [2, P. 87]) and a special case of a result of Rheinboldt [25, Thm. 4.7].

<u>1.3.5 Theorem (Domain Invariance)</u>: *Let $F: R^n \to R^n$ be one-to-one and continuous. Then F is an open mapping.*

<u>1.3.6 Theorem</u>: *Let $F: R^n \to R^n$ be a norm coercive local homeomorphism.*

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

The following result is a corollary of Theorems 1.3.5 and 1.3.6.

1.3.7 Corollary: *Let $F: R^n \to R^n$ be one-to-one, continuous and norm coercive. Then F is a homeomorphism of R^n onto R^n.*

Proof: F is a local homeomorphism by Theorem 1.3.5, and so the result follows from Theorem 1.3.6.

Much of the work involved in establishing Theorem 1.3.6 is in showing F is globally one-to one. Since we assume this in Corollary 1.3.7, we do not need all the power of Theorem 1.3.6. For completeness, we give a direct proof of Corollary 1.3.7.

Direct proof of Corollary 1.3.7: By Theorem 1.3.5, F^{-1} is continuous. Thus, we need only show F is onto. Since, by Theorem 1.3.5, $F(R^n)$ is open, it is sufficient to show $F(R^n)$ is closed. Let $y_k \to y \in R^n$ where $\{y_k\} \subset F(R^n)$. Then there is $\{x_k\} \subset R^n$ such that $F(x_k) = y_k$. Since $\{y_k\}$ is bounded and F is norm coercive, $\{x_k\}$ is bounded. But then a subsequence, $\{x_{m_k}\}$ converges to, say, $x \in R^n$. By the continuity of F, $y = F(x) \in F(R^n)$. Thus $F(R^n)$ is closed.

A uniformly monotone function is one-to-one and norm coercive. Thus Corollary 1.3.7 contains the R^n version fo the following result of Minty [18].

1.3.8 Theorem: *Let H be a real Hilbert space. Suppose $F: H \to H$ is continuous and uniformly monotone. Then F is a homeomorphism of H onto H.*

We now apply Corollary 1.3.7 to mildly nonlinear functions.

1.3.9 Corollary: *Let $F \in M(R^n)$ be continuous and norm coercive and sup-*

pose $[F^D(x,y)]^{-1}$ exists for each x, $y \in R^n$. Then F is a homeomorphism of R^n onto R^n.

Proof: From (1.3.10), we see that F is one-to-one. The result then follows from Corollary 1.3.7.

1.3.10 Corollary: *Let $F \in M(R^n)$ be continuous and suppose $[F^D(x,y)]^{-1}$ exists for each x, $y \in R^n$ and satisfies $\| [F^D(x,y)]^{-1} \| \leqq K < \infty$ independently of x and y. Then F is a homeomorphism of R^n onto R^n.*

Proof: By Corollary 1.3.9, we need only show F is norm coercive. But this follows from

$$\|x\| = \| [F^D(x,0)]^{-1}[F(x) - F(0)] \| \leqq K \| F(x) - F(0) \|$$

1.3.11 Example: Let $F(x) = Ax + \phi(x)$ where $A \in L(R^n,R^n)$ and $\phi \in D(R^n)$.

i) Suppose A is an M-matrix and $\phi \in Mon(R^n)$. Then $\phi^D(x,y) \geqq 0$, and so $\|[F^D(x,y)]^{-1}\|_\infty \leqq \| A^{-1}\|_\infty$. Thus by Corollary 1.3.10, F is a homeomorphism.

ii) Suppose A is symmetric with least eigenvalue μ and $\phi - dI \in Mon(R^n)$ for some $d > -\mu$. Then

$$F^D(x,y) = A - \mu I + \phi^D(x,y) - dI + (d + \mu)I ,$$

and so the least eigenvalue of $F^D(x,y)$ is at least as great as $d + \mu > 0$. Hence $\| [F^D(x,y)]^{-1}\|_2 \leqq \frac{1}{d + \mu}$,and so, by Corollary 1.3.10, F is a homeomorphism.

iii) Suppose A is symmetric with least eigenvalue μ. Let

$$\phi_i(t) \equiv g(t) - \mu t \quad \text{for } 1 \leqq i \leqq n ,$$

where

$$g(t) = \begin{cases} \log(t+1) & \text{if } t \geqq 0 \\ t & \text{if } t \leqq 0 \end{cases}$$

Now, $A - \mu I$ is positive semi-definite. Thus for $x \in R^n$,

$$n\|F(x)\|_\infty \|x\|_\infty \geqq (F(x),x) = ((A-\mu)x,x) + \sum_{i=1}^{n} g(x_i)x_i$$
$$\geqq \sum_{i=1}^{n} |x_i|\log(|x_i|+1)$$
$$\geqq \|x\|_\infty \log(\|x\|_\infty + 1)$$

Hence, $\|F(x)\|_\infty \geqq \frac{1}{n}\log(\|x\|_\infty + 1)$, which shows that F is norm coercive.

Since $\phi_i'(t) > -\mu$, $[F^D(x,y)]^{-1}$ exists for each $x, y \in R^n$. Hence, by Corollary 1.3.9, F is a homeomorphism.

1.3.12 Remark: The functions of Example 1.3.11 ii), iii) can be shown to have unique roots without the use of Theorems 1.3.5 and 1.3.6. For instance, consider the function in ii). Let $g: R^n \to R$ be defined by

$$g(x) = \frac{1}{2}(Ax,x) + \sum_{j=1}^{n} \int_0^{x_j} \phi_j(t)dt$$

Then $[g'(t)]^T = F(x)$. Thus F has a root if g attains its minimum, and, since $F^D(x,y)$ is non-singular for each x and y, the root must be unique. Now

$$g(x) = g(0) + \int_0^1 (F(tx),x)dt$$
$$= g(0) + \int_0^1 (F(tx) - F(0),x)dt + (F(0),x)$$
$$= g(0) + (F(0),x) + \int_0^1 (F^D(tx,0)tx,x)dt$$
$$\geqq g(0) - \|F(0)\|_2\|x\|_2 + \int_0^1 (\mu + d)t\|x\|_2^2 dt$$
$$= g(0) - \|F(0)\|_2\|x\|_2 + \frac{1}{2}(\mu + d)\|x\|_2^2$$

Clearly, then, $g(x) \to +\infty (\|x\|_2 \to \infty)$, which shows that g attains its minimum.

1.4 Bounds on the solution. Let ℓ be defined by (1.2.19) and consider the nonlinear problem (1.3.2). We note that besides being an irreducibly diagonally dominant M-matrix, A_ℓ is positive definite. Let the least eigenvalue of A be μ and suppose $f(P,\cdot) - d \in \mathrm{Mon}(R)$ for each $P \in \Omega$ and some $d > -\mu$. Then by Example 1.3.11 ii), (1.3.2) has a unique solution, u^*. We now derive a priori bounds on u^*. These will be useful in picking a good initial approximation for an iterative process and later in obtaining globally convergent ADI algorithms. We also obtain a priori error bounds which to to zero as the error goes to zero.

In the sequel, for $u \in \mathcal{F}(\Omega)$, let $\underline{u} \in R^n$ be the vector with components $\underline{u}_i = u(P_i)$. Let $x^* = \underline{u}^*$. Then x^* is the unique root of the function, F, given in (1.3.3). Furthermore, assume, for convenience, $f(P,\cdot) \in \mathrm{Mon}(R)$.

Suppose we know a priori that $K_1 \leqq u^*(s,t) \leqq K_2$. Define

$$\hat{f}(P,u) = \begin{cases} f(P,K_1) & \text{if} \quad u \leqq K_1 \\ f(P,u) & \text{if} \quad K_1 \leqq u \leqq K_2 \\ f(P,K_2) & \text{if} \quad K_2 \leqq u \end{cases} \tag{1.4.1}$$

and

$$\hat{\psi}(x) = (\hat{f}(P,x_1),\cdots,\hat{f}(P,x_n))^T \tag{1.4.2}$$

Then u* and x* are the unique solutions of

$$\begin{cases} \ell u(P) = -\hat{f}(P,u(P)) & ; \quad P \in \Omega \\ u(P) = v(P) & ; \quad P \in \Omega' \end{cases} \tag{1.4.3}$$

and

$$\hat{F}(x) \equiv A_\ell x - b_v + \hat{\psi}(x) = 0 \tag{1.4.4}$$

respectively. Thus, we may seek the solution of (1.4.4) instead of that of (1.3.3) and enjoy the added assumption that $\hat{\psi}$ and $\hat{\psi}^D$ are bounded as functions of x and (x,y) respectively. This approach will be used in Chapters IV and V.

From $F(x) = F(x) - F(x^*) = F^D(x,x^*)(x - x^*)$, we have

(1.4.5) $$x - x^* = [F^D(x,x^*)]^{-1}F(x) \quad ,$$

which yields an error bound that goes to zero as the error goes to zero. For instance,

(1.4.6) $$\| x - x^*\|_2 \leqq \frac{1}{\mu + d} \|F(x)\|_2$$

We also have

$$0 \leqq [F^D(x,y)]^{-1} \leqq A_\ell^{-1} \quad ,$$

so that

(1.4.7) $$\| x - x^*\|_p \leqq \| A_\ell^{-1}\|_p \|F(x)\|_p \quad \text{for } 1 \leqq p \leqq \infty$$

A crude two-sided bound on u^* can be obtained in the following way. Suppose $x_0 \in R^n$ satisfies $A_\ell x_0 = b_v$. Then

$$\| x_0\|_\infty \leqq \| v\|_{\Omega',\infty} \quad ,$$

and

$$\|x_0 - x^*\|_\infty \leqq \|A_\ell^{-1}\|_\infty \| \psi(x_0)\|_\infty \quad ,$$

which gives the following nonlinear analog of (1.2.16).

(1.4.8) $$\| u^*\|_{\Omega,\infty} \leqq \| A_\ell^{-1}\|_\infty \sup_{\substack{P\in\Omega \\ |t|\leqq\|v\|_{\Omega',\infty}}} |f(P,t)| + \|v\|_{\Omega',\infty}$$

We can get a sharper estimate from (1.4.5). Suppose

$$F(x) = G(x) - H(x) \quad \text{where } G(x), H(x) \geqq 0$$

Then

(1.4.9) $$-A_\ell^{-1}H(x) \leqq x - x^* \leqq A_\ell^{-1}G(x)$$

The use of (1.4.9) is illustrated in the following examples.

<u>1.4.1 Example</u>: Let $\ell = -\Delta_h$ and $f(P,u) = e^u$, where Ω is as in Example 1.2.5.

Suppose $v(s,t) = \alpha s + \beta t$ is defined on $\bar{\Omega}$ instead of just on Ω' where $\alpha, \beta \geqq 0$. Since $f_u \geqq 0$, a unique solution, u^*, to (1.3.2) exists. Now, $F(0) = -b_v + \psi(0)$ where $\psi(0) = (1,1,\cdots,1)^T \geqq 0$ and $b_v \geqq 0$. Thus

$$-A_\ell^{-1}b_v \leqq -x^* \leqq A_\ell^{-1}\psi(0)$$

Now, $\underline{v} = A_\ell^{-1}b_v$ and $\| A_\ell^{-1}\psi(0)\|_\infty = \|A_\ell^{-1}\|_\infty \leqq \frac{1}{8}$. Thus,

$$(1.4.10) \qquad -\frac{1}{8} \leqq u^*(s,t) \leqq \alpha s + \beta t$$

(1.4.10) can be strengthened independently of α, β. Let $w_1(s,t) = \frac{1}{2}(s^2 - s) \leqq 0$. Then $f(P,w_1) = e^{w_1} \leqq 1$, and $\ell w_1 = -1$. Thus, $F(\underline{w_1}) \leqq 0$. Hence, from (1.4.9), $x^* \geqq \underline{w_1}$. Likewise, $x^* \geqq \underline{w_2}$, where $w_2(s,t) = \frac{1}{2}(t^2 - t)$. Hence, (1.4.10) can be strengthened to

$$(1.4.11) \qquad -\frac{1}{8} \leqq \frac{1}{2}\max(s^2 - s, t^2 - t) \leqq u^*(s,t) \leqq \alpha s + \beta t$$

<u>1.4.2 Example</u>: Let ℓ and v be as in Example 1.4.1, and suppose $f(P,u) = u^{2m+1}$ for some integer $m \geqq 0$. Then a unique solution, u^*, to (1.3.2) exists as above. Now, $\psi(0) = 0$. Thus, as above,

$$(1.4.12) \qquad 0 \leqq u^*(s,t) \leqq \alpha s + \beta t$$

The lower bounds in both (1.4.11) and (1.4.12) are not sharp near the boundary. Let

$$(1.4.13) \qquad \tilde{F}(x) = A_\ell x - b_v + \psi(x^*)$$

Then x^* is the unique root of $\tilde{F}$, and, from (1.4.5),

$$(1.4.14) \qquad x^* = x - A_\ell^{-1}\tilde{F}(x)$$

<u>1.4.3 Example</u>: Let F be as in Example 1.4.1. Now, by (1.4.11), $u^* \leqq \alpha + \beta$.

Thus, $\psi(x^*) \leqq e^{\alpha+\beta}(1,1,\cdots,1)^T$. Let $w_1(s,t) = \frac{1}{2} e^{\alpha+\beta}(s^2 - s) \leqq 0$, and $w_2(s,t) = \frac{1}{2} e^{\alpha+\beta}(t^2 - t) \leqq 0$. Then $\ell w_1(s,t) = -e^{\alpha+\beta}$. Thus, $\ell \underline{w_1} + \psi(x^*) \leqq 0$, and so $\hat{F}(\underline{w_1}) = A_\ell \underline{w_1} - b_v + \psi(x^*) = A_\ell \underline{w_1} - b_{w_1} + \psi(x^*) + b_{\underline{w_1}} - b_v = \underline{\ell w_1} + \psi(x^*) + b_{w_1} - b_v \leqq b_{w_1} - b_v = -b_v$, since $w_1 \leqq 0$. Hence, from (1.4.14), $x^* \geqq \underline{w_1} + A_\ell^{-1} b_v = \underline{w_1} + \underline{v}$. Likewise, $x^* \geqq \underline{w_2} + \underline{v}$, and so

$$(1.4.15) \quad \alpha s + \beta t + \frac{1}{2} e^{\alpha+\beta} \max(s^2-s, t^2-t) \leqq u^*(s,t) \leqq \alpha s + \beta t$$

<u>1.4.4 Example</u>: Let F be as in Example 1.4.2. Let $w_1(s,t) = \frac{1}{2} (\alpha + \beta)^{2m+1}(s^2 - s) \leqq 0$, $w_2(s,t) = \frac{1}{2} (\alpha + \beta)^{2m+1}(t^2 - t) \leqq 0$. Then, as in Example 1.4.3,

$$(1.4.16) \quad \alpha s + \beta t + \frac{1}{2} (\alpha + \beta)^{2m+1} \max(s^2 - s, t^2 - t) \leqq u^*(s,t) \leqq \alpha s + \beta t.$$

We note that (1.4.15) and (1.4.16) are sharper than (1.4.11) and (1.4.12) respectively near the boundary, but probably not in the interior.

<u>1.5 Analogs in the continuous case</u>. In this section, we present results for a uniformly elliptic partial differential operator, L, analogous to the results of the previous sections of this chapter. The main result of this section will give conditions on L, f, ϕ, and $D \subset R^n$ which will guarantee that the mildly nonlinear boundary value problem,

$$\begin{cases} Lu(x) = f(x,u(x)) & ; \quad x \in D \\ u(x) = \phi(x) & ; \quad x \in \partial D \end{cases},$$

has a unique solution.

We present first some notation and definitions.

Let $f: G \subset R^m \to R$. If f is continuous on G, we say $f \in C(G)$. Furthermore, if f is bounded on G, we set $\|f\|_G = \sup_{x \in G} |f(x)|$.

Let $\beta = (i_1,\cdots,i_m)$ and $|\beta| = i_1 + \cdots + i_m$, where the i_j are non-negative integers, and define the operator

$$D^\beta = \frac{\partial^{|\beta|}}{\partial x_1^{i_1}\cdots\partial x_m^{i_m}}$$

If $G \subset R^m$ is open and f is k times continuously differentiable on G, then we say $f \in C^k(G)$. Furthermore, if $D^\beta f$ can be extended to a continuous function on $\bar{G}$ for $0 \leqq |\beta| \leqq k$, then we say $f \in C^k(\bar{G})$.

Let $G \subset R^m$ be open and bounded. If there is a $K < \infty$ such that

$$|f(x) - f(y)| \leqq K|x - y|^\alpha$$

for some $\alpha \in (0,1)$ and for all x, y $\in \bar{G}$, then we say $f \in C_{0,\alpha}(G)$ and set

$$H_{\alpha,G}(f) = \sup_{\substack{x,y\in\bar{G}\\ x\neq y}} \frac{|f(x) - f(y)|}{|x - y|^\alpha}$$

Let $G \subset R^m$ be open and bounded and let $\alpha \in (0,1)$. If $f \in C^k(\bar{G})$ and $D^\beta f \in C_{0,\alpha}(G)$ for $|\beta| = k$, then we say $f \in C_{k,\alpha}(G)$ and set

$$\| f\|_{k,\alpha,G} = \sum_{j=0}^{k} \max_{|\beta|=j} \|D^\beta f\|_G + \max_{|\beta|=k} H_{\alpha,G}(D^\beta f)$$

$G \subset R^m$ is said to be *smooth* if, for each $P \in \partial G$, there is an $i = i_P \in \{1,\cdots,m\}$, an open set $H = H_P$ in R^{m-1} containing the point $\bar{P} = (p_1,\cdots, p_{i-1},p_{i+1},\cdots,p_m)$, and a function $g = g_P \in C_{2,\alpha}(H)$ for some $\alpha \in (0,1)$, such that when $x \in \partial G$ and $\bar{x} = (x_1,\cdots,x_{i-1},x_{i+1},\cdots,x_m) \in H$, ∂D can be expressed in the form $x_i = g(\bar{x})$.

Let $G \subset R^m$ be an open bounded set. Suppose $f \in C_{k,\alpha}(G_1)$ for some fixed $\alpha \in (0,1)$ and each open $G_1 \subset\subset G$. Set

$$\|f\|^*_{k,\alpha,G} = \|f\|_G + \sum_{j=1}^{k} \max_{|\beta|=j} \max_{x \in G} (d_x)^i \cdot |D^\beta f(x)|$$
$$+ \max_{|\beta|=k} \max_{\substack{x,y \in G \\ x \neq y}} (d_{x,y})^{m+\alpha} \cdot \frac{|D^\beta f(x) - D^\beta f(y)|}{|x - y|^\alpha} ,$$

where $d_x = \min_{y \in \partial G} |x - y|$, and $d_{x,y} = \min(d_x, d_y)$. Then, if $\|f\|^*_{k,\alpha,G}$ is finite, we say $f \in C_{k,\alpha}(G)$.

In the sequel, D will be an open bounded set in R^m, and $L: C^2(D) \to C(D)$ will be the uniformly elliptic differential operator given by

$$(1.5.1) \qquad \begin{cases} Lu = \sum_{i,j=1}^{m} a_{ij} u_{x_i x_j} + \sum_{i=1}^{m} b_i u_{x_i} \\ a_{ij} = a_{ji},\ b_i \in C_{0,\alpha}(D) \\ \sum_{i,j=1}^{m} a_{ij}(x) \xi_i \xi_j \geqq a_0 \sum_{i=1}^{m} \xi_i^2 \quad \text{for } x \in D,\ \xi \in R^m \\ a_0 > 0 \text{ and independent of } x \text{ and } \xi \end{cases}$$

We note that we are departing from the notation of the previous sections of this chapter where -L denoted the elliptic operator.

The following maximum principles are the analogs of Theorems 1.2.3 and 1.2.4.

1.5.1 Theorem: *Suppose $u \in C^2(D) \cap C(\bar{D})$ satisfies $Lu \geqq 0$ on D, then*

$$u(x) \leqq \sup_{\partial D} u \quad \text{for } x \in D$$

Proof: See [4, P. 326].

1.5.2 Corollary: *Suppose $u \in C^2(D) \cap C(\bar{D})$ satisfies*

$$\begin{cases} Lu - \gamma u \geqq 0 & \text{in } D \\ u \leqq 0 & \text{on } \partial D \end{cases} ,$$

where γ: $D \to R$ *is non-negative, but not necessarily continuous. Then* $u \leqq 0$ *in* D.

Proof: Let $D_1 = \{x \in D: u(x) > 0\}$. Suppose $D_1 \neq \emptyset$. Then D_1 is open and $u \leqq 0$ on ∂D_1. Now, in D_1, $Lu \geqq \gamma u \geqq 0$. Hence, by Theorem 1.5.1, $u \leqq 0$ in D_1. The contradiction shows $D_1 = \emptyset$ and proves the result.

$G \subset R^m$ is said to be *proper* if G is open and bounded, $G = \bigcup_{i=0}^{\infty} G_i$ where G_i is an open, bounded, smooth set and $G_i \subset G_{i+1}$ for each $i \geqq 0$, and if for each $y \in \partial G$, there exists a *strong barrier function*, i.e., a non-negative function, $w_y \in C^2(D) \cap C(\bar{D})$, which satisfies $w_y(x) = 0 \Leftrightarrow x = y$, and $Lw_y \leqq -1$ in D.

1.5.3 Lemma: *Let* $y \in \partial D$. *If there is a closed sphere* S_y *such that* $S_y \cap \bar{D} = \{y\}$, *then there is a strong barrier function for* y.

Proof: See [4, P. 341].

We note that by Lemma 1.5.3, a rectangular region is proper. In the sequel, we always assume D is proper.

1.5.4 Theorem: *Suppose* $f \in \hat{C}_{0,\alpha}(D)$ *and* $\phi \in C(\partial D)$. *Then there exists a unique solution,* $u \in \hat{C}_{2,\alpha}(D)$, *to*

$$(1.5.2) \qquad \begin{cases} Lu = f & \text{in } D \\ u = \phi & \text{on } \partial D \end{cases}$$

Proof: See [4, P. 340].

For $f \in \hat{C}_{0,\alpha}(D)$ and $\phi \in \hat{C}(\partial D)$, let $w_{f,\phi,D}$ be the solution to (1.5.2). We note that by Theorem 1.5.4, $w_{f,\phi,D}$ exists.

In the sequel, M_D will be defined by

$$M_D = \sup_{\substack{f \in \hat{C}_{0,\alpha}(D) \\ \|f\|_D = 1}} \|w_{f,0,D}\|_D \quad .$$

We see that for $f \in \hat{C}_{0,\alpha}(D)$,

$$\|w_{f,0,D}\|_D \leqq M_D \|f\|_D$$

Let $\hat{L}$: $\{u \in \hat{C}_{2,\alpha}(D): u|_{\partial D} = 0\} \to \hat{C}_{0,\alpha}(D)$ be defined by $\hat{L}u = Lu$. Then Theorem 1.5.4 says $\hat{L}$ is one-to-one and onto. Then $M_D = \|\hat{L}^{-1}\|$ when the domain and range of $\hat{L}$ are considered as subspaces of the Banach space, $C(\bar{D})$ with norm $\|\cdot\|_D$. The following analog of (1.2.38) assures that $M_D < \infty$.

<u>1.5.5 Lemma</u>: *$M_D = \sup_{\bar{D}} w_{-1,0,D}$.*

<u>Proof</u>: The proof follows from the maximum principle. Let $w = w_{-1,0,D}$ and $u = w_{f,0,D}$, where $f \in \hat{C}_{0,\alpha}(D)$ satisfies $\|f\|_D = 1$. It is sufficient to show that $-w \leqq u \leqq w$. But $L(u - w) = f + 1 \geqq 0 \Rightarrow u \leqq w$, and $L(u + w) = f - 1 \leqq 0 \Rightarrow u \geqq -w$. This completes the proof.

The following analog of Theorem 1.2.6 may be used to obtain an explicit bound on M_D.

<u>1.5.6 Corollary</u>: *Suppose $u \in C^2(D) \cap C(\bar{D})$ satisfies $\min_D Lu = b > 0$. Then*

$$M_D \leqq \frac{\max_{\bar{D}} u - \min_{\bar{D}} u}{\min_D Lu}$$

<u>Proof</u>: Let $w = w_{-1,0,D}$, and $u_1 = \frac{1}{b} u$. Then $L(u_1 + w) \geqq 0$. Hence,

$$\frac{1}{b} u + w = u_1 + w \leq \max_{\partial D} (u_1 + w) = \frac{1}{b} \max_{\bar{D}} u .$$

The result follows from this.

As in the discrete case, we can give explicit test functions to bound M_D. Let $w_\alpha(x) = e^{\alpha(x_1 - a)}$ where a is chosen so that $x_1 - a \geq 0$ for $x \in D$. Then

$$Lw_\alpha(x) = \alpha^2 a_{11}(x)\, e^{\alpha(x_1-a)} + \alpha b_1(x)\, e^{\alpha(x_1-a)}$$

But $a_{11}(x) \geq a_0 > 0$. Hence, for some α_0, $Lw_{\alpha_0} \geq 1$ in D. Thus,

$$M_D = \max_{\bar{D}} w_{\alpha_0} - \min_{\bar{D}} w_{\alpha_0}$$

If $b_1(x) \not\equiv 0$ on D, another suitable test function would be similar to the function, v_α, given in (1.2.26).

<u>1.5.7 Lemma</u>: *If $D_1 \subset D$ is proper, then $M_{D_1} \leq M_D$.*

<u>Proof</u>: Let $u = w_{-1,0,D}$ and $v = w_{-1,0,D_1}$. Then $u \geq 0$ in $\bar{D}$. Hence, $u - v \geq 0$ on ∂D_1. But $L(u - v) = 0$ in D_1. So, in D_1, by the maximum principle, $0 \leq v \leq u$. Thus

$$M_{D_1} = \max_{\bar{D}_1} v \leq \max_{\bar{D}} u = M_D$$

We now present the analog of Theorem 1.2.7.

