Technical Report TR-143
GJ-231 and NGL-21-002-008

January 1971

NONLINEAR PROGRAMMING METHODS FOR SOLVING NONLINEAR EIGENVALUE PROBLEMS

by

Steven W. Rauch

UNIVERSITY OF MARYLAND
COMPUTER SCIENCE CENTER
COLLEGE PARK, MARYLAND

Technical Report TR-143
GJ-231 and NGL-21-002-008

January 1971

NONLINEAR PROGRAMMING METHODS FOR SOLVING NONLINEAR EIGENVALUE PROBLEMS

by

Steven W. Rauch

This research was supported in part by the National Science Foundation under Grant GJ-231 and by the National Aeronautics and Space Administration under Grant NGL-21-002-008 to the Computer Science Center of the University of Maryland.

TABLE OF CONTENTS

Introduction

In this paper, we discuss several numerical methods for solving nonlinear eigenvalue problems in R^n of the general form

$$(I1) \qquad f'(x)^T = \lambda g'(x)^T, \; x \in D \subset R^n, \; \lambda \in R^1.$$

During the past decade or so a considerable theory of such variationally-based eigenvalue problems has been developed by many authors, mostly in an infinite dimensional setting. For example, important results, together with extensive bibliographies, can be found in Berger and Berger [1968], Krasnosel'skii [1964], Pimbley [1969], Vainberg [1964], and Vainberg and Aizengendler [1968]. However, in general, the theoretical results do not provide for efficient methods of finding specific numerical solutions of problems of the type (I1).

Applied problems leading to infinite dimensional eigenvalue problems (I1) are usually formulated in terms of certain differential equations, as for example, the equations describing nonlinear vibrations. For the numerical solution of such problems, the differential equation is frequently replaced by a discrete analog and hence we are led in a natural way to finite dimensional eigenvalue problems of the form (I1). If the functionals $f,g:D \subset R^n \to R^1$ are continuously differentiable on some open convex set D, then a theorem of Ljusternik [1934] ensures that a solution of (I1) can be obtained by considering instead the constrained minimization problem

$$\min \{f(x) | x \in D,\ g(x) = 0\}.$$

This is a special case of the nonlinear programming problem

(I2) $$\min \{f(x) | x \in C\}$$

with C defined by

(I3) $$C = \{x \in D | g_j(x) \leq 0,\ j = 1,\ldots,m\}.$$

By far the largest class of methods for solving problems of the form (I2/3) consists of descent type algorithms

$$x^{k+1} = x^k - \omega_k \tau_k s^k,\ f(x^k) \geq f(x^{k+1}),\ x^{k+1} \in C,\ k = 0,1,\ldots,$$

where at each step k, s^k is a suitable direction vector, τ_k a steplength along this direction, and ω_k a relaxation parameter.

The literature on descent methods for constrained as well as unconstrained minimization problems is vast (see, for example, Künzi and Oettli [1969] for constrained problems, and Ortega and Rheinboldt [1970] for unconstrained ones). Of particular interest to us here is a comprehensive convergence theory for such methods in the unconstrained case developed by Elkin [1968].

A central point of this theory is a complete separation of the analysis of the steplength and direction algorithms which allows the combination of many different algorithms into convergence theorems for various descent methods. In brief, if $f: D \subset R^n \to R^1$ is continuously differentiable on some open set D and minimization takes place on a closed component L^0 of a level set on which f is bounded below; then, at the iterate $x^k \in L^0$, a direction s^k is feasible if

(I4) $$f'(x^k)s^k > 0, \; \|s^k\| = 1.$$

Now only the steplength algorithm is used to show that, for any such feasible direction, the next iterate x^{k+1} remains in L^0 and that a basic inequality of the form

(I5) $$f(x^k) - f(x^{k+1}) \geq \mu(f'(x^k)s^k) \geq 0$$

holds where μ is some function with the property that $\mu(t_k) \to 0$ implies that $t_k \to 0$. Under the assumptions on f this already shows that

(I6) $$\lim_{k\to\infty} f'(x^k)s^k = 0.$$

Only at this point do the specific choices of the directions enter the analysis. Without further consideration of the steplength algorithm, that is, entirely on the basis of (I5/6) and the properties of f, it is shown that for certain classes of direction algorithms, (I6) implies the statement

(I7) $$\lim_{k\to\infty} f'(x^k)^T = 0.$$

Finally, it follows essentially only from properties of f that indeed $\{x^k\}$ itself converges to a critical point of f.

We shall extend this convergence theory of Elkin to the case of the constrained minimization problem (I2/3) under rather mild conditions on the constraint functionals g_j. Instead of the closedness of L^0, now only its intersection $L^0 \cap C$ with the constraint set C is assumed to have that property. Correspondingly, the large class of possible directions allowed by (I4) must be suitably restricted to guarantee that the next iterate remains in $L^0 \cap C$. For such a restricted class of feasible directions, the constrained analogs of most standard steplength algorithms can be analyzed in much the same way as in the unconstrained case. This includes the usual minimization step, the Curry [1944] and Altman [1966] algorithms and also algorithms due to Ostrowski [1966], Goldstein [1964/1965/1966] and Armijo [1966]. In each case, the steplength algorithm is shown to guarantee the existence of the next iterate x^{k+1} in $L^0 \cap C$ and, in analogy to (I5), the validity of a more involved estimate of the form

$$f(x^k) - f(x^{k+1}) \geq \Phi(f'(x^k)s^k, \sigma_k, \varepsilon_k) \geq 0.$$

Here σ_k measures essentially the angle made by the feasible direction s^k with the normals to the boundary of the constraint set at x^k, and ε_k is related to the distance of x^k from the boundary. As before, Φ has the property that from the existence of a lower bound for f it follows that at least one of the three limit statements

(I8) $$\lim_{k\to\infty} f'(x^k)s^k = 0, \; \lim_{k\to\infty} \sigma_k = 0, \; \lim_{k\to\infty} \varepsilon_k = 0$$

is valid. Again, only at this point does the specific algorithm for choosing the feasible direction s^k enter the analysis. The central point is now to show that for them, the three quantities $f'(x^k)s^k$, σ_k, and ε_k are related in such a way that from (I8) follows directly the analog of (I7), namely,

$$\lim_{k\to\infty} [f'(x^k)^T - \sum_{j\varepsilon J^k} \nu_j^k g_j'(x^k)^T] = 0$$

with some sequence of numbers $\{\nu_j^k\}$ and index sets $\{J^k\}$, $J^k \subset \{1,\ldots,m\}$. This, together with suitable assumptions on f, then ensures the convergence of the iterates $\{x^k\}$ to a conditional critical point of f on C.

The most important class of direction algorithms considered here concerns directions s^k obtained by normalizing vectors of the form

$$P_q(x^k)p^k + \bar{\alpha} \sum_{j\varepsilon J} \beta_j g_j'(x^k)^T$$

where $P_q(x^k)$ is a certain projection matrix, and $\bar{\alpha}, \beta_j$, $j \varepsilon J \subset \{1,\ldots,m\}$ are suitable coefficients. Specifically, three choices of p^k are examined and, in particular, for the case $p^k = f'(x^k)^T$, the gradient projection method of Rosen [1960] for linear constraints is also obtained. Other choices of p^k are the projected gradient direction transformed by a positive definite matrix, and the coordinate direction forming the smallest angle with the projected gradient direction. The latter is simply the Gauss-Southwell algorithm for the constrained case.

As in the unconstrained case, the separate analysis of steplength and direction choice provides for the combination of many different algorithms into specific descent methods and gives convergence results for all these combinations. In particular, our general theory also provides as corollaries convergence theorems for a method of feasible directions of Zoutendijk and for Rosen's gradient projection method in the case of linear constraints.

In turning again to the eigenvalue problem (I1) we will apply some of these minimization results to obtain numerical solutions for it. Since the natural constraint set C associated with (I1) has no interior, some modifications are needed and we shall give two ways of changing the set C so that our earlier established convergence theory applies to (I1). The first method changes the constraint set into a set resembling an annulus while in the other one, a penalty function is added to f to eliminate one of the two constraints $g(x) \leq 0$ or $-g(x) \leq 0$. In contrast to these approaches we also present an algorithm of Goldstein [1967] for the case $g'(x)^T \equiv Ax$ where $A \in L(R^n)$ is symmetric and positive definite. This method has the basic form

$$x^{k+1} = \frac{x^k - \tau_k A^{-1} f'(x^k)^T}{\|x^k - \tau_k A^{-1} f'(x^k)^T\|_A}, \quad k = 0,1,\ldots, \tag{I9}$$

and we shall extend this result so that the parameter τ_k can be chosen constant throughout the process.

Since the rate of convergence for descent processes is, in general, only linear, it is desirable to terminate the procedure once the iterates are sufficiently close to the solution and to apply then a locally convergent method such as the quadratically convergent Newton process. This constitutes two parts of a complete algorithm for finding entire branches of solutions of (I1). For the final part of this algorithm we follow the suggestion of Pimbley [1969] and solve numerically the initial value problem

$$\frac{dx}{d\lambda} = [f''(x)-\lambda g''(x)]^{-1} g'(x)^T, \quad x(\lambda^*) = x^*, \quad \lambda^* - \alpha \leq \lambda \leq \lambda^* + \alpha$$

in order to compute such a branch of solutions of (I1).

The paper is organized in the following form: Chapter I presents a survey of well-known existence results for the general nonlinear eigenvalue problem

$$Fx = \lambda Gx, \quad x \in D \subset R^n, \quad \lambda \in R^1.$$

We also discuss some aspects of the existence and extendability of continuous branches of eigenvectors and the concept of bifurcation points. Moreover, because of its importance in the later development, a proof of the Ljusternik [1934] theorem is also given.

Chapter II contains the mentioned generalization of the Elkin [1968] convergence theory to the nonlinear programming problem (I2/3) beginning in Section 2.1 with a brief review of the basic concepts of Elkin's theory.

Chapter III applies the results of Chapter II to the eigenvalue problem (I1) and includes also the discussion of the mentioned algorithm (I9) of Goldstein [1967]. Then, the indicated complete algorithm for finding entire branches of solutions of (I1) is applied to a problem related to nonlinear heat generation studied by Joseph [1965] and to a rotating string problem of Kolodner [1955]. Numerical results are then given for these two examples.

CHAPTER I

Survey of Existence Results for Solutions of Nonlinear Eigenvalue Problems in R^n

1.1 Notation

We begin this discussion with a brief description of the notation to be used. R^n is the real n-dimensional linear space of column vectors $x = (x_1,\ldots,x_n)^T$ where T denotes transposition. In particular, e^i, $i = 1,\ldots,n$, are the unit basis vectors in R^n for which the i^{th} component equals one and all others are zero. For $x,y \in R^n$ the Euclidean inner product of x and y is denoted by $x^Ty = \sum_{i=1}^{n} x_iy_i$. In all cases some norm is assumed to be given on R^n. Frequently we will use the Euclidean norm $\|x\| = (x^Tx)^{1/2}$ although most of the results generalize to other norms. The linear space of all real $m \times n$ matrices will be denoted by $L(R^n,R^m)$ and by $L(R^n)$ if $m = n$. The norm on $L(R^n,R^m)$ will always be that which is induced by the vector norm on R^m and R^n. For $A \in L(R^n,R^m)$ we sometimes write $A = (a_{ij})$ where a_{ij}, $i = 1,\ldots,m$, $j = 1,\ldots,n$, are the elements of A. As usual, $A \in L(R^n)$ is symmetric if and only if $(Ax)^Ty = x^TAy$ for all $x,y \in R^n$ and is positive definite if and only if $x^TAx > 0$ for all $x \neq 0$. For an arbitrary set $D \subset R^n$ we write the interior of D as int(D), the closure of D as $\bar{D}$ and the boundary of D as $\dot{D}$. An open interval $(x,y) \subset R^n$ is defined as the set $\{z \in R^n | z = ty + (1-t)x$ for some $t \in (0,1)\}$. The corresponding closed and half-open intervals will be denoted by $[x,y]$ and $(x,y]$ (or $[x,y)$), respectively.

A mapping F with domain $D \subset R^n$ and range in R^m will be denoted by $F:D \subset R^n \to R^m$ and its components by f_i, $i = 1,\ldots,m$. Thus, in particular, $Fx = (f_1(x),\ldots,f_m(x))^T$. We use the standard differentiability concepts: The mapping $F:D \subset R^n \to R^m$ is Gateaux differentiable (G-differentiable) at $x \in \text{int}(D)$ if there is an $A \in L(R^n,R^m)$ such that for any $h \in R^n$,

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

and F is Frechet differentiable (F-differentiable or simply differentiable) if

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

For both cases we denote the unique derivative A by $F'(x)$. Specifically, $F'(x)$ is the Jacobian matrix $(\partial_j f_i(x))$ where $\partial_j f_i(x) \equiv \partial f_i(x)/\partial x_j$, $i = 1,\ldots,m$, $j = 1,\ldots,n$. The following two well-known forms of the mean-value theorem for functionals will be used frequently: If $f:D \subset R^n \to R^1$ has a G-derivative f' and if for any $[x,y] \subset D$, $\psi:[0,1] \to L(R^n,R^1)$, $\psi(t) = f'(tx+(1-t)y)$ is continuous, then

$$f(y) - f(x) = f'(z)(y-x), \text{ for some } z \in (x,y),$$

as well as

$$f(y) - f(x) = \int_0^1 f'(x+t(y-x))(y-x)dt.$$

Finally, we shall at times use various types of convexity for functionals. A functional $f: D \subset R^n \to R^1$ is <u>convex</u> in some convex set $D_0 \subset D$ if for all $x, y \in D_0$ and all $t \in [0,1]$,

$$f(tx+(1-t)y) \leq tf(x) + (1-t)f(y).$$

If D is open and f has a G-derivative f' on D, then f is <u>pseudo-convex</u> if for all $x, y \in D$

$$f(x) < f(y) \text{ implies } f'(y)(y-x) > 0;$$

f is <u>quasi-convex</u> on D_0 if, for any $x, y \in D_0$,

$$f(tx+(1-t)y) \leq \max \{f(x), f(y)\}, \quad \forall t \in (0,1).$$

For a discussion of these various types of convexity, we refer to Elkin [1968] and Ortega and Rheinboldt [1970]. Many of the related results can also be found in Mangasarian [1969].

1.2 Survey of existence results in R^n: The non-potential case

In this section we present a survey of existence results for solutions of the eigenvalue problem

$$(1.2.1) \qquad Fx - \lambda Gx = 0, \; x \in D, \; \lambda \in R^1$$

where $F, G: D \subset R^n \to R^n$ are given mappings on a domain D. Most of these results are known, often in a more general infinite dimensional form, but some minor extensions are also included. The mappings F and G are not necessarily potential operators, that is, they are not necessarily the gradients of some functionals on R^n.

A vector $x \in D$ is said to be an eigenvector of (1.2.1) (or simply of F when $G \equiv I$) corresponding to the eigenvalue $\lambda \in R^1$ if $x \in D$ and $\lambda \in R^1$ constitute a solution of (1.2.1). If $x \neq 0$ and $Gx \neq 0$, then the eigenvector is said to be non-trivial. The set of all $\lambda \in R^1$ for which (1.2.1) has some non-trivial eigenvector $x \in D$ is called the spectrum of (1.2.1). Easy examples show that the spectrum may be empty, discrete or continuous. For ease of notation we sometimes write (x,λ) for a solution $x \in D$, $\lambda \in R^1$ of (1.2.1).

There are several basic questions which are of interest to us here. When can the existence of eigenvalues be guaranteed? If an eigenvalue exists, is its corresponding eigenvector unique? If x is an eigenvector, when is there a continuous curve of solutions of (1.2.1) through x; and if such a curve exists, when is it unique? For a discussion of these questions the following terminology will prove to be useful.

Definition 1.2.1 Let $F,G:D \subset R^n \to R^n$ be defined on the set D. If there exist mappings $x:(t_1,t_2) \subset R^1 \to R^n$ and $\lambda:(t_1,t_2) \subset R^1 \to R^1$ on some open interval (t_1,t_2) with $t_1 < t_2$ such that $(x(t),\lambda(t))$ solves (1.2.1), then x is called a branch of eigenvectors of (1.2.1) in R^n. If x is continuous on (t_1,t_2), then x is a continuous branch of eigenvectors of (1.2.1) in R^n. A branch of eigenvectors of (1.2.1) in R^n is _proper_ if the mapping x is one-one. A point (x^0,λ_0) is a _regular point of (1.2.1)_ if there is a unique proper continuous branch of eigenvectors of (1.2.1) in R^n through x^0, that is, for which $(x(t_0),\lambda(t_0)) = (x^0,\lambda_0)$ for some $t_0 \in (t_1,t_2)$. Any solution (x^0,λ_0) of (1.2.1) which is not a regular point of (1.2.1) is a _branch point of (1.2.1)_.

Note that in the linear case $F \equiv A \in L(R^n)$ symmetric and $G \equiv I$ there is a proper continuous branch of eigenvectors of (1.2.1) in R^n with $\lambda(t) \equiv \lambda_0$ and hence λ is not necessarily one-one. However, if we assume that λ is one-one, then we can parameterise a continuous branch of eigenvectors of (1.2.1) in R^n with λ as the parameter.

Definition 1.2.2 Let $F,G:D \subset R^n \to R^n$ be defined on the open set D and suppose that (x^0,λ_0) solves (1.2.1). If $x_1,x_2:(t_1,t_2) \subset R^1 \to R^n$ are two distinct branches of eigenvectors of (1.2.1) through x^0 such that $\lim_{t\to\lambda_0} (x_1(t)-x_2(t)) = 0$ then λ_0 is called a _bifurcation point of (1.2.1)_.

Frequently in the applications to discretized vibration problems we have $F0 = G0 = 0$. In such cases it usually happens that for small $|\lambda|$, $(0,\lambda)$ is the only solution of (1.2.1) and, beginning with some $\lambda_0 \neq 0$, eigenvectors with small norm appear. More precisely, to each ε and δ there corresponds an eigenvector x of (1.2.1) with associated eigenvalue λ such that

$$0 < \|x\| < \delta, \ |\lambda-\lambda_0| < \varepsilon.$$

Note that this determines a mapping $x_1:(\lambda_0-\varepsilon,\lambda_0+\varepsilon) \subset R^1 \to R^n$ such that $(x_1(\lambda),\lambda)$ solves (1.2.1) and $x_1(\lambda) \neq 0$ for $\lambda \neq \lambda_0$. Therefore the mappings $x_1, x_2(t) \equiv 0$ are two distinct branches of eigenvectors of (1.2.1) in R^n through 0. Hence λ_0 is a bifurcation point of (1.2.1) and $(0,\lambda_0)$ is a branch point of (1.2.1).

We now turn to the existence theory for solving (1.2.1). Some of the proofs use the well-known degree theory for mappings in R^n. For a development of this theory and the proofs of results quoted below, we refer, for example, to Ortega and Rheinboldt [1970]. If $D \subset R^n$ is an open, bounded set and $F:\bar{D} \subset R^n \to R^n$ is continuous, then for any $y \notin F(\dot{D})$ the degree of F at y with respect to D is denoted by deg (F,D,y). The fundamental relation between the degree of F and the solvability of $Fx = y$ in D is given by the following famous result:

Theorem 1.2.3 (Kronecker Existence Theorem) Let $F:\bar{D} \subset R^n \to R^n$ be continuous and D an open, bounded set. If $y \notin F(\dot{D})$ and $\deg(F,D,y) \neq 0$, then the equation $Fx = y$ has a solution in D.

Note therefore that if $y \notin F(\bar{D})$, then $\deg(F,D,y) = 0$. A very important result in degree theory states that under certain conditions the degree of a mapping remains constant under homotopic transformation.

Theorem 1.2.4 (Homotopy Invariance Theorem) Let D be an open, bounded set and $H:\bar{D} \times [0,1] \subset R^{n+1} \to R^n$ a given homotopy. Suppose, further, that $y \in R^n$ satisfies $H(x,t) \neq y$ for all $(x,t) \in \dot{D} \times [0,1]$. Then $\deg(H(\cdot,t),D,y)$ is constant for $t \in [0,1]$.

As a consequence of this, we may obtain the following result.

Theorem 1.2.5 (Poincaré-Bohl) Let $F,G:\bar{D} \subset R^n \to R^n$ be two continuous maps and D an open, bounded set. If $y \in R^n$ is any point such that

$$y \notin \{u \in R^n | u = tFx + (1-t)Gx, \ x \in \dot{D}, \ t \in [0,1]\},$$

then $\deg(G,D,y) = \deg(F,D,y)$.

Proof. Consider the homotopy

$$H:\bar{D} \times [0,1] \subset R^{n+1} \to R^n, \ H(x,t) = tFx + (1-t)Gx.$$

Evidently $H(x,t) \neq y$ for $(x,t) \in \dot{D} \times [0,1]$ and hence the result follows directly from Theorem 1.2.4.

The following result will also be useful:

Theorem 1.2.6 Let $F:D \subset R^n \to R^n$ be continuously differentiable on the open set D, and D_0 an open, bounded set with $\bar{D}_0 \subset D$. Assume that $y \notin F(\dot{D}) \cup F(E(\bar{D}_0))$ where $E(\bar{D}_0) = \{x \in \bar{D}_0 | F'(x) \text{ is singular}\}$. Then either $\Gamma = \{x \in D_0 | Fx = y\}$ is empty and $\deg(F,D_0,y) = 0$, or Γ consists of finitely many points $x^1,\ldots,x^m$ and

$$\deg(F,D_0,y) = \sum_{j=1}^{m} \operatorname{sgn} \det F'(x^j). \tag{1.2.2}$$

We now return to the eigenvalue problem. For the case $G \equiv I$ a general existence result for the problem (1.2.1) is given by Riedrich [1968]. For its proof we require some theorems on the extendability of continuous operators.

The following result is a simplified form of the well-known Tietze extension theorem. (See, for example, Dieudonné [1960]).

Theorem 1.2.7 Let $F:\bar{D} \subset R^n \to R^m$ be continuous on the closed set $\bar{D}$. Then there exists a continuous map $G:R^n \to R^m$ defined on all of R^n such that $G(x) = F(x)$ for $x \in \bar{D}$.

With the help of this, we can prove the following form of a homotopy extension theorem of Granas [1961].

Theorem 1.2.8 (Homotopy Extension Theorem) Let $F,G:\bar{D}_0 \subset \bar{D} \subset R^n \to R^n$ be continuous maps on the closed set $\bar{D}_0$. Suppose that there exists a homotopy $H:\bar{D}_0 \times [0,1] \subset R^{n+1} \to R^n$ for which $H(x,0) = Fx$, $H(x,1) = Gx$ for $x \in \bar{D}_0$ and $H(x,t) \neq 0$ for all $(x,t) \in \bar{D}_0 \times [0,1]$. Moreover, assume that F has a continuous extension $\hat{F}$ to all of $\bar{D}$ such that $\hat{F}x \neq 0$ for $x \in \bar{D}$. Then G has a continuous extension $\hat{G}$ to all of $\bar{D}$ which is homotopic to $\hat{F}$ under a homotopy $\hat{H}$ for which $\hat{H}(x,t) \neq 0$ for all $(x,t) \in \bar{D} \times [0,1]$.

Proof. By hypothesis $Fx = H(x,0) \neq 0$ and $Gx = H(x,1) \neq 0$ for all $x \in \bar{D}_0$. Consider the mapping $H^*:\bar{D}_0 \times [0,1] \cup \bar{D} \times \{0\} \subset R^{n+1} \to R^n$ defined by $H^*(x,0) = \hat{F}x$ for $x \in \bar{D}$ and $H^*(x,t) = H(x,t)$ for $(x,t) \in \bar{D}_0 \times [0,1]$. Then H^* is continuous on the closed set $\bar{D}_0 \times [0,1] \cup \bar{D} \times \{0\}$ and it follows from Theorem 1.2.7 that H^* has a continuous extension $\hat{H}^*$ to all of $\bar{D} \times [0,1]$. By the continuity of $\hat{H}^*$ the set

$$D_1 = \{x \in \bar{D} \mid \hat{H}^*(x,t) = 0 \text{ for some } t \in [0,1]\}$$

is closed and, since $\hat{H}^*(x,t) = H^*(x,t) = H(x,t) \neq 0$ for $(x,t) \in \bar{D}_0 \times [0,1]$, we see that $D_1 \cap \bar{D}_0 = \{\emptyset\}$.

Now consider the continuous functional $\psi:\bar{D} \subset R^n \to [0,1]$,

$$\psi(x) = \inf_{y\in D_1} \|x-y\| / (\inf_{y\in D_1} \|x-y\| + \inf_{y\in \bar{D}_0} \|x-y\|).$$

Evidently $\psi(x) = 0$ for $x \in D_1$ and $\psi(x) = 1$ for $x \in \bar{D}_0$. With its help we define the mapping $\hat{H}:\bar{D} \times [0,1] \subset R^{n+1} \to R^n$ by $\hat{H}(x,t) = \hat{H}^*(x,\psi(x)t)$ and set $\hat{G}x = \hat{H}(x,1)$. By its construction $\hat{H}$ is continuous on $\bar{D} \times [0,1]$ and hence is a homotopy on $\bar{D}$. Moreover,

$$\hat{H}(x,0) = \hat{H}^*(x,0) = H^*(x,0) = \hat{F}x \text{ for } x \in \bar{D},$$

and hence $\hat{F}$ is homotopic to $\hat{G}$ on $\bar{D}$. Finally, if $\hat{H}(x,t) = 0$ for some $(x^0,t_0) \in \bar{D} \times [0,1]$, then also $\hat{H}^*(x^0,\psi(x^0)t_0) = 0$. Therefore, x^0 must lie in the set D_1 which implies that $\psi(x^0) = 0$ and hence that $0 = \hat{H}^*(x^0,0) = H^*(x^0,0) = \hat{F}x^0$ contradicting $\hat{F}x \neq 0$ for $x \in \bar{D}$. This completes the proof.

In order to apply this theorem we need to know when $\hat{F}$ exists. Such a result can be obtained with the help of the following concept introduced by Riedrich [1968].

<u>Definition 1.2.9</u> A mapping $F:D \subset R^n \to R^n$ <u>omits a direction on D</u> if there exists a non-zero $y \in R^n$ such that $F(D)$ does not intersect the ray

$$R(y) = \{z \in R^n | z = ty,\ t \geq 0\}.$$

Note that, since $0 \in R(y)$, an operator which omits a direction on D must be non-zero on D.

Riedrich [1968] showed that if a continuous mapping $F:\bar{D} \subset R^n \to R^n$ omits a direction on the boundary of a compact neighborhood $\bar{U}$ of the origin then F has a continuous extension $\hat{F}$ to all of $\bar{U}$ and $\hat{F}x \neq 0$ for all $x \in \bar{U}$. Using the above homotopy extension theorem we obtain the following result.

Lemma 1.2.10 Let $F:\bar{D}_0 \subset \bar{D} \subset R^n \to R^n$ be continuous. If F omits a direction on $\bar{D}_0$, then F has a continuous extension $\hat{F}$ to all of $\bar{D}$ such that $\hat{F}x \neq 0$ for all $x \in \bar{D}$.

Proof. Since F omits a direction on $\bar{D}_0$, there exists a non-zero $y \in R^n$ such that $Fx \neq ty$ for all $t \geq 0$ and all $x \in \bar{D}_0$. Consider the homotopy

$$H:\bar{D}_0 \times [0,1] \subset R^{n+1} \to R^n,\ H(x,t) = (1-t)Fx - ty.$$

Evidently, the mappings F and $G:\bar{D}_0 \subset \bar{D} \subset R^n \to R^n$, $Gx \equiv -y$ are homotopic on $\bar{D}_0$ and $H(x,t) \neq 0$ for all $(x,t) \in \bar{D}_0 \times [0,1]$. Moreover, G has a continuous extension $\hat{G}$ to all of $\bar{D}$ defined by $\hat{G}x \equiv -y$ for $x \in \bar{D}$. Since $\hat{G}x \neq 0$ for $x \in \bar{D}$, Theorem 1.2.8 now implies that F has a continuous extension $\hat{F}$ to all of $\bar{D}$ such that $\hat{F}x \neq 0$ for all $x \in \bar{D}$.

As a direct consequence of this lemma we can now prove Riedrich's [1968] existence result for the eigenvalue problem (1.2.1).

Theorem 1.2.11 Let D be an open, bounded set containing the origin. If the continuous mapping $F:\dot{D} \subset R^n \to R^n$ omits a direction on $\dot{D}$, then there exists a real number $\lambda > 0$ and an $x^* \in \dot{D}$ such that

$$Fx^* = \lambda x^*.$$

Proof. Since $\dot{D} \subset \bar{D}$, Lemma 1.2.10 implies that F has a continuous extension $\hat{F}$ to all of $\bar{D}$ such that $\hat{F}x \neq 0$ for all $x \in \bar{D}$. Therefore it follows from Theorem 1.2.3 that $\deg(\hat{F},D,0) = 0$.

Now suppose that $\hat{F}x = Fx \neq \lambda x$ for all $x \in \dot{D}$ and $\lambda > 0$. Then we see that the homotopy connecting $\hat{F}$ and $-I$,

$$H:\bar{D} \times [0,1] \subset R^{n+1} \to R^n,\ H(x,t) = t\hat{F}x - (1-t)x,$$

satisfies $H(x,t) \neq 0$ for all $(x,t) \in \dot{D} \times [0,1]$. Hence it follows from Theorems 1.2.4 and 1.2.6 that

$$0 = \deg(\hat{F},D,0) = \deg(-I,D,0) = (-1)^n.$$

This contradiction yields the result.

Corollary 1.2.12 Let $A \in L(R^n)$ be nonsingular. Then under the hypothesis of Theorem 1.2.11 there exists a real number $\lambda > 0$ and an $x^* \in \dot{D}$ such that

$$Fx^* = \lambda Ax^*.$$

Proof. Since $A \in L(R^n)$ is nonsingular, A^{-1} exists and $A^{-1}F$ is continuous on $\dot{D}$. Thus we need only show that $A^{-1}F$ omits a direction on $\mathring{D}$. By hypothesis there exists a $y \neq 0$ such that for no $x \in \mathring{D}$ the relation $Fx = ty$ holds for some $t \geq 0$. Clearly then $A^{-1}F$ omits the direction Ay on $\mathring{D}$, and the result is a direct consequence of Theorem 1.2.11.

The following two examples indicate the necessity of all of the hypotheses of Theorem 1.2.11.

Example 1.2.13 Let $F:\mathring{S}(0,1) \subset R^n \to R^n$ be defined by

$$F(x_1,\dots,x_n) = \begin{cases} (1,0,\dots,0)^T, & \text{if } x = \frac{1}{\sqrt{n}}(1,\dots,1)^T \\ (x_1^2,\dots,x_n^2)^T, & \text{otherwise.} \end{cases}$$

Clearly F is not continuous on $\mathring{S}(0,1)$ since

$$F\left(\frac{1}{\sqrt{n}},\dots,\frac{1}{\sqrt{n}}\right) = (1,0,\dots,0)^T \text{ and } Fx \to \frac{1}{n}(1,\dots,1)^T$$

$$\text{for } x \to \frac{1}{\sqrt{n}}(1,\dots,1)^T, \; \|x\| = 1.$$

However, it is readily verified that F omits the direction $-(1,\dots,1)^T$ on $\mathring{S}(0,1)$. Now suppose that $Fx = \lambda x$ for some $x \in \mathring{S}(0,1)$ and $\lambda > 0$. It then follows that $(x_1^2,\dots,x_n^2)^T = \lambda(x_1,\dots,x_n)^T$ and hence that $\lambda = x_1 = \dots = x_n = \pm\frac{1}{\sqrt{n}}$. Since $Fx \neq \lambda x$ for $x = \frac{1}{\sqrt{n}}(1,\dots,1)^T$, we have $x \neq \frac{1}{\sqrt{n}}(1,\dots,1)^T$ and thus $x = -\frac{1}{\sqrt{n}}(1,\dots,1)^T$ and $\lambda = -\frac{1}{\sqrt{n}} < 0$. Thus there is no $\lambda > 0$ and $x \in \mathring{S}(0,1)$ such that $Fx = \lambda x$. This shows that in Theorem 1.2.11 the continuity condition for F on $\mathring{D}$ cannot, in general, be removed.

Example 1.2.14 Let $F:\dot{S}(0,1) \subset R^{2n} \to R^{2n}$ be defined by

$$F(x_1,\ldots,x_{2n}) = (x_2,-x_1,x_4,-x_3,\ldots,x_{2n},-x_{2n-1})^T.$$

Clearly F is continuous on $\dot{S}(0,1)$. Moreover if $y \neq 0$ is any vector in R^{2n}, then with $t = \|y\|^{-1} > 0$ and $x = \|y\|^{-1}(-y_2,y_1,\ldots,-y_{2n},y_{2n-1})^T$ it is easily seen that $x \in \dot{S}(0,1)$ and $Fx = ty$. Hence F omits no direction on $\dot{S}(0,1)$. Now suppose that there is an $x \in \dot{S}(0,1)$ and a $\lambda > 0$ such that $Fx = \lambda x$. Then

$$(x_2,-x_1,\ldots,x_{2n},-x_{2n-1})^T = \lambda(x_1,\ldots,x_{2n})^T$$

and hence

$$x_{2j-1} = -\lambda x_{2j},\ x_{2j} = \lambda x_{2j-1},\ j = 1,\ldots,n,$$

or

$$x_j = -\lambda^2 x_j,\ j = 1,\ldots,2n. \tag{1.2.3}$$

Since $x \in \dot{S}(0,1)$, there is at least one index i, $1 \leqslant i \leqslant 2n$ such that $x_i \neq 0$. But then (1.2.3) implies that $\lambda^2 = -1$ which contradicts $\lambda > 0$.

Note that in this example the dimension of the space was even. The next result shows that for odd-dimensional spaces the requirement in Theorem 1.2.11 that F omit a direction on $\dot{D}$ can be dropped. The following result is due to Poincaré and Brouwer (for references see, for example, Krasnosel'skii [1964; p. 93]).

Theorem 1.2.15 (Hedgehog Theorem) Let $F:\dot{D} \subset R^{2n-1} \to R^{2n-1}$, $n \geq 1$, be continuous and D an open, bounded set containing the origin. Then there is a real number λ and an $x^* \in \dot{D}$ such that $Fx^* = \lambda x^*$.

Proof. By Theorem 1.2.7 F has a continuous extension $\hat{F}$ to all of $\bar{D}$. Assume that $\hat{F}x = Fx \neq \lambda x$ for all $x \in \dot{D}$ and $\lambda \in R^1$. Consider the two homotopies, $H_i:\bar{D} \times [0,1] \subset R^{2n} \to R^{2n-1}$, $i = 1,2$, $H_1(x,t) = t\hat{F}x + (1-t)x$, $H_2(x,t) = t\hat{F}x - (1-t)x$. Then $H_i(x,t) \neq 0$ for all $(x,t) \in \dot{D} \times [0,1]$, $i = 1,2$, and Theorem 1.2.4 implies that

$$1 = \deg(I,D,0) = \deg(\hat{F},D,0) = \deg(-I,D,0) = (-1)^{2n-1} = -1.$$

Since this is impossible, the proof is complete.

Corollary 1.2.16 Let $A \in L(R^{2n-1})$ be nonsingular. Then under the hypotheses of Theorem 1.2.15 there is a real number λ and an $x^* \in \dot{D}$ such that $Fx^* = \lambda Ax^*$.

Proof. Since $A \in L(R^{2n-1})$ is nonsingular, A^{-1} exists and $A^{-1}F$ is continuous. Hence the result follows trivially from Theorem 1.2.15.

The next result provides us with a general existence result for the problem (1.2.1). It can be found in Krasnosel'skii [1964] and contains Theorem 1.2.15 as a special case. Unfortunately its practical usefulness appears to be limited due to the difficulty in calculating the degree of a mapping.

Theorem 1.2.17 Let $F,G:D \subset R^n \to R^n$ be continuous on the compact set $\bar{D}_0 \subset D$ where D_0 is open and suppose that $Fx \neq 0$, $Gx \neq 0$, for all $x \in \dot{D}_0$. If

$$\deg(F,D_0,0) \neq \deg(-G,D_0,0),$$

then there exists a point $x^* \in \dot{D}_0$ such that $Fx^* = \lambda Gx^*$ with $\lambda > 0$.

Proof. By Theorem 1.2.5 we have

$$0 \in \{u \in R^n | u = tFx - (1-t)Gx,\ x \in \dot{D}_0,\ t \in [0,1]\}.$$

Thus there exists a point $x^* \in \dot{D}_0$ and a number $\hat{t} \in [0,1]$ such that

$$\hat{t}Fx^* - (1-\hat{t})Gx^* = 0.$$

Clearly $\hat{t} \in (0,1)$, for otherwise $Fx^* = 0$ or $Gx^* = 0$ contradicting the hypothesis. Therefore $Fx^* = \lambda Gx^*$ with $\lambda = \frac{1-\hat{t}}{\hat{t}} > 0$.

As a consequence of this result, several corollaries can be phrased which provide information about the eigenvalues of F.

Corollary 1.2.18 Let $F:D \subset R^n \to R^n$ be continuous on the compact set $\bar{D}_0 \subset D$, where D_0 is open and contains the origin. If for some fixed μ, $Fx \neq \mu x$ for all $x \in \dot{D}_0$ and $\deg(\mu I-F,D_0,0) \neq 1$, then F has an eigenvalue $\lambda > \mu$ with a corresponding nontrivial eigenvector on $\dot{D}_0$.

Proof. Taking $G \equiv -I$ in Theorem 1.2.17, we have $\deg(-G,D_0,0) = 1$. Hence it follows that there is an $x^0 \in \dot{D}_0$ and $\lambda_0 > 0$ such that $\mu x^0 - Fx^0 = -\lambda_0 x^0$. Moreover, $0 \notin \dot{D}_0$ implies that $x^0 \neq 0$ and hence $Fx^0 = \lambda x^0$, $\lambda = \lambda_0 + \mu > \mu$.

For the special case $\mu = 0$ we have $Fx \neq 0$ for all $x \in \dot{D}_0$ and $\deg(F,D_0,0) \neq (-1)^n$ implies that F has a positive eigenvalue. Note that in Theorem 1.2.11, the extension $\hat{F}$ of F to all of $\bar{D}_0$ satisfies $0 = \deg(\hat{F},D_0,0) \neq (-1)^n$ and hence Corollary 1.2.18 could be used to prove that theorem.

Corollary 1.2.19 Let $F:D \subset R^n \to R^n$ be continuous on the compact set $\bar{D}_0 \subset D$ with D_0 open and $0 \notin \bar{D}_0$. If for some fixed μ, $Fx \neq \mu x$ for all $x \in \dot{D}_0$ and $\deg(\mu I-F,D_0,0) \neq 0$, then F has eigenvalues λ_1,λ_2 with $\lambda_2 < \mu < \lambda_1$ for which the corresponding nontrivial eigenvectors are on $\dot{D}_0$.

Proof. Since $0 \notin \bar{D}_0$, $\deg(I,D_0,0) = 0$ and hence by Theorem 1.2.17, $\mu x^1 - Fx^1 = -\lambda_0 x^1$, $x^1 \in \dot{D}_0$ with $\lambda_0 > 0$. Thus $Fx^1 = \lambda_1 x^1$, $\lambda_1 = \lambda_0 + \mu > \mu$. Because $\deg(-I,D_0,0) = 0$ also holds, we obtain $\mu x^2 - Fx^2 = \lambda_0' x^2$, $x^2 \in \dot{D}_0$ with $\lambda_0' > 0$. Hence $Fx^2 = \lambda_2 x^2$, $\lambda_2 = \mu - \lambda_0' < \mu$. Clearly $0 \notin \bar{D}_0$ implies that $x^1 \neq 0$ and $x^2 \neq 0$.

We now turn to the question of the existence of proper continuous branches of eigenvectors of (1.2.1) in R^n through a point $x \in R^n$. The discussion will also answer in part some of the uniqueness questions. For results in this area we shall assume the differentiability of F and G although weaker conditions would also be possible. Our discussion follows that given by Krasnosel'skii [1964]. We first prove a perturbation result implicitly contained in Krasnosel'skii [1964].

Theorem 1.2.20 Let $F,G:D \subset R^n \to R^n$ be continuously differentiable on the open set D, and suppose that for some fixed $y \in R^n$ there exists a solution $x^0 \in D$ of $Fx = y$ at which $F'(x^0)$ is nonsingular. Then there exists a ball $\bar{S}(x^0,\delta)$, $\delta > 0$ and a number $\varepsilon_0 > 0$ such that for any ε with $|\varepsilon| \leq \varepsilon_0$ the equation

$$Fx + \varepsilon Gx = y \tag{1.2.4}$$

has a unique solution $x(\varepsilon)$ in $\bar{S}(x^0,\delta)$ and $x(\varepsilon)$ depends continuously on ε with $x(0) = x^0$.

Proof. By assumption there exists an $\alpha > 0$ such that $\|F'(x^0)h\| \geq \alpha\|h\|$ for all $h \in R^n$. Using the openness of D, the continuous differentiability of F and the implicit function theorem, we can choose $\delta > 0$ such that
(i) $\bar{S}(x^0,\delta) \subset D$, (ii) x^0 is the unique solution of $Fx = y$ in $\bar{S}(x^0,\delta)$ and (iii) $\|F'(x)-F'(x^0)\| \leq \frac{\alpha}{2}$ for $x \in \bar{S}(x^0,\delta)$. Now let $\beta > 0$ be such that $\|Fx-y\| \geq \beta$ for all $x \in \dot{S}(x^0,\delta)$ and set

$$\gamma = \sup_{x\in\dot{S}(x^0,\delta)} \|Gx\|, \quad \eta = \sup_{x\in\bar{S}(x^0,\delta)} \|G'(x)\|$$

and $\varepsilon_0 = \frac{1}{2}\min\{\beta/\gamma, \alpha/2\eta\}$. Consider the homotopy

$$H:\bar{S}(x^0,\delta) \times [0,1] \subset R^{n+1} \to R^n, \quad H(x,t) = t(Fx+\varepsilon Gx)+(1-t)Fx$$

with $|\varepsilon| \leq \varepsilon_0$. Then for $(x,t) \in \dot{S}(x^0,\delta) \times [0,1]$ we have

$$\|H(x,t)-y\| \geq |\|Fx-y\|-t|\varepsilon|\|Gx\|| \geq \beta - (\frac{\beta}{2\gamma})\gamma > 0.$$

Therefore, by Theorems 1.2.4 and 1.2.6

$$(1.2.5) \qquad \pm 1 = \deg(F,S(x^0,\delta),y) = \deg(F+\varepsilon G,S(x^0,\delta),y).$$

Hence Theorem 1.2.3 implies that (1.2.4) has at least one solution in $\bar{S}(x^0,\delta)$. Moreover, for any $x \in \bar{S}(x^0,\delta)$ and $h \neq 0$

$$\|(F'(x)+\varepsilon G'(x))h\| \geq |\|F'(x^0)h\|-\|(F'(x^0)-F'(x))h\|-|\varepsilon|\|G'(x)h\||$$

$$\geq (\alpha - \frac{\alpha}{2} - (\frac{\alpha}{4\eta})\eta)\|h\| > 0$$

which shows that $F'(x) + \varepsilon G'(x)$ is nonsingular. But then by the implicit function theorem every solution of (1.2.4) in $\bar{S}(x^0,\delta)$ is isolated. Hence the compactness of $\bar{S}(x^0,\delta)$ implies that there are only finitely many solutions $x^1,\ldots,x^m$ of (1.2.4) in $\bar{S}(x^0,\delta)$.

Now consider the homotopies

$$H_j:\bar{S}(0,\delta) \times [0,1] \subset R^{n+1} \to R^n, \quad j = 1,\ldots,m,$$

$$H_j(h,t) = (1-t)F'(x^0)h + t(F'(x^j)+\varepsilon G'(x^j))h.$$

Then for $(h,t) \in \dot{S}(0,\delta) \times [0,1]$

$$\|H_j(h,t)\| \geq |\|F'(x^0)h\|-t\|(F'(x^j)-F'(x^0))h\|-t|\varepsilon|\|G'(x^j)h\||$$

$$\geq (\alpha - \frac{\alpha}{2}-(\frac{\alpha}{4\eta})\eta)\|h\| > 0$$

which, applying Theorems 1.2.4 and 1.2.6, shows that for $j = 1,\ldots,m$,

$$\deg(F'(x^j)+\varepsilon G'(x^j),S(0,\delta),0) = \deg(F'(x^0),S(0,\delta),0) = \zeta$$

where $\zeta = \pm 1$. Again using Theorem 1.2.6 and (1.2.5) we obtain

$$\pm 1 = \deg(F+\varepsilon G,S(x^0,\delta),y) = \sum_{j=1}^{m} \operatorname{sgn}\det\,(F'(x^j)+\varepsilon G'(x^j))$$
$$= \sum_{j=1}^{m} \deg(F'(x^j)+\varepsilon G'(x^j),S(0,\delta),0) = m\zeta$$

which implies that $m = 1$. Finally, the continuity of F and G guarantees that the unique solution $x(\varepsilon)$ of (1.2.4) in $\bar{S}(x^0,\delta)$, $|\varepsilon| \leq \varepsilon_0$ depends continuously on ε and that $\lim_{\varepsilon\to 0} x(\varepsilon) = x(0) = x^0$.