<u>1.5.8 Lemma</u>: *Let $f \in \hat{C}_{0,\alpha}(D)$ and $\phi \in C(\partial D)$. Then*

$$\| w_{f,\alpha,D} \|_D \leq M_D \| f \|_D + \| \phi \|_{\partial D}$$

<u>Proof</u>: Let $u = w_{f,\alpha,D}$. Then $u = w_{f,0,D} + w_{0,\alpha,D}$. But

$$\| w_{f,0,D} \|_D \leq M_D \| f \|_D ,$$

and

$$\|w_{0,\phi,D}\|_D \leq \|\phi\|_{\partial D}$$

by the maximum principle. The result follows from the triangle inequality.

The following result is the analog of (1.2.14).

1.5.9 Corollary: *Let $L_1u = Lu - \gamma u$, where $\gamma : D \to R$ is non-negative, but not necessarily continuous. Suppose $v \in C^2(D) \cap C(\bar{D})$ satisfies*

$$\begin{cases} L_1 v = f & \text{in } D \\ v = 0 & \text{on } \partial D \end{cases},$$

where $f \in \hat{C}_{0,\alpha}(D)$, then

$$\|v\|_D \leq M_D \|f\|_D$$

Proof: Let $D_1 = \{x \in D : v(x) > 0\}$, and $D_2 = \{x \in D : v(x) < 0\}$. It is sufficient to show $\|v\|_{D_1}, \|v\|_{D_2} \leq M_D \|f\|_D$. Suppose $D_1 \neq \emptyset$. Then

D_1 is open and $v = 0$ on ∂D_1. Let $G_k \subset D_1$ be proper for $k \geq 1$ and satisfy $G_k \subset G_{k+1}$, $\bigcup_{k=0}^{\infty} G_k = D_1$, and $0 \leq v \leq \frac{1}{k}$ on ∂G_k. Let $u_k = w_{f,v,G_k}$. Then $v - u_k = 0$ on ∂G_k, and, in G_k $L(v - u_k) = L_1 v - Lu_k + \gamma v \geq 0$. Hence, in G_k, $0 \leq v \leq u_k$, So, by Lemmas 1.5.8 and 1.5.7,

$$\|v\|_{G_k} \leq M_{G_k} \|f\|_{G_k} + \|v\|_{\partial G_k} \leq M_D \|f\|_D + \frac{1}{k}$$

Letting $k \to \infty$, we get

$$\|v\|_{D_1} \leq M_D \|f\|_D$$

A similar result holds for D_2, and the proof is complete.

We may now prove the analog of (1.4.8).

1.5.10 Lemma: *Suppose $f : \bar{D} \times R \to R$ is continuous and satisfies $f(x,\cdot)$ is* *monotone for each* $x \in D$. *Suppose* $u \in C^2(D) \cap C(\bar{D})$ *satisfies*

$$(1.5.3) \qquad \begin{cases} Lu = f(\cdot, u) & \text{in } D \\ u = \phi & \text{on } \partial D \end{cases}$$

Then

$$\|u\|_D \leq \hat{M} \equiv M_D \max_{|t| \leq \|\phi\|_{\partial D}} \|f(\cdot, t)\|_D + \|\phi\|_{\partial D} .$$

Proof: Let $w = w_{0,\phi,D}$. Then $u - w = 0$ on ∂D, and

$$L(u - w) = f(\cdot,u) - f(\cdot,w) + f(\cdot,w)$$
$$= \gamma(u - w) + f(\cdot,w)$$

where

$$\gamma(x) = \begin{cases} \dfrac{f(x,u(x)) - f(x,w(x))}{u(x) - w(x)} & \text{if } u(x) \neq w(x) \\ 0 & \text{if } u(x) = w(x) \end{cases}$$

Now, since f is monotone in the second argument, $\gamma(x) \geq 0$. Hence, by Corollary 1.5.9,

$$\|u - w\|_D \leq M_D \|f(\cdot,w)\|_D$$

The result then follows from the maximum principle: $\|w\|_D \leq \|\phi\|_{\partial D}$.

1.5.11 Lemma: *Suppose f is as in Lemma 1.5.10, and that* $\phi \in C(\partial D)$. *Then there exists at most one solution,* $u \in C^2(D) \cap C(\bar{D})$, *to (1.5.3).*

Proof: Suppose u_1 and u_2 each satisfy (1.5.3). Then, as in Lemma 1.5.10,

$$L(u_1 - u_2) - \gamma(u_1 - u_2) = 0 \quad \text{in } D$$
$$u_1 - u_2 = 0 \quad \text{on } \partial D ,$$

for some $\gamma = \gamma(x) \geq 0$. Hence, by Lemma 1.5.7, $u_1 = u_2$.

1.5.12 Remark: Consider Lemma 1.5.10. Define $\hat{f}$ by

$$\hat{f}(x,t) = \begin{cases} f(x,-\hat{M}) & \text{if } t \leq -\hat{M} \\ f(x,t) & \text{if } -\hat{M} \leq t \leq \hat{M} \\ f(x,\hat{M}) & \text{if } \hat{M} \leq t \end{cases}$$

Then $\hat{f}$ is bounded and monotone in the second argument. Hence, by Lemma 1.5.11, $u \in C^2(D) \cap C(\bar{D})$ satisfies (1.5.3) if and only if it satisfies

$$\begin{cases} Lu = \hat{f}(\cdot,u) & \text{in } D \\ u = \phi & \text{on } \partial D. \end{cases}$$

In the sequel, we now assume (1.5.1) is satisfied where, in addition,

$$(1.5.4) \qquad \begin{cases} Lu = \sum_{i,j=1}^{m} (a_{ij}u_{x_i})_{x_j} \\ a_{ij} = a_{ji} \in C_{1,\alpha}(D) \end{cases}$$

The existence of solutions to (1.5.3) was considered by Courant [4, P. 369], Parter [22], and Levinson [17] when $L = \Delta$, the Laplacian. Courant proves existence under the assumption that $f = f_1 + f_2$ where f_1, f_2 are C^1 in their arguments, f_1 is bounded and $\partial f/\partial u \geq 0$. If $f_1 = 0$, the solution is, of course, unique. Parter and Levinson prove existence of a solution under assumptions (1.5.7a,b), below, and the assumption that

$$(1.5.5) \qquad \liminf_{|t|\to\infty} \frac{f(x,t)}{t} \geq 0 \quad \text{uniformly for } x \in \bar{D}$$

By use of (1.5.5), it is shown, as in Remark 1.5.12, that f can be replaced by a bounded function, $\hat{f}$. By assuming

$$(1.5.6) \qquad f(x,\cdot) \text{ is monotone for each } x \in \bar{D} \quad ,$$

we obtain this result more easily, and we also assure uniqueness of the solution.

We will now consider the existence of solutions to (1.5.3) under assumptions (1.5.6) and (1.5.7a,b), below.

(1.5.7a) There is a fixed $\alpha \in (0,1)$ such that given $c > 0$, there is a $K(c) < \infty$ such that

$$|f(x,t) - f(y,t)| \leqq K(c)|x - y|^\alpha$$

when $x, y \in \bar{D}$ and $|t| \leqq c$.

(1.5.7b) Given $c > 0$, there is a $K_0(c) < \infty$ such that

$$|f(x,t) - f(x,s)| \leqq K_0(c)|t - s|$$

when $x \in \bar{D}$ and $|t|, |s| \leqq c$.

<u>1.5.13 Theorem</u>: *Let $D \subset R^m$ be proper. Let $f: \bar{D}\times R \to R$ satisfy (1.5.6) and (1.5.7a,b), and suppose $\phi \in C(\partial D)$. Let L satisfy the special case of (1.5.1) given by (1.5.4). Then (1.5.3) has a unique solution, $u \in C^2(D) \cap C(\bar{D})$.*

<u>Proof</u>: By Lemma 1.5.11, we need only show existence. By Remark 1.5.12, we may assume $|f(x,t)| \leqq N < \infty$ for $x \in \bar{D}$ and $t \in R$. The proof, which follows along the lines of that in [4], is presented here in detail for completeness.

Let $u_0 = w_{-N,0,D} \in \hat{C}_{2,\alpha}(D)$ and $v_0 = w_{N,0,D} \in \hat{C}_{2,\alpha}(D)$. Then $v_0 - u_0 = 0$ on ∂D, and, in D,

$$L_1(v_0 - u_0) = 2N \geqq 0$$

Hence, by Corollary 1.5.2, $v_0 \leqq u_0$. Let

$$c = \max [\| u_0 \|_D, \| v_0 \|_D] \quad ,$$

and set

$$K = K(c) \ , \ k = K_0(c)$$

For $u \in C^2(D) \cap C(\bar{D})$, let $L_1 u = Lu - ku$, and define $\{u_j\}$ by

$$L_1u_{j+1} - ku_{j+1} = f(\cdot,u_j) - ku_j \quad \text{in } D$$

$$u_{j+1} = \phi \qquad\qquad \text{on } \partial D$$

By Theorem 1.5.4 and an inductive argument, u_j exists and is in $\hat{C}_{2,\alpha}(D)$ for each $j \geqq 0$. Now $u_1 - u_0 = 0$ on ∂D, and, in D,

$$L_1(u_1 - u_0) = f(\cdot,u_0) - ku_0 + \mathbb{N} + ku_0 \geqq 0$$

Thus, $u_1 \leqq u_0$. Likewise, $u_1 - v_0 = 0$ on ∂D, and, in D,

$$L_1(u_1 - v_0) = f(\cdot,u_0) - ku_0 - \mathbb{N} + kv_0$$
$$\leqq k(v_0 - u_0)$$
$$\leqq 0$$

Thus, $u_1 \geqq v_0$. Suppose

(1.5.8) $$v_0 \leqq u_j \leqq u_{j-1} \leqq u_0 \quad ,$$

for some $j \geqq 1$. Then $u_{j+1} - u_j = 0$ on ∂D, and, in D,

$$L_1(u_{j+1} - u_j) = f(\cdot,u_j) - ku_j - f(\cdot,u_{j-1}) + ku_{j-1}$$
$$\geqq -k|u_j - u_{j-1}| + k(u_{j-1} - u_j)$$
$$= 0$$

Thus, $u_{j+1} \leqq u_j$. Furthermore, $u_{j+1} - v_0 = 0$ on ∂D, and, in D,

$$L_1(u_{j+1} - v_0) = f(\cdot,u_j) - ku_j - \mathbb{N} + kv_0$$
$$\leqq k(v_0 - u_j)$$
$$\leqq 0$$

Thus, $u_{j+1} \geqq v_0$, and (1.5.8) is established by induction. Hence,

$$u_j \downarrow u^* \quad ,$$

for some $u^*: \bar{D} \to R$ satisfying $v_0 \leqq u^* \leqq u_0$.

Let $D_1 \subset\subset D$ be open. Then there exists an open set D_2 such that $D_1 \subset\subset D_2 \subset\subset D$. Now, $\|u_j\|_{D_2} \leqq c$, and

$$\|Lu_j\|_{D_2} \leqq \|f(\cdot,u_{j-1}) - ku_{j-1} + ku_j\|_D$$
$$\leqq N + 2kc$$
$$\equiv M_1 < \infty \quad ,$$

Now, by manipulating formula (3.3) of [3], we see that there is an $M_2 < \infty$, depending only on L, D_2, D_1 and c, such that

$$\|\nabla u_j\|_{D_1} \leqq M_2 \quad \text{for } j \geqq 0$$

Now, $u_j \in \hat{C}_{2,\alpha}(D) \Rightarrow u_j \in C_{2,\alpha}(D_1)$. So, by the interior Schauder estimates, (see [4, P. 332] or [16, P. 110],) there is an $M_3 < \infty$ depending only on L_1, D, D_1, and α, such that, for $j \geqq 0$,

$$\|u_j\|_{2,\alpha,D_1} \leqq M_3 \left[\|Lu_j\|_{0,\alpha,D_1} + \|u_j\|_{D_1}\right]$$

Now, for x, y $\in D_1$,

$$|f(x,u_j(x)) - f(y,u_j(y))| \leqq |f(x,u_j(x)) - f(y,u_j(x))|$$
$$+ |f(y,u_j(x)) - f(y,u_j(y))|$$
$$\leqq K|x - y|^\alpha + k|u_j(x) - u_j(y)|$$
$$\leqq K|x - y|^\alpha + k\|\nabla u_j\|_{D_1} |x - y|^{1-\alpha}|x - y|^\alpha$$
$$\leqq (K + kd^{1-\alpha}M_2)\ |x - y|^\alpha$$
$$\equiv M_4\ |x - y|^\alpha \quad ,$$

where d is the diameter of D. Hence, for $j \geqq 0$,

$$\|f(\cdot,u_j)\|_{0,\alpha,D_1} \leqq N + M_4$$

Now, for x, y $\in D_1$,

$$|u_j(x) - u_j(y)| \leqq \|\nabla u_j\|_{D_1} |x - y|^{1-\alpha}|x - y|^\alpha$$
$$\leqq M_2 d^{1-\alpha}\ |x - y|^\alpha$$

Thus, for $j \geqq 0$,

$$\|u_j\|_{0,\alpha,D_1} \leq c + M_2 d^{1-\alpha}$$

$$\equiv M_5 < \infty$$

So, finally, for $j \geq 0$,

$$\begin{aligned} \|u_j\|_{2,\alpha,D_1} &\leq M_3[\|f(\cdot,u_j)\|_{0,\alpha,D_1} + \|ku_j\|_{0,\alpha,D_1} + c] \\ &\leq M_3[N + M_4 + kM_5 + c] \\ &< \infty \end{aligned}$$

Thus, $\{u_j\}$ and the sequence of 1st and 2nd derivatives are uniformly bounded on D_1, and the sequence of 2nd derivatives is equi-continuous on D_1. Hence, by the Arzela-Ascoli Theorem, there is a subsequence $\{u_{m_j}\}$ of $\{u_j\}$ which converges, necessarily to u^*, in the norm of $C^2(\bar{D}_1)$. Hence, $u^* \in C^2(D_1)$. Now, $u_j \to u^*$ pointwise, and f is continuous. Thus, for $x \in D_1$,

$$\begin{aligned} Lu^*(x) &= \lim_{j\to\infty} Lu_{m_j}(x) \\ &= \lim_{j\to\infty} [f(x,u_{m_j-1}(x)) - ku_{m_j-1}(x) + ku_{m_j}(x)] \\ &= \lim_{j} f(x,u_{m_j-1}(x)) \\ &= f(x,u^*(x)) \end{aligned}$$

Since D_1 is arbitrary, $u^* \in C^2(D)$, and, in D,

$$Lu^* = f(\cdot,u^*)$$

We need only show $u^* \in C(\bar{D})$. It is sufficient to show

$$\lim_{\substack{x\in D \\ x\to y}} [u^*(x) - \phi(y)] = 0$$

when $y \in \partial D$. So let $x \in D$ and $y \in \partial D$. Then, since $u_0 \in C(\bar{D})$,

$$u^*(x) - \phi(y) \leq u_0(x) - \phi(y) \to 0 \ (x \to y)$$

Likewise, since $v_0 \in C(\bar{D})$,

$$u^*(x) - \phi(y) \geqq v_0(x) - \phi(y) \to 0 \ (x \to y)$$

This completes the proof.

A more general form of this problem is considered in [16]. See especially Chapter 4, Section 8 and Chapter 5, Section 6. Furthermore, Theorem 3.1 on page 266 gives an interior bound for ∇u for a much more general L than that given in (1.5.1)/(1.5.4). In particular, by the use of this result, Theorem 1.5.13 can be proved for the non-self-adjoint L of (1.5.1), provided $a_{ij} \in C_{1,\alpha}(D)$.

CHAPTER II

LINEAR ADI METHODS

<u>2.1 Introduction</u>. Consider A_ℓ where ℓ is given by (1.2.19). A_ℓ has a natural splitting, $A_\ell = A_{\ell_H} + A_{\ell_V}$, into "horizontal" and "vertical" parts, where ℓ_H and ℓ_V are given in (1.2.41). We note that A_{ℓ_H} and A_{ℓ_V} are both positive definite.

Generalizing, suppose

$$(2.1.1) \qquad \begin{cases} C = H_1 + V_1 \\ H_1,\ V_1 \in L(R^n, R^n) \text{ are positive semi-definite} \\ \text{One of } H_1 \text{ or } V_1 \text{ is positive definite} \end{cases}$$

Let $\xi \in R^n$ and suppose we wish to find $x^* = C^{-1}\xi$. By (2.1.1), the following iteration is well-defined for $r_k > 0$.

$$(2.1.2) \qquad \begin{cases} x_0 \in R^n \\ [r_k + H_1]\, x_{k+\frac{1}{2}} = [r_k - V_1]\, x_k + \xi \\ [r_k + V_1]\, x_{k+1} = [r_k - H_1]\, x_{k+\frac{1}{2}} + \xi \end{cases}$$

This procedure was first considered by Peaceman and Rachford [23] to approximate the solution of a discretized version of the Dirichlet problem for Laplace's equation on a square. The name "alternating direction implicit", or ADI was given to (2.1.2) because it entails alternately solving along horizontal and vertical mesh lines. In this particular case, $H_1V_1 = V_1H_1$, and, after a suitable permutation, H_1 and V_1 are both tridiagonal matrices, which are relatively easy to invert (see [28, P. 195] or [23].) Thus (2.1.2) is feasible.

Let $\lambda_1, \cdots, \lambda_{n_{H_1}}$ be the distinct eigenvalues of H_1, and let $r_k = \lambda_{k+1}$

for $0 \leqq k \leqq n_{H_1}$. Then, if $H_1V_1 = V_1H_1$, (2.1.2) is a direct method and converges to x^* after n_{H_1} iterations. This is also true if the r_k are successively the n_{V_1} distinct distinct eigenvalues of V_1 (see [28, p. 222] or [23].)

In practical cases, we may not be able to determine the eigenvalues of H_1 or V_1, but we may know that they lie in an interval $[a,b]$ where $a > 0$. If we then apply $\nu \geqq 1$ parameters cyclically, we would try to determine the ν parameters which are in some sense optimal.

Let $e_k = x_{\nu k} - x^*$. Then

$$e_k = \left(\prod_{j=0}^{\nu-1} T_{r_j} \right) e_{k-1}$$

where

$$T_r = [r + V_1]^{-1} [r - H_1] [r + H_1]^{-1} [r - V_1] .$$

Now, when $H_1V_1 = V_1H_1$,

$$T_r = [r - H_1] [r + H_1]^{-1} [r - V_1] [r + V_1]^{-1}$$

Hence,

$$\left\| \prod_{j=0}^{\nu-1} T_{r_j} \right\| \leqq \prod_{j=0}^{\nu-1} \left\|[r_j - H_1] [r_j + H_1]^{-1}\right\| \left\|[r_j - V_1] [r_j + V_1]^{-1}\right\|$$

$$\leqq \prod_{j=0}^{\nu-1} \sup_{\lambda\epsilon\sigma[H]} \left|\frac{r_j - \lambda}{r_j + \lambda}\right| \sup_{\mu\epsilon\sigma[V]} \left|\frac{r_j - \mu}{r_j + \mu}\right| ,$$

where here, and in the rest of the chapter, $\| \cdot \| = \| \cdot \|_2$. Thus

$$\left\| \prod_{j=0}^{\nu-1} T_{r_j} \right\| \leqq \sup_{x\epsilon[a,b]} \prod_{j=0}^{\nu-1} \left|\frac{r_j - x}{r_j + x}\right|^2 < 1 .$$

So, for any positive values of $r_0, \cdots, r_{\nu-1}$, convergence is assured. To enhance convergence, we are led to the problem of minimizing the quantity

$$\sup_{x\epsilon[a,b]} \prod_{j=0}^{\nu-1} \left|\frac{r_j - x}{r_j + x}\right|$$

for $r_0,\cdots,r_{\nu-1} > 0$. The optimal r_j, which lie in the interval $[a,b]$, were given by Wachspress in the case $\nu = 2^k$ (see [28, P. 224] or [30, P. 196]) and by Jordan for any ν (see [30, P. 185].)

If $\nu = 1$, the optimal parameter is $\sqrt{ab}$, and the asymptotic rate of convergence with the optimal ADI parameter is approximately the same as that for SOR with optimal SOR parameter, although the work required for each ADI sweep is approximately twice that required for each SOR sweep. If $\nu > 1$, an asymptotic rate of convergence significantly better than that for SOR can be obtained (see [28, P. 229].)

When $H_1V_1 \neq V_1H_1$, the above analysis fails. If $r_k \equiv r > 0$ is constant, convergence can still be assured. However, convergence cannot be guaranteed for arbitrary positive values of $r_0,\cdots,r_{\nu-1}$. Nevertheless, if "good" parameters for the commutative case are used, rapid convergence is often still obtained. Numerical results indicate that the best parameters are in the interval $[a,b]$ and, indeed, in the lower part of the interval.

In Section 2.2, we present the main convergence results in the noncommutative case. These are pertinent since, in problems with a nonlinear term, the commutative analysis fails. In Section 2.3, a specific ADI iteration for the discretized version of an elliptic boundary value problem is introduced, and in Section 2.4, local convergence results are given for nonlinear versions of this iteration.

We now collect some formulas and inequalities which will be useful later. Suppose $L \in L(R^n,R^n)$ is positive semi-definite and $\sigma[L] \subset [c,d]$ where $c \geqq 0$. Then

$$(2.1.3)\quad \begin{cases} \text{a) } r \geqq 0 \Rightarrow \|[r + L]^{-1}\| \leqq \dfrac{1}{r + c} \\ \text{b) } s \geqq r \geqq \sqrt{cd} \Rightarrow \|[r - L]\,[s + L]^{-1}\| \leqq \dfrac{r - c}{r + c} \\ \text{c) } r \leqq \sqrt{cd} \Rightarrow \|[r - L]\,[r + L]^{-1}\| \leqq \dfrac{d - r}{d + r} \\ \text{d) } r \leqq \dfrac{d + c}{2},\ r \leqq s \Rightarrow \|s - L\| \leqq s - 2r + d \end{cases}$$

We demonstrate (2.1.3d). By assumption, $2r - d \leqq c$. Hence, since $s \geqq \dfrac{(2r - d) + d}{2}$,

$$\|s - L\| \leqq \sup_{c \leqq z \leqq d} |s - z| \leqq \sup_{2r - d \leqq z \leqq d} |s - z| = s - 2r + d$$

Let C, H_1, V_1 satisfy (2.1.1). For $r > 0$, define

$$(2.1.4)\quad \begin{cases} T_r = [r + V_1]^{-1}\,[r - H_1]\,[r + H_1]^{-1}\,[r - V_1] \\ Q_r = 2r\,[r + V_1]^{-1}\,[r + H_1]^{-1} \end{cases}$$

Then, if $\{x_{k/2}\}$ satisfies (2.1.2), the ν^{th} iterate is

$$(2.1.5)\qquad x_\nu = \prod_{i=0}^{\nu-1} T_{r_i} x_0 + \sum_{j=0}^{\nu-1} \left(\prod_{i=j+1}^{\nu-1} T_{r_i} \right) Q_{r_j} \xi ,$$

which, when $r_k \equiv r$ is constant, becomes

$$(2.1.6)\qquad x_\nu = (T_r)^\nu x_0 + \sum_{j=0}^{\nu-1} (T_r)^j Q_r \xi$$

A little algebra and an inductive argument shows

$$(2.1.7)\qquad \prod_{i=0}^{\nu-1} T_{r_i} + \sum_{j=0}^{\nu-1} \left(\prod_{i=j+1}^{\nu-1} T_{r_i} \right) Q_{r_j} C = I$$

Writing $\xi = Cx_0 - [Cx_0 - \xi]$, we see, from (2.1.5) and (2.1.7),

$$(2.1.8)\qquad x_\nu = x_0 - \sum_{j=0}^{\nu-1} \left(\prod_{i=j+1}^{\nu-1} T_{r_i} \right) Q_{r_j} [Cx_0 - \xi]$$

Finally, if $Cx^* = \xi$, then

$$(2.1.9) \qquad x_\nu - x^* = \left(\prod_{i=0}^{\nu-1} T_{r_i} \right) (x_0 - x^*)$$

<u>2.2 Results in the Non-Commutative Case.</u> The following theorem is the basic convergence theorem for ADI in the linear non-commutative case.

<u>2.2.1 Theorem</u>: Let C, H_1, and V_1 satisfy (2.1.1) and $\{x_{k/2}\}$ satisfy (2.1.2) where $r_k \equiv r > 0$ is constant. Then $x_k \to x^* = C^{-1}\xi$.

<u>Proof</u>: Let T_r be defined by (2.1.4). Then, by (2.1.3b,c),

$$\|[r + V_1] T_r [r + V_1]^{-1}\| \leqq \|[r - H_1][r + H_1]^{-1}\| \, \|[r - V_1][r + V_1]^{-1}\| < 1.$$

Hence, $\rho(T_r) < 1$. The result follows, then, from (2.1.9).

In the next two theorems, convergence of (2.1.2) is guaranteed for variable r_k provided the r_k are large enough.

<u>2.2.2 Theorem</u>: Let $H_1, V_1 \in L(R^n, R^n)$ satisfy

$$(2.2.1) \qquad \begin{cases} H_1, V_1 \text{ are positive semi-definite} \\ \sigma[H_1] \subset [a_1, b_1], \quad \sigma[V_1] \subset [\alpha_1, \beta_1] \\ a_1, \alpha_1 \geqq 0, \quad a_1 + \alpha_1 > 0 \\ [H_1 + V_1] x^* = \xi \text{ for some } x^*, \ \xi \in R^n \end{cases}$$

Suppose $r > 0$ satisfies

$$(2.2.2) \qquad \begin{cases} \text{a)} \ r \geqq \dfrac{\beta_1 - \alpha_1}{2} & \text{if } a_1 > 0 \\ \text{b)} \ \dfrac{\beta_1 - \alpha_1}{2} < r \leqq \dfrac{\beta_1 + \alpha_1}{2} & \text{if } a_1 = 0 \end{cases},$$

and $\{r_k\}$ satisfies

$$(2.2.3) \qquad r \leqq r_k \leqq s < \infty \quad \text{for } k \geqq 0$$

Let $\{x_{k/2}\}$ satisfy (2.1.2). Then $x_k \to x^*$.