On the basis of this theorem we can prove the following result of Krasnosel'skii [1964].

Theorem 1.2.21 Given $F,G:D \subset R^n \to R^n$ defined on the open set D, let $x^0 \in D$ be a nontrivial eigenvector of (1.2.1) with corresponding eigenvalue λ_0. Further suppose that F and G are continuously differentiable in some open neighborhood $U \subset D$ of x^0. If $F'(x^0) - \lambda_0 G'(x^0)$ is nonsingular, then the spectrum of (1.2.1) contains some interval $[\lambda_0-\varepsilon,\lambda_0+\varepsilon]$ with $\varepsilon > 0$. Moreover, there exists a ball $\bar{S}(x^0,\delta) \subset U$ such that for each $\lambda \in (\lambda_0-\varepsilon,\lambda_0+\varepsilon)$ there is a unique nontrivial eigenvector $x(\lambda)$ in $\bar{S}(x^0,\delta)$ of (1.2.1) and the mapping $x = x(\lambda), \lambda\in(\lambda_0-\varepsilon,\lambda_0+\varepsilon)$ is a proper continuous branch of eigenvectors of (1.2.1) in R^n through x^0.

Proof. Consider the mapping $H = F - \lambda_0 G$ and its perturbation $H + (\lambda_0-\lambda)G$. Clearly H is continuously differentiable on U and hence by our hypothesis $H'(x^0) = F'(x^0) - \lambda_0 G'(x^0)$ is nonsingular. Therefore Theorem 1.2.20 applies to the perturbed equation $Hx + (\lambda_0-\lambda)Gx = 0$. Consequently a ball $\bar{S}(x^0,\delta) \subset U$, $\delta > 0$, and an $\varepsilon_1 > 0$ can be found such that for all λ with $|\lambda-\lambda_0| \leq \varepsilon_1$ the equation

$$Fx - \lambda Gx = Hx + (\lambda_0-\lambda)Gx = 0$$

has a unique solution $x(\lambda)$ in $\bar{S}(x^0,\delta)$. Thus we have $Fx - \lambda_0 Gx \neq 0$ for all $x \neq x^0$, $x \in \bar{S}(x^0,\delta)$ and hence

$$0 \neq Fx(\lambda) - \lambda_0 Gx(\lambda) = (\lambda-\lambda_0)Gx(\lambda) \text{ for } \lambda \neq \lambda_0,\ |\lambda-\lambda_0| \leq \varepsilon_1$$

which implies that $Gx(\lambda) \neq 0$. Moreover by Theorem 1.2.20 $x(\lambda)$ depends continuously on λ. Therefore since $x^0 \neq 0$, we can choose $\varepsilon \leq \varepsilon_1$ such that $x(\lambda) \neq 0$ for $|\lambda-\lambda_0| < \varepsilon$ and hence $x(\lambda)$ is a nontrivial eigenvector of (1.2.1) for $|\lambda-\lambda_0| < \varepsilon$.

Now suppose $x(\lambda_1) = x(\lambda_2)$ for some $\lambda_1,\lambda_2 \in (\lambda_0-\varepsilon,\lambda_0+\varepsilon)$, then

$$\begin{aligned} 0 &= Fx(\lambda_1) - \lambda_1 Gx(\lambda_1) = Fx(\lambda_2) - \lambda_1 Gx(\lambda_2) \\ &= Fx(\lambda_2) - \lambda_2 Gx(\lambda_2), \end{aligned}$$

and hence

$$0 = (\lambda_1-\lambda_2)Gx(\lambda_2).$$

Thus $Gx(\lambda_2) \neq 0$ implies $\lambda_1 = \lambda_2$ and thus the mapping $x = x(\lambda)$ is also one-one on $(\lambda_0-\varepsilon,\lambda_0+\varepsilon)$. Hence x is a proper branch of eigenvectors of (1.2.1) in R^n through x^0.

One can now easily see that the point (x^0,λ_0) of Theorem 1.2.21 is a regular point of (1.2.1). Indeed under the assumptions of Theorem 1.2.21 with continuously differentiable $F:R^n \to R^n$ and $G \equiv I$, Pimbley [1969] has shown that a proper continuous branch of eigenvectors $x(\lambda)$ of (1.2.1) can be extended maximally. He showed that the extension process could be carried out at least until a value of λ is reached for which $\lambda I - F'(x(\lambda))$ does not have an inverse. Using the same reasoning for the setting described by Theorem 1.2.21, it is possible to show that $x(\lambda)$ can be extended at least until a value of λ is reached for which either $x(\lambda)$ leaves U or $F'(x(\lambda)) - \lambda G'(x(\lambda))$ is singular.

Note that Theorem 1.2.21 and the subsequent discussion directly apply to the problem that is of most interest to us here, namely, when $Gx = Ax$ for $x \in R^n$ and $A \in L(R^n)$ is symmetric and positive definite. We also note that the conditions of Theorem 1.2.21 are not necessary for the existence of a continuous branch. However, in the applications to be discussed in the second half of Chapter III, we shall assume the conditions of Theorem 1.2.21 since other possibilities of guaranteeing the continuation of solutions appear to lead to numerical difficulties due to unbounded terms in a related differential equation and the nonuniqueness of solutions. For some results in this direction, we refer to Pimbley [1969] and Berger and Berger [1968].

For the remainder of the section we consider only the following special case of (1.2.1):

$$(1.2.6) \qquad Fx - \lambda Ax = 0,\ x \in D,\ \lambda \in R^1,$$

where $A \in L(R^n)$ is nonsingular. The next series of results show that even if $F'(x^0) - \lambda_0 A$ is singular there may still exist a branch of eigenvectors of (1.2.6) in R^n through x^0.

Theorem 1.2.22 Let $F:D \subset R^n \to R^n$ be defined on the open set D containing the origin. Suppose that $F0 = 0$, F is differentiable at 0 and $A \in L(R^n)$ is nonsingular. Then all of the bifurcation points of (1.2.6) are eigenvalues of $A^{-1}F'(0)$.

Proof. Suppose that λ_0 is not an eigenvalue of $A^{-1}F'(0)$. Then we can choose $\alpha > 0$ such that

$$\|(A^{-1}F'(0)-\lambda_0 I)x\| \geq \alpha\|x\|, \quad \forall x \in R^n,$$

and $\delta > 0$ such that

$$\|A^{-1}Fx-A^{-1}F'(0)x\| \leq \frac{\alpha}{3}\|x\|, \text{for } \|x\| \leq \delta.$$

Let $\varepsilon = \min\{\delta, \frac{\alpha}{3}\}$, then for $|\lambda-\lambda_0| \leq \varepsilon$, $\|x\| < \varepsilon$,

$$\|A^{-1}Fx-\lambda x\| \geq \big| \|(A^{-1}F'(0)-\lambda_0 I)x\| - \|A^{-1}F'(0)-A^{-1}Fx\| - |\lambda-\lambda_0|\|x\|\big|$$

$$\geq \frac{\alpha}{3}\|x\|.$$

Consequently, there is no nonzero solution of $Fx = \lambda Ax$ with $|\lambda-\lambda_0| \leq \varepsilon$ and $\|x\| \leq \delta$. In other words, λ_0 is not a bifurcation point of (1.2.6).

Although all bifurcation points of (1.2.6) are eigenvalues of $A^{-1}F'(0)$, the converse is not true as an example of Krasnosel'skii [1964] shows. However, a simple modification of a proof given by Krasnosel'skii [1964] yields the following result which we state without proof.

<u>Theorem 1.2.23</u> Let $F:D \subset R^n \to R^n$ be defined on the open set D containing the origin. Suppose that $F0 = 0$, F is differentiable at 0, and $A \varepsilon L(R^n)$ is nonsingular. Then each eigenvalue of $A^{-1}F'(0)$ of odd multiplicity is a bifurcation point of (1.2.6), and to this bifurcation point (if one exists) there corresponds a branch of eigenvectors of $A^{-1}F$ in R^n through 0.

1.3 The Variational Problem

In this section we consider the eigenvalue problem (1.2.1) under the added assumption that F and G are potential operators. For this case a basic result of Ljusternik [1934] assures the existence of eigenvalues. More specifically, we are concerned with finding nontrivial solutions of the problem

$$(1.3.1) \qquad f'(x)^T - \lambda g'(x)^T = 0, \; x \in D, \; \lambda \in R^1$$

where $f,g:D \subset R^n \to R^1$ are given functionals on a domain D. Clearly (1.3.1) is of the same form as (1.2.1) with $F = f'^T$ and $G = g'^T$; in particular, (1.2.6) is obtained for $F = f'^T$ and $g:R^n \to R^1$, $g(x) = \frac{1}{2} x^T Ax - \zeta$, with fixed $\zeta > 0$ and symmetric and nonsingular $A \in L(R^n)$. Hence all of the results of Section 1.2 apply to (1.3.1).

We now consider solving (1.3.1) as a special case of a minimization problem. The following definitions are standard.

Definition 1.3.1 Let $f:D \subset R^n \to R^1$ and a set $C \subset D$ be given. A point $x^0 \in C$ is a relative minimum of f on C if there exists a neighborhood $U(x^0)$ of x^0 in D such that

$$(1.3.2) \qquad f(x) \geq f(x^0), \quad \forall x \in U(x^0) \cap C.$$

Definition 1.3.2 Let $f:D \subset R^n \to R^1$ and $G:D \subset R^n \to R^m$ be defined on the open set D and define a set $C \subset D$ by

$$(1.3.3) \qquad C = \{x \in D | Gx = 0\}.$$

Further suppose that f and G are differentiable at $x^0 \in C$. Then x^0 is a conditional critical point of f on C if there exists a $b \in R^m$ such that

$$(1.3.4) \qquad f'(x^0)^T = G'(x^0)^T b.$$

The following result of Ljusternik [1934] characterizes the relative minima of f on C for a special class of sets C. Our proof follows Ljusternik and Sobolev [1955].

Theorem 1.3.3 Let $f:D \subset R^n \to R^1$ be defined on the open set D. Suppose that $G:D \subset R^n \to R^m$, $1 \leq m < n$ is continuously differentiable on D and let C be given by (1.3.3). If x^0 is a relative minimum of f on C and if f is differentiable at x^0 and rank $G'(x^0) = m$, then x^0 is a conditional critical point of f on C.

Proof. Consider the linear space

$$T = \{u \in R^n | G'(x^0)u = 0\} \tag{1.3.5}$$

and its orthogonal complement $T^{\perp}$ and denote the orthogonal projection from R^n onto T by P. Let $U(x^0)$ be as in Definition 1.3.1 and choose a fixed nonzero $u \in T$ arbitrarily. Then there exists a vector $y \in R^n$ such that $Py = u$ and, for sufficiently small $t_0 > 0$, we have, by the openness of $U(x^0)$, that $x_t = x^0 + ty \in U(x^0)$ whenever $0 \leq |t| < t_0$. Moreover, $x_t - x^0$ can be written uniquely as

$$x_t - x^0 = P(x_t - x^0) + v, \ v \in T^{\perp}$$

and therefore

$$x_t = x^0 + tu + v, \ v \in T^{\perp}, \ 0 \leq |t| < t_0. \tag{1.3.6}$$

We consider the mapping

$$\hat{G}(t,v) = Gx_t = G(x^0+tu+v),\ v \in T^{\perp},\ 0 \leq |t| < t_0.$$

It follows from (1.3.5) that, for any $h \in T^{\perp}$, $\partial_2\hat{G}(0,0)h = G'(x^0)h = 0$ implies $h = 0$ and hence that $\partial_2\hat{G}(0,0)$ is nonsingular on $T^{\perp}$. Since $\hat{G}(0,0) = Gx^0 = 0$, the implicit function theorem implies that there exists a $\delta > 0$ and a function $v:[-\delta,\delta] \subset R^1 \to R^n$ such that

$$x_t = x^0 + tu + v(t) \in U(x^0), \quad \forall t \in [-\delta,\delta],$$

$$v(0) = 0,\ v(t) \in T^{\perp}$$

and

$$\hat{G}(t,v(t)) = G(x^0+tu+v(t)) = 0, \quad \forall t \in [-\delta,\delta].$$

Therefore $x_t \in U(x^0) \cap C$, $\forall t \in [-\delta,\delta]$. Moreover, v is differentiable at $t = 0$ and so by the chain rule we obtain

(1.3.7) $$0 = G'(x^0)(u+v'(0)) = G'(x^0)v'(0).$$

Now, $v(0) = 0$ implies that $\lim_{t\to 0} v(t)/t = v'(0)$. Also, it follows from $v(t) \in T^{\perp}$, $|t| \leq \delta$ that $w^Tv(t) = 0$, $\forall w \in T$, $|t| \leq \delta$ and hence that $w^Tv'(0) = 0$, $\forall w \in T$, that is, $v'(0) \in T^{\perp}$. Together with (1.3.5) and (1.3.7) we see that $v'(0) = 0$. Consequently, for all sufficiently small $\varepsilon > 0$ we have

(1.3.8) $$\|v(t)\|/|t| = \|v(t)-v(0)-v'(0)t\|/|t| \leq \varepsilon,$$

$$0 < |t| \leq \delta(\varepsilon) \leq \delta.$$

Now consider the one-dimensional mapping $\psi:[-\delta,\delta] \subset R^1 \to R^1$, $\psi(t) = f(x^0+tu+v(t))$. Then

$$\frac{1}{|t|} |\psi(t)-\psi(0)-(f'(x^0)u)t|$$

$$\leq \frac{1}{|t|} \{|f(x^0+tu+v(t))-f(x^0)-f'(x^0)(tu+v(t))| + |f'(x^0)v(t)|\}$$

$$\leq \|tu+v(t)\|^{-1} |f(x^0+tu+v(t))-f(x^0)-f'(x^0)(tu+v(t))| [\|u\|+\|v(t)\|/|t|]$$

$$+ \|f'(x^0)\| \|v(t)\|/|t|.$$

Therefore, it follows from (1.3.8) and the differentiability of f at x^0 that

$$\psi'(0) = f'(x^0)u.$$

Because $x_t = x^0 + tu + v(t) \in U(x^0) \cap C$, $\forall t \in [-\delta,\delta]$, we now conclude that $f(x_t) \geq f(x^0)$, $\forall t \in [-\delta,\delta]$. Thus $t = 0$ is a local minimum of ψ and hence $\psi'(0) = 0$.

Hence $f'(x^0)u = 0$ and, since u was an arbitrary nonzero element of T, we must have $f'(x^0)^T \in T^\perp$. Therefore, since $T^\perp$ is spanned by the row vectors of $G'(x^0)$, there must be a vector $b \in R^m$ such that (1.3.4) holds.

As a simple corollary we obtain the following result for (1.3.1).

<u>Corollary 1.3.4</u> Let $f:D \subset R^n \to R^1$ be given on the open set D. Suppose that $g:D \subset R^n \to R^1$ is continuously differentiable on D and define the set C by

$$C = \{x \in D | g(x) = 0\}.$$

Let $x^0 \in C$ be a relative minimum of f on C and assume that f is differentiable at x^0. If $g'(x^0)^T \neq 0$, then there exists a real number λ such that

$$f'(x^0)^T = \lambda g'(x^0)^T.$$

If $A \in L(R^n)$ is symmetric and nonsingular and $g: R^n \to R^1$, $g(x) = \frac{1}{2} x^T Ax - \zeta$ where $\zeta > 0$ is fixed, then $C = \{x \in D | \frac{1}{2} x^T Ax = \zeta\}$, $g'(x)^T \neq 0$ on C and $A^{-1}f'(x^0)^T = \lambda x^0$. Because of its importance later on we formulate this case as a separate corollary.

Corollary 1.3.5 Let $f: D \subset R^n \to R^1$ be given on the open set D. Suppose that $A \in L(R^n)$ is symmetric and nonsingular and let $C = \{x \in D | \frac{1}{2} x^T Ax = \zeta\}$ with fixed $\zeta > 0$. If $x^0 \in C$ is a relative minimum of f on C and f is differentiable at x^0, then there is a real number λ such that

$$A^{-1}f'(x^0)^T = \lambda x^0 \tag{1.3.9}$$

We again consider the existence of a proper branch of eigenvectors through x^0. It turns out that for the variational problem (1.3.1) we can obtain a result similar to Theorem 1.2.21 under simpler assumptions. The following theorem can be found in Krasnosel'skii [1964].

Theorem 1.3.6 Let $f: D \subset R^n \to R^1$ be differentiable on $\bar{S}(0,\rho_0) \subset D$ where D is open and $\rho_0 > 0$. Then $f'^T: \bar{S}(0,\rho_0) \subset R^n \to R^n$ has a nontrivial eigenvector on each $\dot{S}(0,\rho)$, $0 < \rho \leq \rho_0$. Moreover, with $A = I$, (1.3.9) has a proper branch of eigenvectors in $\bar{S}(0,\rho_0)$.

Proof. Since $\dot{S}(0,\rho)$, $0 < \rho \leq \rho_0$ is compact and f is continuous on $\bar{S}(0,\rho_0)$, there is an $x(\rho) \in \dot{S}(0,\rho)$ such that $f(x(\rho)) = \inf_{x\in\dot{S}(0,\rho)} f(x)$. Hence Corollary 1.3.5, with $A = I$, implies that $f'(x(\rho))^T = \lambda(\rho)x(\rho)$ with $x(\rho) \in \dot{S}(0,\rho)$ for some $\lambda(\rho) \in R^1$. Clearly $\rho_1 \neq \rho_2$ implies $x(\rho_1) \neq x(\rho_2)$ and hence the mapping $x = x(\rho)$, $\rho \in (0,\rho_0)$, is a proper branch of eigenvectors of (1.3.9) in $\bar{S}(0,\rho_0)$.

This leads to the following corollary which, in a Hilbert space setting, is due to Golomb [1934].

Corollary 1.3.7 Let $f:R^n \to R^1$ be differentiable and suppose that $A \in L(R^n)$ is symmetric and positive definite. Then $A^{-1}f'^T:R^n \to R^n$ has a nontrivial eigenvector on each $\dot{S}(0,\rho)$, $0 < \rho < \infty$. Moreover, (1.3.9) has a proper branch of eigenvectors in R^n.

Proof. Clearly the functional $h:R^n \to R^1$, $h(x) = f(A^{-1/2}x)$ is well-defined and differentiable. Thus, by Theorem 1.3.6 h'^T has an eigenvector on each $\dot{S}(0,\rho)$, $0 < \rho < \infty$. That is, there is a $y(\rho) \in \dot{S}(0,\rho)$ and $\lambda(\rho) \in R^1$ such that $h'(y(\rho))^T = \lambda(\rho)y(\rho)$. Using the chain rule it is readily verified that

$$h'(y(\rho))^T = A^{-1/2}f'(A^{-1/2}y(\rho))^T = \lambda(\rho)y(\rho).$$

Hence $A^{-1}f'(A^{-1/2}y(\rho))^T = \lambda(\rho)A^{-1/2}y(\rho)$ and, setting $x(\rho) = A^{-1/2}y(\rho)$, we obtain $A^{-1}f'(x(\rho))^T = \lambda(\rho)x(\rho)$.

Finally, we close this section with a result of Krasnosel'skii [1964] on bifurcation points which we state again without proof.

Theorem 1.3.8 Let $F:D \subset R^n \to R^n$ be defined on the open set D containing the origin. Suppose that $F0 = 0$ and that F is differentiable in a neighborhood of 0. Further, assume that F is the gradient of a functional $f:D \subset R^n \to R^1$. Then every eigenvalue of $F'(0)$ is a bifurcation point of (1.3.9) with $A = I$, and hence (1.3.9) has a branch of eigenvectors in R^n through 0.

CHAPTER II

A Convergence Theory for a Class of Nonlinear Programming Problems

2.1 Preliminaries

Consider the nonlinear eigenvalue problem

$$(2.1.1)\qquad f'(x)^T - \lambda g'(x)^T = 0,\ x \in D,\ \lambda \in R^1,$$

where $f,g:D \quad R^n \to R^1$ are continuously differentiable on the open set D. As a result of Corollary 1.3.4 (Ljusternik [1934]) we may replace (2.1.1) by a constrained minimization problem of the form

$$(2.1.2)\qquad \min \{f(x) \mid g(x) = 0,\ x \in D\},$$

and thus we are led to consider methods for solving such problems.

In this chapter we will study a general class of nonlinear programming problems irrespective of its application to our non-linear eigenvalue problem, and we begin with a short review of basic results on unconstrained minimization. Suppose that the functional $f:D \subset R^n \to R^1$ is to be minimized and that f has a uniformly continuous derivative on the open set D. Let x^0 be any point in D and define $L^0(f(x^0))$ to be the connected component of the level set

$$(2.1.3)\qquad L(f(x^0)) = \{x \in D \mid f(x) \leqslant f(x^0)\}$$

containing x^0. We shall usually write simply L^0 instead of $L^0(f(x^0))$. Assume now that f is bounded below on L^0 and that

L^0 is closed. The latter assumption implies that L^0 is a proper subset of D.

Recently Elkin [1968] obtained convergence results for minimization sequences of the form

(2.1.4) $$x^{k+1} = x^k - \omega_k \tau_k p^k, \; k = 0,1,\ldots$$

(2.1.5) $$f(x^k) \geq f(x^{k+1}), \; k = 0,1,\ldots \; .$$

Here p^k is a suitable search direction, τ_k a basic steplength, and ω_k a relaxation factor. Elkin's analysis is based on the observation that the final convergence statement

(2.1.6) $$\lim_{k\to\infty} x^k = x^*, \; f'(x^*)^T = 0$$

can be proved by showing the validity of a sequence of intermediate conditions, some of which depend essentially only on the steplength algorithm and some only on the procedure for selecting the direction.

Suppose that $x^k \in L^0$ and consider any direction p^k for which

(2.1.7) $$f'(x^k)p^k > 0, \; \|p^k\| = 1.$$

Then it is easily seen that $f(x^k) \geq f(x^k - \tau p^k)$ and hence that $x^k - \tau p^k \in L^0$ for $\tau \in [0,\delta]$ with some $\delta > 0$. This is the basis for showing, for certain steplength algorithms, that (2.1.5) holds and $x^{k+1} \in L^0$. As a next step, it is proved for these steplength algorithms, under the condition (2.1.7), that

(2.1.8) $$\lim_{k\to\infty} f'(x^k)p^k = 0.$$

Now the particular algorithm for choosing p^k is taken into account in order to conclude from (2.1.8) that

(2.1.9) $$\lim_{k\to\infty} f'(x^k)^T = 0.$$

Finally additional conditions on f, and sometimes also another intermediate condition such as

(2.1.10) $$\lim_{k\to\infty} (x^k - x^{k+1}) = 0,$$

allow the proof of (2.1.6) from (2.1.9).

For the verification of condition (2.1.8), Elkin [1968] introduced the concept of forcing functions. In order to extend his results to the constrained case, we generalize the definition to higher dimensions.

Definition 2.1.1 A functional $\Phi: [0,\infty)\times\ldots\times[0,\infty) \subset R^m \to R^1$ is a forcing function (F-function) of m variables if for any m sequences $\{t_i^k\} \subset [0,\infty)$, $i = 1,\ldots,m$

(2.1.11) $$\lim_{k\to\infty} \Phi(t_1^k,\ldots,t_m^k) = 0 \text{ implies } \lim_{k\to\infty} t_i^k = 0 \text{ for at least one } i,\ 1 \leqslant i \leqslant m.$$

In the case $m = 1$ we will usually call Φ simply an F-function. Any function $\mu: [0,\infty) \to [0,\infty)$ which is bounded away from zero on

$[\alpha,\infty)$ for $\alpha > 0$, or which is isotone and satisfies $\mu(t) > 0$ for $t > 0$ is an F-function. Moreover, the sum, product, and composition of any two (isotone) F-functions is again an (isotone) F-function. Clearly if $\Phi:[0,\infty)\times\ldots\times[0,\infty) \subset R^m \to R^1$ is an F-function of m variables, then $\mu:[0,\infty) \to [0,\infty)$, $\mu(t) = \Phi(t,\ldots,t)$ is an F-function. Furthermore, the product $\Phi(t_1,\ldots,t_m) = \phi_1(t_1)\cdots\phi_m(t_m)$, of m F-functions $\phi_i:[0,\infty) \to [0,\infty)$, $i = 1,\ldots,m$, is an F-function of m variables, and if Φ_1,Φ_2 are F-functions of m variables, so is Φ, $\Phi(t_1,\ldots,t_m) = \min\{\Phi_1(t_1,\ldots,t_m),\Phi_2(t_1,\ldots,t_m)\}$.

The validity of (2.1.7) for a specific steplength algorithm is always obtained by first demonstrating that

(2.1.12) $$f(x^k) - f(x^{k+1}) \geq \mu(f'(x^k)p^k),\ k = 0,1,\ldots$$

for some F-function μ. This leads us to the following generalization of the "principle of sufficient decrease " (Elkin [1968]).

Lemma 2.1.2 Suppose that $f:D \subset R^n \to R^1$ is G-differentiable and bounded below on some set $D_0 \subset D$ and that a given sequence $\{x^k\}$ remains in D_0. Suppose that there are m associated sequences $\{t_i^k\} \subset [0,\infty)$, $i = 1,\ldots,m$, such that

(2.1.13) $$f(x^k) - f(x^{k+1}) \geq \Phi(t_1^k,\ldots,t_m^k),\ k = 0,1,\ldots,$$

for some F-function Φ of m variables. Then $\lim_{k\to\infty} t_i^k = 0$ for at least one i, $1 \leq i \leq m$.

Proof. Since f is bounded below on D_0 and (2.1.13) implies that $f(x^k) \geq f(x^{k+1})$, it follows that $\lim_{k\to\infty} (f(x^k)-f(x^{k+1})) = 0$, and hence, by the definition of F-functions of m variables, that $\lim_{k\to\infty} t_i^k = 0$ for at least one i, $1 \leq i \leq m$.

Note that for the case $m = 1$ this lemma applies immediately to (2.1.12) with $t_1^k = f'(x^k)p^k$ and hence under the conditions of Lemma 2.1.2, (2.1.12) implies that $\lim_{k\to\infty} f'(x^k)p^k = 0$. Connected with this result is a "comparison principle" also due to Elkin [1968]. Suppose that we have two different steplength algorithms I and II, and that at a point x^k the application of I and II yields x_I^{k+1} and x_{II}^{k+1}, respectively. If $\{x_I^k\}$ satisfies (2.1.13), then in order to obtain (2.1.13) for $\{x_{II}^k\}$ it suffices to show that

$$f(x_{II}^{k+1}) \leq f(x_I^{k+1}),$$

for we have

$$f(x^k) - f(x_{II}^{k+1}) \geq f(x^k) - f(x_I^{k+1}) \geq \Phi(t_1^k,\ldots,t_m^k).$$

We shall see that also in the constrained case this principle will be useful for proving results for several well-known steplength algorithms.

As mentioned earlier, in order to conclude (2.1.9) from (2.1.8) the particular choice of direction p^k must be analyzed. A very important class of directions $\{p^k\}$, the so-called gradient related directions, are defined by the inequality

(2.1.14) $$f'(x^k)p^k \geq \mu(\|f'(x^k)^T\|), \quad k = 0,1,\ldots$$

where μ is some F-function. For these directions clearly (2.1.9) is a direct consequence of (2.1.8). On the other hand, there are other classes of directions, such as those used in univariate relaxation methods, for which (2.1.14) does not hold. In that case (2.1.9) must be deduced from (2.1.8) with the help of specific properties of the sequence $\{p^k\}$ and some additional assumptions about f.

For several steplength algorithms, the condition (2.1.10) is a consequence of the proof of (2.1.8). In other cases it can be proved with the help of the following two concepts due to Elkin [1968]. We use the terminology of Ortega and Rheinboldt [1970].

Definition 2.1.3 A functional $f:D \subset R^n \to R^1$ is hemivariate on a set $D_0 \subset D$ if it is not constant on any line segment of D_0; that is, if there exist no distinct points $x,y \in D_0$ such that

$$(1-t)x + ty \in D_0 \text{ and } f([1-t]x+ty) = f(x), \quad \forall t \in [0,1].$$

Definition 2.1.4 Given $f:D \subset R^n \to R^1$, a sequence $\{x^k\}$ in some set $D_0 \subset D$ is strongly downward in D_0 if

$$(1-t)x^k + tx^{k+1} \in D_0, \quad \forall t \in [0,1]$$

and

$$f(x^k) \geq f([1-t]x^k+tx^{k+1}) \geq f(x^{k+1}), \quad \forall t \in [0,1].$$

The indicated result of Elkin [1968] concerning (2.1.10) can now be stated as follows (see Ortega and Rheinboldt [1970]).

Lemma 2.1.5 Suppose that the functional $f:D \subset R^n \to R^1$ is continuous and hemivariate on a compact set $D_0 \subset D$. Then every strongly downward sequence $\{x^k\} \subset D_0$ satisfies (2.1.10).

Finally for the proof of the actual convergence statement (2.1.6) a result of Ostrowski [1966] plays a central role. Briefly it states that, if $f:D \subset R^n \to R^1$ is continuously differentiable on a compact set $D_0 \subset D$ and $\{x^k\} \subset D_0$ is any sequence which satisfies (2.1.9), then the set of critical points Ω of f in D_0 is not empty and $\lim_{k\to\infty} [\inf_{x\in\Omega}\|x^k-x\|] = 0$. In particular, if Ω consists of only one point x^*, then (2.1.6) holds. We shall obtain a minor extension of this result in Section 2.2. It may also be mentioned that if Ω consists of only a finite number of points and if, in addition to (2.1.9), (2.1.10) holds, then we may still conclude (2.1.6). For the proof we refer to Ortega and Rheinboldt [1970].

We close this section with a discussion of two important F-functions which will be needed in proving many of the results in the succeeding sections.

Definition 2.1.6 Let $f:D \subset R^n \to R^1$ be continuously differentiable on D and $D_0 \subset D$. Then the function $\omega:[0,\infty) \to [0,\infty)$,

$$\omega(t) = \sup \{\|f'(x)^T - f'(y)^T\| \,|\, x,y \in D_0,\ \|x-y\| \leq t\},$$

is called the modulus of continuity of $f':D \subset R^n \to L(R^n,R^1)$ on D_0.

Clearly, ω is isotone and $\omega(0) = 0$. Moreover, if D_0 is also convex and f' is uniformly continuous on D_0, then the modulus of continuity of f' on D_0 is well-defined and uniformly continuous on $[0,\infty)$ (see Ortega and Rheinboldt [1970]). If f' is not constant on the convex set D_0, then it can be shown that the function $\eta:[0,\infty) \to [0,\infty)$,

$$\eta(t) = \int_0^1 \omega(ts)ds \tag{2.1.15}$$

is well-defined and strictly isotone and hence is an F-function.

Definition 2.1.7 Let $f:D \subset R^n \to R^1$ be continuously differentiable and assume that on some set $D_0 \subset D$

$$\alpha = \sup \{\|f'(x)-f'(y)\| \,|\, x,y \in D_0\} > 0.$$

Then the mapping $\delta:[0,\infty) \to [0,\infty)$,

$$\delta(t) = \begin{cases} \inf \{\|x-y\| \,|\, x,y \in D_0, \|f'(x)-f'(y)\| \geq t\}, & \text{if } t \in [0,\alpha) \\ \lim\limits_{s\to\alpha^-} \delta(s), & \text{if } t \in [\alpha,\infty) \end{cases}$$

is the reverse modulus of continuity of $f':D \subset R^n \to L(R^n,R^1)$ on D_0.

As in the case of ω, clearly δ is isotone and $\delta(0) = 0$. It can be shown that if f' is uniformly continuous, then $\delta(t) > 0$ for all $t > 0$ and hence δ is an F-function (see Ortega and Rheinboldt [1970]).

2.2 The Constrained Minimization Problem

We turn now to methods for the solution of a certain class of nonlinear programming problems. In general, any such problem has the form

$$(2.2.1) \qquad \min \{f(x) \mid x \in C\}$$

where $f:D \subset R^n \to R^1$ is a given functional and $C \subset D$ a specific constraint set. In this section we discuss the conditions which will be placed on f and the constraint set C.

The underlying assumptions on f are essentially the same as those made at the beginning of Section 2.1. Specifically, let $f:D \subset R^n \to R^1$ be differentiable on the open set D. Moreover, for a given set C and $x^0 \in C$, denote again by $L^0(f(x^0))$ the connected component of the level set $L(f(x^0))$ defined by (2.1.3). Then we shall assume that f is bounded below on $L^0 \cap C$.

For the remainder of this chapter C will always stand for the constraint set

$$(2.2.2) \qquad C = \{x \in D \mid g_j(x) \leq 0,\ j \in J_0\}$$

where $g_j:D \subset R^n \to R^1$, $j \in J_0 = \{1,\ldots,m\}$, are given functionals on the open set D. The precise properties of the g_j, $j \in J_0$, will always be specified in each result. In most cases we will need continuous differentiability. However, without exception we shall always assume the three conditions:

(2.2.3)

(i) $int(C) = \{x \in D | g_j(x) < 0, \ \forall j \in J_0\} \neq \{\emptyset\}$

(ii) Every point of the set $\{x \in C | g_j(x) = 0 \text{ for some } j \in J_0\}$ is an accumulation point of $int(C)$

(iii) $L^0 \cap C$ is closed.

The second condition excludes situations of the type shown in Fig. 1(a), while the third excludes those of the type shown in Fig. 1(b).

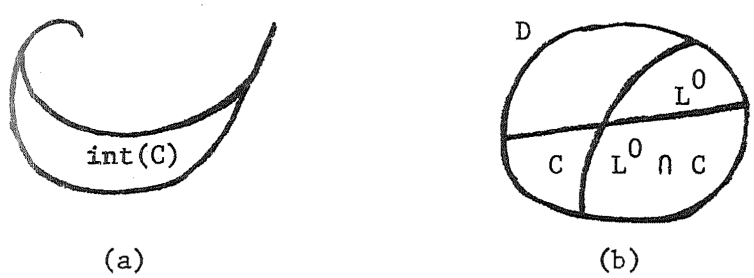


Fig. 1.

For any $x \in C$ and $\varepsilon \geq 0$ we define the index set

$$J(x,\varepsilon) = \{j \in J_0 | -\varepsilon \leq g_j(x) \leq 0\}$$

and use the standard notation $|J(x,\varepsilon)|$ for the cardinality of $J(x,\varepsilon)$. For any index set $J \subset J_0$ we denote by J^N or J^L the subsets of indices $j \in J$ for which g_j is a nonlinear or linear functional, respectively. Many of our results will be based on the following regularity condition for C.

Definition 2.2.1 Let $g_j : D \subset R^n \to R^1$, $j \in J_0$ be differentiable on the open set D. The constraint set C is regular if there exists an $\bar{\varepsilon} > 0$ independent of $x \in C$ such that

$$\text{(2.2.4)} \qquad \text{rank } (g_j'(x)^T | j \in J(x,\bar{\varepsilon})) = q < n, \quad \forall x \in C$$

whenever $q = |J(x,\bar{\varepsilon})| > 0$.

We note that if C is compact then we need only assume (2.2.4) with $\bar{\varepsilon} = 0$ since the existence of an $\bar{\varepsilon} > 0$ can then be proved directly. Further, our particular regularity assumption is mainly for convenience of proof and similar results can be obtained under other types of assumptions. For example, in the case of convex constraints, that is, when the functionals g_j, $j \in J_0$ are convex, Zoutendijk [1960] assumes that for each $j \in J_0^N$ there is an $x^j \in C$ such that $g_j(x^j) < 0$. Another regularity condition which is more general than ours is given by Altman [1964] as follows: If $\sum_{j \in J(x,0)} \nu_j g_j'(x)^T = 0$, $\nu_j \geq 0$, then $\nu_j = 0$ for $j \in J^N(x,0)$.

If C is regular in the sense of Definition 2.2.1, then it can be shown that for any $x \in C$ there is a nonzero $s \in R^n$ and a $\tau_1 > 0$ such that

$$x - \tau s \in C, \quad \forall \tau \in [0,\tau_1].$$

Any such vector s is called a feasible direction at x for the constraint set C. Indeed, a small modification of the proof of Lemma 2.4.5 shows that there is a nonzero $s \in R^n$ such that

$$g_j'(x)s > 0, \ j \in J^N(x,0), \ g_j'(x)s \geq 0, \ j \in J^L(x,0),$$

and hence, by the mean-value theorem, that s is feasible.

We now rephrase Definition 1.3.2 in terms of the constraint set C.

Definition 2.2.2 Let $f:D \subset R^n \to R^1$ and $g_j:D \subset R^n \to R^1$, $j \in J_0$ be differentiable on the open set D. A point x of the constraint set C is a conditional critical point of f on C if there exist real numbers ν_j, $j \in J(x,0) \subset J_0$ called multipliers, such that

$$(2.2.5) \qquad f'(x)^T = \begin{cases} \sum_{j \in J(x,0)} \nu_j g_j'(x)^T, & \text{if } J(x,0) \neq \{\emptyset\} \\ 0 & \text{, otherwise.} \end{cases}$$

Note that without a regularity condition on C a point $x \in C$ could be a conditional critical point of every functional in R^n, for example, if the vectors $g_j'(x)^T$, $j \in J(x,0)$ are linearly dependent.

Two important results which characterize solutions of (2.2.1) in terms of conditional critical points were given by Kuhn and Tucker [1951]. We present them here without proof in a generalized form given by Mangasarian [1969].

Theorem 2.2.3 (Kuhn-Tucker Sufficient Optimality Theorem) Let $f:D \subset R^n \to R^1$ be pseudo-convex and differentiable, and $g_j:D \subset R^n \to R^1$, $j \in J_0$ be quasi-convex and differentiable on the open convex set D. Assume that for some $x^* \in C$ there exist multipliers $\nu_j, j \in J(x^*,0)$ such that

$$f'(x^*)^T = \sum_{j \in J(x^*,0)} \nu_j g_j'(x^*)^T, \quad \nu_j \leq 0.$$

Then x^* is a relative minimum of f on C.

Since the level sets of quasi-convex functionals are convex, the constraint set C is, in this case, convex. As a consequence, the usual existence and uniqueness results for minima of f on C remain the same as in the unconstrained case. For a discussion of the characterizations of minima of f in terms of the various types of convexity, see Elkin [1968], Mangasarian [1969], or Ortega and Rheinboldt [1970].

In order for the conditions of Theorem 2.2.3 to be necessary for a point x^* to be a relative minimum of f on C we must place an added qualification on the constraint set C. A common one was given by Kuhn and Tucker [1951].

Suppose that $g_j: D \subset R^n \to R^1$, $j \in J_0$ are differentiable functionals on the open set D. Then the Kuhn-Tucker constraint qualification is satisfied at $x^0 \in C$ if for each $u \in R^n$ with $g_j'(x^0)u \leqslant 0$, $j \in J(x^0,0)$ there exists a function $x:[0,\hat{\delta}] \to R^n$ which is differentiable at 0 and satisfies $x(0) = x^0$, $x(t) \in C$ for all $t \in [0,\hat{\delta}]$, with some $\hat{\delta} > 0$, as well as $\left(\frac{dx}{dt}\right)_{t=0} = \lambda u$ for some $\lambda > 0$.

Theorem 2.2.4 (Kuhn-Tucker Necessary Optimality Theorem) Let $f:D \subset R^n \to R^1$ and $g_j:D \subset R^n \to R^1$, $j \in J_0$ be differentiable on the open set D and suppose that x^* is a relative minimum of f on the constraint set C. Assume further that the Kuhn-Tucker constraint qualification holds at x^*. Then there exist multipliers ν_j, $j \in J(x^*,0)$ such that

$$f'(x^*)^T = \sum_{j \in J(x^*,0)} \nu_j g_j'(x^*)^T, \quad \nu_j \leqslant 0, \ j \in J(x^*,0).$$

The following lemma allows us to apply this result to our particular problem.

Lemma 2.2.5 Let $g_j : D \subset R^n \to R^1$, $j \in J_0$ be continuously differentiable on the open set D and suppose that C is regular. Then the Kuhn-Tucker constraint qualification holds at every point in C.

Proof. Let x be any point in C. If $x \in \mathrm{int}(C)$, then $J(x,0)$ is empty and the result is trivial. Suppose therefore that $J(x,0)$ is not empty and that $u \in R^n$ is any nonzero vector such that $g_j'(x)u \leq 0$, $j \in J(x,0)$. Let $G_1 : D \subset R^n \to R^{q_1}$ and $G_2 : D \subset R^n \to R^{q_2}$, $q_1 + q_2 = |J(x,0)|$ be defined by

$$G_1 x = (g_j(x)),\ j \in J(x,0),\ g_j'(x)u = 0$$

$$G_2 x = (g_j(x)),\ j \in J(x,0),\ g_j'(x)u < 0.$$

Since C is regular, it follows that rank $G_1'(x) = q_1$, and furthermore, that $u \in T = \{y \in R^n | G_1'(x)y = 0\}$. Hence, by the first part of the proof of Theorem 1.3.3 we see that for some $\delta > 0$ there is a differentiable arc

$$x : [0,\delta] \subset R^1 \to R^n,\ x(t) = x + tu + v(t),\ v(t) \in T^{\perp},\ v(0) = v'(0) = 0$$

such that

$$G_1(x(t)) = 0, \quad \forall t \in [0,\delta],$$

and clearly,

$$x(0) = x \text{ and } \left(\frac{dx(t)}{dt} \right)_{t=0} = u.$$

Because $G_2'(x),u < 0$ it follows from the continuity of g_j', $j \in J_0$ that there is a $\delta_1 \in (0,\delta]$ such that

$$G_2'(x(t))u < 0 \text{ for all } t \in [0,\delta_1].$$

Now consider the mapping

$$b:[0,\delta_1] \subset R^1 \to R^{q_2},\ b(t) = G_2(x(t)).$$

Clearly b is differentiable and thus, by the mean-value theorem, it follows that, for $t \in (0,\delta_1]$,

$$b(t) = b(0) + b'(\hat{t})t,\ \hat{t} \in (0,t)$$

$$= 0 + G_2'(x(\hat{t}))ut < 0.$$

Finally, from the continuity of g_j, $j \in J_0$ it is evident that there is a $\hat{\delta} \in (0,\delta_1]$ such that

$$g_j(x(t)) \leq 0,\ j \in J_0 \sim J(x,0),\quad \forall t \in [0,\hat{\delta}].$$

Therefore $x(t) \in C$, for all $t \in [0,\hat{\delta}]$ and the Kuhn-Tucker constraint qualification holds.

We now consider iterative methods of the form

(2.2.6) $$x^{k+1} = x^k - \omega_k \tau_k s^k,\ k = 0,1,\ldots$$

(2.2.7) $$f(x^k) \geq f(x^{k+1}),\ k = 0,1,\ldots,$$

for solving (2.2.1). Our interest is in generalizing the Elkin theory for unconstrained minimization to the constrained case, and accordingly we shall follow the same basic procedure as described in Section 2.1. Since the constraint set C may not contain any zeros of $f'(x)^T$, it is evident that (2.1.6) and (2.1.9) need to be modified appropriately and as a consequence of Theorem 2.2.4 we are led to use

(2.2.8) $$\lim_{k\to\infty} x^k = x^*, \quad f'(x^*)^T - \sum_{j\in J(x^*,0)} \nu_j^* g_j'(x^*)^T = 0$$

for some multipliers ν_j^*, $j \in J(x^*,0)$ and

(2.2.9) $$\lim_{k\to\infty} [f'(x^k)^T - \sum_{j\in J^k} \nu_j^k g_j'(x^k)^T] = 0$$

for some sequence of multipliers $\{\nu_j^k, j \in J^k \subset J_0\}_{k=0}^{\infty}$, respectively. This leads us to the following minor extension of Ostrowski's result mentioned in Section 2.1.

Lemma 2.2.6 Let $f: D \subset R^n \to R^1$ and $g_j: D \subset R^n \to R^1$, $j \in J_0$ be continuously differentiable on a compact set $C_0 \subset C$ and suppose that $\{x^k\} \subset C_0$ is any sequence which satisfies (2.2.9) for some bounded sequence of multipliers $\{\nu_j^k, j \in J^k \subset J_0\}_{k=0}^{\infty}$. Then, for some index set $J \subset J_0$ and multipliers ν_j, $j \in J$, the set

$$\Omega = \{x \in C_0 \mid f'(x)^T - \sum_{j\in J} \nu_j g_j'(x)^T = 0\}$$

of conditional critical points of f on C is not empty and

$$\lim_{k\to\infty} [\inf_{x\in\Omega} \|x^k - x\|] = 0.$$

In particular, if Ω consists of a single point x^*, then (2.2.8) holds.