<u>Proof</u>: Let T_{r_k} be defined by (2.1.4). Suppose (2.2.2a) holds. Then by

(2.1.3b,c) and (2.2.3), there is a $\delta < 1$ and independent of k such that

$$\|[r_k - H_1][r_k + H_1]^{-1}\| \leqq \delta$$

So, by (2.1.3a,d),

$$\|T_{r_k}\| \leqq \|[r_k + V_1]^{-1}\| \cdot \delta \cdot \|r_k - V_1\| \leqq \frac{1}{r_k + \alpha_1} \cdot \delta \cdot (r_k - (\beta_1 - \alpha_1) + \beta_1) = \delta.$$

Suppose (2.2.2b) holds. Then by (2.1.3b,c),

$$\|[r_k - H_1][r_k + H_1]^{-1}\| \leqq 1$$

So, by (2.1.3a,d), as above,

$$\|T_{r_k}\| \leqq \frac{r_k - 2r + \beta_1}{r_k + \alpha_1} \leqq \frac{s - 2r + \beta_1}{s + \alpha_1} < 1$$

So, in either case, $\|T_{r_k}\|$ is bounded uniformly below 1. By (2.1.9), then, $x_k \to x^*$.

<u>2.2.3 Theorem</u>: Let H_1, $V_1 \in L(R^n, R^n)$ be M-matrices. Let ξ, x_0, $x^* \in R^n$ satisfy

$$[H_1 + V_1]\, x^* = \xi,\ [H_1 + V_1]\, x_0 \geqq \xi,\ x^* \leqq x_0 \quad ,$$

where $[H_1 + V_1]$ is non-singular. Set

$$K = \max_{1 \leqq i \leqq n} \max\, (h_{ii}, v_{ii})$$

where $H_1 = (h_{ij})$ and $V_1 = (v_{ij})$. Let $\{x_{k/2}\}$ be defined by (2.1.2) where

$$K \leqq r_k \leqq s < \infty \quad .$$

Then $x_{k/2} \downarrow x^*$.

<u>Proof</u>: See Theorem 5.2.4 of which this is a special case.

We note that if $[H_1 + V_1]$ is itself an M-matrix, then

$$[H_1 + V_1]\, x_0 \geqq \xi \Rightarrow x_0 \geqq x^* \quad .$$

We note also that Theorem 2.2.3 does not assume any symmetry conditions. Thus, it would apply, for example, to some discrete version of the boundary value problem on $D = \{(s,t) : 0 < s,t < 1\}$:

$$(2.2.4)\quad \begin{cases} Lu \equiv au_{ss} + bu_{tt} + cu_s + du_t + eu = f \quad ; \quad (s,t) \in D \\ \qquad\qquad\qquad\qquad\qquad u = v \quad ; \quad (s,t) \in \partial D \\ a,\ b,\ c,\ d,\ e,\ f \in C(\bar{D});\ a \geqq a_0 > 0,\ b \geqq b_0 > 0,\ e \leqq 0 \end{cases}$$

Usually we would like to take the r_k smaller than allowed by Theorems 2.2.2 and 2.2.3. The following theorem and remarks allow us to pick the r_k as small as we wish, but they impose other conditions.

<u>2.2.4 Theorem (Pearcy)</u>: Let $H_1,\ V_1 \in L(R^n,R^n)$ satisfy (2.2.1). Suppose

$$(2.2.5a)\qquad \max(a_1,\alpha_1) \leqq r_{\nu-1} \leqq r_{\nu-2} \leqq \cdots \leqq r_0 \leqq \min(b_1,\beta_1)$$

and

$$(2.2.5b)\qquad r_{i\nu+j} = r_j \quad \text{for } 0 \leqq i < \infty \text{ and } 0 \leqq j \leqq \nu-1 \qquad ,$$

where

$$(2.2.5c)\qquad \nu > \frac{\log \dfrac{2\alpha_1}{\beta_1 + \alpha_1}}{\log \dfrac{(b_1 - a_1)(\beta_1 - \alpha_1)}{(b_1 + a_1)(\beta_1 + \alpha_1)}}$$

Let $\{x_{k/2}\}$ satisfy (2.1.2). Then $x_k \to x^*$.

<u>Proof</u>: See [24] or [30, Thm. 6.8].

<u>2.2.5 Remark</u>: In Theorem 2.2.4, (2.2.5a,c) can be replaced with

$$(2.2.6a)\qquad 0 < r_{\nu-1} \leqq r_{\nu-2} \leqq \cdots \leqq r_0$$

and

$$(2.2.6c)\quad \frac{\max(r_0-\alpha_1,\beta_1-r_0)}{r_{\nu-1}+\alpha_1} \prod_{j=0}^{\nu-1} \left[\max\left(\frac{r_j-\alpha_1}{r_j+\alpha_1},\frac{\beta_1-r_j}{\beta_1+r_j}\right)\cdot\max\left(\frac{r_j-a_1}{r_j+a_1},\frac{b_1-r_j}{b_1+r_j}\right)\right] < 1.$$

<u>2.2.6 Remark</u>: In Theorem 2.2.4, (2.2.5) can be replaced by

$$(2.2.7) \qquad 0 < r \leqq r_{k+1} \leqq r_k \leqq r_0 \quad \text{for } k \geqq 0$$

2.3 An Application. Let D be a bounded region in R^2 and consider the problem

$$(2.3.1) \qquad \begin{cases} -(pu_s)_s - (qu_t)_t + \sigma u = -f \quad ; \quad (s,t) \in D \\ \xi u + \eta \frac{\partial u}{\partial n} = \gamma \quad ; \quad (s,t) \in \partial D \\ p,\ q,\ \sigma \in C(\bar{D});\ p > 0,\ q > 0,\ \sigma \geqq 0 \\ \xi,\ \eta \in C(\partial D);\ \xi \geqq 0,\ \eta \geqq 0,\ \xi + \eta > 0 \end{cases} .$$

If a rectangular, but not necessarily uniform, mesh is imposed on D, we can derive a difference approximation to (2.3.1) which results in the matrix problem

$$(2.3.2) \qquad [H + V + \Sigma]\, x = \xi \quad ,$$

where H, V, $\Sigma \in L(R^n,R^n)$ for some n, Σ is non-negative diagonal, and H and V are, after a suitable permutation, direct sums of tridiagonal Stieltjes matrices (see [28, Section 6.3].)

If, for some $c \in R$, we set

$$H_1 = H + c\Sigma,\ V_1 = V + (1-c)\Sigma \quad ,$$

we obtain from (2.1.2) the following iteration considered in the case $c = \frac{1}{2}$ by Varga [28].

$$(2.3.3) \qquad \begin{cases} x_0 \in R^n . \\ [r_k + H + c\Sigma]\, x_{k+\frac{1}{2}} = [r_k - V - (1-c)\Sigma]\, x_k + \xi \\ [r_k + V + (1-c)\Sigma]\, x_{k+1} = [r_k - H - c\Sigma]\, x_{k+\frac{1}{2}} + \xi \end{cases} .$$

More generally, suppose H, V, $\Sigma \in L(R^n,R^n)$ are symmetric with eigenvalues in the ranges $[a,b]$, $[\alpha,\beta]$, $[s,t]$ respectively. For M a symmetric matrix, let $\mu(M)$ be the least eigenvalue of M. Then using the fact that for symmetric M and N, $\mu(M) + \mu(N) \leqq \mu(M + N)$, it can be shown that

$H_1 = H + c\Sigma$ and $V_1 = V + (1-c)\Sigma$ satisfy (2.1.1) when one of the conditions of Table 2.3.1 is satisfied.

2.3.1 Table:

a	α	s	t	c	additional conditions
+	+	+,0		$[-a/t, 1+\alpha/t]$	
0	+	+		$[0, 1+\alpha/t]$	
0	+	0		$[0, 1+\alpha/t)$	
+	0	+		$[-a/t, 1]$	
+	0	0		$(-a/t, 1]$	
0	0	+		$[0,1]$	
+,0	+,0	$(-(a+\alpha), 0)$		$[0,1] \cap [1+\alpha/s, -a/s]$	
+	+	$[-\alpha, 0)$	0,-	$[1+\alpha/s, 0)$	
"	"	"	+	$(1+\alpha/s, 0)$	$-a/t = 1+\alpha/s$
"	"	"	"	$[1+\alpha/s, 0) \cap [-a/t, 0)$	$-a/t \neq 1+\alpha/s$
"	"	$[-a, 0)$	0,-	$(1, -a/s]$	
"	"	"	+	$(1, -a/s)$	$-a/s = 1+\alpha/t$
"	"	"	"	$(1, -a/s] \cap (1, 1+\alpha/t]$	$-a/s \neq 1+\alpha/t$
+	0	$(-a, 0)$	0	$[1, -a/s)$	
"	"	"	-	$[1, -a/s]$	
0	+	$(-\alpha, 0)$	0	$(1+\alpha/s, 0]$	
"	"	"	-	$[1+\alpha/s, 0]$	

The choice of $c = \frac{1}{2}$ in (2.3.3) is a reasonable one, but it may not be optimal. Suppose that $\Sigma = \lambda I$ for some $\lambda > 0$, so that $s = t = \lambda$, and suppose that one condition from Table 2.3.1 is satisfied. Let T_r be defined by (2.1.4). When we try to minimize $\rho(T_r)$ for $r > 0$, we are led to the min max problem,

$$\min_{\substack{r>0 \\ -\frac{a}{\lambda}\leq c\leq 1+\frac{\alpha}{\lambda}}} \max_{a\leq x\leq b} \left|\frac{r - x - c\lambda}{r + x + c\lambda}\right| \max_{\alpha\leq x\leq \beta} \left|\frac{r - x - (1-c)\lambda}{r + x + (1-c)\lambda}\right|$$

The optimal parameters, r_0 and c_0, are given by

$$r_0 = \sqrt{(a + c_0\lambda)(b + c_0\lambda)} = \sqrt{(\alpha + (1-c_0)\lambda)(\beta + (1-c_0)\lambda)}$$

$$c_0 = \frac{(\alpha + \lambda)(\beta + \lambda) - ab}{\lambda(a + b + \alpha + \beta + 2\lambda)}$$

We note that c_0 may not satisfy $0 \leq c_0 \leq 1$, but it must satisfy

$$-\frac{a}{\lambda} < c_0 < 1 + \frac{\alpha}{\lambda}$$

We note also that the choice, $c = \frac{1}{2}$, is optimal when $(a,b) = (\alpha,\beta)$.

2.4 Local Convergence of Some ADI Iterations in the Nonlinear Case.

Let $F, H_1, V_1: R^n \to R^n$ satisfy $F = H_1 + V_1$. We are interested in finding a solution to the equation

(2.4.1) $$F(x) = 0 \quad .$$

If $H_1, V_1 \in C^1(R^n)$ and $[F'(x)]^{-1}$ exists for each $x \in R^n$, then $\{x_k\}$ is well defined by the Newton iteration,

(2.4.2) $$F'(x_k)\, x_{k+1} = F'(x_k)\, x_k - F(x_k) \quad ,$$

and if $\lim x_k = x$ exists, then $F(x) = 0$.

We may try to solve (2.4.2) for x_{k+1} by performing one or more ADI sweeps of the form (2.1.2). If we apply the same ν parameters at each Newton stage, we have, formally, the N-ν step ADI iteration:

$$(2.4.3)\begin{cases} x_0 \in R^n \\ [r_j + H_1'(x_k)]\, x_k^{j+\frac{1}{2}} = [r_j - V_1'(x_k)]\, x_k^j + F'(x_k)\, x_k - F(x_k) \\ [r_j + V_1'(x_k)]\, x_k^{j+1} = [r_j - H_1'(x_k)]\, x_k^{j+\frac{1}{2}} + F'(x_k)\, x_k - F(x_k) \\ x_k^0 = x_k\ ,\ x_{k+1} = x_k^{\nu} \end{cases}$$

Alternately, we may try to solve (2.4.1) by applying a nonlinear version of (2.1.2) directly. If we apply ν parameters cyclically, we have, formally, the <u>ν step ADI iteration</u>:

$$(2.4.4)\begin{cases} x_0 \in R^n \\ r_j x_k^{j+\frac{1}{2}} + H_1(x_k^{j+\frac{1}{2}}) = r_j x_k^j - V_1(x_k^j) \\ r_j x_k^{j+1} + V_1(x_k^{j+1}) = r_j x_k^{j+\frac{1}{2}} - H_1(x_k^{j+\frac{1}{2}}) \\ x_k^0 = x_k\ ,\ x_{k+1} = x_k^{\nu} \end{cases}$$

Each of the equations in (2.4.4) is nonlinear, and hence, we may try to approximate the iterates by taking one Newton step during each half sweep. We have, formally, the <u>ν step ADI-N iteration</u>:

$$(2.4.5)\begin{cases} x_0 \in R^n \\ [r_j + H_1'(x_k^j)]\, x_k^{j+\frac{1}{2}} = [r_j + H_1'(x_k^j)]\, x_k^j - F(x_k^j) \\ [r_j + V_1'(x_k^{j+\frac{1}{2}})]\, x_k^{j+1} = [r_j + V_1'(x_k^{j+\frac{1}{2}})]\, x_k^{j+\frac{1}{2}} - F(x_k^{j+\frac{1}{2}}) \\ x_k^0 = x_k\ ,\ x_{k+1} = x_k^{\nu} \end{cases}$$

In order to guarantee that (2.4.3)--(2.4.5) are well defined, we assume, analogously to (2.1.1), that $H_1'(x)$ and $V_1'(x)$ are positive semi-definite for $x \in R^n$ and that one of $H_1'(x)$ or $V_1'(x)$ is uniformly positive definite on R^n. Then, if $r_j > 0$ for $0 \leqq j \leq \nu - 1$, (2.4.3)--(2.4.5) are well defined. This is immediate for (2.4.3) and (2.4.5) and follows for (2.4.4) from the fact that a differentiable function $G\colon R^n \to R^n$, which satisfies $(G'(x)\xi,\xi) \geq c\|\xi\|^2$ for all $x,\xi \in R^n$ and some $c > 0$, is a homeomorphism. Indeed, under these assumptions, F itself is a homeomorphism, and so (2.4.1) has a unique solution, x^*. In the remainder of this section, x^* will be the root of F.

Let $\{y_k\} \subset R^n$ satisfy $y_{k+1} = h(y_k)$ for $k \geq 0$, where $h \in C^1(R^n)$. Suppose h has a fixed point, y^*, and that $\rho(h'(y^*)) < 1$. Then, there is a norm $\|\cdot\|^*$, an $\epsilon > 0$, and a $\delta < 1$, such that $\|h'(y)\|^* \leq \delta$ when $\|y - y^*\|^* \leq \epsilon$. Hence, if $\|y_k - y^*\|^* \leq \epsilon$, then $\|y_{k+1} - y^*\|^* = \|h(y_k) - h(y^*)\| \leq \max_{y\epsilon[y_k,y^*]} \|h'(y)\|^* \, \|y_k - y^*\|^* \leq \delta\|y_k - y^*\|^*$. Thus, the iteration is locally convergent to y^*, and the quantity $\rho(h'(y^*))$ gives some measure of the rate of convergence. Suppose $\{y_k\}$, $\{z_k\} \subset R^n$ satisfy $y_{k+1} = h(y_k)$ and $z_{k+1} = g(z_k)$ for $k \geq 0$ where $h, g \in C^1(R^n)$ have a common fixed point, y^*. For the purposes of this paper, we will say that these two iterations have the same <u>aysmptotic rate of convergence to y^*</u> if $\rho(h'(y^*)) = \rho(g'(y^*))$. For a more precise discussion of this idea, see [21].

We now consider the relative asymptotic rates of convergence of (2.4.3)--(2.4.5).

Define $T_r(x)$ and $Q_r(x)$ by (2.1.4) where H_1 and V_1 are replaced by $H_1'(x)$ and $V_1'(x)$ respectively. By (2.1.8), we see that (2.4.3) is given by

$$x_{k+1} = h_1(x_k) \quad , \quad k \geq 0 \qquad ,$$

where

$$h_1(x) = x - \sum_{j=0}^{\nu-1} \left(\prod_{i=j+1}^{\nu-1} T_{r_i}(x)\right) Q_{r_j}(x)\, F(x)$$

Now, if $H_1, V_1 \in C^2(R^n)$, the $h_1 \in C^1(R^n)$, and, for $\xi \in R^n$,

$$h_1'(\xi) = [I - \sum_{j=0}^{\nu-1} \left(\prod_{i=j+1}^{\nu-1} T_{r_i}(x)\right) Q_{r_j}(x)\, F'(x)]\, \xi$$

$$- \frac{d}{dx}[\sum_{j=0}^{\nu-1} \left(\prod_{i=j+1}^{\nu-1} T_{r_i}(x)\right) Q_{r_j}(x)]\, \xi\, F'(x)$$

But $F(x^*) = 0$. So, using (2.1.7),

$$h_1'(x^*) = I - \sum_{j=0}^{\nu-1} \left(\prod_{i=j+1}^{\nu-1} T_{r_i}(x^*) \right) Q_{r_j}(x^*) F'(x^*)$$
$$= \prod_{i=0}^{\nu-1} T_{r_i}(x^*)$$

Now consider (2.4.4). Let $g_{i/2}: R^n \to R^n$, $0 \leqq i \leqq 2\nu - 1$, be defined by

$$g_j(x) = (r_j + H_1)^{-1} (r_j x - V_1(x)) \quad , \quad 0 \leqq j \leqq \nu - 1$$
$$g_{j+\frac{1}{2}}(x) = (r_j + V_1)^{-1} (r_j x - H_1(x)) \quad , \quad 0 \leqq j \leqq \nu - 1$$

Then (2.4.4) is given by

$$x_{k+1} = h(x_k) \quad , \quad k \geqq 0 \qquad ,$$

where

$$h(x) = g_{\nu-\frac{1}{2}} \circ g_{\nu-1} \circ \cdots \circ g_{\frac{1}{2}} \circ g_0(x)$$

Now, if $G_1: R^n \to R^n$ is a diffeomorphism and $G_2: R^n \to R^n$ is differentiable, then

$$G_2'(x) = \frac{d}{dx} [G_1(G_1^{-1}(G_2(x)))]$$
$$= G_1'(G_1^{-1}(G_2(x))) \cdot \frac{d}{dx} [G_1^{-1}(G_2(x))]$$

Thus,

$$\frac{d}{dx} G_1^{-1}(G_2(x)) = [G_1'(G_1^{-1}(G_2(x)))]^{-1} G_2'(x) \qquad ,$$

and so,

$$g_j'(x) = [r_j + H_1'(g_j(x))]^{-1} [r_j - V_1'(x)]$$
$$g_{j+\frac{1}{2}}'(x) = [r_j + V_1'(g_{j+\frac{1}{2}}(x))]^{-1} [r_j - H_1'(x)]$$

Now, $g_{i/2}(x^*) = x^*$ for $0 \leqq i \leqq 2\nu - 1$. Hence,

$$g_{j+\frac{1}{2}}'(x^*) \cdot g_j'(x^*) = T_{r_j}(x^*)$$

Furthermore,

$$g_{i/2} \circ g_{i/2 - \frac{1}{2}} \circ \cdots \circ g_0(x^*) = x^* \quad \text{for } 0 \leqq i \leqq 2\nu - 1$$

Hence,

$$h'(x^*) = g'_{\nu-\frac{1}{2}}(x^*) \cdot g'_{\nu-1}(x^*) \cdots g'_0(x^*)$$

$$= \prod_{i=0}^{\nu-1} T_{r_i}(x^*)$$

Now consider (2.4.5). Let $f_{1/2}: R^n \to R^n$, $0 \leqq i \leqq 2\nu - 1$, be defined by

$$f_j(x) = x - [r_j + H_1'(x)]^{-1} F(x) \quad , \quad 0 \leqq j \leqq \nu - 1$$

$$f_{j+\frac{1}{2}}(x) = x - [r_j + V_1'(x)]^{-1} F(x) \quad , \quad 0 \leqq j \leqq \nu - 1$$

Then (2.4.5) is given by

$$x_{k+1} = h_2(x_k) \quad , \quad k \geqq 0 \quad ,$$

where

$$h_2(x) = f_{\nu-\frac{1}{2}} \circ f_{\nu-1} \circ \cdots \circ f_{\frac{1}{2}} \circ f_0 \quad .$$

Now, if H_1, $V_1 \in C^2(R^n)$, the $f_{i/2} \in C^1(R^n)$ for $0 \leqq i \leqq 2\nu - 1$. Let $\xi \in R^n$. Then, for $0 \leqq j \leqq \nu - 1$,

$$f_j'(x)\ \xi = [I - [r_j + H_1'(x)]^{-1} F'(x)]\ \xi$$

$$- \frac{d}{dx}[[r_j + H_1'(x)]^{-1}]\ \xi\ F(x)$$

Thus, since $F(x^*) = 0$,

$$f_j'(x^*) = I - [r_j + H_1'(x^*)]^{-1} [H_1'(x^*) + V_1'(x^*)]$$

$$= [r_j + H_1'(x^*)]^{-1} [r_j - V_1'(x^*)]$$

Likewise, for $0 \leqq j \leqq \nu - 1$,

$$f'_{j+\frac{1}{2}}(x^*) = [r_j + V_1'(x^*)]^{-1} [r_j - H_1'(x^*)] \quad ,$$

and so,

$$f'_{j+\frac{1}{2}}(x^*) \cdot f_j'(x^*) = T_{r_j}(x^*)$$

Proceeding as above, we find

$$h_2'(x^*) = \prod_{i=0}^{\nu-1} T_{r_i}(x^*) \quad .$$

Thus, we see that when H_1, $V_1 \in C^2(R^n)$, (2.4.3), (2.4.4), and (2.4.5) have identical asymptotic rates of convergence, and indeed, near the solution, the three iterations behave very nearly alike. (2.4.4) involves the inversion of nonlinear functions and is usually not practical. (2.4.3) requires one function evaluation and one derivative evaluation per cycle, while (2.4.5) requires 2ν function evaluations and 2ν derivative evaluations per cycle. Thus, in terms of work requirement, (2.4.3) seems to be far superior to (2.4.5)--at least locally.

Consider (2.3.1). If f depends on u as well as the space variables, s and t, (2.3.2) becomes

$$F(x) \equiv Hx + Vx + \varphi(x) = 0 \qquad (2.4.6)$$

where $\varphi \in D(R^n)$. Motivated by (2.3.3), we may consider the following special cases of (2.4.3)--(2.4.5).

<u>N-ν step ADI</u>:

$$(2.4.7)\quad \begin{cases} x_0 \in R^n \\ [r_j + H + c\varphi'(x_k)]\, x_k^{j+\frac{1}{2}} = [r_j - V - (1-c)\varphi'(x_k)]\, x_k^j + F'(x_k)x_k - F(x_k) \\ [r_j + V + (1-c)\varphi'(x_k)]\, x_k^{j+1} = [r_j - H - c\varphi'(x_k)]\, x_k^{j+\frac{1}{2}} + F'(x_k)x_k - F(x_k) \\ x_k^0 = x_k \,,\ x_{k+1} = x_k^\nu \end{cases}$$

<u>ν step ADI</u>:

$$(2.4.8)\quad \begin{cases} x_0 \in R^n \\ r_j x_k^{j+\frac{1}{2}} + Hx_k^{j+\frac{1}{2}} + c\varphi(x_k^{j+\frac{1}{2}}) = r_j x_k^j - Vx_k^j - (1-c)\varphi(x_k^j) \\ r_j x_k^{j+1} + Vx_k^{j+1} + (1-c)\varphi(x_k^{j+1}) = r_j x_k^{j+\frac{1}{2}} - Hx_k^{j+\frac{1}{2}} - c\varphi(x_k^{j+\frac{1}{2}}) \\ x_k^0 = x_k \,,\ x_{k+1} = x_k^\nu \end{cases}$$

ν step ADI-N:

$$(2.4.9)\begin{cases} x_0 \in R^n \\ [r_j + H + c\varphi'(x_k^j)]\, x_k^{j+\frac{1}{2}} = [r_j + H + c\varphi'(x_k^j)]\, x_k^j - F(x_k^j) \\ [r_j + V + (1-c)\varphi'(x_k^{j+\frac{1}{2}})]\, x_k^{j+1} = [r_j + V + (1-c)\varphi'(x_k^{j+\frac{1}{2}})]\, x_k^{j+\frac{1}{2}} - F(x_k^{j+\frac{1}{2}}) \\ x_k^0 = x_k \,,\; x_{k+1} = x_k^\nu . \end{cases}$$

From the results of Section 2.2, we have the following theorem.

2.4.1 Theorem: Let $F: R^n \to R^n$ be defined by (2.4.6) where H and V are positive semi-definite, $\varphi \in C^2(R^n)$, $\varphi'(x^*)$ is symmetric, and $F(x^*) = 0$. Suppose $\sigma(H) \subset [a,b]$, $\sigma(V) \subset [\alpha,\beta]$, $\sigma(\varphi'(x^*)) \subset [s,t]$ and that one condition from Table 2.3.1 is satisfied. Define

$$a_1 = \begin{cases} a + cs & , c \geq 0 \\ a + ct & , c \leq 0 \end{cases} \quad , \qquad b_1 = \begin{cases} b + ct & , c \geq 0 \\ b + cs & , c \leq 0 \end{cases}$$

$$\alpha_1 = \begin{cases} \alpha + (1-c)s & , c \leq 1 \\ \alpha + (1-c)t & , c \geq 1 \end{cases} \quad , \qquad \beta_1 = \begin{cases} \beta + (1-c)t & , c \leq 1 \\ \beta + (1-c)s & , c \geq 1 \end{cases} .$$

Then methods (2.4.7)--(2.4.9) are locally convergent to x^* if one of the following conditions is satisfied.

i) $r_j \equiv r > 0$ is constant ($\nu = 1$.) (Theorem 2.2.1.)

ii) $\dfrac{\beta_1 - \alpha_1}{2} < r_j$ for $0 \leq j \leq \nu - 1$. (Theorem 2.2.2.)

iii) (2.2.5) holds. (Theorem 2.2.4.)

iv) (2.2.5b) and (2.2.6a,c) hold. (Remark 2.2.5.)

CHAPTER III

NONLINEAR ADI ITERATIONS

In this chapter, we consider the nonlinear iteration (2.4.4) and obtain convergence results. Specifically, let X be a real Hilbert space, and suppose $F = H + V$, where $F, H, V: X \to X$ are monotone. Then, by Theorem 1.3.8, the following nonlinear ADI iteration is well defined for $r_k, s_k > 0$.

$$(3.1.1) \qquad \begin{cases} x_0 \in X \\ s_k x_{k+\frac{1}{2}} + H(x_{k+\frac{1}{2}}) = s_k x_k - V(x_k) \\ r_k x_{k+1} + V(x_{k+1}) = r_k x_{k+\frac{1}{2}} - H(x_{k+\frac{1}{2}}) \end{cases}$$

We assume F has a unique root, x^*. If F is uniformly monotone, this is guaranteed. We now consider conditions under which $x_{k/2} \to x^*$, where $\{x_{k/2}\}$ satisfies (3.1.1). Many of the convergence results in the linear non-commutative case carry over to the nonlinear case--the nonlinearity tends to destroy the special properties of the commutative case. The positive definite conditions in the linear case will be replaced by monotonicity conditions (as they could be in the linear case,) and the boundedness conditions--from the linearity of the operators in finite dimensions--will be replaced by Lipschitz conditions.

We will use the following lemma which is a slight extension of a result by Kellogg [15]. We use the notation of Definitions 1.1.1 and 1.1.2.

<u>3.1.1 Lemma</u>: Let X be a real Hilbert space. Let $B: X \to X$ be monotone and continuous. Define $T: X \to X$ by

$$T(x) = (r - B)\circ(s + B)^{-1}(x)$$

where $0 < r_0 \le r, s \le s_0$. Then

$$\|T(x) - T(y)\| \leqq \max\left(1, \frac{r}{s}\right) \|x - y\|$$

Furthermore, if $B \in Lip_b$ and either

$$\text{i) } r < s$$

or

$$\text{ii) } r \leqq s, \text{ and } B \in Mon_b \qquad ,$$

then, given a bounded set D, there is a $\delta_D < 1$ such that $T \in Lip(D, \delta_D)$. Moreover, if ii) holds, δ_D can be chosen to depend only on r_0, s_0 and D.

<u>Proof</u>: Since r, s > 0, T is defined. Given x, y $\in$ X, let

$$\bar{x} = (s + B)^{-1}(x) \ , \ \bar{y} = (s + B)^{-1}(y)$$

Then

$$\frac{\|T(x) - T(y)\|^2}{\|x - y\|^2} = \frac{\|(r - B)(\bar{x}) - (r - B)(\bar{y})\|^2}{\|(s + B)(\bar{x}) - (s + B)(\bar{y})\|^2}$$

$$= \frac{(r(\bar{x} - \bar{y}) - (B(\bar{x}) - B(\bar{y})), r(\bar{x} - \bar{y}) - (B(\bar{x}) - B(\bar{y})))}{(s(\bar{x} - \bar{y}) + (B(\bar{x}) - B(\bar{y})), s(\bar{x} - \bar{y}) + (B(\bar{x}) - B(\bar{y})))}.$$

Thus,

(3.1.2)

$$\frac{\|T(x) - T(y)\|}{\|x - y\|} = \left[\frac{r^2\|\bar{x} - \bar{y}\|^2 - 2r(B(\bar{x}) - B(\bar{y}), \bar{x} - \bar{y}) + \|B(\bar{x}) - B(\bar{y})\|^2}{s^2\|\bar{x} - \bar{y}\|^2 - 2r(B(\bar{x}) - B(\bar{y}), \bar{x} - \bar{y}) + \|B(\bar{x}) - B(\bar{y})\|^2}\right]^{\frac{1}{2}} .$$

So, by the monotonicity of B,

$$(3.1.3) \qquad \frac{\|T(x) - T(y)\|}{\|x - y\|} \leqq \left[\frac{r^2\|\bar{x} - \bar{y}\|^2 + \|B(\bar{x}) - B(\bar{y})\|^2}{s^2\|\bar{x} - \bar{y}\|^2 + \|B(\bar{x}) - B(\bar{y})\|^2}\right]^{\frac{1}{2}}$$

Hence,

$$\|T(x) - T(y)\| \leqq \max\left(1, \frac{r}{s}\right) \|x - y\|$$

Now, suppose $B \in Lip_b$. Let D be a bounded set, and set

$$D_1 = \{ (s + B)^{-1}(w) : w \in D, r_0 \leqq s \leqq s_0 \}$$

Suppose $\|w\| \leqq M$ for $w \in D$. Let $z \in D_1$. Then there is a $w \in D$ and an $s \in [r_0, s_0]$ such that $(s + B)(z) = w$. Now, by the monotonicity of B,

$$M\|z\| \geqq \|w\| \ \|z\| \geqq (w, z) = ((s + B)(z), z)$$

$$= s\|z\|^2 + (B(z) - B(0), z - 0) + (B(0), z)$$

$$\geqq s\|z\|^2 - \|B(0)\| \ \|z\|$$

Hence,

$$\|z\| \leqq \frac{1}{s_0}[M + \|B(0)\|] \equiv M_1 \qquad ,$$

and M_1 is independent of r and s. Thus, there is a β independent of r and s such that $B \in Lip(D_1, \beta)$. Now, suppose i) holds. Then, from (3.1.3),

$$\frac{\|T(x) - T(y)\|}{\|x - y\|} \leqq \left[\frac{r^2 + \beta^2}{s^2 + \beta^2}\right]^{\frac{1}{2}} \equiv \delta_D < 1$$

Now, suppose ii) holds. Then there is an $\alpha > 0$ and independent of r and s such that $B \in Mon(D_1, \alpha)$. So, from (3.1.2),

$$\frac{\|T(x) - T(y)\|}{\|x - y\|} \leqq \left[\frac{r^2 - 2r\alpha + \beta^2}{s^2 - 2s\alpha + \beta^2}\right]^{\frac{1}{2}} \leqq \left[\frac{s_0^2 - 2r_0\alpha + \beta^2}{s_0^2 + 2r_0\alpha + \beta^2}\right]^{\frac{1}{2}} \equiv \delta_D < 1 \quad ,$$

and δ_D is independent of r and s. This completes the proof.

Consider (3.1.1). Each of the equations is nonlinear. Thus rather than solve each of them exactly, one might solve them incompletely by applying a finite number of sweeps of an appropriate inner iterative procedure, e.g. a Newton or Picard procedure. Since the inner iterations do not yield the exact solutions of (3.1.1), the actual iteration is of the form,

$$(3.1.4)\quad \begin{cases} x_0 \in X \\ s_k x_{k+\frac{1}{2}} + H(x_{k+\frac{1}{2}}) = s_k x_k - V(x_k) + \epsilon_k \\ r_k x_{k+1} + V(x_{k+1}) = r_k x_{k+\frac{1}{2}} - H(x_{k+\frac{1}{2}}) + \epsilon_{k+\frac{1}{2}} \end{cases} ,$$

where ϵ_k and $\epsilon_{k+\frac{1}{2}}$ are, in effect, defined by (3.1.4). Lemma 3.1.1 will be used to obtain convergence results for (3.1.4). We first establish some additional lemmas.

3.1.2 Lemma: Let X be a Banach space and suppose $h, h_k : X \to X$ satisfy

i) given a bounded set D, there is a $\delta_D < 1$ such that

$$h_k \in \text{Lip}(D, \delta_D) \text{ for } k \geq 0 ,$$

ii) $h(x^*) = x^* \in X$,

iii) $\epsilon_k \equiv \|h_k(x^*) - h(x^*)\| \to 0 \ (k \to \infty)$,

iv) $\{T_k(z)\}$ is bounded for some $z \in X$, where

$$T_k = h_k \circ h_{k-1} \circ \cdots \circ h_0$$

Suppose $\{x_k\} \subset X$ satisfies

$$x_{k+1} = h_k(x_k) + \eta_k \text{ for } k \geq 0 ,$$

where

v) $\sum\limits_{k=0}^{\infty} \|\eta_k\| < \infty$

Then $x_k \to x^*$.

Proof: The proof is an application of Theorem 2 of [19]. We give a direct proof. By i), $h_k \in \text{Lip}(X, 1)$. So

$$\begin{aligned} \|x_{k+1} - T_k(z)\| &\leq \|h_k(x_k) - h_k(T_{k-1}(z))\| + \|\eta_k\| \\ &\leq \|x_k - T_{k-1}(z)\| + \|\eta_k\| \end{aligned}$$

Hence,

$$\|x_{k+1} - T_k(z)\| \leq \|x_0 - z\| + \sum_{j=0}^{\infty} \|\eta_k\|$$

Thus, by iv) and v), $\{x_k\}$ is bounded. Let $D = \{x_k\} \cup \{x^*\}$. Then

Proof: F^{-1} exists since $\alpha > 0$. Let $x, y \in D_1$. Then

$$\|x - y\| \, \|F^{-1}(x) - F^{-1}(y)\| \geq (x - y, F^{-1}(x) - F^{-1}(y))$$

$$= (F(F^{-1}(x)) - F(F^{-1}(x)), F^{-1}(x) - F^{-1}(y))$$

$$\geq \alpha \, \|F^{-1}(x) - F^{-1}(y)\|^2$$

Thus,

$$\|F^{-1}(x) - F^{-1}(y)\| \leq \frac{1}{\alpha} \|x - y\| \quad \text{for } x, y \in D_1$$

3.1.5 Definition: Let X be a Banach space. Let $\{T_k\}$ be a sequence of maps from X to X. Then $\{T_k\}$ is equicontinuous at $x_0 \in X$ if, given $\epsilon > 0$, there is a $\delta > 0$ such that

$$\|x_0 - y\| \leq \delta \Rightarrow \|T_k(x_0) - T_k(y)\| \leq \epsilon \text{ for } k \geq 0$$

3.1.6 Lemma: Let X be a Banach space. Let $x_0 \in X$, $\{x_k\} \subset X$, and T_k, $T: X \to X$, $k \geq 0$, satisfy

i) $x_k \to x_0$,

ii) $\{T_k\}$ is equicontinous at x_0 ,

iii) $T_k(x_0) \to T(x_0)$.

Then $T_k(x_k) \to T(x_0)$.

Proof: We have

$$\|T_k(x_k) - T(x_0)\| \leq \|T_k(x_k) - T_k(x_0)\| + \|T_k(x_0) - T(x_0)\| \quad .$$

The first term on the left goes to zero by i) and ii), and the second term goes to zero by iii).

The following two lemmas, which apply to the proof of Theorem 3.1.9, are stated separately to keep the proof of Theorem 3.1.9 as clear as possible.

<u>3.1.7 Lemma</u>: Let X be a real Hilbert space. Let H, V: X → X be continuous and monotone, and suppose

$$(3.1.5)\qquad \begin{cases} 0 < r \leqq s_{k+1} \leqq r_k \leqq s_k \leqq s < \infty \quad \text{for } k \geqq 0 & , \\ r_k,\ s_k \to r & . \end{cases}$$

Define

$$(3.1.6)\quad \begin{cases} T_H^k = (r_k - H)\circ(s_k + H)^{-1} & , \quad T_H = (r - H)\circ(r + H)^{-1} \\ T_V^k = (s_{k+1} - V)\circ(r_k + V)^{-1} & , \quad T_V = (r - V)\circ(r + V)^{-1} \\ h_k = T_V^k \circ T_H^k & , \quad h = T_V \circ T_H \end{cases}$$

Then $T_H^k \to T_H$, $T_V^k \to T_V$, and $h_k \to h$ pointwise on X.

<u>Proof</u>: Let $x \in X$. By Lemma 3.1.6, to show $T_H^k(x) \to T_H(x)$, it is sufficient to show

i) $(s_k + H)^{-1}(x) \to (r + H)^{-1}(x)$,

ii) $\{(r_k - H)\}$ is equicontinuous at $(r + H)^{-1}(x)$,

iii) $(r_k - H)\circ(r + H)^{-1} \to (r - H)\circ(r + H)^{-1}(x)$.

To show i), let $z_k = (s_k + H)^{-1}(x)$ and $z = (r + H)^{-1}(x)$. Then, as in the proof of Lemma 3.1.1, there is an M independent of k such that $\|z_k\| \leqq M$ for $k \geqq 0$. Thus,

$$\begin{aligned} 0 = (z_k - z,\ x - x) &= (z_k - z,\ (s_k + H)(z_k) - (r + H)(z)) \\ &= r\ \|z_k - z\|^2 + (s_k - r)\ (z_k - z,\ z_k) + (z_k - z,\ H(z_k) - H(z)) \\ &\geqq r\ \|z_k - z\|^2 - (s_k - r)\ \|z_k - z\|\ \|z_k\| \end{aligned}$$

Hence,

$$\|z_k - z\| \leqq \frac{1}{r}\ |s_k - s|\ M \quad \to \quad 0$$

This establishes i).

Now, for any $y, z \in X$

$$\begin{aligned} \|(r_k - H)(y) - (r_k - H)(z)\| &\leqq r_k\ \|y - z\| + \|H(y) - H(z)\| \\ &\leqq s\ \|y - z\| + \|H(y) - H(z)\| \end{aligned}$$

Thus, since H is continuous, $\{(r_k - H)\}$ is equicontinuous at each point of X. This establishes ii). Since $r_k \to r$, iii) follows immediately. Thus, $T_H^k \to T_H$ pointwise. That $T_V^k \to T_V$ pointwise follows in the same way.

Now to show $h_k(x) \to h(x)$, again by Lemma 3.1.6, it is sufficient to show

i) $T_H^k \to T_H(x)$,

ii) $\{T_V^k\}$ is equicontinuous at $T_H(x)$,

iii) $T_V^k(T_H(x)) \to T_V(T_H(x))$.

But i) and iii) follow as above, and ii) follows because, from Lemma 3.1.1, $T_V^k \in \mathrm{Lip}(X, 1)$ for $k \geqq 0$.

<u>3.1.8 Lemma</u>: Let X be a real Hilbert space. Let H: X → X be monotone and continuous. Suppose (3.1.5) is satisfied. Let $y_k \to y \in X$. Then $(s_k + H)^{-1}(y_k) \to (r + H)^{-1}(y)$.

<u>Proof</u>: By Lemma 3.1.6, we need only show

ii) $\{(s_k + H)^{-1}\}$ is equicontinuous at y ,

iii) $(s_k + H)^{-1}(y) \to (r + H)^{-1}(y)$.

But, by Lemma 3.1.4, $(s_k + H)^{-1} \in \mathrm{Lip}(X, \frac{1}{r})$, which establishes ii). Condition iii) is established as in the proof of Lemma 3.1.7.

The following theorem is an extension of a result of Kellogg [15].

<u>3.1.9 Theorem</u>: Let X be a real Hilbert space. Let F, H, V: X → X be continuous and satisfy $F = H + V$, $F(x^*) = 0$ for some $x^* \in X$, $H \in \mathrm{Mon}_b \cap \mathrm{Lip}_b$, and $V \in \mathrm{Mon}(X,0)$. Suppose that (3.1.5) holds and that $\{x_{k/2}\}$ satisfies (3.1.4) where

(3.1.7) $$\sum_{k=0}^{\infty} \|\epsilon_{k/2}\| < \infty$$

Then $x_{k/2} \to x^*$.

Proof: Use the definitions of (3.1.6). Further define

$$G_k = h_k \circ h_{k-1} \circ \cdots \circ h_0, \quad y_k = (s_k + H)(x_{k+\frac{1}{2}}), \quad y^* = (r + H)(x^*).$$

Now,

$$y_{k+1} = (s_{k+1} + H)(x_{k+\frac{3}{2}}) = (s_{k+1} - V)(x_{k+1}) + \epsilon_{k+1}$$

$$= (s_{k+1} - V)\circ(r_k + V)^{-1}((r_k - H)(x_{k+\frac{1}{2}}) + \epsilon_{k+\frac{1}{2}}) + \epsilon_{k+1}$$

$$= T_V^k(T_H^k(y_k) + \epsilon_{k+\frac{1}{2}}) + \epsilon_k$$

But, by Lemma 3.1.1, $T_V^k \in Lip(X, 1)$. Hence,

$$y_{k+1} = T_V^k \circ T_H^k(y_k) + \zeta_{k+\frac{1}{2}} + \epsilon_{k+1} ,$$

where

$$\|\zeta_{k+\frac{1}{2}}\| \leqq \|\epsilon_{k+\frac{1}{2}}\|$$

Let $\eta_k = \zeta_{k+\frac{1}{2}} + \epsilon_{k+1}$. Then

(3.1.8) $$y_{k+1} = h_k(y_k) + \eta_k$$

We now show (3.1.8) satisfies the conditions of Lemma 3.1.2.

i): Let D be a bounded set. By Lemma 3.1.1, there is a $\delta_D < 1$ and independent of k such that $T_H^k \in Lip(D, \delta_D)$ for $k \geqq 0$. But, also by Lemma 3.1.1, $T_V^k \in Lip(X, 1)$. Hence, $h_k \in Lip(D, \delta_D)$.

ii): Since $F(x^*) = 0$, we have $(r + H)(x^*) = (r - V)(x^*)$ and $(r + V)(x^*) = (r - H)(x^*)$. Hence,

$$(r + H)(x^*) = (r - V) \circ (r + V)^{-1} \circ (r - H) \circ (r + H)^{-1} \circ (r + H)(x^*),$$

i.e., $y^* = h(y^*)$.

iii): By Lemma 3.1.7, $h_k \to h$ pointwise. In particular, $h_k(y^*) \to h(y^*)$.

iv): Let $z = (s_0 - V)(x^*)$. Then $G_k(z) = (s_{k+1} - V)(x^*)$. So $\{G_k(z)\}$ is bounded by $s\|x^*\| + \|V(x^*)\|$.

v): $\sum_{k=0}^{\infty} \|\eta_k\| \leq \sum_{k=0}^{\infty} \|\epsilon_{k+\frac{1}{2}}\| + \|\epsilon_k\| < \infty$

Thus, by Lemma 3.1.2, $y_k \to y^*$. But $x_{k+\frac{1}{2}} = (s_k + H)^{-1}(y_k)$. So, by Lemma 3.1.8, $x_{k+\frac{1}{2}} \to (r + H)^{-1}(y^*) = x^*$. Now,

$$(r_k + V)(x_{k+1}) = (r_k - H)(x_{k+\frac{1}{2}}) + \epsilon_{k+\frac{1}{2}}$$
$$= (r_k + V)(x_{k+\frac{1}{2}}) - F(x_{k+\frac{1}{2}}) + \epsilon_{k+\frac{1}{2}}$$

Thus, $x_{k+1} = x_{k+\frac{1}{2}} + \xi_{k+\frac{1}{2}}$, where, by Lemma 3.1.4,

$$\|\xi_{k+\frac{1}{2}}\| = \|x_{k+1} - x_{k+\frac{1}{2}}\|$$
$$= \|(r_k + V)^{-1}((r_k + V)(x_{k+\frac{1}{2}}) - F(x_{k+\frac{1}{2}}) + \epsilon_{k+\frac{1}{2}}) - (r_k + V)^{-1}((r_k + V)(x_{k+\frac{1}{2}}))\|$$
$$\leq \frac{1}{r} \|-F(x_{k+\frac{1}{2}}) + \epsilon_{k+\frac{1}{2}}\| .$$

But, $\epsilon_{k+\frac{1}{2}} \to 0$, and, since $x_{k+\frac{1}{2}} \to x^*$, by the continuity of F, $F(x_{k+\frac{1}{2}}) \to 0$. Hence, $\xi_{k+\frac{1}{2}} \to 0$, and so $x_{k+1} \to x^*$. This completes the proof.

Kellogg proved Theorem 3.1.9 under the assumptions that $r_k = s_k = r$ and $\epsilon_{k/2} = 0$.

If the Lipschitz and monotonicity restrictions on H hold uniformly on X, (3.1.7) can be weakened somewhat.

<u>3.1.10 Corollary</u>: Assume the conditions of Theorem 3.1.9 except (3.1.7). Assume also that

$$H \in \mathrm{Lip}(X, b) \cap \mathrm{Mon}(X, a) \quad \text{where } a > 0 \text{ and } b < \infty$$

and that $\epsilon_{k/2} \to 0$. Then $x_k \to x^*$.

<u>Proof</u>: The proof is the same as that for Theorem 3.1.9 except that we use Corollary 3.1.3.

Usually, we would like to apply a finite number of parameters in a cyclic manner. The following theorem, which is an extension of Pearcy's result, Theorem 2.2.5, allows us to do this.

3.1.11 Theorem: Let X be a real Hilbert space. Let F, H, V: $X \to X$ be continuous and satisfy $F = H + V$, $F(x^*) = x^*$, $H \in \mathrm{Lip}(X, b) \cap \mathrm{Mon}(X, a)$, and $V \in \mathrm{Lip}(X, \beta) \cap \mathrm{Mon}(X, \alpha)$, where

(3.1.9) $\quad b < \infty,\ a > 0,\ \beta \leqq \infty,\ \alpha \geqq 0 \quad \text{or} \quad b \leqq \infty,\ a \geqq 0,\ \beta < \infty,\ \alpha > 0$

Let

$$\delta_H = \begin{cases} \left[\dfrac{r^2 - 2ra + b^2}{r^2 + 2ra + b^2}\right]^{\frac{1}{2}}, & b < \infty \\ 1, & b = \infty \end{cases}$$

$$\delta_V = \begin{cases} \left[\dfrac{r^2 - 2r\alpha + \beta^2}{r^2 + 2r\alpha + \beta^2}\right]^{\frac{1}{2}}, & \beta < \infty \\ 1, & \beta = \infty \end{cases},$$

and $\delta = \delta_H \cdot \delta_V$. Suppose

$$\begin{cases} 0 < r = r_{\nu-1} \leqq s_{\nu-1} \leqq r_{\nu-2} \leqq \cdots \leqq r_0 \leqq s_0 = s \\ r_{k\nu+j} = r_j\ ,\quad s_{k\nu+j} = s_j \quad \text{for} \quad 0 \leqq j \leqq \nu - 1 \text{ and } k \geqq 0 \end{cases},$$

where $\nu \geqq 1$ and

(3.1.10)
$$\nu - 1 > \frac{\log \frac{r}{s\delta_H}}{\log \delta}$$

Suppose, finally, that $\{x_{k/2}\}$ satisfies (3.1.4). Then $x_{k/2} \to x^*$.

Proof: By (3.1.9), $\delta < 1$. Thus, ν is well defined and finite. Define T_H^k and T_V^k as in (3.1.6). Set

$$g = T_H^{\nu-1} \circ T_V^{\nu-2} \circ T_H^{\nu-2} \circ \cdots \circ T_V^0 \circ T_H^0 \circ (s_0 - V) \circ (r_{\nu-1} + V)^{-1},$$

$y_k = (r_{\nu-1} + V)(x_{\nu k})$ for $k \geqq 0$, and $y^* = (r_{\nu-1} + V)(y^*)$. Then

$$y_{k+1} = g(y_k) + \eta_k,$$

where

$$\|\eta_k\| \leq \|\epsilon_{(\nu-1)k}\| + \|\epsilon_{(\nu-1)k+\frac{1}{2}}\| + \cdots + \|\epsilon_{\nu k+\frac{1}{2}}\| ,$$

and

$$g(y^*) = y^* .$$

Now, by Lemma 3.1.1, $T_H^j \in \mathrm{Lip}(X, \delta_H)$ and $T_V^j \in \mathrm{Lip}(X, \delta_V)$ for $0 \leq j \leq \nu - 2$, and $(s_0 - V) \circ (r_{\nu-1} + V)^{-1} \in \mathrm{Lip}(X, \frac{s}{r})$. Thus, $g \in \mathrm{Lip}(X, \delta^*)$, where $\delta^* = \delta_H \cdot \delta^{\nu-1} \frac{s}{r}$. Now, provided $\nu \geq 1$ satisfies (3.1.10), $\delta^* < 1$, and so g is a uniform contraction on X. Now, $\epsilon_{k/2} \to 0$ implies $\eta_k \to 0$. Thus, as in the proof of Lemma 3.1.2, $y_k \to y^*$. I.e.,

$$(r_{\nu-1} + V)(x_{\nu k}) \to (r_{\nu-1} + V)(x^*) .$$

But, $(r_{\nu-1} + V)$ is continuous and uniformly monotone. Hence, $(r_{\nu-1} + V)^{-1}$ is continuous. Thus, $x_{\nu k} \to x^*$ $(k \to \infty)$. Then, by the technique at the end of Theorem 3.1.9,

$$\lim_{k\to\infty} x_{(\nu-1)+\frac{1}{2}} = \lim_{k\to\infty} x_{(\nu-1)+1} = \lim_{k\to\infty} x_{\nu k} = x^*$$

This completes the proof.

CHAPTER IV

PICARD-ADI ITERATIONS

Let $F: R^n \to R^n$ satisfy (2.4.6) and have a root, x^*. In this chapter, we will consider two-level Picard-ADI iterations for approximating x^*. This type of iteration was considered by Douglas [5] and Gunn [12], [13].

Generally speaking, Picard type iterations for finding the root of F are not globally convergent unless the growth of φ is sufficiently restricted. However, in some cases, we can replace the problem of finding the root of F with the problem of finding a root of a related function, $F_0 = A + \varphi_0$. where the growth of φ_0 is sufficiently restricted. See, e.g., the discussion in Section 1.4. Even if the growth of φ is not sufficiently restricted on all of R^n, if some bound on $\|x_0 - x^*\|$ is known, we may be able to formulate a Picard type iterative procedure which will converge to x^* from x_0.

In Section 4.1, we give some preliminary results concerning general two-level iterations, define Picard type iterations and give some examples of specific Picard type iterations.

In Section 4.2, we consider multistep Picard-ADI iterations, i.e., iterations in which the inner iteration is composed of several ADI sweeps. In Section 4.3, we consider single step Picard-ADI and ADI-Picard iterations.