<u>Proof</u>. Since C_0 is compact, $\{x^k\}$ has a convergent subsequence $\{x^{k_i}\}$ with, say, $\lim_{i\to\infty} x^{k_i} = x$. Since $\{J^{k_i}\} \subset J_0$ and $\{\nu_j^{k_i}, j \in J^{k_i}\}$ is bounded, we may refine $\{x^{k_i}\}$ so that $J^{k_i} = J$, $\forall\, i \geq 0$, and $\lim_{i\to\infty} \nu_j^{k_i} = \nu_j$, $j \in J$, for some limit set J of $\{J^{k_i}\}$. Then, by the continuity of f' and g_j', $j \in J_0$, $f'(x)^T - \sum_{j\in J} \nu_j g_j'(x)^T = 0$, and hence, $x \in \Omega$. Now, let $\delta_k = \inf_{x\in\Omega} \|x^k - x\|$ and suppose that $\lim_{i\to\infty} \delta_{k_i} = \delta$. Then, since $\{x^{k_i}\}$ must have a convergent subsequence whose limit is in Ω, it follows that $\delta = 0$, which proves the result.

We turn now to the choice of the direction vectors. Since we want the iterates of (2.2.6) to remain in $L^0 \cap C$, clearly the choice of the direction sequence $\{p^k\}$ of (2.1.4) must be suitably restricted. In fact, in the following sections we will consider directions, denoted by s^k, for which it is not only guaranteed that the iterates remain in $L^0 \cap C$ but for which it is possible to obtain an appropriate estimate on the amount of decrease at each step. For this purpose we will need to place another condition upon C.

Let $t_1 = -\max_{j\in J_0} \inf_{x\in C} g_j(x)$, then $t_1 > 0$ since $\mathrm{int}(C) \neq \{\emptyset\}$. Define the mapping $\gamma: [0,\infty) \to [0,\infty)$ by

$$(2.2.10) \quad \gamma(t) = \begin{cases} \min_{j\in J_0} \inf\{\|x-y\| \,|\, x,y \in C,\ g_j(x) = 0,\ g_j(y) = -t\}, & \text{if } t \in [0,t_1) \\ \lim_{s\to t_1^-} \gamma(s), \text{ if } t \in [t_1,\infty). & \end{cases}$$

It follows from the continuity of g_j, $j \in J_0'$ that γ is well-defined, nondecreasing and $\gamma(0) = 0$. The next result gives conditions for γ to be an F-function.

Lemma 2.2.7 Let $g_j : D \subset R^n \to R^1$, $j \in J_0$ be continuous on the open set D. If either

$$(2.2.11) \quad \begin{cases} |g_j(x)-g_j(y)| \leq \beta \|x-y\|, \ \forall x,y \in C \text{ such that} \\ g_j(x) = 0,\ g_j(y) = -t \text{ for } j \in J_0 \text{ and } t \in [0,t_1], \end{cases}$$

or if C is compact, then $\gamma : [0,\infty) \to [0,\infty)$ defined by (2.2.10) is an F-function.

Proof. Suppose that $0 < t \leq t_1$ and that (2.2.11) holds. Then for $j \in J_0$, and $x,y \in C$ such that $g_j(x) = 0$ and $g_j(y) = -t$, we have

$$\|x-y\| \geq \beta^{-1} |g_j(x)-g_j(y)| = t\beta^{-1} > 0$$

independent of x and y. Hence this inequality holds also for the infimum, and since J_0 is finite, it follows that $\gamma(t) > 0$.

Now suppose that C is compact and that $\gamma(t) = 0$ for some $t \in (0,t_1]$. Since J_0 is finite, there exist a $j \in J_0$ and sequences $\{x^k\}$, $\{y^k\} \subset C$ such that $g_j(x^k) = 0$ and $g_j(y^k) = -t$ for all k and $\lim_{k\to\infty} x^k = x^* \in C$, $\lim_{k\to\infty} y^k = y^* \in C$ as well as $\lim_{k\to\infty} \|x^k-y^k\| = 0$. Therefore $x^* = y^*$ and from the continuity of g_j

$$0 = g_j(x^*) = g_j(y^*) = -t < 0,$$

contradicting the choice of $t > 0$. Hence γ is a nondecreasing function for which $t > 0$ implies $\gamma(t) > 0$ and thus γ is an F-function.

Since we will need γ to be an F-function in order to obtain estimates of the type (2.1.13), we introduce the following definition.

Definition 2.2.8 Let $g_j : D \subset R^n \to R^1$, $j \in J_0$ be continuous on the open set D. Then the constraint set C is admissible if $\gamma : [0,\infty) \to [0,\infty)$ defined by (2.2.10) is an F-function.

In the subsequent sections of this chapter we will present our generalization of Elkin's convergence theory as discussed in Section 2.1. As a result, we will be able to generate some new algorithms for solving (2.2.1) and to obtain complete convergence proofs for some well-known methods, namely, Zoutendijk's [1960; procedure P1, p. 73] method of feasible directions and Rosen's [1960] gradient projection method with linear constraints. In their convergence proofs the separation of steplength and direction analysis is not made clear and hence the applicability of other steplength algorithms to their methods does not follow immediately. We should note that Cannon, Cullum and Polak [1970] have also given proofs of these methods which proceed in essentially the same manner as Zoutendijk [1960] and Rosen [1960]. Topkis and Veinott [1967] have obtained some results on the separation of steplength and direction analysis for minimizations over a class of convex sets. Their results contain a proof of Zoutendijk's method for the special case of convex polyhedral constraints; but these results are not as general as ours because the steplength analysis used requires that the direction sequence $\{s^k\}$ have a uniformly feasible subsequence in the following sense: If $\{x^k\}$ and $\{s^k\}$ are the sequences of (2.2.6), then it is assumed that there are subsequences $\{x^{k_j}\}$ and $\{s^{k_j}\}$ such that $\lim_{j\to\infty} x^{k_j} = x^*$, $\lim_{j\to\infty} s^{k_j} = s^*$ and that for some $\delta > 0$, $x^{k_j} - \tau s^{k_j} \in C$ for all $j \geq 0$ and all $\tau \in (0,\delta]$. This condition is rather strong since in general the limiting direction of a subsequence converging to a solution of the problem need not even be feasible as the following example in R^2 shows.

Example 2.2.9 Let $f,g:R^2 \to R^1$ be given by $f(x_1,x_2) = -x_2$ and $g(x_1,x_2) = \|x\|_2^2 - 1$. Then applying Zoutendijk's [1960] method (P1) to the problem

$$\min \{f(x) \mid g(x) \leq 0\}$$

produces a sequence of directions whose limiting direction is tangent to the unit ball at $(0,1)^T$. Clearly such a direction is not feasible.

We will replace the condition of uniform feasibility by certain weaker assumptions which allow greater freedom in the choice of direction and steplength. Moreover, by eliminating the convexity requirements our results apply to a larger class of nonlinear programming problems.

2.3 Basic Steplength Analysis

Let x^0 be some initial point in C and $x \in L^0(f(x^0)) \cap C \equiv L^0 \cap C$ any given point. In this section we will show that for certain well-known steplength algorithms and any suitably chosen $s \in R^n$ with $f'(x)s \geq 0$, $\|s\| = 1$, the next iterate $x - \omega\tau s$ is contained in $L^0 \cap C$ and satisfies

$$(2.3.1) \qquad f(x) - f(x-\omega\tau s) \geq \Phi(f'(x)s,\sigma,\varepsilon)$$

with some $\sigma > 0$ and $\varepsilon > 0$. Here ω is a relaxation parameter, τ is the steplength determined by the particular algorithm and $\Phi: [0,\infty) \times [0,\infty) \times [0,\infty) \subset R^3 \to [0,\infty)$ is some F-function of three variables. Among the steplength algorithms considered here will be the constrained analogues of a minimization procedure, and of the Curry [1944] and Altman [1966] algorithms. We will also investigate how certain steplength algorithms such as the Curry one-step Newton method and the methods due to Ostrowski [1966], Goldstein [1964], [1965], [1966] and Armijo [1966] can be modified to apply to constrained minimization. Our presentation will follow that of Ortega and Rheinboldt [1970].

We first discuss the setting in which a suitable direction $s \in R^n$ will be chosen. Clearly we want s to be feasible at $x \in C$, in other words, we need to know that there exists a $\tau_1 > 0$ such that $x - \tau s \in C$ for $\tau \in [0,\tau_1]$. Let $\mu_j: [0,\infty) \to [0,\infty)$, $j \in J_0^N$ be given F-functions with $\mu_j(0) = 0$, and, for any $x \in C$, index set $J \subset J_0$ and $\sigma \in [0,\infty)$, consider the set

$$(2.3.2) \quad K(x,J,\sigma) = \{s \in R^n \,\big|\, \|s\| = 1,\ g_j'(x)s \geq \mu_j(\sigma),\ j \in J^N,$$

$$g_j'(x)s \geq 0,\ j \in J^L\}.$$

Then the following lemma of Zoutendijk [1960] provides a necessary condition for feasibility.

Lemma 2.3.1 Let $g_j: D \subset R^n \to R^1$, $j \in J_0$ be continuously differentiable on the open set D. If $s \in R^n$, $\|s\| = 1$ is feasible at $x \in C$; then necessarily $s \in K(x,J(x,0),0)$.

Proof. If $J(x,0) = \{\emptyset\}$, then clearly $K(x,J(x,0),0) = R^n$ and the result is trivial. Suppose therefore that $J(x,0)$ is not empty and that $s \notin K(x,J(x,0),0)$, $\|s\| = 1$. Then for at least one $j \in J(x,0)$ we have $g_j'(x)s < 0$ and by the continuity of g_j' there is an interval $[0,\tau_1)$, $\tau_1 > 0$ such that

$$g_j'(x-\tau s)s < 0, \quad \forall\, \tau \in [0,\tau_1).$$

Thus, by the mean-value theorem, it follows that for any $\tau \in (0,\tau_1)$,

$$g_j(x-\tau s) = g_j(x) - \tau g_j'(x-\tau_2 s)s$$

$$= -\tau g_j'(x-\tau_2 s)s > 0, \quad \tau_2 \in (0,\tau)$$

or $x - \tau s \notin C$ which contradicts the feasibility of s.

In general, the condition $s \in K(x,J(x,0),0)$ is not sufficient for s to be feasible at $x \in C$. Moreover, it does not suffice to know only the feasibility of s at $x \in C$ in order to guarantee a sufficient decrease in the value of f at each step. For that we need to obtain, for each $x \in C$ and any suitably chosen feasible direction s at x, an estimate for the quantity $\tau_1 > 0$ in the definition of feasibility. This leads to the following result which plays a key role in the remainder of this chapter. As discussed in Section 2.1, ω_j, $j \in J_0$ denotes the modulus of continuity of g_j', $j \in J_0$ on D and η_j, $j \in J_0$, the corresponding quantities of (2.1.15).

Lemma 2.3.2 Let $g_j : D \subset R^n \to R^1$, $j \in J_0$, have uniformly continuous derivatives on the open convex set D and assume that the constraint set C is admissible. Suppose that $x \in C$ and $s \in K(x, J(x,\varepsilon), \sigma)$ for some $\varepsilon > 0$ and $\sigma > 0$. If $\tau^* = \sup \{\tau \geq 0 | x - ts \in C,\ t \in [0,\tau)\}$, then either

$$\tau^* \geq \hat{\Phi}(\sigma,\varepsilon) > 0$$

where $\hat{\Phi} : [0,\infty) \times [0,\infty) \to [0,\infty)$ is a forcing function of two variables which depends only on C, or

$$\tau^* = \hat{\tau} = \sup \{t \geq 0 | x - ts \in D\} > 0.$$

Proof. If J_0 is empty, then $C = D$ and $\tau^* = \hat{\tau}$. Suppose therefore that J_0 is not empty. For any $j \in J^N(x,\varepsilon)$ the modulus of continuity ω_j of g_j' on D is not identically zero and hence η_j defined by (2.1.15) is a continuous, strictly isotone F-function with $\eta_j(0) = 0$. Consequently, $\eta_j^{-1} : [0,\infty) \to [0,\infty)$ exists and has the same properties as η_j. Since D is open, clearly $\hat{\tau} = \sup \{t \geq 0 | x - ts \in D\} > 0$ and for any $\tau \in [0,\hat{\tau})$ the mean-value theorem shows that

$$g_j(x-\tau s) - g_j(x) + \tau g_j'(x)s$$

$$= \int_0^1 [-\tau g_j'(x-\zeta\tau s)s + \tau g_j'(x)s]d\zeta$$

$$= \tau\int_0^1 [g_j'(x) - g_j'(x-\zeta\tau s)]sd\zeta$$

$$\leqslant \tau\int_0^1 \omega_j(\zeta\tau)d\zeta = \tau\eta_j(\tau).$$

Now $s \in K(x,J(x,\varepsilon),\sigma)$ and $j \in J^N(x,\varepsilon)$ imply that $g_j'(x)s \geqslant \mu_j(\sigma)$ and hence, together with $g_j(x) \leqslant 0$, we see that

$$g_j(x-\tau s) \leqslant \tau(\eta_j(\tau)-\mu_j(\sigma)), \quad \forall\tau \in [0,\hat{\tau}),\ j \in J^N(x,\varepsilon).$$

Since $\sigma > 0$, it follows that

$$\tau_j = \sup\{\tau \in [0,\hat{\tau}) | \eta_j(\tau) \leqslant \mu_j(\sigma)\},\ j \in J^N(x,\varepsilon)$$

is well-defined and positive. Moreover, by definition we have either $\tau_j = \eta_j^{-1}(\mu_j(\sigma))$ or $\tau_j = \hat{\tau}$ and therefore

$$g_j(x-\tau s) \leqslant 0, \quad \forall\tau \in [0,\tau_j),\ j \in J^N(x,\varepsilon).$$

Now consider any index $j \in J^L(x,\varepsilon)$. Since $s \in K(x,J(x,\varepsilon),\sigma)$ and $g_j(x) \leqslant 0$, it follows that $h_j^T s \geqslant 0$, $h_j \equiv g_j'(x)^T$, and, again by the mean-value theorem, that

$$g_j(x-\tau s) = g_j(x) - \tau h_j^T s \leqslant 0, \quad \forall\tau \in [0,\hat{\tau}),\ j \in J^L(x,\varepsilon).$$

Finally, for all the remaining indices j, we clearly have

$$g_j(x-\tau s) \leq 0, \quad \forall \tau \in [0,\gamma(\varepsilon)],\ j \in J_0 \sim J(x,\varepsilon).$$

Since at least one of the sets $J^N(x,\varepsilon)$, $J^L(x,\varepsilon)$ and $J_0 \sim J(x,\varepsilon)$ is not empty, we obtain altogether

$$(2.3.3) \qquad g_j(x-\tau s) \leq 0, \quad \forall \tau \in [0, \min\{\hat{\tau}, \hat{\Phi}(\sigma,\varepsilon)\}), \quad \forall j \in J_0$$

where $\hat{\Phi}:[0,\infty) \times [0,\infty) \to [0,\infty)$ is given by

$$(2.3.4) \qquad \hat{\Phi}(t_1,t_2) = \begin{cases} \min\{\min_j \{\eta_j^{-1}(\mu_j(t_1)) \mid j \in J_0^N\}, \gamma(t_2)\}, & \text{if } J_0^N \neq \{\emptyset\} \\ \gamma(t_2), & \text{otherwise} \end{cases}$$

If J_0^N is not empty, $\min_j \{\eta_j^{-1}(\mu_j(\cdot)) \mid j \in J_0^N\} = \eta_{j_0}^{-1}(\mu_{j_0}(\cdot))$ is clearly an F-function, and by the admissibility assumption also γ is an F-function and therefore $\hat{\Phi}$ is in all cases an F-function of two variables. Clearly $\sigma > 0$, $\varepsilon > 0$ implies that $\hat{\Phi}(\sigma,\varepsilon) > 0$ and hence, by (2.3.3), that $x - \tau s \in C$ for $\tau \in [0,\min\{\hat{\tau},\hat{\Phi}(\sigma,\varepsilon)\})$. Therefore, we have shown that either $\tau^* \geq \hat{\Phi}(\sigma,\varepsilon) > 0$ or $\tau^* = \hat{\tau} > 0$ and thus that $\min\{\hat{\tau},\hat{\Phi}(\sigma,\varepsilon)\}$ is an estimate on the quantity τ_1 in the definition of feasibility.

This result is basic in proving the following generalization of Elkin's [1968] fundamental lemma.

Lemma 2.3.3 Let $f:D \subset R^n \to R^1$ be continuously differentiable and assume that $g_j:D \subset R^n \to R^1$, $j \in J_0$ have uniformly continuous derivatives on the open convex set D and that C is admissible. For any $x \in L^0 \cap C$ and some $\sigma > 0$, $\varepsilon > 0$ let $s \in K(x,J(x,\varepsilon),\sigma)$ be such that $f'(x)s > 0$. Then $L^0 \cap C$ contains an open interval $(x, x-\tau_0 s)$ with some $\tau_0 > 0$. Moreover, if $\tau_1 > 0$ is any number such that

(2.3.5) $\quad f(x-\tau s) < f(x), \quad \forall\ x - \tau s \in (x, x-\tau_1 s] \cap L^0 \cap C$

then $[x, x-\tau_1 s] \cap C \subset L^0 \cap C$.

Proof. By Lemma 2.3.2

$$\tau^* = \sup \{\tau \geq 0 \mid x - ts \in C,\ t \in [0,\tau)\} \geq \min \{\hat{\tau}, \hat{\Phi}(\sigma, \varepsilon)\} > 0$$

where $\hat{\Phi}$ is given by (2.3.4). Since D is open and $f'(x)s > 0$, the continuity of f' implies that there is a $\tau_0 \in (0, \tau^*)$ such that $f'(x-\tau s)s > 0$ for all $\tau \in [0, \tau_0)$. By the mean-value theorem we conclude that $f(x-\tau s) < f(x)$ for all $x - \tau s \in (x, x-\tau_0 s)$ and hence that (2.3.5) holds with $\tau_1 = \min \{\tau_0, \hat{\Phi}(\sigma, \varepsilon)\} > 0$.

Suppose that the closed set $[x, x-\tau_1 s] \cap L^0 \cap C$ is a proper subset of $[x, x-\tau_1 s] \cap C$. Then $[x, x-\tau_1 s] \cap L^0 \cap C = [x,y]$ for some $y \in \dot{L}^0$, $y \neq x$ and hence, by (2.3.5), $f(y) < f(x)$. But if $y \in \dot{L}^0$ and $x \in L^0 \cap C$, then $f(y) \geq f(x) > f(y)$ which is a contradiction.

This result contains that of Elkin [1968] for $C \equiv D$ except that Elkin did not require D to be convex. We use that condition to guarantee in Lemma 2.3.2 the continuity of the moduli of continuity ω_j of g_j', $j \in J_0^N$ on D. If it is known beforehand that the ω_j, $j \in J_0^N$ are already continuous, then the convexity assumption can be dropped. The requirement that C be admissible is, in general, not very restrictive since, for example, the class of functionals satisfying (2.2.11) is fairly large. Although the compactness assumption for C is slightly restrictive, a small modification of Lemma 2.2.7 would show that it is sufficient to assume only $L^0 \cap C$ to be compact, provided that the definition of γ is suitably modified.

We now begin our discussion of the steplength procedures. Throughout the remainder of this section we will assume that at a given point $x \in L^0 \cap C$, a direction

$$s \in K(x,J(x,\varepsilon),\sigma), \qquad f'(x)s \geq 0 \tag{2.3.6}$$

has been chosen for some $\sigma > 0$ and $\varepsilon > 0$. The basic steplength algorithm for constrained minimization, used by Zoutendijk [1960] and Rosen [1960] is the minimization on level sets without leaving the constraint set. This algorithm is specified as follows:

$$\text{(2.3.7)} \quad \begin{cases} \text{At } x \in L^0 \cap C \text{ choose } \tau \text{ so that} \\ \\ f(x-\tau s) = \min \{f(x-t's) \,|\, [x,x-t's] \subset L^0(f(x)) \cap C\}. \end{cases}$$

In order to obtain an estimate of the form (2.3.1) we follow Elkin [1968] and investigate first the following constrained version of the Curry-Altman algorithm:

For fixed $\alpha \in [0,1)$ and with

$$\tau^* = \sup \{t \geq 0 \,|\, x - t's \in C,\ t' \in [0, t)\} \tag{2.3.8}$$

set $\tau' = 0$ if $f'(x)s = 0$ and otherwise

(2.3.9) $\tau' = \sup \{t \geq 0 | f'(x-t's)s > \alpha f'(x)s \text{ for } t' \in [0,t)\}$.

Then define τ by

(2.3.10) $\tau = \min \{\tau^*, \tau'\}$.

Note that if $C \equiv D$ and $\alpha = 0$, then (2.3.10) is the Curry [1944] steplength while for $\alpha > 0$ it is the Altman [1966] steplength. Although the case $\alpha > 0$ is of little practical value, it is useful in providing results for other algorithms.

Theorem 2.3.4 Suppose that $f:D \subset R^n \to R^1$ and $g_j:D \subset R^n \to R^1$, $j \in J_0$ are continuously differentiable on the open convex set D, and that C is admissible and $L^0 \cap C$ is compact. Assume that $x \in L^0 \cap C$ and that $s \in R^n$ satisfies (2.3.6), and with given $\alpha \in [0,1)$ and $\hat{\omega} \in (0,1]$ let τ be chosen by (2.3.10) and $\omega \in [\hat{\omega},1]$. Then $x - \omega\tau s \in L^0 \cap C$,

(2.3.11) $f(x) \geq f([1-t]x + t(x-\omega\tau s)) \geq f(x-\omega\tau s), \quad \forall t \in [0,1]$,

as well as

$$f(x) - f(x-\omega\tau s) \geq \Phi(f'(x)s,\sigma,\varepsilon)$$

for some forcing function Φ of three variables which depends only on C and the steplength algorithm.

Proof. If $f'(x)s = 0$, then $\tau = 0$ and the conclusion of the theorem is trivial. Therefore assume that $f'(x)s > 0$. Since D is convex and $L^0 \cap C \subset D$ compact, we may enclose $L^0 \cap C$ in a compact convex set D_0. Then the moduli of continuity ω_j of g_j', $j \in J(x,\varepsilon)$ on D_0 are well-defined and continuous on $[0,\infty)$. Hence Lemma 2.3.2 implies that either $\tau^* \geq \hat{\Phi}(\sigma,\varepsilon) > 0$, where $\hat{\Phi}$ is given by (2.3.4), or $\tau^* = \hat{\tau} = \sup \{t \geq 0 | x-ts \in D\} > 0$, and Lemma 2.3.3 ensures the existence of a $\tau_0 > 0$ such that $[x,x-\tau_0 s] \subset L^0 \cap C$. Since $f'(x)s > 0$, the number τ' given by (2.3.9) is positive. Moreover, if $x - ts \in (x,x-\tau' s] \cap L^0 \cap C$, then by the mean-value theorem

$$f(x) - f(x-ts) = tf'(x-t_1 s)s > \alpha tf'(x)s > 0$$

for some $t_1 \in (0,t)$, and thus (2.3.5) holds. Therefore, by Lemma 2.3.3, $[x,x-\tau' s] \cap C \subset L^0 \cap C$ and hence, since $\omega \leq 1$ and $\tau \leq \tau^*$, we have $[x,x-\omega\tau s] \cap C = [x,x-\omega\tau s]$ and $x - \omega\tau s \in L^0 \cap C$.

Now, by the definition of τ it follows that

$$f'(x-ts)s > \alpha f'(x)s > 0, \quad \forall t \in [0,\omega\tau)$$

and hence $f(x-ts)$ is monotone decreasing on $[0,\omega\tau]$ which implies that (2.3.11) holds.

For the proof of the last part of the theorem assume first that $\alpha > 0$. Then the mean-value theorem shows that for some $t_2 \in (0,\omega\tau)$,

$$(2.3.12) \quad f(x) - f(x-\omega\tau s) = \omega\tau f'(x-t_2 s)s \geq \tau\hat{\omega}\alpha f'(x)s.$$

We estimate a lower bound for τ. If $f' \equiv$ constant on $L^0 \cap C$, then we must have $\tau = \tau^*$ and by Lemma 2.3.2 either $\tau^* \geq \hat{\Phi}(\sigma,\varepsilon)$ or $\tau^* = \hat{\tau}$.

Otherwise the reverse modulus of continuity δ of f' on $L^0 \cap C$ is an F-function. Now, by (2.3.10) either $\tau = \tau^*$ or

(2.3.13) $(1-\alpha)f'(x)s = f'(x)s - f'(x-\tau s)s \leq \|f'(x)-f'(x-\tau s)\|$.

In the first case it follows again by Lemma 2.3.2 that either $\tau^* \geq \hat{\Phi}(\sigma,\varepsilon)$ or $\tau^* = \hat{\tau}$. Since $[x,x-\tau s] \subset L^0 \cap C \subset D$, evidently $\tau' < \hat{\tau}$ and hence, if $\tau^* = \hat{\tau}$, then (2.3.13) holds. Therefore, either $\tau \geq \hat{\Phi}(\sigma,\varepsilon)$ or (2.3.13) applies. In the latter case the definition of δ implies that

$$\tau \geq \delta((1-\alpha)f'(x)s)$$

and hence (2.3.12) can be continued to

(2.3.14) $f(x) - f(x-\omega\tau s) \geq \Phi(f'(x)s,\sigma,\varepsilon)$

where $\Phi(t_1,t_2,t_3) = \alpha\hat{\omega}t_1 \min\{\delta((1-\alpha)t_1),\hat{\Phi}(t_2,t_3)\}$. Clearly, Φ is an F-function of three variables.

Now suppose that $\alpha = 0$, and, together with τ, consider a corresponding steplength $\bar{\tau}$ obtained from (2.3.10) with $\alpha = \frac{1}{2}$. Then evidently $f(x-\omega\tau s) \leq f(x-\omega\bar{\tau}s)$ and hence, by the comparison principle of Section 2.1 (2.3.14) holds with $\Phi(t_1,t_2,t_3) = \frac{1}{2}\hat{\omega}t_1 \min\{\delta(\frac{1}{2}t_1),\hat{\Phi}(t_2,t_3)\}$.

We now turn to the steplength choice (2.3.7).

Theorem 2.3.5 Let $f:D \subset R^n \to R^1$ and $g_j:D \subset R^n \to R^1$, $j \in J_0$, be continuously differentiable on the open convex set D. Assume that C is admissible and that $L^0 \cap C$ is compact. Suppose that $x \in L^0 \cap C$ and that $s \in R^n$ satisfies (2.3.6) and that τ is chosen by (2.3.7).

Then $x - \tau s \in L^0 \cap C$ and (2.3.11) holds with $\omega = 1$ as well as

$$f(x) - f(x-\tau s) \geq \Phi(f'(x)s, \sigma, \varepsilon)$$

with some F-function Φ of three variables dependent only on C and the steplength algorithm.

Proof. Since $L^0 \cap C$ is compact, there is at least one τ such that (2.3.7) holds and hence $x - \tau s \in L^0 \cap C$. Now let $\bar{\tau}$ be obtained from (2.3.10) with $\alpha = 0$. Then, by (2.3.7) we clearly have $f(x-\tau s) \leq f(x-\bar{\tau}s)$ and hence, by the comparison principle of Section 2.1, we obtain

$$f(x) - f(x-\tau s) \geq \Phi(f'(x)s, \sigma, \varepsilon)$$

where $\Phi(t_1,t_2,t_3) = \frac{1}{2} t_1 \min\{\delta(\frac{1}{2} t_1), \hat{\Phi}(t_2,t_3)\}$ is an F-function of three variables. Finally, since $[x,x-\tau s] \subset L^0 \cap C$ we obtain (2.3.11).

The next result is a generalization of an algorithm of Ostrowski [1966; Theorem 27.1] to the constrained minimization case.

Theorem 2.3.6 Let $f: D \subset R^n \to R^1$ and $g_j: D \subset R^n \to R^1$, $j \in J_0$, be continuously differentiable on the open convex set D and assume that C is admissible. Suppose that $L^0 \cap C$ is compact and that

$$(2.3.15) \quad \|f'(x)-f'(y)\| \leq \beta\|x-y\|, \quad \forall x,y \in L^0 \cap C.$$

Assume that $x \in L^0 \cap C$ and that $s \in R^n$ satisfies (2.3.6), and for given $\hat{\omega} \in (0,1]$ let

$$(2.3.16) \qquad \tau = \frac{1}{\beta} f'(x)s$$

and

$$(2.3.17) \qquad \begin{cases} \omega = \tau^*/\tau, \text{ if } \tau^*/\tau < \hat{\omega} \\ \hat{\omega} \leq \omega \leq \min \{2-\hat{\omega}, \tau^*/\tau\}, \text{ otherwise} \end{cases}$$

where τ^* is specified by (2.3.8). Then $x - \omega\tau s \in L^0 \cap C$ and

$$f(x) - f(x-\omega\tau s) \geq \Phi(f'(x)s, \sigma, \varepsilon)$$

where Φ is an F-function of three variables dependent only on C and $\hat{\omega}$.

Proof. For $f'(x)s = 0$ the result is trivial; hence assume that $f'(x)s > 0$. Then, by Lemma 2.3.3, there exists an open interval $(x, x-\tau_0 s) \subset L^0 \cap C$, $\tau_0 > 0$. Moreover it follows readily from (2.3.15) that for $[x, x-ts] \subset L^0 \cap C$,

$$(2.3.18) \qquad f(x) - f(x-ts) \geq tf'(x)s - \frac{1}{2}\beta t^2.$$

By (2.3.16) and (2.3.17) we have $\omega\tau \leq \tau^*$ and hence $x - ts \in C$ for all $t \in [0, \omega\tau]$. If now $x - ts \in (x, x-\omega\tau s] \cap L^0 \cap C$, then by (2.3.18)

$$f(x) - f(x-ts) \geq t(f'(x)s - \frac{1}{2}\beta t) \geq t(\beta\tau - \frac{1}{2}\beta\omega\tau)$$

$$\geq t\beta\tau(1 - \frac{1}{2}(2-\hat{\omega})) = \frac{1}{2} t\tau\beta\hat{\omega} > 0$$

and hence, by Lemma 2.3.3, $[x, x-\omega\tau s] = [x, x-\omega\tau s] \cap C \subset L^0 \cap C$, that is,

$x - \omega\tau s \in L^0 \cap C$.

Again using (2.3.18) we have

$$(2.3.19) \qquad f(x) - f(x-\omega\tau s) \geq \omega\tau f'(x)s - \frac{\beta}{2}(\omega\tau)^2$$

$$= \frac{\omega}{\beta}(f'(x)s)^2 - \frac{\omega^2}{2\beta}(f'(x)s)^2$$

$$= \frac{1}{2\beta}\,\omega(2-\omega)(f'(x)s)^2.$$

If $\omega \geq \hat{\omega}$, then we can continue this inequality to

$$(2.3.20) \qquad f(x) - f(x-\omega\tau s) \geq \frac{1}{2\beta}\,\hat{\omega}(2-\hat{\omega})(f'(x)s)^2$$

since $\omega(2-\omega) = 1 - (1-\omega)^2 \geq 1 - (1-\hat{\omega})^2 = \hat{\omega}(2-\hat{\omega})$.

On the other hand, if $\omega < \hat{\omega}$, then $\omega\tau = \tau^* < \tau$. Now by Lemma 2.3.2 either $\tau^* \geq \hat{\Phi}(\sigma,\varepsilon)$ where $\hat{\Phi}$ is given by (2.3.4), or $\tau^* = \hat{\tau} = \sup\{t \geq 0 \mid x-ts \in D\}$. In the latter case $x - \omega\tau s \in L^0 \cap C \subset D$ implies that $\omega\tau < \tau^*$, which means that $\omega \geq \hat{\omega}$ and hence that (2.3.20) holds. Otherwise, by (2.3.16) and (2.3.18)

$$f(x) - f(x-\tau^* s) \geq \tau^*(f'(x)s - \frac{1}{2}\beta\tau^*)$$

$$> \tau^*(f'(x)s - \frac{1}{2}\beta\tau) = \frac{1}{2}\tau^* f'(x)s \geq \frac{1}{2}\hat{\Phi}(\sigma,\varepsilon)f'(x)s.$$

Altogether we therefore find that

$$f(x) - f(x-\omega\tau s) \geq \Phi(f'(x)s,\sigma,\varepsilon)$$

where $\Phi(t_1,t_2,t_3) = \frac{1}{2}t_1 \min\{\frac{\hat{\omega}}{\beta}(2-\hat{\omega})t_1, \hat{\Phi}(t_2,t_3)\}$ is clearly an F-function of three variables.

Another possible choice of steplength is obtained by taking one Newton step towards the solution of the Curry-type equation

$$f'(x-\tau s)s = 0$$

and by using a suitable relaxation parameter to ensure that $x - \omega\tau s \in L^0 \cap C$ (see Ortega and Rheinboldt [1970]). The next result is a generalization of this process to the constrained minimization case.

Theorem 2.3.7 Let $f:D \subset R^n \to R^1$ be twice continuously differentiable and $g_j:D \subset R^n \to R^1$, $j \in J_0$ once continuously differentiable on the open convex set D and assume that C is admissible. Moreover, suppose that $L^0 \cap C$ is compact and that there exist constants η_0, η_1, $0 < \eta_1 \leq \eta_2$, such that

(2.3.21) $$\eta_0\|h\|^2 \leq f''(x)hh \leq \eta_1\|h\|^2, \quad \forall x \in L^0 \cap C, \quad \forall h \in R^n.$$

Further, assume that $x \in L^0 \cap C$ and that $s \in R^n$ satisfies (2.3.6), and for given $\hat{\omega} \in (0,1]$ let

(2.3.22) $$\tau = f'(x)s/f''(x)ss$$

and

(2.3.23) $$\begin{cases} \omega = \tau^*/\tau, \text{ if } \tau^*/\tau < \hat{\omega} \\ \hat{\omega} \leq \omega \leq \bar{\omega} = \min\{(2/\gamma_1)-\hat{\omega}, \tau^*/\tau\}, \text{ otherwise} \end{cases}$$

where τ^* is again specified by (2.3.8) and

$$\gamma_1 = \sup \{\frac{f''(x-ts)ss}{f''(x)ss} \ | t > 0, \ f(x-t_1 s) < f(x), \ \forall t_1 \in (0,t]\}.$$

Then $x - \omega\tau s \in L^0 \cap C$ and

$$f(x) - f(x-\omega\tau s) \geq \Phi(f'(x)s,\sigma,\varepsilon)$$

where Φ is an F-function of three variables which depends only on C and the constants $\hat{\omega}$ and η_1.

<u>Proof</u>. Since $x \in L^0 \cap C$, τ and γ_1 are well-defined. For $f'(x)s = 0$ the result is trivial, hence assume that $f'(x)s > 0$ and therefore that $\tau > 0$. Then Lemma 2.3.3 implies the existence of an open interval $(x,x-\tau_0 s) \subset L^0 \cap C$, $\tau_0 > 0$. Moreover, if $x - ts \in (x,x-\omega\tau s] \cap L^0 \cap C$, then by the mean-value theorem and because of $\gamma_1 \geq 1$,

$$f(x) - f(x-ts) = tf'(x)s - \frac{1}{2} t^2 f''(x-t's)ss, \ t' \in (0,t)$$

$$\geq tf'(x)s\{1- \frac{1}{2} [(2/\gamma_1)-\hat{\omega}]f''(x-t's)ss/f''(x)ss\}$$

$$\geq \frac{1}{2} \hat{\omega} tf'(x)s > 0.$$

Since (2.3.22) and (2.3.23) imply that $\omega\tau \leq \tau^*$, Lemma 2.3.3 ensures that $[x,x-\omega\tau s] = [x,x-\omega\tau s] \cap C \subset L^0 \cap C$. Therefore, $x - \omega\tau s \in L^0 \cap C$ and

$$f(x) - f(x-\omega\tau s) \geq \frac{1}{2} \hat{\omega}\omega\tau f'(x)s. \tag{2.3.24}$$

If $\omega \geq \hat{\omega}$, then using (2.3.21) and (2.3.22) we can continue (2.3.24) to

$$f(x) - f(x-\omega\tau s) \geq \frac{1}{2} \hat{\omega}^2 (\frac{1}{\eta_1})(f'(x)s)^2. \tag{2.3.25}$$

On the other hand, if $\omega < \hat{\omega}$, then $\omega\tau = \tau^* < \tau$. By Lemma 2.3.2 we have either $\tau^* \geq \hat{\Phi}(\sigma,\varepsilon)$ where $\hat{\Phi}$ is given by (2.3.4) or $\tau^* = \hat{\tau} = \sup\{t \geq 0 | x - ts \in D\}$. In the latter case $x - \omega\tau s \in L^0 \cap C \subset D$ implies that $\omega\tau < \tau^*$ and hence necessarily $\omega \geq \hat{\omega}$, that is, (2.3.25) holds. Otherwise it follows from (2.3.24) that

$$f(x) - f(x-\tau^* s) \geq \frac{1}{2}\hat{\omega}\tau^* f'(x)s \geq \frac{1}{2}\hat{\omega}\hat{\Phi}(\sigma,\varepsilon)f'(x)s$$

and this together with (2.3.25) implies that

$$f(x) - f(x-\omega\tau s) \geq \Phi(f'(x)s,\sigma,\varepsilon)$$

where $\Phi(t_1,t_2,t_3) = \frac{1}{2}\hat{\omega}t_1 \min\{\frac{\hat{\omega}}{\eta_1}t_1, \hat{\Phi}(t_2,t_3)\}$ is clearly an F-function of three variables.

Until now we have considered only steplength algorithms which specify the steplength precisely. Another approach is to choose the steplength arbitrarily from an interval of permissible values. Along this line we investigate here the "Goldstein range" proposed by Goldstein [1964], [1965], [1966], and extend it, together with the results of Armijo [1966], Elkin [1968], and Ortega and Rheinboldt [1970], to the constrained case. The following algorithm is considered:

For $f'(x)s = 0$ set $\alpha_1 = 0$, otherwise, let $\alpha_1 > 0$ be such that

(2.3.26) $$\zeta_1\alpha_1 f'(x)s \leq f(x) - f(x-\alpha_1 s) \leq \zeta_2\alpha_1 f'(x)s$$

where $0 < \zeta_1 \leq \zeta_2 < 1$ are fixed numbers.

The next lemma, whose proof can be found in Ortega and Rheinboldt [1970; Theorem 8.3.2] concerns the solvability of the inequalities (2.3.26).

Lemma 2.3.8 (Goldstein [1964]) Let $f:D \subset R^n \to R^1$ be continuously differentiable on the open set D and suppose that $x \in D$, $s \in R^n$ and $t > 0$ satisfy $f'(x)s > 0$, $[x,x-ts] \subset D$ and

$$f(x) - f(x-ts) < \zeta_1 tf'(x)s$$

with some $\zeta_1 \in (0,1)$. Then for any $\zeta_2 \in [\zeta_1,1)$ there is an $\omega \in (0,1)$ such that (2.3.26) holds with $\alpha_1 = \omega t$.

We note that the next two results require the set D to be convex even in the unconstrained case.

Theorem 2.3.9 Let $f:D \subset R^n \to R^1$ and $g_j:D \subset R^n \to R^1$, $j \in J_0$, be continuously differentiable on the open convex set D, and assume that C is admissible and $L(f(x^0)) \cap C$ is compact. Suppose further that $x \in L(f(x^0)) \cap C$, and (2.3.6) holds for $s \in R^n$ and that $\tau \geq 0$ is any number satisfying (2.3.26) with $\alpha_1 = \tau$ for fixed $0 < \zeta_1 \leq \zeta_2 < 1$. Let τ^* be given by (2.3.8) and $\omega^* \in (0,1]$ be such that (2.3.26) holds with $\alpha_1 = \omega^*\tau^*$. Moreover, specify ω by

$$(2.3.27) \qquad \omega = \begin{cases} 1, \text{ if } \tau \leq \tau^* \\ \tau^*/\tau, \text{ if } \tau^* < \tau \text{ and } f(x)-f(x-\tau^* s) \geq \zeta_1 \tau^* f'(x)s \\ \omega^*\tau^*/\tau, \text{ if } \tau^* < \tau \text{ and } f(x)-f(x-\tau^* s) < \zeta_1 \tau^* f'(x)s. \end{cases}$$

Then $x - \omega\tau s \in L(f(x)) \cap C$ and

$$f(x) - f(x-\omega\tau s) \geq \Phi(f'(x)s,\sigma,\varepsilon)$$

for some F-function Φ of three variables dependent only on C and the steplength algorithm.

Proof. For $f'(x)s = 0$ the result is again trivial. For $f'(x)s > 0$ it follows from $x \in L(f(x^0)) \cap C$ and Lemma 2.3.3 that there exists a $\tau > 0$ such that (2.3.26) holds for $\alpha_1 = \tau$. Then, from (2.3.27) we have $\omega\tau \leqslant \tau^*$ and

$$f(x) - f(x-\omega\tau s) \geqslant \zeta_1 \omega\tau f'(x)s > 0$$

and hence $x - \omega\tau s \in L(f(x)) \cap C$. Moreover, by the convexity of D, $x - ts \in D, \quad \forall t \in [0,\omega\tau]$.

Now, by Lemma 2.3.2, either $\tau^* \geqslant \hat{\Phi}(\sigma,\varepsilon)$ where $\hat{\Phi}$ is given by (2.3.4) or $\tau^* = \hat{\tau} = \sup\{t \geqslant 0 | x - ts \in D\}$. In the latter case $x - \omega\tau s \in L(f(x)) \cap C \subset D$ implies that $\omega\tau < \tau^*$ and hence that the second condition in (2.3.27) does not apply. If the second condition holds, then $\tau^* \geqslant \hat{\Phi}(\sigma,\varepsilon)$ and

$$(2.3.28) \qquad f(x)-f(x-\tau^* s) \geqslant \zeta_1 \tau^* f'(x)s \geqslant \zeta_1 \hat{\Phi}(\sigma,\varepsilon) f'(x)s.$$

On the other hand, if either the first or third condition of (2.3.27) applies, consider the modulus of continuity $\bar{\omega}$ of f' on some compact convex set $D_0 \subset D$ containing $L(f(x^0)) \cap C$. Clearly $\bar{\omega}$ is well-defined and continuous on $[0,\infty)$ and hence so is $\eta(t) = \int_0^1 \bar{\omega}(t\xi)d\xi$. Moreover, using the mean-value theorem it is easily seen that

$$f(x) - f(x-\omega\tau s) \geqslant \omega\tau f'(x)s - \omega\tau\eta(\omega\tau).$$

Therefore, by (2.3.26) with $\alpha_1 = \omega\tau$ we have

$$\zeta_2 \omega\tau f'(x)s \geq \omega\tau f'(x)s - \omega\tau\eta(\omega\tau)$$

and hence

$$\hat{\eta}(\omega\tau) \geq \eta(\omega\tau) \geq (1-\zeta_2)f'(x)s. \tag{2.3.29}$$

Here $\hat{\eta}:[0,\infty) \to [0,\infty)$ is any strictly isotone function such that $\hat{\eta}(t) \geq \eta(t)$, for all $t \geq 0$, and therefore that $\hat{\eta}^{-1}$ exists and is a strictly isotone F-function. Thus (2.3.26) and (2.3.29) imply that

$$f(x) - f(x-\omega\tau s) \geq \zeta_1 \omega\tau f'(x)s \geq \zeta_1 f'(x)s\ \hat{\eta}^{-1}((1-\zeta_2)f'(x)s).$$

Together with (2.3.28) this shows that

$$f(x) - f(x-\omega\tau s) \geq \Phi(f'(x)s,\sigma,\varepsilon)$$

where $\Phi(t_1,t_2,t_3) = \zeta_1 t_1 \min\{\hat{\eta}^{-1}((1-\zeta_2)t_1),\hat{\Phi}(t_2,t_3)\}$ is clearly an F-function of three variables.

The choice of $\omega\tau$ by (2.3.27) is not constructive but is the basis for the following algorithm:

Goldstein-Armijo algorithm: Constrained case

Let μ be a fixed F-function and $\bar{\alpha} \in (0,1)$ and $\beta > 1$ given constants. Let $x \in L(f(x^0)) \cap C$ and $s \in R^n$ satisfy (2.3.6).

I. Basic steplength selection: If $f'(x)s = 0$, set $\tau = 0$; otherwise let $\tau > 0$ be any real number such that $\tau \geq \mu(f'(x)s)$.

II. Choice of relaxation parameter: If $\tau \leq \tau^*$ and

(2.3.30) $$f(x) - f(x-\tau s) \geq \bar{\alpha}\tau f'(x)s$$

set $\omega = 1$. If $\tau > \tau^*$ or if (2.3.30) is not satisfied, let ω be the largest number in the sequence $\{\beta^{-j}\}_{j=1}^{\infty}$ such that $\omega\tau \leq \tau^*$ and

(2.3.31) $$f(x) - f(x-\omega\tau s) \geq \bar{\alpha}\omega\tau\, f'(x)s$$

where τ^* is given by (2.3.8).