In this chapter, $\|\cdot\|$, in R^n, will denote $\|\cdot\|_2$

4.1 Preliminary Results.

4.1.1 Lemma: Let X be a Banach space. Let $\{x_k\} \subset X$ and $x^* \in X$, and suppose x_k, $x^* \in D_k \subset X$ for $k \geqq 0$. Let $h_k: D_k \to X$, satisfy $h_k(x^*) = x^*$ and $h_k \in \mathrm{Lip}(D_k, \delta)$ for $k \geqq 0$, where $\delta < 1$ independently of k. Suppose, finally, that

$$\|x_{k+1} - h_k(x_k)\| \leqq \eta \, \|x_k - h_k(x_k)\| \quad \text{for } k \geqq 0 \qquad ,$$

where

$$\eta < \frac{1 - \delta}{1 + \delta}$$

independently of k. Then $x_k \to x^*$.

<u>Proof</u>: Let $\delta^* = \eta(1 + \delta) + \delta$. Then $\delta^* < 1$. Now,

$$\|x_0 - x^*\| \leqq (\delta^*)^0 \|x_0 - x^*\|$$

Suppose

$$\|x_k - x^*\| \leqq (\delta^*)^k \|x_0 - x^*\|$$

Then

$$\begin{aligned} \|x_{k+1} - x^*\| &\leqq \|x_{k+1} - h_k(x_k)\| + \|h_k(x_k) - h_k(x^*)\| \\ &\leqq \eta \, (\|x_k - x^*\| + \|h_k(x^*) - h_k(x_k)\|) + \|h_k(x_k) - h_k(x^*)\| \\ &\leqq (\eta \, (1 + \delta) + \delta) \, \|x_k - x^*\| \qquad . \end{aligned}$$

Hence, by induction, $\|x_k - x^*\| \leqq (\delta^*)^k \, \|x_0 - x^*\|$ for $k \geqq 0$. This completes the proof.

Suppose we wish to find the common fixed point of the functions, h_k, of Lemma 4.1.1. If h_k is difficult to evaluate, we may consider a two-level iteration in which the theoretical outer iteration is given by $x_{k+1} = h_k(x_k)$. If the error in the inner iteration is reduced by a factor of η, Lemma 4.1.1 guarantees convergence of the two level scheme.

Let $F(x) = Ax + \varphi(x)$, where $\varphi: R^n \to R^n$, and $A \in L(R^n, R^n)$ is non-singular. Suppose $F(x^*) = 0$. Then $Ax^* = -\varphi(x^*) = Ax^* - F(x^*)$, which suggests the following iteration, commonly called the Picard iteration for approximating x^*.

$$Ax_{k+1} = Ax_k - F(x_k)$$

Motivated by this, we make the following definition.

<u>4.1.2 Definition:</u> Let $F(x) = Ax + \varphi(x)$, where $\varphi\colon R^n \to R^n$, and $A \in L(R^n,R^n)$. If $\sigma \notin \sigma[A]^{(*)}$, the iteration,

$$[\sigma + A]\, x_{k+1} = [\sigma + A]\, x_k - F(x_k) \qquad ,$$

is called a <u>Picard iteration</u>. The parameter, σ, is called the <u>Picard parameter</u>.

Definition 4.1.2 is generalized to the following.

<u>4.1.3 Definition:</u> Let $F\colon R^n \to R^n$, and let $C \in L(R^n,R^n)$ be non-singular. The iteration,

$$(4.1.1) \qquad Cx_{k+1} = Cx_k - F(x_k)$$

is called an <u>iteration of Picard type</u>.

Let $F \in M(R^n)$, and let $C \in L(R^n,R^n)$ be non-singular. Let

$$h(x) = C^{-1}[Cx - F(x)]$$

Then

$$h(x) - h(y) = C^{-1}[C - F^D(x,y)]\,(x - y)$$

Thus h is a contraction on $D \subset R^n$ if there is a $\delta < 1$ such that

$$\|C^{-1}[C - F^D(x,y)]\| \leq \delta \text{ for } x,\ y \in D$$

Generalizing, we have the following lemma.

(*) This double usage of σ--as a parameter and as the spectrum set--should cause no confusion.

4.1.4 Lemma: Let $F \in F(R^n, R^n)$ have the root, x^*. Let $\{C_k\}$ be a sequence of non-singular matrices. Let $x_0 \in R^n$, and suppose $\{x_k\} \subset R^n$ satisfies

$$C_k x_{k+1} = C_k x_k - F(x_k) \quad \text{for } k \geqq 0 \tag{4.1.2}$$

Suppose that for some $\delta < 1$ and independent of k,

$$\|C_k^{-1}[C_k - F^D(x_k, x^*)]\| \leqq \delta \quad \text{for } k \geqq 0 \tag{4.1.3}$$

Then $x_k \to x^*$.

Proof: We need only note that

$$x_{k+1} - x^* = C_k^{-1}[C_k - F^D(x_k, x^*)]\,(x_k - x^*)$$

4.1.5 Remark: We note that if $F \in M(R^n)$ is differentiable, then

$$F^D(x, y) = \int_0^1 F'(y + t(x - y))dt \tag{4.1.4}$$

Thus, (4.1.3) is satisfied if

$$\|C_k^{-1}[C_k - F'(x)]\| \leqq \delta \quad \text{for } x \in [x^*, x_k] \text{ and } k \geqq 0$$

We now give some further examples of Picard type iterations.

4.1.6 Example: Let $F \in M(R^n)$ be differentiable. Consider (4.1.1). If $C = F'(x_0)$, we have the simplified Newton iteration,

$$F'(x_0)\, x_{k+1} = F'(x_0)\, x_k - F(x_k) \quad .$$

Let $x_0, y_0 \in R^n$. If $C = F^D(x_0, y_0)$, we have the following discrete simplified Newton iteration,

$$F^D(x_0, y_0)\, x_{k+1} = F^D(x_0, y_0)\, x_k - F(x_k) \quad .$$

(4.1.2) includes a variety of methods. If $C_k = F'(x_k)$, we have the Newton iteration,

$$F'(x_k)\, x_{k+1} = F'(x_k)\, x_k - F(x_k)$$

If $C_k = F^D(x_k, x_{k-1})$, we have the following secant iteration,

$$F^D(x_k, x_{k-1})\, x_{k+1} = F^D(x_k, x_{k-1})\, x_k - F(x_k) \quad .$$

<u>4.2 Multi-step Two Level Iterations</u>: Let $F \in M(R^n)$ have the root, x^*. Let $\{C_k\} \subset L(R^n,R^n)$ be non-singular. Then we may consider the iteration (4.1.2) for approximating x^*. If C_k is not easily inverted, we may apply an inner iteration and use Lemma 4.1.2. For instance, suppose the C_k satisfy, for $k \geqq 0$,

$$(4.2.1) \quad \begin{cases} C_k = H_k + V_k \; ; \quad H_k, V_k \in L(R^n,R^n) \text{ are symmetric} \\ \sigma[H_k] \subset [a_k, b_k] \; , \quad \sigma[V_k] \subset [\alpha_k, \beta_k] \\ 0 < a^* \leqq a_k \leqq b_k \leqq b^* \quad , \quad 0 < \alpha^* \leqq \alpha_k \leqq \beta_k \leqq \beta^* \\ a^*, b^*\ \alpha^*, \beta^* \text{ are independent of } k \end{cases} \quad .$$

Then an ADI inner iteration could be applied. Following (2.1.7), define

$$\begin{cases} T_{k,r} = [r + V_k]^{-1} [r - H_k] [r + H_k]^{-1} [r - V_k] \\ Q_{k,r} = 2r\, [r + V_k]^{-1} [r + H_k]^{-1} \end{cases}$$

Then, from (2.1.8) and (2.1.11), we have the two-level <u>Picard-multi-step ADI iteration</u>.

$$(4.2.2) \quad x_{k+1} = \prod_{i=0}^{\nu_k-1} T_{k,r_{k,i}}\, x_k + \sum_{j=0}^{\nu_k-1} \Big(\prod_{i=j+1}^{\nu_k-1} T_{k,r_{k,i}} \Big) Q_{k,r_{k,j}} [C_k x_k - F(x_k)]$$

$$= x_k - \sum_{j=0}^{\nu_k-1} \Big(\prod_{i=j+1}^{\nu_k-1} T_{k,r_{k,i}} \Big) Q_{k,r_{k,j}} F(x_k) \quad ,$$

where ν_k is the number of ADI sweeps employed during the k^{th} stage and $\{r_{k,0}, \cdots, r_{k,\nu_k-1}\}$ is the ADI parameter sequence applied during the k^{th} stage. Using Lemma 4.1.2, we have the following result for (4.2.2).

4.2.1 Theorem: Let $F \in M(R^n, R^n)$ have the root, x^*. Let $\{C_k\}$ satisfy (4.2.1). Let $x_0 \in R^n$, and suppose $\{x_k\}$ is defined by (4.2.2) for some sequence $\{\nu_k\}$ of positive integers and some collection $\{r_{k,j}\}$ of positive parameters. Suppose (4.1.3) is satisfied for some $\delta < 1$ and independent of k. Finally, suppose

$$(4.2.3) \qquad \left\| \prod_{i=0}^{\nu_k - 1} T_{k,r_{k,i}} \right\| \leqq \eta \quad \text{for } k \geqq 0 ,$$

for some η independent of k and satisfying

$$\eta < \frac{1 - \delta}{1 + \delta}$$

Then, $x_k \to x^*$.

Proof: Let

$$h_k(x) = C_k^{-1}[C_k x - F(x)] .$$

Then $h_k(x^*) = x^*$. Let $D_k = \{x_k, x^*\}$. Then $h_k \in \mathrm{Lip}(D_k, \delta)$. Now, by (2.1.9),

$$x_{k+1} - h_k(x_k) = \prod_{i=0}^{\nu_k - 1} T_{k,r_{k,i}} (x_k - h_k(x_k))$$

Thus, by (4.2.3),

$$\|x_{k+1} - h_k(x_k)\| \leq \eta \, \|x_k - h_k(x_k)\|$$

Hence, by Lemma 4.1.1, $x_k \to x^*$.

4.2.2 Remark: By (4.2.1), we know that if r and s are independent of k and

$$0 < r \leqq r_{k,j+1} \leqq r_{k,j} \leqq s \quad \text{for } 0 \leqq j \leqq \nu_k - 2 \text{ and } k \geqq 0 ,$$

or, if $H_k V_k = V_k H_k$ for $k \geqq 0$, only

$$0 < r \leqq r_{k,j} \leqq s \quad \text{for } 0 \leqq j \leqq \nu_k - 1 \text{ and } k \geqq 0 ,$$

then there is an $M = M(\delta)$, which is independent of k and the parameter sequence, such that (4.2.3) holds whenever $\nu_k \geqq M$ for $k \geqq 0$. In the commutative case, this is the basic result, and in the non-commutative case, this is, essentially, Pearcy's result, Theorem 2.2.4. Thus, we can bound, a priori, the number of inner iterations and still guarantee convergence of (4.2.2).

<u>4.2.3 Example</u>: Let $F \in M(R^n)$ satisfy

$$(4.2.4) \quad \begin{cases} F(x) = Ax + \varphi(x), \quad F(x^*) = 0 \\ A = H + V; \quad H, V \in L(R^n, R^n) \text{ are positive semi-definite} \\ \sigma[H] \subset [a, b], \ \sigma[V] \subset [\alpha, \beta]; \quad \alpha, a \geqq 0, \quad a + \alpha > 0 \\ \varphi \in D(R^n) \end{cases}$$

Suppose also that

$$(4.2.5) \qquad \begin{cases} mI \leqq \varphi^D(x,y) \leqq MI \quad \text{for } x, y \in R^n \\ -(a + \alpha) < m \end{cases} .$$

In [5], Douglas considered (4.2.2) under the conditions $C_k \equiv C = A + \sigma I$ (for some suitable ADI splitting of $A + \sigma I$) where $\sigma = \frac{1}{2}(M + m)$ and $HV = VH$. Indeed, in this case,

$$\|C^{-1}[C - F^D(x,y)]\| \leqq \|C^{-1}\| \, \|\sigma - \varphi^D(x,y)\| \leqq \frac{M - m}{M + m + 2(a + \alpha)} < 1$$

Thus, Theorem 4.2.1 can be applied. Douglas obtained convergence of (4.2.2) but did not show this could be done with an a priori bound on ν_k.

In the previous example, it was assumed that $\varphi^D(x, y)$ is globally bounded from above. By the discussion of Section 1.4, this is no real restriction. Nevertheless, in the next example, we do not make this assumption but assume, instead, that $\varphi^D(x, y)$ is bounded from above on bounded sets.

<u>4.2.4 Example</u>: Let F satisfy (4.2.4) and suppose that, given a bounded set, $D \subset R^n$, there is an $M(D)$ such that

$$\begin{cases} mI \leqq \varphi^D(x,\ y) \leq M(D)I & \text{for } x,\ y \in D \\ -(a + \alpha) < m \text{ independently of } D \end{cases}$$.

Let $\{x_k\}$ be generated by (4.2.2). We will pick C_k successively such that (4.1.3) is satisfied.

Now,

$$\|[F^D(x_k,\ x^*)]^{-1}\| \leqq q \equiv \frac{1}{a + \alpha + m}$$

Thus, as in Section 1.4,

$$\|x_k - x^*\| \leqq q\ \|F(x_k)\| \equiv s_k$$

Let

$$D_k = \bar{S}(x_k,\ s_k) \qquad ,$$

and let

$$m_k\ (M_k) = \inf_{\substack{1 \leqq i \leqq n \\ x,\ y \in D_k}} (\sup)\ \varphi^D_{ii}(x,\ y)$$

Now suppose $C_k = A + \sigma_k$ (with an ADI splitting, $C_k = [H + c\sigma_k] + [V + (1-c)\sigma_k],$) where $\sigma_k \geqq m$. Then

$$\|C_k^{-1}\ [C_k - F^D(x,\ y)]\| \leq \frac{1}{a + \alpha + \sigma_k}\ \|\sigma_k - \varphi^D(x,\ y)\|$$

Now, for $x,\ y \in D_k$,

$$\|\sigma_k - \varphi^D(x,\ y)\| \leqq \begin{cases} M_k - \sigma_k & \text{for } \sigma_k \leqq \dfrac{m_k + M_k}{2} \\ \sigma_k - m_k & \text{for } \sigma_k \geqq \dfrac{m_k + M_k}{2} \end{cases}$$

Now, $x_k,\ x^* \in D_k$. Thus (4.1.3) holds if

$$\frac{M_k - \sigma_k}{a + \alpha + \sigma_k} \leqq \delta < 1, \text{ and } \sigma_k \leqq \frac{m_k + M_k}{2} \qquad ,$$

or

$$\frac{\sigma_k - m_k}{a + \alpha + \sigma_k} \leqq \delta < 1, \text{ and } \sigma_k \geqq \frac{m_k + M_k}{2}$$

These conditions reduce to

$$(4.2.6) \qquad \frac{M_k - \delta(a + \alpha)}{1 + \delta} \leqq \sigma_k \leqq \frac{m_k + \delta(a + \alpha)}{1 - \delta} ,$$

where δ must satisfy

$$(4.2.7) \qquad \frac{M_k - m_k}{M_k + m_k + 2(a + \alpha)} \leqq \delta < 1$$

We now show that δ can be picked to satisfy (4.2.7). Suppose (4.2.6) and (4.2.7) are satisfied where δ is replaced by δ_k, and suppose (4.2.3) is satisfied where η is replaced by $\eta_k < (1-\delta_k)/(1+\delta_k)$. Then, if $\{x_k\}$ is given by (4.2.2),

$$\|x_{k+1} - x^*\| \leqq (\eta_k(1 + \delta_k) + \delta_k)\|x_k - x^*\| \quad \text{for } k \geqq 0$$

Hence,

$$\|x_{k+1} - x^*\| \leqq \|x_0 - x^*\| \leqq s_0 \quad \text{for } k \geqq 0$$

Thus, $\{x_k\} \subset \bar{S}(x^*, s_0) \subset \bar{S}(x_0, 2s_0) \equiv D$. Let

$$N = \max_{x \in D} \|F(x)\|$$

Then

$$\|x_k - x^*\| \leqq q \cdot N \equiv s ,$$

and

$$D_k \subset \bar{S}(x_k, s)$$

Thus

$$\bigcup_{k=0}^{\infty} D_k \subset \bar{S}(x_0, s + 2s_0) \equiv D^*$$

Now, let

$$m^* \; (M^*) = \inf_{\substack{1 \leqq i \leqq n \\ x,\, y \in D^*}} (\sup) \; \varphi_{ii}^D(x, y)$$

Then $m \leqq m^* \leqq m_k \leqq M_k \leqq M^*$, and so,

$$\frac{M_k - m_k}{M_k + m_k + 2(a + \alpha)} \leqq \frac{M^* - m^*}{M^* + m^* + 2(a + \alpha)} < 1$$

Thus, if δ satisfies

$$\frac{M^* - m^*}{M^* + m^* + 2(a + \alpha)} \leqq \delta < 1, \tag{4.2.8}$$

(4.2.7) will be satisfied. Now, if we try to minimize $\|C_k^{-1}[C_k - F^D(x_k,x^*)]\|$,

we are led to

$$\min \left[\min_{m_k \leqq \sigma_k \leqq \frac{1}{2}(M_k+m_k)} \frac{M_k - \sigma_k}{a + \alpha + \sigma_k}, \min_{\frac{1}{2}(M_k+m_k) \leqq \sigma_k} \frac{\sigma_k - m_k}{a + \alpha + \sigma_k} \right]$$

$$= \frac{M_k - m_k}{M_k + m_k + 2(a + \alpha)} ,$$

and the minimum is achieved at

$$\sigma_k = \tfrac{1}{2}(M_k + m_k)$$

Thus, if M_k and m_k change much during the iteration, we might wish to consider a variable σ_k. If we pick $\sigma_k = \frac{1}{2}(M_k + m_k)$, then (4.2.6) is satisfied for any δ satisfying (4.2.8)

If we wish to employ a single Picard parameter throughout the iteration, it is sufficient to choose σ such that

$$\frac{M^* - (a + \alpha)}{2} < \sigma ,$$

for then (4.2.6) will be satisfied (for $\sigma_k = \sigma$) for some $\delta < 1$. In particular, we may pick, as Douglas did, $\sigma = \frac{1}{2}(M^* + m^*)$.

Under more restrictive conditions than in Theorem 4.2.1, we can obtain convergence of (4.2.2) without requiring a condition like (4.2.3). The following result is essentially Theorem 1 of [12].

4.2.5 Theorem: Let $F \in M(R^n)$ have the root, x^*. Suppose

(4.2.9) $\quad F^D(x, y) - C^*$ is positive semi-definite for $x, y \in R^n$,

where C^* is positive definite. Suppose

(4.2.10) $\quad C - F^D(x, y)$ is positive semi-definite for $x, y \in R^n$,

where $C = H_1 + V_1$ and $H_1, V_1 \in L(R^n, R^n)$ are positive definite and commute. For $r > 0$, let T_r and Q_r be defined by (2.1.4). Let $x_0 \in R^n$, and suppose $\{x_k\} \subset R^n$ satisfies

$$x_{k+1} = \prod_{i=0}^{\nu-1} T_{r_i} x_k + \sum_{j=0}^{\nu-1} \left(\prod_{i=j+1}^{\nu-1} T_{r_i} \right) Q_{r_j} \, [Cx_k - F(x_k)] ,$$

where $r_i > 0$ for $0 \leq i \leq \nu - 1$. Then $x_k \to x^*$.

4.2.6 Remark: Condition (4.2.3) is not necessary in Theorem 4.2.5 in that $\nu \geq 1$ can be chosen arbitrarily.

In terms of Example 4.2.3, assumption (4.2.9) and (4.2.10) imply (4.2.5). Furthermore, if $\sigma \geq M$, (4.2.9) and (4.2.10) are satisfied.

4.2.7 Remark: If F is differentiable, then, by (4.1.4), conditions (4.2.9) and (4.2.10) will be satisfied if $F^D(x, y)$ is replaced by $F'(x)$.

The proof of Theorem 4.2.5 depends heavily on the commutivity of H_1 and V_1. However, even in the non-commutative case, if $r_{k,i} \equiv r_k$ is constant within each ADI cycle, convergence independent of ν_k, the number of ADI iterations at the k^{th} stage, can be obtained.

4.2.8 Theorem: Let $F \in M(R^n)$ have the root, x^*. Let $\{C_k\} \subset L(R^n, R^n)$ satisfy (4.2.1). Let $\{r_k\}$ be a sequence of positive parameters and $\{\nu_k\}$ a sequence of positive integers. Let $x_0 \in R^n$, and suppose $\{x_k\} \subset R^n$

satisfies

$$x_{k+1} = (T_{k,r_k})^{\nu_k} x_k + \sum_{j=0}^{\nu_k-1} (T_{k,r_k})^j Q_{k,r_k} [C_k x_k - F(x_k)]$$

Suppose there is a $\delta < 1$ and independent of k such that

(4.2.11) $\|C_k - F^D(x,y)\| \leqq \delta(a + \alpha)$ for $x, y \in [x^*, x_k]$ and $k \geqq 0$

Suppose

(4.2.12) $$\max\left(\sqrt{a_k \cdot b_k}, \frac{\alpha_k + \beta_k}{2}\right) \leqq r_k \leqq s < \infty$$

Then $x_k \rightarrow x^*$.

<u>Proof:</u> We have

$$x_{k+1} - x^* = E_k(x_k)(x_k - x^*) \qquad ,$$

where

$$E_k(x) = (T_{k,r})^{\nu_k} + \sum_{j=0}^{\nu_k-1} (T_{k,r})^j Q_{k,r} [C_k - F^D(x,x^*)]$$

It is sufficient to show that $\|E_k(x_k)\| \leqq \delta^*$ for $x \in [x^*, x_k]$ and for some $\delta^* < 1$ and independent of k. Let

$$\eta_k = (r_k - a_k)(r_k - \alpha_k) \text{ and } \zeta_k = (r_k + a_k)(r_k + \alpha_k)$$

Then, by (4.2.11), (4.2.12), and (2.1.3), for $x \in [x^*, x_k]$,

$$\|E_k(x)\| \leqq \left(\frac{\eta_k}{\zeta_k}\right)^{\nu_k} + 2r_k \sum_{j=0}^{\nu_k-1} \left(\frac{\eta_k}{\zeta_k}\right)^j \frac{\delta\,(a_k + \alpha_k)}{\zeta_k}$$

$$= \left(\frac{\eta_k}{\zeta_k}\right)^{\nu_k} + \left[1 - \left(\frac{\eta_k}{\zeta_k}\right)^{\nu_k}\right] \frac{2r_k\,(a_k + \alpha_k)\,\delta}{\zeta_k - \eta_k}$$

Now, $\zeta_k - \eta_k = 2r_k\,(a_k + \alpha_k)$. Hence,

$$\|E_k(x)\| \leqq \left(\frac{\eta_k}{\zeta_k}\right)^{\nu_k} (1 - \delta) + \delta$$

Let

$$\eta = (r - a^*)(r - \alpha^*) \text{ and } \zeta = (r + a^*)(r + \alpha^*)$$

Then

$$\left(\frac{\eta_k}{\zeta_k}\right)^{\nu_k} \leqq \frac{\eta_k}{\zeta_k} \leqq \frac{\eta}{\zeta} < 1 \qquad ,$$

and so,

$$\|E_k(x)\| \leqq \frac{\eta}{\zeta}(1 - \delta) + \delta \equiv \delta^* < 1$$

This completes the proof.

4.2.8 Remark: In Theorem 4.2.8, if $H_k V_k = V_k H_k$ for $k \geqq 0$, (4.2.12) can be weakened to

$$\max(\sqrt{a_k \cdot b_k}, \sqrt{\alpha_k \cdot \beta_k}) \leqq r_k \leqq s < \infty$$

In terms of Example 4.2.4, $H_k V_k = V_k H_k$ if $HV = VH$

. If $r_k \equiv r$ and $V_k \equiv V_1$ are constant, we see, by the similarity transformation $z_k = (r + V_1)x_k$, that (4.2.12) can be weakened to

(4.2.13) $$\max\left(\max_k \sqrt{a_k \cdot b_k}, \sqrt{\alpha_1 \cdot \beta_1}\right) \leqq r$$

In terms of Example 4.2.4, $V_k \equiv V_1$ if $\sigma_k \equiv \sigma$ or if $c = 1$. If $\sigma_k \equiv \sigma$, then $H_k \equiv H_1$ and (4.2.13) can be easily satisfied. If σ_k is chosen as in Example 4.2.4, then $\sigma_k \leqq M^*$, and so $a_k \leqq a + \sigma_k \leqq a + M^*$, and, likewise, $b \leqq b + M^*$. Thus, (4.2.13) can be satisfied a priori.

Suppose F is as in Example 4.2.3 and is differentiable. Consider the Newton and secant methods of Example 4.1.7. In these cases, (4.2.11) becomes, effectively,

$$\|\varphi'(x) - \varphi'(y)\| \leqq \delta\,(a + \alpha + m)$$

for x and y in some set containing $\bigcup_{k=0}^{\infty} [x^*, x_k]$. This is a very severe restriction and holds only if x_0 is close enough to x^*. However, Theroem 4.2.8 may still be usable.