Note that if C is admissible and D is open then it follows from Lemma 2.3.3 that $x - \beta^{-j}\tau s \in C$ for sufficiently large j, and from Lemma 2.3.8 that (2.3.31) can be satisfied.

Theorem 2.3.10 Let $f:D \subset R^n \to R^1$ and $g_j:D \subset R^n \to R^1$, $j \in J_0$ be continuously differentiable on the open convex set D and assume that C is admissible and that $L(f(x^0)) \cap C$ is compact. Suppose that $x \in L(f(x^0)) \cap C$, $s \in R^n$ satisfies (2.3.6), and that τ and ω are selected by the Goldstein-Armijo algorithm. Then $x - \omega\tau s \in L(f(x)) \cap C$ and

$$f(x) - f(x-\omega\tau s) \geq \Phi(f'(x)s,\sigma,\varepsilon)$$

where Φ is an F-function of three variables dependent only on C and the steplength algorithm.

Proof. Assume that $f'(x)s > 0$, then (2.3.30) and (2.3.31) imply that $f(x) - f(x-\omega\tau s) > 0$ while by the definition ω, $\omega\tau \leq \tau^*$; and hence $x - \omega\tau s \in L(f(x)) \cap C$.

If (2.3.30) holds, then

(2.3.32) $$f(x) - f(x-\omega\tau s) \geq \bar{\alpha}f'(x)s\mu(f'(x)s).$$

On the other hand, suppose that (2.3.31) holds. Then by the definition of ω either $\beta\omega\tau > \tau^*$ or

$$(2.3.33) \qquad f(x) - f(x-\beta\omega\tau s) < \bar{\alpha}\beta\omega\tau f'(x)s.$$

Let δ denote the reverse modulus of continuity of f' on $L(f(x)) \cap C$. We may assume that f' is not constant on $L(f(x^0)) \cap C$ and hence that δ is an F-function. Suppose that $\beta\omega\tau > \tau^*$ and consider the steplength $\bar{\tau}$ obtained by the Curry-Altman algorithm with $\alpha = 0$. Then, by Theorem 2.3.4, $[x, x-\bar{\tau}s] \subset L^0 \cap C$ and $\bar{\tau} \leq \tau^*$, and as in that theorem either $\bar{\tau} \geq \hat{\Phi}(\sigma,\varepsilon)$, where $\hat{\Phi}$ is given by (2.3.4), or

$$f'(x)s = f'(x)s - f'(x-\bar{\tau}s)s$$

$$\leq \|f'(x) - f'(x-\bar{\tau}s)\|.$$

In the latter case $\bar{\tau} \geq \delta(f'(x)s)$ and hence

$$\bar{\tau} \geq \min\{\delta(f'(x)s), \hat{\Phi}(\sigma,\varepsilon)\}.$$

Using $\beta\omega\tau > \tau^* \geq \bar{\tau}$ we see that

$$f(x) - f(x-\omega\tau s) \geq \bar{\alpha}\omega\tau f'(x)s$$

$$(2.3.34) \qquad \geq \frac{\bar{\alpha}}{\beta} f'(x)s \min\{\delta(f'(x)s), \hat{\Phi}(\sigma,\varepsilon)\}.$$

Finally, suppose that $\beta\omega\tau \leq \tau^*$ and that (2.3.33) holds. If $x - \beta\omega\tau s \notin L(f(x)) \cap C$, then we again have $\beta\omega\tau > \bar{\tau}$ and (2.3.34) holds. Otherwise, because of (2.3.33) there is a $t'' \in (0,\beta\omega\tau)$ such that

$$f'(x-t''s)s = \frac{1}{\beta\omega\tau}\,(f(x)-f(x-\beta\omega\tau s)) < \bar{\alpha} f'(x)s$$

and hence that

$$(1-\bar{\alpha})f'(x)s < \|f'(x) - f'(x-t''s)\|.$$

Therefore,

$$\beta\omega\tau \geq t'' \geq \delta((1-\bar{\alpha})f'(x)s)$$

and, by (2.3.31)

(2.3.35) $$f(x) - f(x-\omega\tau s) \geq \frac{\bar{\alpha}}{\beta}\, f'(x)s\delta((1-\bar{\alpha})f'(x)s).$$

Now (2.3.32), (2.3.34) and (2.3.35) together give

$$f(x) - f(x-\omega\tau s) \geq \Phi(f'(x)s,\sigma,\varepsilon)$$

where $\Phi(t_1,t_2,t_3) = \frac{\bar{\alpha}t_1}{\beta} \min\,\{\mu(t_1)\beta,\delta(t_1),\delta((1-\bar{\alpha})t_1),\hat{\Phi}(t_2,t_3)\}$

is clearly an F-function of three variables.

2.4 Zoutendijk's Method of Feasible Directions

Zoutendijk [1960] considered iterations of the form

(2.4.1) $$x^{k+1} = x^k - \tau_k s^k,\; f(x^{k+1}) \leq f(x^k),\; k = 0,1,\ldots$$

for a special choice of feasible directions $\{s^k\}$ and a steplength sequence $\{\tau_k\}$ given by (2.3.7). In this section we will obtain a simple convergence proof of the method as a consequence of the basic estimate (2.3.1). We will also show that, under suitable assumptions on f and g_j, $j \in J_0$, all of the steplength results of Section 2.3 can be used in connection with Zoutendijk's method. Before proceeding with this, we require the following well-known result of Gordan [1873] on linear inequalities which we state without proof.

Lemma 2.4.1 Let A be an $m \times n$ matrix. Then the system of inequalities $Ax > 0$, $x \in R^n$ is inconsistent if and only if $A^T u = 0$ for some $u \in R^m$, $u \leq 0$ such that $u_j < 0$ for at least one j, $1 \leq j \leq m$.

In other words, $Ax > 0$ is inconsistent if and only if the rows of A are linearly dependent with coefficients of one sign.

We now turn to the particular choice of a feasible direction s at $x \in C$ considered by Zoutendijk. For this the following notation will be useful: Assume that $f:D \subset R^n \to R^1$ and $g_j:D \subset R^n \to R^1$, $j \in J_0$ are G-differentiable on the open set D. Let $\tilde{\mu}_0$, $\tilde{\mu}_j:[0,\infty) \to [0,\infty)$, $j \in J_0^N$ be given continuous, strictly isotone F-functions for which $\tilde{\mu}_0(0) = \tilde{\mu}_j(0) = 0$ and, for $j = 0$, $j \in J_0^N$, define $\mu_j:R^1 \to R^1$,

$$\mu_j(t) = \begin{cases} t, & \text{if } t \in (-\infty,0) \\ \tilde{\mu}_j(t), & \text{if } t \in [0,\infty). \end{cases}$$

For any $x \in D$, $\sigma \in R^1$ and fixed index set $J \subset J_0$ consider the set

$$(2.4.2) \qquad \hat{K}(x,J,\sigma) = K(x,J,\sigma) \cap \{s \in R^n | f'(x)s - \mu_0(\sigma) \geq 0\}$$

where $K(x,J,\sigma)$ is given by (2.3.2). Define the mapping $\sigma(\cdot,J):D \subset R^n \to R^1$ by

(2.4.3) $\sigma(x,J) = \sup \{\sigma \in R^1 | \hat{K}(x,J,\sigma) \neq \{\emptyset\}\}$.

Some of the properties of $\hat{K}(x,J,\sigma)$ and $\sigma(x,J)$ are summarized in the following sequence of lemmas. We note that on the basis of the definition (2.3.2) of the sets $K(x,J,\sigma)$ the linear constraints do not affect the value of $\sigma(x,J)$.

Lemma 2.4.2 Let $f:D \subset R^n \to R^1$ and $g_j:D \subset R^n \to R^1$, $j \in J_0$ be G-differentiable on the open set D. Let x be any point in D and $J \subset J_0$ a fixed index set. Then $\sigma(x,J)$ given by (2.4.3) is well-defined and

$$\{\sigma \in R^1 | \hat{K}(x,J,\sigma) \neq \{\emptyset\}\} = (-\infty, \sigma(x,J)],$$

in particular, $\hat{K}(x,J,\sigma(x,J)) \neq \{\emptyset\}$.

Proof. The set $\hat{K}(x,J,\sigma)$ is closed and contained in the unit sphere and hence is compact. Therefore, using the assumed properties of μ_0, μ_j, $j \in J^N$, we see that $\sigma(x,J)$ is well-defined and that $\hat{K}(x,J,\sigma) \neq \{\emptyset\}$ for all $\sigma \in (-\infty, \sigma(x,J))$. Now let $\{\sigma_k\}_{k=0}^{\infty}$ be any sequence in $(-\infty, \sigma(x,J))$ such that $\lim_{k\to\infty} \sigma_k = \sigma(x,J)$. Clearly $\hat{K}(x,J,\sigma_k) \neq \{\emptyset\}$ for any k and hence we may choose vectors $s^k \in \hat{K}(x,J,\sigma_k)$, $k = 0,1,\ldots$. It is no restriction to assume that $\lim_{k\to\infty} s^k = s^*$, $\|s^*\| = 1$ and hence it follows that

$$g_j'(x)s^k \geq \mu_j(\sigma_k),\ j \in J^N,\ g_j'(x)s^k \geq 0,\ j \in J^L$$

$$f'(x)s^k \geq \mu_0(\sigma_k),\ \|s^k\| = 1,\ k = 0,1,\ldots .$$

Thus, going to the limit and using the continuity of μ_0, μ_j, $j \in J^N$ we have

$$g_j'(x)s^* \geq \mu_j(\sigma(x,J)),\ j \in J^N,\ g_j'(x)s^* \geq 0,\ j \in J^L$$

$$f'(x)s^* \geq \mu_0(\sigma(x,J)),\ \|s^*\| = 1.$$

This completes the proof of the lemma.

Lemma 2.4.3 Let $f:D \subset R^n \to R^1$ and $g_j:D \subset R^n \to R^1$, $j \in J_0$ be G-differentiable on the open set D. Let $x \in D$ be given and $J_1 \subset J_2 \subset J_0$ fixed index sets. Then

$$\hat{K}(x,J_2,\sigma) \subset \hat{K}(x,J_1,\sigma),\quad \sigma \in R^1$$

and

$$\sigma(x,J_2) \leq \sigma(x,J_1).$$

Proof. Suppose that $s \in \hat{K}(x,J_2,\sigma(x,J_2))$. Then

$$g_j'(x)s - \mu_j(\sigma(x,J_2)) \geq 0,\ j \in J_2^N,\ g_j'(x)s \geq 0,\ j \in J_2^L,$$

$$f'(x)s - \mu_0(\sigma(x,J_2)) \geq 0,$$

and clearly $J_1 \subset J_2$ implies that these relations hold also for $j \in J_1$. Hence by (2.4.3), $\sigma(x,J_2) \leq \sigma(x,J_1)$. If $\sigma > \sigma(x,J_2)$, then $\hat{K}(x,J_2,\sigma) = \{\emptyset\}$ and we are done. Therefore assume that $\sigma \leq \sigma(x,J_2)$ and $s \in \hat{K}(x,J_2,\sigma)$. Then

$$g_j'(x)s - \mu_j(\sigma) \geq 0,\ j \in J_2^N,\ g_j'(x)s \geq 0,\ j \in J_2^L,$$

$$f'(x)s - \mu_0(\sigma) \geq 0,$$

and hence, since $J_1 \subset J_2$, these relations also hold for $j \in J_1$ which implies that $s \in \hat{K}(x,J_1,\sigma)$.

The next lemma will play a key role in our convergence proof of Zoutendijk's method.

Lemma 2.4.4 Let $f:D \subset R^n \to R^1$ and $g_j:D \subset R^n \to R^1$, $j \in J_0$ be continuously differentiable on some compact subset $D_0 \subset D$. If $J \subset J_0$ is some fixed index set, then $\sigma(\cdot,J):D_0 \subset R^n \to R^1$ is uniformly continuous on D_0.

Proof. Let x^* be any point in D_0 and suppose that $\sigma(\cdot,J)$ is not continuous at x^*. Then there is a sequence $\{x^k\} \subset D_0$ with $\lim_{k\to\infty} x^k = x^* \in D_0$ and an $\varepsilon > 0$ such that for any $k_0 \geq 0$

$$|\mu_j(\sigma(x^k,J))-\mu_j(\sigma(x^*,J))| \geq \varepsilon \text{ for some } k \geq k_0,\ j = 0,\ j \in J^N.$$

Since D_0 is compact, it follows from the continuity of f' and g_j', $j \in J$ that the sequence $\{\sigma(x^k,J)\}$ is bounded. Therefore we may pick a subsequence $\{x^{k_i}\} \subset \{x^k\}$ such that $\lim_{i\to\infty} x^{k_i} = x^*$, $\lim_{i\to\infty} \sigma(x^{k_i},J) = \sigma^* \in R^1$ and

$$\text{(2.4.4)} \qquad |\mu_j(\sigma(x^{k_i},J))-\mu_j(\sigma(x^*,J))| \geq \varepsilon,\ \forall i,\ j = 0,\ j \in J^N$$

Now choose $s^{k_i} \in \hat{K}(x^{k_i},J,\sigma(x^{k_i},J))$; then we may refine all of the subsequences so that also $\lim_{i\to\infty} s^{k_i} = \bar{s}$. Hence for each k_i we have

$$g_j'(x^{k_i})s^{k_i} - \mu_j(\sigma(x^{k_i},J)) \geq 0,\ j \in J^N,\ g_j'(x^{k_i})s^{k_i} \geq 0,\ j \in J^L,$$

$$f'(x^{k_i})s^{k_i} - \mu_0(\sigma(x^{k_i},J)) \geq 0,$$

and thus, by the continuity of f' and g_j', $j \in J$ on D_0 and the continuity of μ_0, μ_j, $j \in J^N$ it follows that

$$g_j'(x^*)\bar{s} - \mu_j(\sigma^*) \geq 0,\ j \in J^N,\ g_j'(x^*)\bar{s} \geq 0,\ j \in J^L,$$

$$f'(x^*)\bar{s} - \mu_0(\sigma^*) \geq 0,\ \|\bar{s}\| = 1.$$

This implies that $\sigma^* \leq \sigma(x^*,J)$ and consequently, by (2.4.4), that $\mu_j(\sigma^*) \leq \mu_j(\sigma(x^*,J)) - \varepsilon$, $j = 0$, $j \in J^N$. Therefore, for all $i \geq i_0$ we have

$$\mu_j(\sigma(x^{k_i},J)) < \mu_j(\sigma(x^*,J)) - \frac{\varepsilon}{2},\ j = 0,\ j \in J^N.$$

Choose $s^* \in \hat{K}(x^*,J,\sigma(x^*,J))$. Then

$$f'(x^*)s^* > (\mu_0(\sigma(x^*,J)) - \frac{\varepsilon}{4}),\ g_j'(x^*)s^* > (\mu_j(\sigma(x^*,J)) - \frac{\varepsilon}{4}),\ j \in J^N,$$

$$g_j'(x^*)s^* \geq 0,\ j \in J^L.$$

By the continuity of f' and g_j', $j \in J_0$, there exists a $\delta > 0$ such that for $\|x-x^*\| \leq \delta$

$$|f'(x)s^* - f'(x^*)s^*| \leq \frac{\varepsilon}{4}$$

and

$$|g_j'(x)s^* - g_j'(x^*)s^*| \leq \frac{\varepsilon}{4},\ j \in J^N.$$

Hence,

$$f'(x)s^* \geq f'(x^*)s^* - |f'(x)s^* - f'(x^*)s^*| \geq f'(x^*)s^* - \frac{\varepsilon}{4}$$

and similarly

$$g_j'(x)s^* \geq g_j'(x^*)s^* - \frac{\varepsilon}{4},\ j \in J^N.$$

Thus for $\|x-x^*\| \leq \delta$ we have

$$f'(x)s^* \geq \mu_0(\sigma(x^*,J)) - \frac{\varepsilon}{4} - \frac{\varepsilon}{4} = \mu_0(\sigma(x^*,J)) - \frac{\varepsilon}{2},$$

$$g_j'(x)s^* \geq \mu_j(\sigma(x^*,J)) - \frac{\varepsilon}{4} - \frac{\varepsilon}{4} = \mu_j(\sigma(x^*,J)) - \frac{\varepsilon}{2},\ j \in J^N.$$

Since $\lim_{i\to\infty} x^{k_i} = x^*$, $i_1 > i_0$ can be chosen such that $\|x^{k_i}-x^*\| \leq \delta$, for all $i \geq i_1$ and hence that

$$f'(x^{k_i})s^* \geq \mu_0(\sigma(x^*,J)) - \frac{\varepsilon}{2},\ g_j'(x^{k_i})s^* \geq \mu_j(\sigma(x^*,J)) - \frac{\varepsilon}{2},\ j \in J^N,$$

$$g_j'(x^{k_i})s^* \geq 0,\ j \in J^L.$$

Because of $\mu_j(\sigma(x^*,J)) - \frac{\varepsilon}{2} > \mu_j(\sigma(x^{k_i},J))$, $j = 0$, $j \in J^N$, this contradicts (2.4.3). Therefore, there exists no sequence $\{x^k\} \subset D_0$ such that $\lim_{k\to\infty} x^k = x^*$ and $\lim_{k\to\infty} \sigma(x^k,J) \neq \sigma(x^*,J)$. Hence $\sigma(\cdot,J)$ is continuous on D_0, and since D_0 is compact, it is also uniformly continuous.

For regular constraint sets it is now possible to characterize the conditional critical points of f on C in terms of $\hat{K}(x,J,\sigma(x,J))$ and $\sigma(x,J)$. We first obtain the following lemma.

Lemma 2.4.5 Let $f:D \subset R^n \to R^1$ and $g_j:D \subset R^n \to R^1$, $j \in J_0$ be differentiable on the open set D. Let x^* be any point in C and $J \subset J_0$ a fixed index set such that the gradients $g_j'(x^*)^T$, $j \in J$ are linearly independent. If $\sigma(x^*,J) \leq 0$, then

$$(2.4.5) \qquad \nu_0 f'(x^*)^T + \sum_{j \in J} \nu_j g_j'(x^*)^T = 0, \ \nu_0 < 0, \ \nu_j \leq 0, \ j \in J.$$

Proof. By (2.4.3), $\sigma(x^*,J) \leq 0$ implies that the system

$$f'(x^*)s \geq \mu_0(\sigma), \ g_j'(x^*)s \geq \mu_j(\sigma), \ j \in J^N, \ g_j'(x^*)s \geq 0, \ j \in J^L,$$

is inconsistent for $\sigma > 0$. Since μ_0, μ_j, $j \in J^N$ are continuous, strictly isotone functions with $\mu_0(0) = \mu_j(0) = 0$, also the system

$$(2.4.6) \qquad f'(x^*)s > 0, \ g_j'(x^*)s > 0, \ j \in J$$

is inconsistent and therefore, by Lemma 2.4.1, we have

$$\nu_0 f'(x^*)^T + \sum_{j \in J} \nu_j g_j'(x^*)^T = 0, \ \nu_0 \leq 0, \ \nu_j \leq 0, \ j \in J,$$

where at least one of these numbers is negative. For $\nu_0 = 0$ the gradients $g_j'(x^*)^T$, $j \in J$, would be linearly dependent, against assumption; hence $\nu_0 < 0$ and (2.4.5) holds.

The next result is essentially due to Zoutendijk [1960].

Theorem 2.4.6 Let $f:D \subset R^n \to R^1$ and $g_j:D \subset R^n \to R^1$, $j \in J_0$, be differentiable on the open set D and assume that the constraint set C is regular. Then $x^* \in C$ is a conditional critical point of f on C with nonpositive multipliers if and only if

$\sigma(x^*,J(x^*,0)) \leq 0$.

Proof. Suppose that $\sigma(x^*,J(x^*,0)) \leq 0$; then by the regularity of C, it follows from Lemma 2.4.5 that x^* satisfies (2.2.5) with $\nu_j \leq 0$, $j \in J(x^*,0)$, and hence that x^* is a conditional critical point of f on C.

Conversely, suppose that $x^* \in C$ is a conditional critical point of f on C with nonpositive multipliers and that $\sigma(x^*,J(x^*,0)) > 0$. Then, choosing $s^* \in \hat{K}(x^*,J(x^*,0),\sigma(x^*,J(x^*,0)))$ we would have

$$f'(x^*)s^* \geq \mu_0(\sigma(x^*,J(x^*,0))) > 0, \; g_j'(x^*)s^* \geq 0, \; j \in J^L(x^*,0)$$

$$g_j'(x^*)s^* \geq \mu_j(\sigma(x^*,J(x^*,0))) > 0, \; j \in J^N(x^*,0).$$

But since (2.2.5) holds with $\nu_j \leq 0$, $j \in J(x^*,0)$, this leads to

$$0 < f'(x^*)s^* = \sum_{j \in J(x^*,0)} \nu_j g_j'(x^*)s^* \leq 0$$

which is a contradiction. Therefore, necessarily $\sigma(x^*,J(x^*,0)) \leq 0$.

The last result related $f'(x)s$ and $\sigma(x,J(x,0))$; however, in addition to these two quantities the basic estimate (2.3.1) also contains the number ε which controls the proximity of x to the boundary of C. The following lemma brings this ε into relation with the other quantities.

Lemma 2.4.7 Let $\psi:[0,\infty) \to [0,\infty)$ be any step function such that $\psi(0) = \psi(t) > 0$ for all $t \in [0,t_0]$ with $t_0 > 0$ and $\psi(t)$ = constant for all $t \in [t_1,\infty)$ with $t_1 > t_0$. Further assume that $\hat{\mu}:[0,\infty) \to [0,\infty)$

is any isotone function for which $\hat{\mu}(0) = 0$ and $\hat{\mu}(t) > 0$ for $t > 0$. Then, for given $t_2 > 0$, the quantity

$$\varepsilon = \frac{1}{2} \sup \{t \in [0,t_2] \,|\, \hat{\mu}(\psi(t_3)) - t_3 \geq 0,\ t_3 \in [0,t]\}$$

is well-defined and positive, and $\psi(\varepsilon) > 0$.

<u>Proof</u>. Since $\hat{\mu}$ is isotone and $\hat{\mu}(t) > 0$ when $t > 0$, it follows that $\hat{\mu}(\psi(t)) > 0$ for $t \in [0,t_0]$ and clearly $\hat{\mu}(\psi(t))$ = constant for $t \in [t_1,\infty)$. Consequently the set $\{t \in [0,t_2] \,|\, \hat{\mu}(\psi(t_3)) - t_3 \geq 0,$ $t_3 \in [0,t]\} \neq \{\emptyset\}$ and $\varepsilon \geq \frac{1}{2} \min \{t_0, t_2, \psi(0)\} > 0$ is well-defined. Finally, from $\hat{\mu}(\psi(\varepsilon)) \geq \varepsilon$ follows that $\psi(\varepsilon) > 0$.