4.2.9 Example: Consider the discrete boundary value problem,

$$(4.2.14) \qquad \begin{cases} \Delta_h u(P) = e^{u(P)} & , \quad P \in \Omega \\ u(P) = v(P) \quad \gamma s + \zeta t & , \quad P = (s,t) \in \Omega' \end{cases} ,$$

where Δ_h and Ω are given in Example 1.2.5. Let u^* be the solution of (4.2.14). We have seen that

$$-\frac{1}{8} \leqq u^*(P) \leqq \gamma + \zeta$$

Thus, u^* is the unique solution of

$$\begin{cases} \Delta_h u(P) = f(u(P)) & , \ P \in \Omega \\ u(P) = v(P) & , \ P \in \Omega' \end{cases} ,$$

where

$$f(u) = \begin{cases} e^{-\frac{1}{8}} + e^{-\frac{1}{8}}(u + \frac{1}{8}) & , \quad u \leqq -\frac{1}{8} \\ e^u & , \quad -\frac{1}{8} \leqq u \leqq \gamma + \zeta \\ e^{\gamma+\zeta} + e^{\gamma+\zeta}(u - (\gamma + \zeta)) & , \quad \gamma + \zeta \leqq u \end{cases}$$

Let F be given by (1.3.3) where $\ell = -\Delta_h$. Then the root, x^*, of F is $x^* = \underline{u}^*$. Furthermore,

$$\max_{x,y \in R^n} \|F'(x) - F'(y)\| \leqq \max_{-1/8 \leqq s,t \leqq \gamma+\zeta} |e^s - e^t| = e^{\gamma+\zeta} - e^{-\frac{1}{8}} ,$$

and

$$m \geqq e^{-\frac{1}{8}} .$$

Now, if A_ℓ is split as in (1.2.40), then $a = \alpha \approx \pi^2$. Hence, (4.2.11) is satisfied (for small enough h) if

$$e^{\gamma+\zeta} - e^{-1/8} < 2\pi^2 + e^{-1/8}$$

Now, $\log(2\pi^2 + 2e^{-1/8}) > 3.06$. Hence, (4.2.11) is satisfied (for small enough h) if

$$\gamma + \zeta < 3.06$$

4.3 Single Step Two Level Iterations. In this section, we consider two level iterations in which the inner iteration is carried only one step. The procedure may be either a Picard type iteration coupled with an inner ADI iteration or a nonlinear ADI iteration coupled with an inner Picard iteration.

Suppose F satisfies (4.2.4). For convenience, we will usually assume (4.2.5), though this can be weakened. If we attempt to solve the equation, $F(x) = 0$, by coupling a Picard iteration with an inner ADI iteration, we obtain the Picard-ADI iteration.

$$(4.3.1)\quad \begin{cases} [r_k + H + c\sigma_k]\, x_{k+\frac{1}{2}} = [r_k - V - (1-c)\sigma_k]\, x_k + \sigma_k x_k - \varphi(x_k) \\ [r_k + V + (1-c)\sigma_k]\, x_{k+1} = [r_k - H - \sigma_k]\, x_{k+\frac{1}{2}} + \sigma_k x_k - \varphi(x_k), \end{cases}$$

where c is an appropriate scalar, r, is the ADI parameter, σ_k is the Picard parameter, and the Picard matrix, $C_k = A + \sigma_k$, has the ADI splitting, $C_k = [H + c\sigma_k] + [V + (1-c)\sigma_k]$.

Alternately, we may couple a nonlinear ADI iteration with an inner Picard iteration and obtain the ADI-Picard iteration.

$$(4.3.2)\quad \begin{cases} [r_k + H + c\sigma_k]\, x_{k+\frac{1}{2}} = [r_k - V - (1-c)\sigma_k]\, x_k + \sigma_k x_k + \varphi(x_k) \\ [r_k + V + (1-c)\sigma_k]\, x_{k+1} = [r_k - H - c\sigma_k \;\; x_{k+\frac{1}{2}}] + \sigma_k x_{k+\frac{1}{2}} + \varphi(x_{k+\frac{1}{2}})\,. \end{cases}$$

In (4.3.1) and (4.3.2), r_k and σ_k could be updated to $r_{k+\frac{1}{2}}$ and $\sigma_{k+\frac{1}{2}}$ in the second equation, but we will not consider this.

From Theorem 4.2.8 and the estimates of Example 4.2.4, we can derive some conditions under which (4.3.1) will converge. By a straightforward proof, it can be shown that (4.3.2) converges under practically the same conditions.

Condition (4.2.12) of Theorem 4.2.8 is rather stringent since we would like to take the ADI parameters smaller than allowed. If $\sigma_k \equiv \sigma$ is held constant, we can relax (4.2.12) and obtain a result analogous to Pearcy's result.

4.3.1 Theorem: Let F satisfy (4.2.4) where (4.2.5) holds. Suppose

$$\frac{M + m}{2} = \sigma \geqq 0 \tag{4.3.3}$$

Let $0 \leqq c \leqq 1$, and set

$$\begin{cases} a_1 = a + c\sigma , & b_1 = b + c\sigma \\ \alpha_1 = \alpha + (1-c)\sigma, & \beta_1 = \beta + (1-c)\sigma \end{cases}$$

Then let

$$r = \max \left(\sqrt{a_1 \cdot b_1} , \sqrt{\alpha_1 \cdot \beta_1} \right) \tag{4.3.4}$$

Let

$$\delta = \frac{M - m}{2(a_1 + \alpha_1)} = \frac{M - m}{M + m + 2(a + \alpha)} < 1 \tag{4.3.5}$$

Now, set

$$q = \frac{s - a_1}{s + a_1} \cdot \frac{s - \alpha_1}{s + \alpha_1} + \frac{2s\delta\,(a_1 + \alpha_1)}{(s + a_1)(r + \alpha_1)} ,$$

and

$$\delta^* = 1 - (1 - \delta)\,\frac{2r\,(a_1 + \alpha_1)}{(s + a_1)(r + \alpha_1)} < 1$$

Suppose $\{x_{k/2}\}$ satisfies (4.3.1) where $\sigma_k \equiv \sigma$, and $\{r_k\}$ satisfies

$$\begin{cases} r \leqq r_{\nu-1} \leqq r_{\nu-2} \leqq \cdots \leqq r_0 \leqq s \\ r_{i\nu+j} = r_j \quad \text{for } 0 \leqq j \leqq \nu - 1 \text{ and } i \geqq 0 \end{cases} \tag{4.3.6} ,$$

where

$$\nu > 1 + \left| \frac{\log q}{\log \delta^*} \right| \tag{4.3.7}$$

Then $x_k \to x^*$.

<u>Proof</u>: Let $H_1 = H + c\sigma$ and $V_1 = V + (1-c)\sigma$. Let

$$e_{k/2} = x_{k/2} - x^* \quad \text{and} \quad \Delta_k = \sigma - \varphi^D(x_k, x^*) \text{ for } k \geqq 0$$

Then, by (4.3.3),

$$\|\Delta_k\| \leqq \frac{M - m}{2} = \delta\,(a_1 + \alpha_1)$$

Now,

$$\begin{cases} [r_k + H_1]\, e_{k+\frac{1}{2}} = [r_k - V_1]\, e_k + \Delta_k e_k \\ [r_k + V_1]\, e_{k+1} = [r_k - H_1]\, e_{k+\frac{1}{2}} + \Delta_k e_k \end{cases}$$

Let $e_k^* = [r_{k-1} + V_1]\, e_k$ for $k \geqq 0$, where $r_{-1} \equiv r_{\nu-1}$. Then,

$$e_{(i+1)\nu} = \left(\prod_{j=0}^{\nu-1} E_{i,j} \right) e_{i\nu} \qquad ,$$

where,

$$E_{i,j} = [r_j - H_1]\,[r_j + H_1]^{-1}\,[r_j - V_1]\,[r_{j-1} + V_1]^{-1} + 2r_j\,[r_j + H_1]^{-1}\,\Delta_{i\nu+j}\,[r_{j-1} + V_1]^{-1}$$

Now, by (4.3.4) and (4.3.6),

$$\|[r_0 - V_1]\,[r_{\nu-1} + V_1]^{-1}\| \leqq \frac{s - \alpha_1}{r + \alpha_1}$$

Thus, for $i \geqq 0$

$$\|E_{i,0}\| \leqq q_0 \equiv \frac{r_0 - a_1}{r_0 + a_1}\,\frac{s - \alpha_1}{r + \alpha_1} + \frac{2r_0\delta\,(a_1 + \alpha_1)}{(r_0 + a_1)(r_{\nu-1} + \alpha_1)}$$

$$\leqq \frac{s - a_1}{s + a_1}\,\frac{s - \alpha_1}{r + \alpha_1} + \frac{2s\delta\,(a_1 + \alpha_1)}{(s + a_1)(r + \alpha_1)}$$

$$= q$$

Now, for $1 \leqq j \leqq \nu - 1$, by (4.3.4) and (4.3.6),

$$\|[r_j - V_1][r_{j-1} + V_1]^{-1}\| \leqq \frac{r_j - \alpha_1}{r_j + \alpha_1}$$

and

$$\|[r_{j-1} + V_1]^{-1}\| \leqq \frac{1}{r_{j-1} + \alpha_1} \leqq \frac{1}{r_j + \alpha_1}$$

Thus, for $1 \leqq j \leqq \nu - 1$, and $i \geqq 0$,

$$\|E_{i,j}\| \leqq \frac{r_j - a_1}{r_j + a_1} \frac{r_j - \alpha_1}{r_j + \alpha_1} + \frac{2r_j\delta\,(a_1 + \alpha_1)}{(r_j + a_1)(r_j + \alpha_1)}$$

$$= 1 - (1 - \delta)\frac{2r_j\,(a_1 + \alpha_1)}{(r_j + a_1)(r_j + \alpha_1)}$$

$$\delta_j \leqq 1 - (1 - \delta)\frac{2r\,(a_1 + \alpha_1)}{(s + a_1)(r + \alpha_1)}$$

$$= \delta*$$

Thus, for $i \geqq 0$,

$$\|e^*_{(i+1)\nu}\| \leqq q\,(\delta*)^{\nu-1}\,\|e^*_{i\nu}\|$$

But, by (4.3.7),

$$q\,(\delta*)^{\nu-1} < 1 \qquad .$$

Hence, $e_{i\nu} \to 0$ $(i \to \infty)$. But, clearly, there is a $K < \infty$ and independent of i and j, such that

$$\|e^*_{i\nu+j}\| \leqq K\,\|e^*_{i\nu}\|$$

Hence, $e^*_k \to 0$, and so $e_k \to 0$. This completes the proof.

<u>4.3.2 Remark</u>: In Theorem (4.3.1, (4.3.7) can be replaced with

(4.3.8) $$q_0 \prod_{j=1}^{\nu-1} \delta_j < 1 \qquad ,$$

where q_0 and δ_j are as defined in the proof of Theorem 4.3.1.

4.3.3 Remark: In Theorem 4.3.1, (4.3.6) and (4.3.7) can be replaced with

(4.3.9) $$r \leqq r_{k+1} \leqq r_k \leqq s \quad \text{for } k \geqq 0$$

4.3.4 Corollary: Let the conditions of Theorem 4.3.1 be satisfied except redefine

$$q = \frac{(s - a_1)(s - \alpha_1)}{(s + a_1)(r + \alpha_1)} + \frac{\delta\,(a_1 + \alpha_1)(s - \alpha_1)}{(r + a_1)(r + \alpha_1)} + \frac{\delta\,(a_1 + \alpha_1)(s - a_1)}{(s + a_1)(r + \alpha_1)} + \frac{\delta^2\,(a_1 + \alpha_1)^2}{(r + a_1)(r + \alpha_1)}$$

and

$$\delta^* = \left[1 - (1 - \delta)\,\frac{a_1 + \alpha_1}{r + \alpha_1}\right]\left[1 - (1 - \delta)\,\frac{a_1 + \alpha_1}{r + a_1}\right] < 1 \quad ,$$

and let $\{x_{k/2}\}$ satisfy (4.3.2). Then $x_k \to x^*$.

Proof: The proof follows along the lines of that of Theorem 4.3.1. We now have

$$\begin{cases} [r_k + H_1]\, e_{k+\frac{1}{2}} = [r_k - V_1]\, e_k + \Delta_k e_k \\ [r_k + V_1]\, e_{k+1} = [r_k - H_1]\, e_{k+\frac{1}{2}} + \Delta_{k+\frac{1}{2}} e_{k+\frac{1}{2}} \end{cases}$$

Thus,

$$e^*_{(i+1)\nu} = \left(\prod_{j=0}^{\nu-1} E_{i,j}\right) e^*_{i\nu} \quad ,$$

where

$$\begin{aligned} E_{i,j} = \; & [r_j - H_1]\,[r_j + H_1]^{-1}\,[r_j - V_1]\,[r_{j-1} + V_1]^{-1} \\ & + \Delta_{i\nu+j+\frac{1}{2}}\,[r_j + H_1]^{-1}\,[r_j - V_1]\,[r_{j-1} + V_1]^{-1} \\ & + [r_j - H_1]\,[r_j + H_1]^{-1}\,\Delta_{i\nu+j}\,[r_{j-1} + V_1]^{-1} \\ & + \Delta_{i\nu+j+\frac{1}{2}}\,[r_j + H_1]^{-1}\,\Delta_{i\nu+j}\,[r_{j-1} + V_1]^{-1} \quad . \end{aligned}$$

Now, as in the proof of Theorem 4.3.1,

$$\|E_{i,0}\| \leq q_0 \equiv \frac{(r_0 - a_1)(s - \alpha_1)}{(r_0 + a_1)(s + \alpha_1)} + \frac{\delta\,(a_1 + \alpha_1)(s - \alpha_1)}{(r_0 + a_1)(r + \alpha_1)} + \frac{\delta\,(a_1 + \alpha_1)(r_0 - a_1)}{(r_{\nu-1} + \alpha_1)(r_0 + a_1)} + \frac{\delta^2\,(a_1 + \alpha_1)^2}{(r_0 + a_1)(r_{\nu-1} + \alpha_1)}$$

$$\leq q \qquad ,$$

and, for $1 \leq j \leq \nu - 1$ and $i \geq 0$,

$$\|E_{i,j}\| \leq \frac{(r_j - a_1)(r_j - \alpha_1)}{(r_j + a_1)(r_j + \alpha_1)} + \frac{\delta\,(a_1 + \alpha_1)(r_j - \alpha_1)}{(r_j + a_1)(r_j + \alpha_1)} + \frac{\delta\,(a_1 + \alpha_1)(r_j - a_1)}{(r_j + \alpha_1)(r_j + a_1)} + \frac{\delta^2\,(a_1 + \alpha_1)^2}{(r_j + a_1)(r_j + \alpha_1)}$$

$$= \left[1 - (1 - \delta)\frac{a_1 + \alpha_1}{r_j + \alpha_1}\right]\left[1 - (1 - \delta)\frac{a_1 + \alpha_1}{r_j + a_1}\right]$$

$$\equiv \delta_j \leq \delta^* \qquad .$$

So $\|e^*_{(i+1)\nu}\| \leq q\,(\delta^*)^{\nu-1}\,\|e^*_{i\nu}\|$, and the result follows as in the proof of Theorem 4.3.1.

<u>4.3.5 Remark</u>: In Corollary 4.3.4, (4.3.7) can be replaced with (4.3.8), where q_0 and δ_j are defined as in the proof of Theorem 4.3.4

<u>4.3.6 Remark</u>: Remark 4.3.3 holds for Corollary 4.3.4 also.

Let $\sigma_k \equiv \sigma$ be constant, and set $c = \frac{1}{2}$. The (4.3.2) becomes

$$(4.3.10) \qquad \begin{cases} [s_k + H]\, x_{k+\frac{1}{2}} = [s_k - V]\, x_k - \varphi(x_k) \\ [s_k + V]\, x_{k+1} = [s_k - H]\, x_{k+\frac{1}{2}} - \varphi(x_{k+\frac{1}{2}}) \end{cases} ,$$

where $s_k = r_k + \frac{1}{2}\sigma$. We now establish a result for (4.3.10) which is sharp in the linear case.

<u>4.3.7 Theorem:</u> Let F satisfy (4.2.4), where (4.2.5) holds, and, in addition

$$- 2 \min (a, \alpha) < m$$

Let

$$K = \frac{M - m}{4} \max \left[\frac{b + \frac{1}{2}M}{a + \frac{1}{2}m} , \frac{\beta + \frac{1}{2}M}{\alpha + \frac{1}{2}m} \right] ,$$

and suppose

(4.3.11) $$\frac{M + m}{4} + K < s \leqq s_{k+1} \leqq s_k \quad \text{for } k \geqq 0 .$$

Suppose $\{x_{k/2}\}$ satisfies (4.3.10). Then $x_k \to x^*$.

<u>Proof:</u> Let $e_k = x_k - x^*$. Then

$$e_{k+1} = [s_k + V]^{-1} \prod_{j=0}^{k} (E_{H,j} E_{V,j}) e_0^* ,$$

where

$$e_0^* = [s_0 - H - \varphi^D(x_{\frac{1}{2}}, x^*)] [s_0 + H]^{-1} [s_0 - V - \varphi^D(x_0, x^*)] e_0 ,$$

and

$$\begin{cases} E_{H,j} = [s_j - H - \varphi^D(x_{j+\frac{1}{2}}, x^*)] [s_j + H]^{-1} \\ E_{V,j} = [s_j - V - \varphi^D(x_j, x^*)] [s_{j-1} + V]^{-1} \end{cases} .$$

Now, $\|[s_k + V]^{-1}\|$ is uniformly bounded. Thus, it is sufficient to show that there are $\delta_H, \delta_V < 1$ and independent of k, such that

$$\|E_{H,k}\| \leqq \delta_H \quad , \quad \|E_{V,k}\| \leqq \delta_V \quad \text{for } k \geqq 0$$

Now, since V is positive definite and $s_{k-1} \geqq s_k$,

$$\|[s_{k-1} + V] x\|^2 \geqq \|[s_k + V] x\|^2 \quad \text{for } x \in R^n$$

Hence

$$\|E_{V,k}\|^2 \geqq \sup_{x \neq 0} \frac{\|[s_k - V - \varphi^D(x_k, x^*)]\, x\|^2}{\|[s_{k-1} + V]\, x\|^2}$$

$$\leqq \sup_{x \neq 0} \frac{\|[s_k - V - \varphi^D(x_k, x^*)]\, x\|^2}{\|[s_k + V]\, x\|^2}$$

$$= \sup_{x \neq 0} \frac{\|[R_k - V_k]\, x\|^2}{\|[R_k + V_k]\, x\|^2} ,$$

where

$$\begin{cases} R_k = s_k - \frac{1}{2}\varphi^D(x_k, x^*) \\ V_k = V + \frac{1}{2}\varphi^D(x_k, x^*) \end{cases}$$

Thus,

$$\text{(4.3.12)} \qquad \|E_{V,k}\|^2 \leqq \sup_{x \neq 0} \frac{\|R_k x\|^2 - 2(R_k x, V_k x) + \|V_k x\|^2}{\|R_k x\|^2 + 2(R_k x, V_k x) + \|V_k x\|^2}$$

Now, for $\|x\| = 1$,

$$\text{(4.3.13)} \qquad (\alpha + \tfrac{1}{2}m)^2 \leqq \|R_k x\|^2 + \|V_k x\|^2 \leqq (s_0 - \tfrac{1}{2}m)^2 + (\beta + \tfrac{1}{2}M)^2$$

Suppose $s_k = r_k + \frac{1}{2}\sigma$ where $\sigma = \frac{1}{2}(M + m)$. Then, for $\|x\| = 1$,

$$(R_k x, V_k x) = \tfrac{1}{2}([\sigma - \varphi^D(x_k, x^*)]\, x, V_k x) + (r_k x, V_k x)$$

$$\geqq -\tfrac{1}{2}\|\sigma - \varphi^D(x_k, x^*)\| \, \|V_k\| + r_k(\alpha + \tfrac{1}{2}m)$$

$$\geqq (\alpha + \tfrac{1}{2}m)\left(r_k - \frac{M + m}{4} \; \frac{\beta + \frac{1}{2}M}{\alpha + \frac{1}{2}m}\right)$$

$$\geqq (\alpha + \tfrac{1}{2}m)\,(s - K)$$

Thus, for $\|x\| = 1$,

$$\text{(4.3.14)} \qquad 0 < (\alpha + \tfrac{1}{2}m)(s - K) \leqq (R_k x, V_k x) \leqq (\beta + \tfrac{1}{2}M)(s_0 - \tfrac{1}{2}m)$$

Now, by (4.3.12)--(4.3.14), there is a $\delta_V < 1$ and independent of k such that $\|E_{V,k}\| \leqq \delta_V$ for $k \geqq 0$. A similar result holds for $E_{H,k}$ and the proof is complete.

4.3.8 Remark: If $\varphi(x) = \sigma x + \gamma$ for some $\gamma \in R^n$ and some $\sigma > 0$, then the linear theory says (4.3.10) converges to x* if

$$\tfrac{1}{2}\sigma < s \leqq s_{k+1} \leqq s_k \quad \text{for } k \geqq 0 \qquad .$$

But, this is exactly what (4.3.11) reduces to.

CHAPTER V

NEWTON-ADI ITERATIONS

In this chapter, we consider iterations (2.4.3) and (2.4.5) in the case $\nu = 1$. The Newton-ADI iteration takes the form,

$$x_{k+1} = x_k - 2r_k\,[r_k + V_1'(x_k)]^{-1}\,[r_k + H_1'(x_k)]^{-1}\,F(x_k) \qquad (5.0.1)$$

and the ADI-Newton iteration takes the form,

$$\begin{cases} x_{k+\frac{1}{2}} = x_k - [r_k + H_1'(x_k)]^{-1}\,F(x_k) \\ x_{k+1} = x_{k+\frac{1}{2}} - [r_{k+\frac{1}{2}} + V_1'(x_{k+\frac{1}{2}})]^{-1}\,F(x_{k+\frac{1}{2}}) \end{cases} \qquad (5.0.2)$$

The methods of Chapter IV can be used to obtain convergence results for (5.0.1) and (5.0.2) under the assumption that $F'(x)$ does not vary too much over a certain set. In this chapter, we will not need such an assumption.

In Section 5.1, we consider convergence results based on contraction principles, and in Section 5.2, we consider convergence results based on monotonic principles. In Section 5.3, we present a counter-example to a certain assertion about the Newton-ADI iteration. Finally, in Section 5.4, we present some numerical results.

5.1 Contractive Results. Suppose $F: D \subset R^n \to R^n$, D convex, satisfies the following conditions uniformly on D.

$$\begin{cases} F \in C^2(D), \quad \|F(x)\| \leqq N, \quad \|F''(x)\| \leqq M, \\ F'(x) \text{ is positive definite}, \quad \sigma[F'(x)] \subset [\mu,\infty),\ \mu > 0 \end{cases},$$

where, in this chapter, $\|\cdot\| = \|\cdot\|_2$. These conditions are not enough to insure that the Newton iteration function,

$$N_0(x) = x - [F'(x)]^{-1}\,F(x) ,$$

is a contraction on D. However, we can apply a parameter in a way similar to the way in which the ADI parameter is applied and get a Newton type iteration function which is contractive on D.

Let $A: R^n \to L(R^n, R^n)$ be differentiable. Then, it can be shown that, if $A(x)$ is nonsingular, then $\frac{d}{dx}[A^{-1}(x)]$ exists and, for $\xi \in R^n$,

$$\frac{d}{dx}[A^{-1}(x)]\,\xi = -A^{-1}(x)\,A'(x)\,\xi\,A^{-1}(x)$$

Now, for $r > 0$, let

$$N_r(x) = x - [r + F'(x)]^{-1} F(x)$$

Then, for $\xi \in R^n$,

$$N_r'(x)\,\xi = [I - [r + F'(x)]^{-1} F'(x)]\,\xi$$

$$+ [r + F'(x)]^{-1} F''(x)\,\xi\,[r + F'(x)]^{-1} F(x)$$

The first term equals $r\,[r + F'(x)]^{-1}\,\xi$. Thus, on D,

$$\|N_r'(x)\| \leqq \frac{r}{r + \mu} + \frac{MN}{(r + \mu)^2}$$

$$= 1 - \frac{1}{(r + \mu)^2}\,(\mu(r + \mu) - MN)$$

Hence, if $\epsilon > 0$, and

$$r \geqq \frac{\epsilon + MN}{\mu} - \mu ,$$

then, on D,

$$\|N_r'(x)\| \leqq 1 - \frac{\epsilon}{(r + \mu)^2} < 1$$

Thus, since D is convex, by (1.3.7), N_r is a uniform contraction on D.

By a similar analysis, we can determine conditions on the ADI parameter which will guarantee that the Newton-ADI iteration function is a contraction on a given convex set.

Suppose $F, H_1, V_1: D \subset R^n \to R^n$ satisfy

$$(5.1.1)\qquad \begin{cases} \text{a) } F = H_1 + V_1 \\ \text{b) } H_1,\ V_1 \in C^2(D) \\ \text{c) } H_1'(x) \text{ and } V_1'(x) \text{ are positive semi-definite on } D \end{cases}$$

Then, for $r > 0$, we can define the following Newton-ADI iteration function.

$$(5.1.2)\qquad h_r(x) = x - 2r\,[r + V_1'(x)]^{-1}\,[r + H_1'(x)]^{-1}\,F(x)$$

For simplicity, we shall say

$$(5.1.3)\qquad (a_D,\ b_D,\ \alpha_D,\ \beta_D,\ M_D^H,\ M_D^V,\ N_D) \in \text{Bound }(D) \quad ,$$

if the following hold uniformly on D.

$$\begin{cases} \sigma[H_1'(x)] \subset [a_D,\ b_D],\ \sigma[V_1'(x)] \subset [\alpha_D,\ \beta_D] \\ a_D,\ \alpha_D \geqq 0,\ a_D + \alpha_D > 0 \\ \|F(x)\| \leqq N_D,\ \|H_1''(x)\| \leqq M_D^H\ ,\ \|V_1''(x)\| \leqq M_D^V \end{cases}$$

<u>5.1.1 Lemma</u>: Let $F,\ H_1,\ V_1\colon D \subset R^n \to R^n$ satisfy (5.1.1), and let $h_r\colon D \to R^n$ be defined by (5.1.2). Suppose (5.1.3) holds. Let $0 < K < 1$ and $\epsilon \geqq 0$, and suppose

$$(5.1.4)\qquad r \geqq \max\left[\frac{\epsilon + M_D^H N_D}{K(a_D + \alpha_D)} - a_D,\ \frac{\epsilon + M_D^V N_D}{(1-K)(a_D + \alpha_D)} - \alpha_D\right] \quad ,$$

and, in addition,

$$(5.1.5)\qquad r \geqq \max\left[\sqrt{a_D \cdot b_D}\ ,\frac{\alpha_D + \beta_D}{2}\right] \quad .$$

Then, on D,

$$\|h_r'(x)\| \leqq 1 - \frac{2r\epsilon}{(r + a_D)(r + \alpha_D)}\left[\frac{1}{r + a_D} + \frac{1}{r + \alpha_D}\right] \quad .$$

<u>Proof</u>: Let

$$T_r(x) = [r + V_1'(x)]^{-1}\,[r - H_1'(x)]\,[r + H_1'(x)]^{-1}\,[r - V_1'(x)]$$

Then,

$$I - 2r\,[r + V_1'(x)]^{-1}\,[r + H_1'(x)]^{-1}\,F'(x) = T_r(x)$$

So, for $\xi \in R^n$,

$$\begin{aligned} h_r'(x)\,\xi = {} & T_r(x)\,\xi \\ & + 2r\,[r + V_1'(x)]^{-1}\,V_1''(x)\,\xi\,[r + V_1'(x)]^{-1}\,[r + H_1'(x)]^{-1}\,F(x) \\ & + 2r\,[r + V_1'(x)]^{-1}\,[r + H_1'(x)]^{-1}\,H_1''(x)\,\xi\,[r + H_1'(x)]^{-1}\,F(x) \;. \end{aligned}$$

So, by (2.1.3), (5.1.3), and (5.1.5), on D,

$$\|h_r'(x)\| \leqq \frac{(r - a_D)(r - \alpha_D)}{(r + a_D)(r + \alpha_D)} + \frac{2r}{(r + a_D)(r + \alpha_D)}\left[\frac{M_D^V\,N_D}{r + \alpha_D} + \frac{M_D^H\,N_D}{r + a_D}\right].$$

But,

$$(r - a_D)(r - \alpha_D) = (r + a_D)(r + \alpha_D) - 2r\,(a_D + \alpha_D)$$

Hence,

$$\begin{aligned} \|h_r'(x)\| & \leqq 1 - \frac{2r}{(r + a_D)(r + \alpha_D)}\left[(a_D + \alpha_D) - \frac{M_D^V\,N_D}{r + \alpha_D} - \frac{M_D^H\,N_D}{r + \alpha_D}\right] \\ & \leqq 1 - \frac{2r}{(r + a_D)(r + \alpha_D)}\left[\frac{K\,(a_D + \alpha_D)(r + \alpha_D) - M_D^V\,N_D}{r + \alpha_D} \right. \\ & \qquad \left. + \frac{(1-K)(a\;\; + \alpha_D)(r + a_D) - M_D^H\,N_D}{r + a_D}\right]. \end{aligned}$$

The result now follows from (5.1.4).

<u>5.1.2 Remark</u>: In order to make the restriction, (5.1.4), as weak as possible, we would pick K such that

$$\frac{\epsilon + M_D^H\,N_D}{K\,(a_D + \alpha_D)} - a_D = \frac{\epsilon + M_D^V\,N_D}{(1-K)(a_D + \alpha_D)} - \alpha_D \qquad ,$$

provided that this holds for some $0 < K < 1$.

<u>5.1.3 Iterative Procedure</u>: Let $F,\ H_1,\ V_1\colon D \subset R^n \to R^n$ satisfy (5.1.1),

and suppose (5.1.3) holds Let x_0, $x^* \in R^n$, where $F(x^*) = 0$, and suppose $\bar{S}(x^*, \|x_0 - x^*\|) \subset D$. Let $0 < K < 1$ and $\epsilon > 0$. Define x_k, for $k \geqq 1$, successively as follows. If x_k has been defined and $[x^*, x_k] \subset D$, determine a_k, b_k, etc., such that

$$(5.1.6)\quad \begin{cases} (a_k, b_k, \alpha_k, \beta_k, M_k^H, M_k^V, N_k) \in \text{Bound}([x^*, x_k]) \\ a_k,\ \alpha_k \geqq a_D,\ \alpha_D \text{ respectively} \\ b_k,\ \beta_k,\ M_k^H,\ M_k^V,\ N_k \leqq b_D,\ \beta_D,\ M_D^H,\ M_D^V,\ N_D \text{ respectively} \end{cases}$$

Set

$$(5.1.7)\qquad r_k = \max\left[\sqrt{a_k \cdot b_k},\ \frac{\alpha_k + \beta_k}{2},\ \frac{\epsilon + M_k^H N_k}{K\,(a_k + \alpha_k)} - a_k,\ \frac{\epsilon + M_k^V N_k}{(1-K)(a_k + \alpha_k)} - \alpha_k\right],$$

and define

$$x_{k+1} = h_{r_k}(x_k),$$

where h_{r_k} is defined by (5.1.2).

<u>5.1.4 Theorem</u>: Consider Iterative Procedure 5.1.3. Then $[x^*, x_k] \subset D$ for $k \geqq 0$ and, hence, $\{x_k\}$ is well defined. Furthermore, $x_k \to x^*$.

<u>Proof</u>: Let

$$\begin{cases} r = \max\,(a_D, \alpha_D) > 0 \\ s = \max\left[b_D,\ \beta_D,\ \dfrac{\epsilon + M_D^H N_D}{K\,(a_D + \alpha_D)} - a_D,\ \dfrac{\epsilon + M_D^V N_D}{(1-K)(a_D + \alpha_D)} - \alpha_D\right] \\ \delta = 1 - \dfrac{2r\epsilon}{(s + b_D)(s + \beta_D)}\left[\dfrac{1}{r + b_D} + \dfrac{1}{r + \beta_D}\right] < 1 \end{cases}$$

Now, x_0 is defined, and $\|x_0 - x^*\| \leqq \delta^0 \|x_0 - x^*\|$. Suppose

$$(5.1.8)\qquad x_k \text{ is defined, and } \|x_k - x^*\| \leqq \delta^k \|x_0 - x^*\|.$$

Then, $[x^*, x_k] \subset D$ and so x_{k+1} is defined, and

$$\|x_{k+1} - x^*\| = \|h_{r_k}(x_k) - h_{r_k}(x^*)\|$$
$$\leq \max_{x \in [x^*, x_k]} \|h'_{r_k}(x)\| \; \|x_k - x^*\| \quad .$$

Now, since $[x^*, x_k] \subset D$, (5.1.9) is possible, and $r \leq r_k \leq s$. Thus, by Lemma 5.1.1,

$$\max_{x \in [x^*, x_k]} \|h'_{r_k}(x)\| \leq 1 - \frac{2r_k \epsilon}{(r_k + a_k)(r_k + \alpha_k)} \left[\frac{1}{r_k + a_k} + \frac{1}{r_k + \alpha_k} \right]$$
$$\leq \delta \quad .$$

Thus,

$$\|x_{k+1} - x^*\| \leq \delta \; \|x_k - x^*\| \leq \delta^{k+1} \; \|x_0 - x^*\| \quad ,$$

and so $[x^*, x_{k+1}] \subset D$. Thus (5.1.8) is established by induction for all $k \geq 0$. Hence, $x_k \to x^*$. This completes the proof.

5.1.5 Example: Suppose F satisfies (4.2.4) where $\varphi \in C^2(R^n)$ and $\varphi'(x) \geq 0$ on R^n. Set

$$H_1(x) = Hx + c\varphi(x) \quad \text{and} \quad V_1(x) = Vx + (1-c)\varphi(x) \quad ,$$

where $0 \leq c \leq 1$, and consider Iterative Procedure 5.1.4. Now,

$$\|[F'(x)]^{-1}\| \leq \frac{1}{a + \alpha}$$

So, from (1.4.5),

$$\|x - x^*\| \leq \frac{1}{a + \alpha} \|F(x)\| \quad \text{for } x \in R^n$$

Hence,

$$S(x^*, \|x^* - x_0\|) \subset S(x_0, 2\|x^* - x_0\|) \subset S\left(x_0, \frac{2}{a + \alpha} \|F(x_0)\|\right) \equiv D \quad ,$$

and

$$[x^*, x_k] \subset S\left(x_k, \frac{1}{a + \alpha} \|F(x_k)\|\right) \equiv D_k$$

Set

$$\begin{cases} a_D = a + c \min\limits_{\substack{1 \leq i \leq n \\ x \in D}} \varphi_i'(x_i) & , \quad b_D = b + c \max\limits_{\substack{1 \leq i \leq n \\ x \in D}} \varphi_i'(x_i) \\ \alpha_D = \alpha + (1-c) \min\limits_{\substack{1 \leq i \leq n \\ x \in D}} \varphi_i'(x_i) & , \quad \beta_D = \beta + (1-c) \max\limits_{\substack{1 \leq i \leq n \\ x \in D}} \varphi_i'(x_i) \\ M_D^H = c \max\limits_{\substack{1 \leq i \leq n \\ x \in D}} \varphi_i''(x_i) & , \quad M_D^V = (1-c) \max\limits_{\substack{1 \leq i \leq n \\ x \in D}} \varphi_i''(x_i) \\ N_D = \max\limits_{x \in D} \|F(x)\| & \end{cases} .$$

Then, (5.1.3) holds. Define a_k^*, b_k^*, etc., as above with respect to D_k instead of D. Then set

$$\begin{cases} a_k = \max(a_D,\ a_k^*), \quad \alpha_k = \max(\alpha_D,\ \alpha_k^*) \\ b_k = \min(b_D,\ b_D^*) \text{ and similarily for } \beta_k,\ M_k^H,\ M_k^V, \text{ and } N_k \end{cases}$$

Then, since $[x^*, x_k] \subset D$, (5.1.6) holds.

Now, since $x_k \to x^*$, $N_k \to 0$. Hence, (5.1.7) becomes, eventually,

$$r_k = \max\left[\sqrt{a_k \cdot b_k}\ ,\ \frac{\alpha_k + \beta_k}{2}\right] \tag{5.1.9}$$

We would like to be able to choose r_k smaller than allowed by (5.1.9). If c = 1 in Example 5.1.5, or, more generally, if $V_1'(x) \equiv V^*$ is constant, we may eventually do this.

<u>5.1.6 Lemma</u>: Let F, H_1, V_1: $D \subset R^n \to R^n$ satisfy (5.1.1), where $V_1'(x) \equiv V^*$ is constant. Suppose $(a_D, b_D, \alpha, \beta, M_D^H, 0, N_D) \in$ Bound (D). For r > 0, define g_r: $[r + V^*]^{-1}(D) \to R^n$ by

$$g_r(y) = [r + V^*]^{-1}\, h_r([r + V^*]^{-1} y) \tag{5.1.10}$$,

where h_r is defined by (5.1.2). Suppose $\epsilon \geq 0$ and let

$$r \geqq \frac{\epsilon + M_D^H N_D}{a_D + \alpha} - a_D \quad ,$$

and, in addition,

(5.1.11) $$r \geqq \max \, [\sqrt{a_D \cdot b_D}, \sqrt{\alpha \cdot \beta}\,]$$

Then, for $y \in D^* = [r + V^*](D)$,

$$\|g_r'(y)\| \leqq 1 - \frac{2r\epsilon}{(r + a_D)^2 (r + \alpha)}$$

Proof: We have

$$g_r'(y) = [r + V^*]\, h_r'([r + V^*]^{-1} y)\, [r + V^*]^{-1}$$

Thus, for $\xi \in R^n$,

$$g_r'(y)\,\xi = S_r(x)\,\xi + 2r\,[r + H_1'(x)]^{-1} H_1''(x)\,[r + V^*]^{-1} \xi\,[r + H_1'(x)]^{-1} F(x) \quad ,$$

where

$$S_r(x) = [r - H_1'(x)]\,[r + H_1'(x)]^{-1}\,[r - V^*]\,[r + V^*]^{-1} \quad ,$$

and

$$x = [r + V^*]^{-1} y \in D \quad .$$

The proof now follows from the estimates of Lemma 5.1.1 with (5.1.11) being sufficient instead of (5.1.5).

5.1.7 Iterative Procedure: Let $F, H_1, V_1 : D \subset R^n \to R^n$ satisfy (5.1.1) where $V_1'(x) \equiv V^*$ is constant. Suppose $(a_D, b_D, \alpha, \beta, M_D^H, 0, N_D) \in$ Bound (D). Let $r = \max\,(a_D, \sqrt{\alpha\beta}\,)$, and set $\eta = (r + \beta)/(r + \alpha) \geqq 1$. Let $x_0, x^* \in R^n$, where $F(x^*) = 0$, and suppose $\bar{S}(x^*, \eta\,\|x_0 - x^*\|) \subset D$. Let $\epsilon > 0$, and define x_k, for $k \geqq 1$, successively as follows. If x_k has been defined and $[x^*, x_k] \subset D$, we can determine a_k, b_k, etc., such that

(5.1.12) $$\begin{cases} (a_k, b_k, \alpha, \beta, M_k^H, 0, N_k) \in \text{Bound}([x^*, x_k]) \\ a_k \geqq a_D \\ b_k, M_k^H, N_k \leqq b_D, M_D^H, N_D \text{ respectively} \end{cases} \quad .$$

Assume we can pick r_k such that

$$(5.1.13) \qquad \max\left[\sqrt{a_k \cdot b_k}\ ,\ \sqrt{\alpha\cdot\beta}\ ,\ \frac{\epsilon + M_k^H N_k}{a_k + \alpha} - a_k\right] \leqq r_k \leqq r_{k-1} \quad .$$

Then define

$$x_{k+1} = h_{r_k}(x_k) \quad ,$$

where h_{r_k} is defined by (5.1.2).

5.1.8 Theorem: Consider Iterative Procedure 5.1.7. If (5.1.13) is satisfied at each stage, then $\{x_k\}$ is well defined (i.e., $[x^*, x_k] \subset D$ for $k \geqq 0$,) and $x_k \to x^*$.

Proof: Let

$$\begin{cases} s = \max\left[b_D,\ \beta,\ \dfrac{\epsilon + M_D^H N_D}{a_D + \alpha} - a_D\right] \\ \delta = 1 - \dfrac{2r\epsilon}{(s + b_D)^2(r + \alpha)} < 1 \end{cases}$$

Now, x_0 is defined and $[x^*, x_0] \subset D$. Suppose $x_0, \cdots, x_k$ are defined and satisfy $[x^*, x_j] \subset D$ for $0 \leqq j \leqq k$. Then, x_{k+1} is defined. For $0 \leqq j \leqq k + 1$, let

$$y_j = [r_j + V^*]\, x_j \quad \text{and} \quad y_j^* = [r_j + V^*]\, x^*$$

Then, for $0 \leqq j \leqq k$,

$$y_{j+1} - y_{j+1}^* = [r_{j+1} + V^*]\,[r_j + V^*]^{-1}\,[g_{r_j}(y_j) - g_{r_j}(y_j^*)] \quad ,$$

where g_{r_j} is defined by (5.1.10). Now, since $r_{j+1} \leqq r_j$,

$$\|[r_{j+1} + V^*]\,[r_j + V^*]^{-1}\| \leqq 1$$

Furthermore,

$$[y_j^*, y_j] = [r_j + V^*]\,[x^*, x_j] \subset [r_j + V^*](D) \quad ,$$

Thus, by Lemma 5.1.6 and (5.1.12),

$$\|y_{j+1} - y^*_{j+1}\| \leqq 1 \cdot \max_{y \in [y^*_k, y_k]} \|g'_{r_j}(y)\| \; \|y_j - y^*_j\|$$

$$\leqq \quad 1 - \frac{2r_j \epsilon}{(r_j + a_j)^2 (r_j + \alpha)} \quad \|y_j - y^*_j\|$$

Hence,

$$\|y_{k+1} - y^*_{k+1}\| \leqq \delta^{k+1} \; \|y_0 - y^*_0\|$$

But,

$$x_{k+1} - x^* = [r_{k+1} + V^*]^{-1} \; [y_{k+1} - y^*_{k+1}] \quad ,$$

and

$$y_0 - y^*_0 = [r_0 + V^*] \; (x_0 - x^*)$$

Thus,

$$\|x_{k+1} - x^*\| \leqq \|[r_{k+1} + V^*]^{-1}\| \; \delta^{k+1} \; \|r_0 + V^*\| \; \|x_0 - x^*\| \quad .$$

$$\leqq \eta \; \delta^{k+1} \; \|x_0 - x^*\| \quad .$$

Thus, since $\delta < 1$, $[x^*, x_{k+1}] \subset D$. So, by induction, $\{x_k\}$ is well defined. Furthermore, since $\|x_0 - x^*\| \leqq \eta \; \delta^0 \; \|x_0 - x^*\|$, it is shown by induction that

$$\|x_k - x^*\| \leqq \eta \; \delta^{k+1} \; \|x_0 - x^*\| \quad \text{for } k \geqq 0 \quad ,$$

and, hence, $x_k \to x^*$. This completes the proof.

Consider Iterative Procedure 5.1.7. It appears that D depends on η, which depends on r, which depends on D, and this might make D impossible to determine. However, if $\alpha > 0$, it is sufficient to pick $\eta = \frac{\beta}{\alpha}$. Alternately, suppose F, H_1, and V_1 are as in Example 5.1.6, where c = 1. Then $\sigma[H_1'(x)] \subset [a, \infty)$ and $\sigma[V^*] = \sigma[V] \subset [\alpha, \beta]$ for all $x \in R^n$. Thus, for any $D \subset R^n$, we may pick $a_D = a$. Then $r = \max(a, \alpha)$ and $\eta = \frac{r + \beta}{r + \alpha}$ are independent of D.

Suppose $y \in S(x^*, \eta \, \|x_0 - x^*\|)$. Then

$$\|y - x_0\| \leqq \|y - x^*\| + \|x_0 - x^*\| \leqq (\eta + 1)\ \|x_0 - x^*\|$$

But, by (1.4.5),

$$\|x_0 - x^*\| \leq \frac{1}{a + \alpha}\ \|F(x_0)\|$$

Hence, we may take

$$D = \bar{S}(\ x_0, \frac{\eta + 1}{a + \alpha}\ \|F(x_0)\|)$$

A more serious restriction is condition (5.1.13), since it may happen that

$$r_{k-1} < \max\left[\sqrt{a_k \cdot b_k}\ ,\ \sqrt{\alpha \cdot \beta}\ ,\ \frac{\epsilon + M_k^H N_k}{a_k + \alpha} - a_k\right]$$

However, if F is as above, we can guarantee that (5.1.13) can be satisfied. Suppose $x_0, \cdots, x_k$ have been determined. For $0 \leqq j \leqq k$, define

$$D_j^* = \bar{S}(x_j, \frac{\eta + 1}{a + \alpha}\ \|F(x_j)\|) \quad \text{and} \quad D_j = \bigcup_{i=0}^{j} D_j^* \quad .$$

Then, for $1 \leqq j \leqq k$, $D_j \subset D_{j-1}$. Hence, we can determine b_j, M_j^H, and N_j such that

$$(5.1.14)\begin{cases}(a,\ b_j,\ \alpha,\ \beta,\ M_j^H,\ 0,\ N_j) \ \epsilon \ \text{Bound}\ (D_j)\ ,\ 0 \leqq j \leqq k \\ b_j,\ M_j^H,\ N_j \leqq b_{j-1},\ M_{j-1}^H,\ N_{j-1} \ \text{respectively}\quad ,\ 1 \leqq j \leqq k\end{cases}$$

Thus, if

$$(5.1.15) \qquad r_k = \max\left[\sqrt{a \cdot b_k}\ ,\ \sqrt{\alpha \cdot \beta}\ ,\ \frac{\epsilon + M_k^H N_k}{a + \alpha} - a\right] \quad ,$$

(5.1.13) is satisfied. Hence, if $[x^*, x_j] \subset D_j$ for $0 \leqq j \leqq k$, Iterative Procedure 5.1.7 can be carried out. Clearly $[x^*, x_0] \subset D_0$. Suppose $[x^*, x_i] \subset D_i$ for $0 \leqq i \leqq j < k$. Then, as in the proof of Theorem 5.1.9,

$$\|x_{j+1} - x^*\| \leqq \eta\ \|x_i - x^*\| \text{ for } 0 \leqq i \leqq j$$

Hence, for $0 \leqq i \leqq j$,

$$[x^*,\ x_{j+1}] \subset S(x^*,\ \eta\ \|x_i - x^*\|) \subset S(x_i,\ (\eta + 1)\ \|x_i - x^*\|) \subset D_i^* \quad .$$

But, $[x^*, x_{j+1}] \subset D^*_{j+1}$. Hence, $[x^*, x_{j+1}] \subset D_{j+1}$. Thus, by induction, $[x^*, x_j] \subset D_j$ for $0 \leqq j \leqq k$. Therefore, Iterative Procedure 5.1.7 can be carried out, and, by Theorem 5.1.8, $x_k \to x^*$.

Now, since $x_k \to x^*$, it is clear that $D_k \to \{x^*\}$. Thus, $N_k \to 0$, and so (5.1.15) becomes, eventually,

$$r_k = \max [\sqrt{a \cdot b_k} , \sqrt{\alpha \cdot \beta}] \tag{5.1.16}$$

Suppose $\alpha > 0$ and let $t \in (0, \sqrt{\alpha\beta}]$. Determine $\alpha^* \in (0, \alpha]$ such that $\sqrt{\alpha^* \cdot \beta} = t$. Let $\eta = (t + \beta)/(t + \alpha^*)$ and $D_0 = \bar{S}(x_0, \frac{\eta + 1}{a + \alpha} \|F(x_0)\|)$. Determine b_0 as above, and then determine $a^* \in [0, a]$ such that $\sqrt{a^* \cdot b_0} \leqq t$. Then apply Iterative Procedure 5.1.8 as above. Condition (5.1.15) then becomes, eventually, $r_k = t$.

We note that it is not necessary to use a^* and α^* in the estimate $\|x - x^*\| \leqq \frac{1}{a + \alpha} \|F(x)\|$, since any bound for $\max_{x \in [x^*, x]} \|[F'(x)]^{-1}\| \, \|F(x)\|$ is valid here. Indeed, the estimate, $\|[F'(x)]^{-1}\| \leqq \frac{1}{a + \alpha}$, itself may be improved. See, e.g., the discussion at the end of Section 1.2.

In practice, we may begin Iterative Procedure 5.1.7 using a and α. When N_k becomes so small that (5.1.15) becomes (5.1.16), we may redefine the current iterate to be x_0 and begin Iterative Procedure 5.1.7 again using a^* and α^*. In this way, we can eventually bring r_k down to any desired fixed positive number.

<u>5.1.9 Example</u>: Let F be as in Example 1.4.1. Let H and V be the matrices corresponding to the horizontal and vertical differences respectively. Let $\varphi(x) = \psi(x) - b_v$, and set $H_1(x) = Hx + \varphi(x)$ and $V_1(x) = Vx$. Suppose we have determined $a, b, \alpha, \beta > 0$ such that $\sigma[H] \subset [a,b]$ and $\sigma[V] \subset [\alpha,\beta]$. (These bounds will depend on the region, $D \subset R^2$ of (1.3.1). In order to avoid confusion between this set and the set

$D \subset R^n$, which is assumed to contain $\bar{S}(x^*, \|x_0 - x^*\|)$, we shall call the latter set G in this example.) Let $x_0 \in R^n$, and set $G = \bar{S}(x_0, \rho)$ for some $\rho > 0$. Suppose, finally, that $F(x^*) = 0$ for some $x^* \in G$. We will now determine a_G, b_G, α, β, M_G^H, and N_G such that $(a_G, b_G, \alpha, \beta, M_G^H, 0, N_G) \in$ Bound (G). Let

$$d^0 \; (d^1) = \min_{\substack{1 \leq i \leq n \\ x_i - \rho \leq s \leq x_i + \rho}} (\max) \; e^s$$

Then we may take $a_G = a + d^0$, $b_G = b + d^1$, and $M_G^H = d^1$. Now, for $y \in G$,

$$\|y - x^*\| \leq 2\rho \quad \text{and} \quad \|F'(y)\| \leq b + \beta + d^1$$

Furthermore,

$$\|F(y)\| = \|F(y) - F(x^*)\| \leq \max_{\xi \in [x^*, y]} \|F'(\xi)\| \, \|y - x^*\|$$

Thus, we may take $N_G = 2\rho \, (b + \beta + d^1)$.

An iterative procedure similar to 5.1.3 can be defined for the ADI-N iteration of (5.0.2). The following lemma corresponds to Lemma 5.1.1.

<u>5.1.10 Lemma</u>: Let $F, H_1, V_1 : D \subset R^n \to R^n$ satisfy (5.1.1), and suppose (5.1.3) holds. For $r > 0$, let

$$h_{H,r}(x) = x - [r - H_1'(x)]^{-1} F(x)$$

$$h_{V,r}(x) = x - [r - V_1'(x)]^{-1} F(x)$$

i) Let $\epsilon \geq 0$. If

$$r \geq \max \left[\frac{\alpha_D + \beta_D}{2} , \frac{\epsilon + M_D^H N_D}{a_D + \alpha_D} - a_D \right] ,$$

then, on D,

$$\|h_{H,r}'(x)\| \leq 1 - \frac{\epsilon}{(r + a_D)^2}$$

ii) Let $\epsilon \geqq 0$. If

$$r \geqq \max\left[\frac{a_D + b_D}{2}, \frac{\epsilon + M_D^V N_D}{a_D + \alpha_D} - \alpha_D\right],$$

then, on D,

$$\|h'_{V,r}(x)\| \leqq 1 - \frac{\epsilon}{(r + \alpha_D)^2}$$

Proof: The proof is similar to that of Lemma 5.1.1.

2.1 Monotonic Results. We first state the following two definitions and a lemma which is a special case of Theorem 4.1 of [20].

5.2.1 Definition: Let $F: D \subset R^n \to R^n$ be differentiable on the convex set D. F is order-convex on D if

$$F(x) - F(y) \leqq F'(x)\ (x - y)$$

whenever $x, y \in D$ satisfy $x \leqq y$ or $y \leqq x$.

Order-convexity can be defined, of course, for non-differentiable functions, but, for our purposes, this definition will be sufficient.

5.2.2 Definition: Let $A \in L(R^n,R^n)$. Then $B \in L(R^n,R^n)$ is a subinverse of A if $AB \leqq I$ and $BA \leqq I$.

5.2.3 Lemma: Let $F: D \subset R^n \to R^n$ be differentiable of D. Suppose $[x^*, x_0]^* \subset D$ where x^* is the unique root of F in $[x^*, x_0]^*$, $F(x_0) \geqq 0$, and $x^* \leqq x_0$. Suppose F is order-convex on $[x^*, x_0]^*$. Let $\{x_k\}$ satisfy

$$x_{k+1} = x_k - B_k F(x_k) \quad \text{for } k \geqq 0,$$

where $B_k \in L(R^n,R^n)$ is a non-negative subinverse of $F'(x_k)$. Then

$x_k \downarrow y^* \in [x^*, x_0]^*$. If, in addition, $B_k \geqq B \geqq 0$ for $k \geqq 0$, where $B \in L(R^n, R^n)$ is non-singular. Then, $x_k \downarrow x^*$.

We now apply Lemma 5.4.3 to (5.0.1) and (5.0.2).

<u>5.2.4 Theorem</u>: Let $F, H_1, V_1 : [x^*, x_0]^* \to R^n$, where

$$(5.2.1) \qquad \begin{cases} x^* \text{ is the unique root of } F \text{ in } [x^*, x_0]^* \\ x^* \leqq x_0, \text{ and } F(x_0) \geqq 0 \end{cases} .$$

Suppose

$$(5.2.2) \qquad \begin{cases} \text{a) } F = H_1 + V_1;\ H_1, V_1 \in C^1([x^*, x_0]^*) \\ \text{b) } F \text{ is order-convex on } [x^*, x_0]^* \\ \text{c) } [r + H_1'(x)] \text{ and } [r + V_1'(x)] \text{ are M-matrices for } r > 0 \\ \qquad \text{and } x \in [x^*, x_0]^* \end{cases} .$$

Let $H_1'(x) = (h_{ij}(x))$ and $V_1'(x) = (v_{ij}(x))$, and set

$$d(x) = \max_{1 \leqq i \leqq n} \max\, [h_{ii}(x), v_{ii}(x)]$$

i) Suppose $\{x_k\}$ is defined by (5.0.1), where

$$(5.2.3) \qquad d(x_k) \leqq r_k \leqq s < \infty \quad \text{for } k \geqq 0$$

Then, $x_k \downarrow x^*$.

ii) Suppose $\{x_{k/2}\}$ is defined by (5.0.2), where

$$\begin{cases} \max_{1 \leqq i \leqq n} v_{ii}(x_k) \leqq r_k \leqq s < \infty \quad \text{for } k \geqq 0 \\ \max_{1 \leqq i \leqq n} h_{ii}(x_{k+\frac{1}{2}}) \leqq r_{k+\frac{1}{2}} \leqq s < \infty \quad \text{for } k \geqq 0 \end{cases}$$

Then, $x_{k/2} \downarrow x^*$.

<u>Proof</u>: i) Suppose $x \in [x^*, x_0]^*$ and $d(x) \leqq r \leqq s$. Let

$$B_r(x) = 2r\, [r + V_1'(x)]^{-1}\, [r + H_1'(x)]^{-1}$$

By Lemma 5.2.3, we need only verify

a) $B_r(x)$ is a non-negative subinverse of $F'(x)$,

and

b) there is a non-negative, non-singular $B \in L(R^n, R^n)$, which is independent of x, such that $B_r(x) \geqq B(x)$.

By (5.2.2c), $B_r(x) \geqq 0$. Now, a little algebra shows

$$I - B_r(x)\, F'(x) = [r + V_1'(x)]^{-1}\, [r + H_1'(x)]^{-1}\, [r - H_1'(x)]\, [r - V_1'(x)] .$$

Thus, by (5.2.2c) and the fact that $r \geqq d(x)$, $B_r(x)\, F'(x) \leqq I$. Likewise, $F'(x)\, B_r(x) \leqq I$. Thus, a) is verified.

Let $d_1 = \max_{x \in [x^*, x_0]^*} d(x) < \infty$. Then,

$$r + V_1'(x) \leqq (s + d_1)\, I ,$$

and so,

$$[r + V_1'(x)]^{-1} \leqq \frac{1}{s + d_1}\, I$$

Likewise,

$$[r + H_1'(x)]^{-1} \leqq \frac{1}{s + d_1}\, I$$

Now, by (5.2.1c),

$$d_0 = \min_{x \in [x^*, x_0]^*} d(x) > 0$$

Now, $r \geqq d_0$. Hence,

$$B_r(x) \geqq \frac{2d_0}{(s + d_1)^2}\, I \equiv B ,$$

and B is non-negative and non-singular. This verifies b) and completes the proof of i). The proof of ii) is similar.

We note that Theorem 5.2.4 does not assume any symmetry conditions of $F'(x)$. Thus, it would apply, for example to some discrete versions of the boundary value problem,

$$\begin{cases} Lu = f(s,t,u,u_s,u_t) & ; \quad (s,t) \in D \\ u = v & ; \quad (s,t) \in \partial D \end{cases} ,$$

where L is given by (2.2.4). The condition, (5.2.2c), may impose some restrictions on the values of f_u, f_{u_s}, and f_{u_t} and also on the discretizations of u_s and u_t being employed.

<u>5.2.5 Example</u>: Let $F\colon R^n \to R^n$ satisfy $F(x) = Hx + Vx + \varphi(x)$, where $H, V \in L(R^n,R^n)$ are M-matrices and $\varphi \in D(R^n)$ satisfies $\varphi_i \in C^2(R)$, $\varphi_i'(t) \geqq 0$, and $\varphi_i''(t) \geqq 0$ for $0 \leqq i \leqq n$ and $t \in R$. If

$$H_1(x) = Hx + c\varphi(x) \quad \text{and} \quad V_1(x) = Vx + (1-c)\varphi(x) ,$$

where $0 \leqq c \leqq 1$, then (5.2.2) is satisfied for any x^*, $x_0 \in R^n$. Let $H = (h_{ij})$ and $V = (v_{ij})$. Then (5.2.3) becomes

$$\max_{1 \leqq i \leqq n} \max\, [h_{ii} + c\varphi_i'((x_k)_i), v_{ii} + (1-c)\varphi_i'((x_k)_i)] \leqq r_k \leqq s < \infty .$$

<u>5.2.6 Remark</u>: Suppose $C \in L(R^n,R^n)$ is non-singular and has the splitting, $C = H_1 + V_1 + B$. In [30], Wachspress considers an ADI iteration of the form,

$$\begin{cases} [r_k + H_1 + B]\, x_{k+\frac{1}{2}} = [r_k - V_1]\, x_k + \xi \\ [r_k + V_1 + B]\, x_{k+1} = [r_k - H_1]\, x_{k+\frac{1}{2}} + \xi \end{cases} ,$$

for approximating $C^{-1}\xi$ where $\xi \in R^n$. Using this iteration in tandem with an outer Newton iteration for the F of Example 5.2.5, where, for some $c \in R$,

$$H_1(x) = Hx + (1-c)\varphi(x),\ V_1(x) = Vx + (1-c)\varphi(x), \text{ and } B(x) = (2c-1)\varphi(x),$$

we obtain the following iteration,

$$[r_k + H + c\varphi'(x_k)]\, x_{k+\frac{1}{2}} = [r_k - V - (1-c)\varphi'(x_k)]\, x_k \quad + F'(x_k)\, x_k - F(x_k)$$

$$[r_k + V + c\varphi'(x_k)]\, x_{k+1} = [r_k - H - (1-c)\varphi'(x_k)]\, x_{k+\frac{1}{2}} + F'(x_k)\, x_k - F(x_k),$$

which can be put in the form,

$$x_{k+1} = x_k - [r_k + V + c\varphi'(x_k)]^{-1}\, [2r_k + (2c-1)\varphi'(x_k)]\, [r_k + H + c\varphi'(x_k)]^{-1}\, F(x_k).$$

If (5.2.1) holds, then Lemma 5.2.3 will guarantee convergence of $\{x_k\}$ to x^* if $\frac{1}{2} \leqq c \leqq 1$ and

$$\max_{1 \leqq i \leqq n} \max\, [h_{ii} + (1-c)\varphi_i'((x_k)_i),\ v_{ii} + (1-c)\varphi_i'((x_k)_i)] \leqq r_k \leqq s < \infty \quad .$$

We note that by picking $c = 1$, this is independent of k. Thus, we can determine a priori a sequence, $\{r_k\}$ of acceleration parameters and still guarantee convergence. The choice of $c = 1$ is not unattractive, since it corresponds to putting all of $\varphi'(x_k)$ into the matrices to be inverted.

5.2.7 Example: Let F, H, V and φ be as in Example 5.1.9. Then H and V are M-matrices, and $\varphi_i'(t) \geqq 0$ and $\varphi_i''(t) \geqq 0$ for $1 \leqq i \leqq n$ and $t \in R$. Now, F has a unique root, x^*, and $F'(x)$ is an M-matrix for each $x \in R^n$. Thus, by (1.4.5), $F(x_0) \geqq 0$ implies $x_0 \geqq x^*$. Hence, Theorem 5.2.4 can be applied if we can find $x_0 \in R^n$ such that $F(x_0) \geqq 0$. Let $x_0 = \underline{v}$. Then, $[H + V]\, x_0 - b_v = \underline{\ell v} = 0$. Hence

$$F(x_0) = (e^{(x_0)_i}) > 0.$$

Thus, x_0 is a suitable starting vector.

5.3 A Counterexample: Let F: $R^n \to R^n$ be convex and satisfy $F \in C^1(R^n)$, $F'(x)$ is an M-matrix for each x, and F has a unique root, x^*. Then the Newton iteration converges to x^* for any starting vector, x_0. Furthermore, suppose $F = H_1 + V_1$, where $H_1'(x)$ and $V_1'(x)$ are uniformly positive definite. Then, in the linear case $(H_1(x) = Hx + \xi_1,\ V_1(x) = Vx + \xi_2,\ \xi_1 + \xi_2 = \xi,)$ the ADI iteration (2.1.2) converges for all fixed $r_k = r > 0$ and for all x_0. In this case, the ADI iteration coincides with the Newton-ADI iteration (5.0.1).

The question naturally arises: given the above assumptions on F, H_1, and V_1 (except the linearity,) does the Newton-ADI iteration converge globally for all fixed $r > 0$? The following counterexample shows that these assumptions are not sufficient. In particular, it indicates that some assumption on the nature of the splitting, $F = H_1 + V_1$, is necessary.

<u>3.1.1 Counterexample</u>: Let $c_0 \in (0,\frac{1}{2})$ be the solution of $c(1-c) = \frac{1}{8}$. Then suppose

$$\begin{cases} 0 \leqq c \leqq c_0 \\ \gamma = 8c(1-c) \\ a > 1/(1-\gamma) + (1/(1-\gamma) + (1-\gamma)^2)^{\frac{1}{2}} \end{cases}$$

Let

$$\begin{cases} b_0 = \dfrac{-(a+1)^2}{2a} \\ b_1 = \dfrac{(a+1+4ca)(a+1+4(1-c)a)}{2a} - 2 \end{cases}$$

Then, it is straightforward to verify that

$$0 < b_1 - b_0 < 4a$$

Thus, there is a convex, non-decreasing function $\varphi \in C^2(R^n)$ which satisfies

$$\varphi(0) = b_0, \quad \varphi(1) = b_1, \quad \varphi'(0) = 0, \quad \varphi'(1) = 4a$$

Let $F, H_1, V_1 : R \to R$ be defined by

$$\begin{cases} F(x) = 2x + \varphi(x) \\ H_1(x) = x + c\varphi(x) \\ V_1(x) = x + (1-c)\varphi(x) \end{cases} .$$

Then, F, H_1, V_1 Satisfy the conditions given above. The Newton-ADI iteration takes the form

(5.3.1) $$x_{k+1} = x_k - \frac{2rF(x_k)}{[r + V_1'(x_k)][r + H_1'(x_k)]}$$

It can be verified immediately that if $r = a$ and $x_0 = 1$, then $x_k = 0$ or 1 depending on whether k is even or odd. Thus, the iteration does not converge to the root of F.

It is interesting to note that if $F \in C^2(R)$ is convex and stricly increasing with a root, x^*, then (5.3.1) converges globally to x^* for all $r > 0$ if $H_1 = V_1 = \frac{1}{2}F$ (i.e., $c = \frac{1}{2}$ in the counterexample.) This can be demonstrated by comparing (5.3.1) with the Newton iteration. This indicates, as noted above, that any global Newton-ADI convergence theorem for all $r > 0$ would have to include assumptions on the nature of the splitting of F.

5.4 Numerical Results. Let Ω be as in Example 1.2.5. The following problem was considered.

(5.4.1) $$\begin{cases} \Delta_h u(P) = e^{u(P)} & ; \quad P \in \Omega \\ u(P) = v(P) & ; \quad P \in \Omega' \\ v(s,t) = s + 2t & , \quad h = .1 \end{cases} .$$

Let H and V be the matrices corresponding to the horizontal and vertical differences respectively. Then H and V commute and have the same eigenvalues, $\lambda_1 \leqq \cdots \leqq \lambda_n$, where

$$\begin{cases} h^2 a = h^2\alpha = h^2\lambda_1 = 4 \sin^2 \dfrac{\pi}{2(N+1)} \\ h^2 b = h^2\beta = h^2\lambda_n = 4 \sin^2 \dfrac{N\pi}{2(N+1)} \\ h = {}^1/_{N+1} \quad , \; n = N^2 \end{cases}$$

(see [28, P. 214].)

Let $F(x) = H_1(x) + V_1(x)$ where

$$\begin{cases} H_1(x) = Hx + \tfrac{1}{2}\varphi(x) \\ V_1(x) = Vx + \tfrac{1}{2}\varphi(x) \\ \varphi(x) = (e^{x_i}) - b_v \end{cases}$$

F has a unique root, x^* (see Example 1.4.1.) Let $x_0 = \underline{v}$. Then $[H + V]\, x_0 - b_v = 0$. Hence,

$$\|F(x_0)\| = [\sum_{i=1}^{n} e^{2(x_0)_i}]^{\frac{1}{2}}$$

But, $0 \leqq (x_0)_i \leqq 3$. Thus,

$$\|F(x_0)\| \leqq \sqrt{n}\, e^3 = 9e^3$$

Now, $S(x^*, \|x_0 - x^*\|) \subset S(x_0, \rho) \equiv D$, where

$$\rho = \frac{1}{a + \alpha} \|F(x_0)\| \leqq \frac{9e^3}{a + \alpha} .$$

But, $a + \alpha \approx 2\pi^2$. Thus, for small enough h, $\rho < 12$. By Example 5.1.9, we may take

$$\begin{cases} a_D = \alpha_D = a + e^{-12}, \; b_D = \beta_D = b + e^{15} \\ M_D^H = M_D^V = \tfrac{1}{2}e^{15} \quad , \; N_D = 26 + e^{15} \end{cases}$$

We note that these bounds can be considerably improved if F is replaced by the function (also called F) defined in Example 4.2.10.

The Newton-ADI iteration of (5.0.1) was employed to solve (5.4.1).

The initial vector was $x_0 = \underline{v}$.

The convergence criterion was $\|x_k - x_{k-1}\|_2 \leqq 10^{-6}$.

(This does not give an absolute error bound. Since, by (1.4.7), $\|x - x^*\|_2 \leqq \|[H + V]^{-1}\| \, \|F(x)\|$, a better convergence criterion would be $\|F(x_k)\| \leqq \gamma$ for some suitable γ. We note that $\|[H + V]^{-1}\| \approx 2\pi^2$.)

The results, when $r_k \equiv r$ is constant, are given in Table 5.4.1. By (1.4.10), $x^* \leqq x_0$, and the diagonal entries of h^2H and h^2V are all equal to 2. Hence, Theorem 5.2.4 guarantees monotonic convergence if

$$h^2 r \geqq 2 + h^2 \max_{t \leqq 3} e^t = 2 + h^2 e^3 \quad .$$

However, monotonic convergence was obtained for even smaller values of r. This is indicated in Table 5.4.1.

The number of iterations is plotted against $h^2 r$ in Graph 5.4.3. We note that the graph is approximately linear above the optimal parameter but more sharply decreasing below the optimal parameter. This phenomenon was also noted in the linear case and in other similar nonlinear cases.

The results when several parameters were used cyclically are given in Table 5.4.2. The parameters used were the 2^{ν} Wachspress optimal parameters for H and V (see [28, P. 224].)

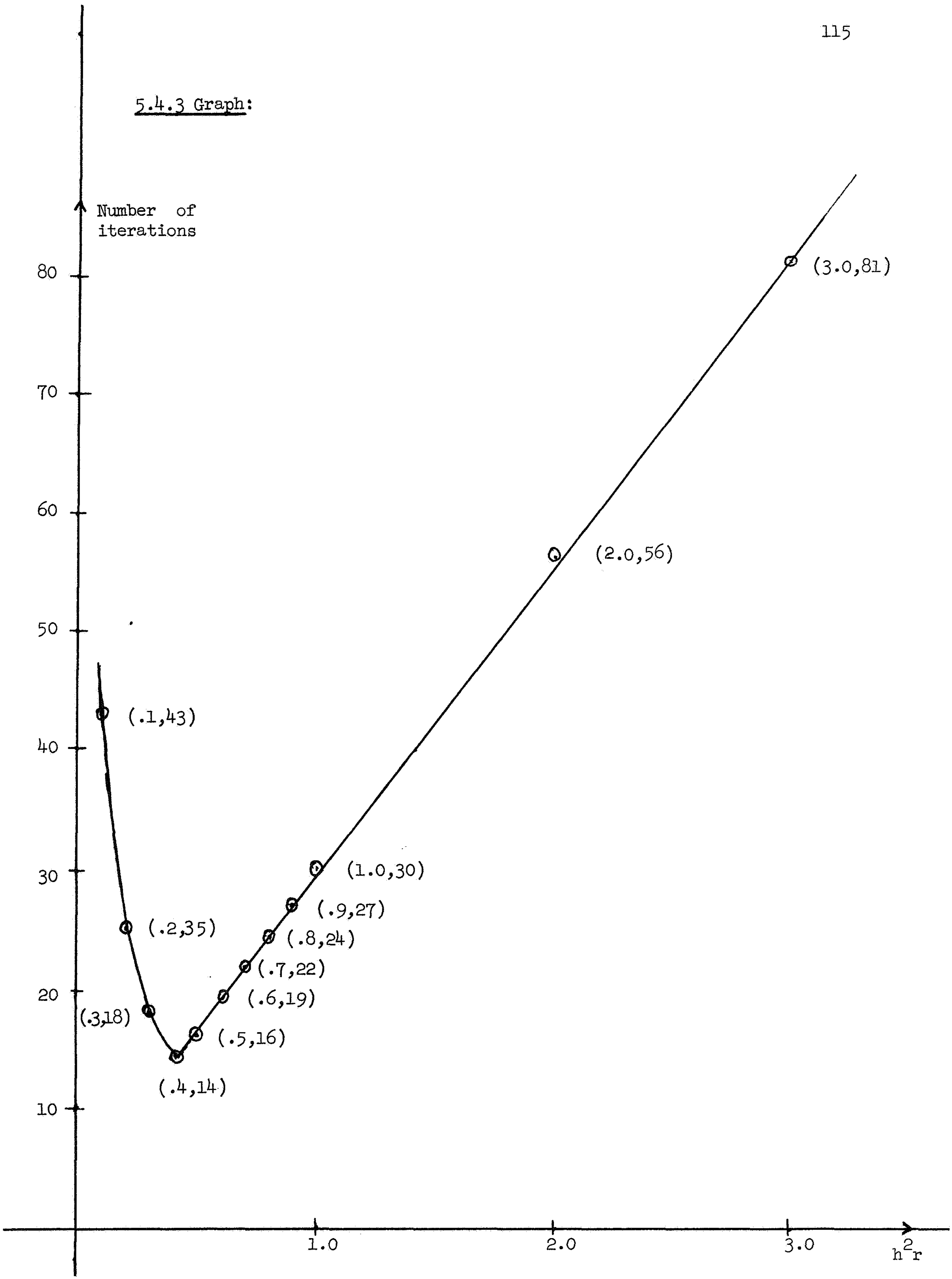
5.4.1 Table:

h^2r	Number of iterations	Convergence monotonic ?
.1	43	No
.2	25	"
.3	18	"
.4	14	"
.5	16	Yes
.6	19	"
.7	22	"
.8	24	"
.9	27	"
1.0	30	"
2.0	56	"
3.0	81	"

5.4.2 Table:

Number of parameters used cyclically	Values of h^2r_k (to 3 places)	Number of iterations
1	.619	19
2	.188, 2.04	11
4	.118, .335, 1.14, 3.23	9
8	.103, .146, .249, .454 .841, 1.54, 2.62, 3.71	11
16	.099, .109, .130, .164 .216, .288, .390, .529 .722, .979, 1.33, 1.77 2.32, 2.93, 3.50, 3.85	12

5.4.3 Graph:


Number of iterations
80
70
60
50
40
30
20
10
1.0
2.0
3.0
h^2r
(.1,43)
(.2,35)
(.3,18)
(.4,14)
(.5,16)
(.6,19)
(.7,22)
(.8,24)
(.9,27)
(1.0,30)
(2.0,56)
(3.0,81)

REFERENCES

1. L. Bers, *On mildly nonlinear partial differential equations of elliptic type*, J. Res. Nat. Bur. Standards 51B (1953), 229-236.

2. L. Bers, *Topology*, Courant Institute of Mathematical Sciences, New York University, New York, 1956-1957.

3. J. H. Bramble and L. E. Payne, *Bounds for derivatives in elliptic boundary value problems*, Pacific J. Math. 14 (1964), 777-782.

4. R. Courant, *Partial differential equations*, Methods of Mathematical Physics, Vol. II, Interscience, New York, 1962.

5. J. Douglas, Jr., *Alternating direction iteration for mildly nonlinear elliptic difference equations*, Numer. Math. 3 (1961), 92-98.

6. J. Douglas, Jr., *Alternating direction methods for three space variables*, Numer. Math. 4 (1962), 41-63.

7. J. Douglas, Jr., R. B. Kellogg, and R. S. Varga, *Alternating direction methods for n space variables*, Math. Comp. 17 (1963), 279-282.

8. J. Douglas, Jr. and H. H. Rachford, *On the numerical solution of heat conduction problems in two and three space dimensions*, J. Soc. Indust. Appl. Math. 3 (1955), 28-41.

9. E. G. D'yakonov, *On some difference schemes for solutions of boundary problems*, Z. Vycisl. Mat. Mat. Fiz. 2 (1962), 57-79 (Russian.)

10. G. E. Forsythe and W. R. Wasow, *Finite-difference methods for partial differential equations*, John Wiley & Sons, New York, 1960.

11. J. Guittet, *Methodes de directions alternees*, Rev. Francaise Traitment Information Chiffres 9 (1966), 95-107.

12. J. E. Gunn, *The numerical solution of* $\nabla \cdot a \nabla u = f$ *by a semi-explicit alternating-direction iterative technique*, Numer. Math. 6 (1964), 181-184.

13. J. E. Gunn, *On the two stage iterative method of Douglas for mildly nonlinear elliptic difference equations*, Numer. Math. 6 (1964), 243-249.

14. R. B. Kellogg, *Another alternating-direction-implicit method*, J. Soc. Indust. Appl. Math. 11 (1963), 976-979.

15. R. B. Kellogg, *A nonlinear alternating direction method*, to appear in Math. Comp.

16. O. A. Ladyzhenskaya and N. N. Ural'tseva, *Linear and quasilinear elliptic equations*, transl. Scripta Technica, Inc., Academic Press, New York, 1968.

17. N. Levinson, *Dirichlet problem for* $u = f(P,u)$, J. Math. Mech. 12 (1963), 567-574.

18. G. J. Minty, *Monotone (nonlinear) operators in Hilbert spaces*, Duke Math. J. 29 (1962), 341-346.

19. J. M. Ortega and W. C. Rheinboldt, *On a class of approximate iterative processes*, Arch. Rational Mech. Anal. 23 (1967), 352-365.

20. J. M. Ortega and W. C. Rheinboldt, *Monotone iterations for nonlinear equations with application to Gauss-Seidel methods*, SIAM J. Numer. Anal. 4 (1967), 171-190.

21. J.M. Ortega and M. L. Rockoff, *Nonlinear difference equations and Gauss-Seidel type iterative methods*, SIAM J. Numer. Anal. 3 (1966), 497-513.

22. S. V. Parter, *Mildly non-linear elliptic partial differential equations and their numerical solution. I*, Numer. Math. 7 (1965), 113-128.

23. D. W. Peaceman and H. H. Rachford, Jr., *The numerical solution of parabolic and elliptic differential equations*, J. Soc. Indust. Appl. Math. 3 (1955), 28-41.

24. C. Pearcy, *On convergence of alternating direction procedures*, Numer. Math. 4 (1962), 172-176.

25. W. C. Rheinboldt, *Local mapping relations and global implicit function theorems*, Tech. Rep. 68-62, Computer Science Center, University of Maryland, College Park, 1968, to appear in Trans. Amer. Math. Soc.

26. J. T. Schwartz, *Nonlinear functional analysis*, Courant Institute of Mathematical Sciences, New York University, New York, 1963-1964.

27. J. Spanier, *Alternating direction methods applied to hear conduction problems*, Mathematical Methods for Digital Computers, Vol. 2, John Wiley & Sons, New York, 1967, 215-245.

28. R. S. Varga, *Matrix iterative analysis*, Prentice-Hall, Englewood Cliffs, N. J., 1962.

29. M. M. Vainberg, *Variational methods for the study of nonlinear operators*, transl. A. Feinstein, Holden-Day, San Francisco, 1964.

30. E. L. Wachspress, *Iterative solution of elliptic systems and applications to the neutron diffusion equations of reactor physics*, Prentice-Hall, Englewood Cliffs, N. J., 1966.