We note that Lemma 2.4.3 shows that for any $x \in C$ such that $\sigma(x,J(x,0)) > 0$ the mapping

$$\psi : [0,\infty) \to [0,\infty),\ \psi(\delta) = \begin{cases} \sigma(x,J(x,\delta)), & \text{if } \sigma(x,J(x,\delta)) \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

is an antitone step-function which satisfies the hypothesis of Lemma 2.4.7. Hence the ε of Lemma 2.4.7 is well-defined and positive, and $\sigma(x,J(x,\varepsilon)) > 0$. Moreover, because of the antitonicity of $\hat{\mu}(\psi(\cdot))$, the definition of ε simplifies to

$$\text{(2.4.7)} \qquad \varepsilon = \frac{1}{2} \sup \{\delta \geq 0 \,|\, \hat{\mu}(\psi(\delta)) - \delta \geq 0\}.$$

Furthermore, if $\{x^k\}$ is any sequence in C for which the corresponding sequences $\{\varepsilon_k\}$ and $\{\sigma(x^k,J(x^k,\varepsilon_k))\}$ are determined by (2.4.7) and are positive, then clearly $\lim_{k\to\infty} \varepsilon_k = 0$ if and only if $\lim_{k\to\infty} \sigma(x^k,J(x^k,\varepsilon_k)) = 0$. We now characterize Zoutendijk's choice of feasible directions. Suppose that $f: D \subset R^n \to R^1$ and $g_j: D \subset R^n \to R^1$, $j \in J_0$, are continuously differentiable and that C is regular. Moreover, let $\hat{\mu}: [0,\infty) \to [0,\infty)$ be as specified in Lemma 2.4.7. Consider any point x in the compact set $L^0 \cap C$. If $\sigma(x,J(x,0)) \leq 0$, then by Theorem 2.4.6 x is a conditional critical point of f on C and the iteration is terminated. If $\sigma(x,J(x,0)) > 0$, then the Zoutendijk direction algorithm has the form

$$\text{(2.4.8)} \quad \begin{cases} \text{(i) choose } \varepsilon \text{ by (2.4.7)} \\ \text{(ii) choose } s \in \hat{K}(x,J(x,\varepsilon),\sigma(x,J(x,\varepsilon))). \end{cases}$$

We observe that Zoutendijk [1960] considered only the case $\mu_0(t) \equiv t$, $\mu_j(t) \equiv \alpha_j t$, $\alpha_j > 0$, $j \in J_0^N$. Furthermore, our choice of ε is somewhat different from that of Zoutendijk [1960] although the effect is the same. Finally, Zoutendijk discusses various normalizations of s including our case $\|s\| = 1$. For ease of notation we have restricted ourselves to $\|s\| = 1$ although all of our results would remain valid for any other normalization as long as the generated sequence of directions remains bounded away from zero and infinity.

We now turn to the question of the convergence of Zoutendijk's method to a conditional critical point of f on C. In order to simplify the proof, we first isolate two lemmas; the first one concerns the remaining relationships between the quantities $f'(x)s$, $\sigma(x,J(x,\varepsilon))$ and ε as they are generated by the algorithm.

Lemma 2.4.8 Let $f:D \subset R^n \to R^1$ and $g_j:D \subset R^n \to R^1$, $j \in J_0$ be continuously differentiable on the open set D and assume that C is regular. Let $\{x^k\}$ be any convergent sequence in a compact subset $C_0 \subset C$ with $\lim_{k\to\infty} x^k = x^*$. Furthermore, let $\{\varepsilon_k\}$ be any sequence of nonnegative numbers such that $\lim_{k\to\infty} \varepsilon_k = 0$ and assume that the sequence $\{\sigma(x^k,J(x^k,\varepsilon_k))\}$ is nonnegative and satisfies $\lim_{k\to\infty} \sigma(x^k,J(x^k,\varepsilon_k)) = 0$. Then for any sequence $\{s^k\}$ with $s^k \in \hat{K}(x^k,J(x^k,\varepsilon_k),\sigma(x^k,J(x^k,\varepsilon_k)))$ and $\lim_{k\to\infty} s^k = s^*$, we have

$$\lim_{k\to\infty} f'(x^k)s^k = f'(x^*)s^* = 0.$$

Proof. Since $s^k \in \hat{K}(x^k,J(x^k,\varepsilon_k),\sigma(x^k,J(x^k,\varepsilon_k)))$, we have $f'(x^k)s^k \geq \mu_0(\sigma(x^k,J(x^k,\varepsilon_k))) \geq 0$ and hence $\lim_{k\to\infty} f'(x^k)s^k = f'(x^*)s^* \geq 0$. Suppose that $f'(x^*)s^* > 0$. Because J_0 is finite, $\{J(x^k,\varepsilon_k)\}$ has a limit set J^* and we may choose a subsequence $\{x^{k_i}\} \subset \{x^k\}$ such that $\lim_{i\to\infty} x^{k_i} = x^*$, $\lim_{i\to\infty} s^{k_i} = s^*$, $\lim_{i\to\infty} \varepsilon_{k_i} = 0$ and $J(x^{k_i},\varepsilon_{k_i}) = J^*$, $i \geq 0$ as well as $\lim_{i\to\infty} \sigma(x^{k_i},J^*) = 0$. By Lemma 2.4.4 $\sigma(\cdot,J^*)$ is continuous at x^* and hence evidently $\sigma(x^*,J^*) = 0$. Therefore, because of the continuity of f',g_j', $j \in J_0$, and μ_0,μ_j, $j \in J_0^N$ and because $\mu_0(0) = \mu_j(0) = 0$, $j \in J_0^N$, we have

$$f'(x^*)s^* > 0, \quad g_j'(x^*)s^* \geq 0, \quad j \in J^*.$$

Now by Lemma 2.4.5, $\sigma(x^*,J^*) = 0$ implies that (2.4.5) holds and hence the assumption $f'(x^*)s^* > 0$ leads to

$$0 > \nu_0 f'(x^*)s^* = \sum_{j\varepsilon J^*} (-\nu_j)g_j'(x^*)s^* \geq 0 .$$

This is a contradiction and therefore $f'(x^*)s^* = 0$.

The next lemma will allow us to conclude that $\sigma(x^*,J(x^*,0)) \leq 0$ for a limit point x^* of $\{x^k\}$.

Lemma 2.4.9 Let $g_j : D \subset R^n \to R^1$, $j \varepsilon J_0$ be continuous on the open set D. Consider any convergent sequence $\{x^k\} \subset C$ with $\lim_{k\to\infty} x^k = x^* \varepsilon C$ and any convergent sequence $\{\varepsilon_k\}$ of nonnegative numbers with $\lim_{k\to\infty} \varepsilon_k = 0$. Then there exists an integer k_0 such that

$$J(x^k,\varepsilon_k) \subset J(x^*,0), \quad \forall k \geq k_0.$$

In particular, all of the limit sets of $\{J(x^k,\varepsilon_k)\}$ are contained in $J(x^*,0)$.

Proof. It follows from the continuity of g_j, $j \varepsilon J_0$, that there is an $\hat{\varepsilon} > 0$ such that

$$J(x^*,0) = J(x^*,\varepsilon), \quad 0 \leq \varepsilon \leq \hat{\varepsilon}.$$

Choose k_0 such that for all $k \geq k_0$, $\varepsilon_k \leq \frac{\hat{\varepsilon}}{2}$ and

$$|g_j(x^k)-g_j(x^*)| \leq \frac{\hat{\varepsilon}}{2}, \quad \forall j \varepsilon J_0.$$

Then for any $j \varepsilon J(x^k,\varepsilon_k)$, $k \geq k_0$ we have

$$0 \geq g_j(x^*) \geq g_j(x^k) - |g_j(x^k)-g_j(x^*)| \geq -\varepsilon_k - \frac{\hat{\varepsilon}}{2} \geq -\hat{\varepsilon}.$$

Hence $j \in J(x^*,\hat{\varepsilon}) = J(x^*, 0)$ which completes the proof.

We are now prepared to prove the main convergence result.

Theorem 2.4.10 (Zoutendijk's method of feasible directions; Procedure P1) Let $f:D \subset R^n \to R^1$ and $g_j:D \subset R^n \to R^1$, $j \in J_0$ be continuously differentiable on the open convex set D and assume that the constraint set C is both regular and admissible. Suppose that for $x^0 \in C$ the set $L^0(f(x^0)) \cap C \equiv L^0 \cap C$ is compact, that f is hemivariate on $L^0 \cap C$, and that the conditional critical points of f on C in $L^0 \cap C$ are isolated. Consider the iteration (2.4.1) where at x^k, ε_k and s^k are chosen by (2.4.8) and τ_k by (2.3.7). Then $\lim_{k\to\infty} x^k = x^*$, with $x^* \in \Omega$, where Ω is the set of conditional critical points of f on C with nonpositive multipliers.

Proof. Suppose that $x^0,\ldots,x^k$, $k \geq 0$ are already well-defined and contained in $L^0 \cap C$ and satisfy (2.3.11) with $\omega = 1$ and (2.3.1) with $\varepsilon = \varepsilon_{k-1}$ and $\sigma = \sigma(x^{k-1},J(x^{k-1},\varepsilon_{k-1}))$. If $\sigma(x^k,J(x^k,0)) \leq 0$, then, by Theorem 2.4.6, $x^k \in \Omega$; otherwise, $\sigma(x^k,J(x^k,0)) > 0$ and, by Lemma 2.4.7, $\sigma(x^k,J(x^k,\varepsilon_k)) > 0$. Therefore, with $s^k \in \hat{K}(x^k,J(x^k,\varepsilon_k),\sigma(x^k,J(x^k,\varepsilon_k)))$, we have

$$f'(x^k)s^k \geq \mu_0(\sigma(x^k,J(x^k,\varepsilon_k))) > 0, \ \|s^k\| = 1. \tag{2.4.9}$$

Hence, by Theorem 2.3.5, $x^{k+1} \in L^0 \cap C$ and

(2.4.10) $$f(x^k) - f(x^{k+1}) \geq \Phi(f'(x^k)s^k, \sigma(x^k, J(x^k,\varepsilon_k)), \varepsilon_k)$$

with some F-function Φ of three variables dependent only on C and the steplength algorithm. Therefore, either $\sigma(x^i, J(x^i, 0)) \leq 0$ for some $i \geq 0$, and $x^i \in \Omega$ or, by induction, the entire sequence $\{x^k\}$ is well-defined, lies in $L^0 \cap C$ and satisfies (2.4.9) and (2.4.10) for each $k \geq 0$. In the latter case, it follows from Lemma 2.1.2 that at least one of the following three limit statements applies:

(2.4.11) $$\lim_{k\to\infty} f'(x^k)s^k = 0, \quad \lim_{k\to\infty} \sigma(x^k, J(x^k, \varepsilon_k)) = 0, \quad \lim_{k\to\infty} \varepsilon_k = 0.$$

By construction we have $\lim_{k\to\infty} \varepsilon_k = 0$ if and only if $\lim_{k\to\infty} \sigma(x^k, J(x^k, \varepsilon_k)) = 0$. Moreover, since μ_0 is an F-function on $[0,\infty)$, (2.4.9) shows that $\lim_{k\to\infty} \sigma(x^k, J(x^k, \varepsilon_k)) = 0$ whenever $\lim_{k\to\infty} f'(x^k)s^k = 0$ while Lemma 2.4.8 ensures the converse. Therefore, if any one of the relations in (2.4.11) holds, all three are valid and in particular $\lim_{k\to\infty} \sigma(x^k, J(x^k, \varepsilon_k)) = 0$. Now, $\{x^k\}$ is contained in the compact set $L^0 \cap C$ and J_0 is finite. Hence $\{x^k\}$ has a limit point x^* in $L^0 \cap C$ and $\{J(x^k, \varepsilon_k)\}$ has a limit set J^* which, by Lemma 2.4.9 is contained in $J(x^*, 0)$. Therefore we may choose a subsequence $\{x^{k_i}\} \subset \{x^k\}$ such that $\lim_{i\to\infty} x^{k_i} = x^*$, $J(x^{k_i}, \varepsilon_{k_i}) = J^*$, $\forall\, i$ and $\lim_{i\to\infty} \sigma(x^{k_i}, J^*) = 0$. Then, by Lemma 2.4.4, $\sigma(\cdot, J^*)$ is continuous on $L^0 \cap C$ and hence $\sigma(x^*, J^*) = 0$. Moreover, since $J^* \subset J(x^*, 0)$, it follows from Lemma 2.4.3 that

$$\sigma(x^*, J(x^*, 0)) \leq \sigma(x^*, J^*) = 0.$$

Thus, by Theorem 2.4.6, x^* is a conditional critical point of f on C with nonpositive multipliers. Finally, the induction proof also shows that $\{x^k\}$ satisfies (2.3.11) with $\omega = 1$ for each $k \geq 0$ and hence that the sequence $\{x^k\}$ is strongly downward on $L^0 \cap C$. Therefore, since f is hemivariate on $L^0 \cap C$, it follows from Lemma 2.1.5 that $\lim_{k\to\infty} \|x^k - x^{k+1}\| = 0$. Hence, since the conditional critical points of f on C in $L^0 \cap C$ are isolated, Lemma 2.2.6 gives $\lim_{k\to\infty} x^k = x^*$ as desired.

We note that any of the steplength algorithms of Section 2.3 could be used as long as the corresponding hypotheses on f are assumed.

2.5 Direction Algorithms

After considering in the preceding section the Zoutendijk choice of feasible directions, we will now turn to some other methods for obtaining feasible directions s at $x \in C$. In all cases, s will be chosen such that $f'(x)s > 0$ and

$$(2.5.1) \qquad s \in K(x, J(x,\varepsilon), \sigma)$$

with suitable $\sigma > 0$ and $\varepsilon > 0$, and more specifically, s will depend on a certain class of projection matrices. We begin with a brief survey of relevant results about such projection matrices without proofs. These results are well-known, and we refer to Householder [1964] for a general discussion or to Rosen [1960] for their relations to nonlinear programming.

Let $g_j : D \subset R^n \to R^1$, $j \in J_0$ be continuously differentiable on the open set D and assume that C is regular. For any given $x \in C$, and index set $J \subset J_0$, we order the $q = |J|$ vectors $g_j'(x)^T$, $j \in J$

in some as yet unspecified fashion, and denote the resulting sequence by $n_1(x),\ldots,n_q(x)$. Since C is regular, it is possible to assume that these vectors are linearly independent and hence that the $n \times i$ matrix

$$(2.5.2) \qquad N_i(x) = (n_1(x),\ldots,n_i(x)),\ i = 1,\ldots,q$$

has rank $i < n$. Then

$$(2.5.3) \qquad P_i(x) = I - N_i(x)[N_i(x)^T N_i(x)]^{-1} N_i(x)^T,\ i = 1,\ldots,q,\ x \in C$$

is the orthogonal projection from R^n onto the orthogonal complement of the column space $N_i(x)$:

$$P_i(x): R^n \to \text{span}\ \{n_1(x),\ldots,n_i(x)\}^{\perp},\ i = 1,\ldots,q,\ x \in C.$$

We add the convention $P_0(x) \equiv I$ to simplify notation. Note that $N_i(x)$ and hence also $P_i(x)$ depend not only on $x \in C$ but also on the index set J. Rosen [1960] discusses a procedure for computing the sequence of projection matrices by using the Sherman-Morrison formula. Since in our later application there will be only one constraint, we do not require this procedure; and therefore we do not go into further details here.

For fixed $x \in C$ and any index set $J \subset J_0$ such that rank $N_i(x) = i$, $i = 1,\ldots,q = |J|$, the following properties of the projection matrices are well-known or immediate.

$$(2.5.4)\quad \begin{cases} \text{(i) } P_i(x),\ i = 1,\dots,q \text{ is symmetric and idempotent} \\ \text{(ii) } P_i(x)y = 0 \text{ if and only if } y \in \text{span}\,\{n_1(x),\dots,n_i(x)\}, \\ \qquad\qquad i = 1,\dots,q \\ \text{(iii) } P_i(x) = P_i(x)P_{i-1}(x) = P_{i-1}(x)P_i(x),\ i = 1,\dots,q. \end{cases}$$

As a direct consequence of (2.5.4) (i) and (ii), we have in particular

$$(2.5.5)\qquad P_i(x)n_k(x) = 0,\ k = 1,\dots,i,\ i = 1,\dots,q$$

$$(2.5.6)\qquad y^T P_i(x)y = \|P_i(x)y\|_2^2,\quad \forall y \in R^n,\ i = 1,\dots,q.$$

By the orthogonal projection property of $P_i(x)$ there exist, for any $y \in R^n$, vectors $v \in R^i$, $i = 1,\dots,q$, such that

$$(2.5.7)\qquad y = (I-P_i(x))y + P_i(x)y = N_i(x)v + P_i(x)y,\ i = 1,\dots,q.$$

Hence

$$(2.5.8)\qquad P_{i-1}(x)y = v_i P_{i-1}(x)n_i(x) + P_i(x)y,\ i = 1,\dots,q$$

and, from (2.5.5) and (2.5.7),

$$(2.5.9)\qquad v = [N_i(x)^T N_i(x)]^{-1} N_i(x)^T y,\ i = 1,\dots,q.$$

Moreover, (2.5.5) together with (2.5.8) imply that

$$(2.5.10)\qquad \|P_q(x)y\|_2 \leq \|P_{q-1}(x)y\|_2 \leq \dots \leq \|P_1(x)y\|_2 \leq \|y\|_2,\quad \forall y \in R^n.$$

Finally, if $C_0 \subset C$ is a compact subset of C, then

(2.5.11) $\|(N_i(x)^T N_i(x))^{-1}\| \leq \kappa, \quad \forall x \in C_0, \; i = 1,\ldots,q,$

and it follows from (2.5.5) and (2.5.9) with $y = P_{q-1}(x)n_q(x)$ that

(2.5.12) $\|P_{i-1}(x)n_i(x)\| \geq \kappa^{-1/2} > 0, \quad \forall x \in C_0, \; i = 1,\ldots,q.$

For the special case $y = f'(x)^T$, $x \in C$ and $i = q$ in (2.5.9), we write

(2.5.13) $r(x) = [N_q(x)^T N_q(x)]^{-1} N_q(x)^T f'(x)^T \equiv (r_1(x),\ldots,r_q(x))^T.$

Then Mangasarian [1963] has shown the equivalence of the next result with the Kuhn-Tucker theorem.

Theorem 2.5.1 Let $f: D \subset R^n \to R^1$ and $g_j: D \subset R^n \to R^1$, $j \in J_0$ be continuously differentiable on the open set D and suppose that C is regular. Then $x^* \in C$ is a conditional critical point of f on C if and only if

(2.5.14) $P_q(x^*)f'(x^*)^T = 0, \qquad q = |J(x^*,0)|.$

Moreover, if f is convex, then x^* is a minimum of f on C if and only if (2.5.14) and

(2.5.15) $r(x^*) \leq 0$

hold.

Proof. Suppose that $P_q(x^*)f'(x^*)^T = 0$. Then it follows from (2.5.4) (ii) that $f'(x^*)^T \in \text{span}\,\{g_j'(x^*)^T, \; j \in J(x^*,0)\}$, and hence that there exist numbers ν_0, ν_j, $j \in J(x^*,0)$ not all zero, such that

$$\nu_0 f'(x^*)^T + \sum_{j \in J(x^*,0)} \nu_j g_j'(x^*)^T = 0.$$

If $\nu_0 = 0$, then $g_j'(x^*)^T$, $j \in J(x^*,0)$ would be linearly dependent, in contradiction of the regularity of C. Hence $\nu_0 \neq 0$ and by (2.2.5) x^* is a conditional critical point of f on C.

Conversely, if x^* is a conditional critical point of f on C, then (2.2.5) holds and thus by (2.5.4) (ii)

$$P_q(x^*)f'(x^*)^T = \sum_{j \in J(x^*,0)} \nu_j P_q(x^*) g_j'(x^*)^T = \sum_{i=1}^{q} \hat{\nu}_i P_q(x^*) n_i(x^*) = 0$$

where $\hat{\nu}_i$, $i = 1,\ldots,q$ is a renumbering of the ν_j. Hence (2.5.14) holds at x^*. For the last part of the theorem we need to observe only that if x^* is a conditional critical point of f on C then $P_q(x^*)f'(x^*)^T = 0$ and

$$N_q(x^*)r(x^*) = (I-P_q(x^*))f'(x^*)^T = f'(x^*)^T$$

and thus $r(x^*)$ simply represents the multipliers ν_j, $j \in J(x^*,0)$. Hence, the result is exactly the Kuhn-Tucker theorem.

It is important to point out that if f is not convex then (2.5.14) and (2.5.15) together are not, in general, sufficient for a point $x^* \in C$ to be a minimum of f on C although they are necessary conditions for this. Indeed, since C is regular, Lemma 2.2.5 implies that the Kuhn-Tucker constraint qualification holds, and hence the necessity of (2.5.14/15) is a consequence of Theorems 2.2.4 and 2.5.1.

We now consider the question of how a direction s might be chosen so that with the steplength algorithms of Section 2.3, the iteration converges to points $x^* \varepsilon C$ for which (2.5.14) holds and possibly also (2.5.15). As mentioned before, s is supposed to satisfy (2.5.1) for certain positive ε and σ and in Section 2.4 we saw that the three quantities $f'(x)s$, ε and σ have to be suitably related in order to conclude convergence on the basis of the estimate (2.3.1). We first discuss the specification of ε. Suppose that $f:D \subset R^n \to R^1$ and $g_j:D \subset R^n \to R^1$, $j \varepsilon J_0$ are continuously differentiable on the open set D and that C is regular. If $\bar{\varepsilon} > 0$ is the constant of Definition 2.2.1 and x is any given point in C, then we use the following procedure for choosing ε.

Algorithm for ε

(i) If $\|P_q(x)f'(x)^T\| = 0$, $q = |J(x,0)|$, set $\varepsilon = 0$.

(ii) Otherwise, with fixed isotone F-function $\hat{\mu}:[0,\infty) \to [0,\infty)$ for which $\hat{\mu}(0) = 0$ and $\hat{\mu}(t) > 0$ when $t > 0$, let $\psi:[0,\infty) \to [0,\infty)$,

(2.5.16) $$\psi(\delta) = \begin{cases} \|P_q(x)f'(x)^T\|, \ q = |J(x,\delta)|, \text{ if } \delta \ \varepsilon \ [0,\bar{\varepsilon}] \\ \psi(\bar{\varepsilon}), \text{ if } \delta \ \varepsilon \ [\bar{\varepsilon},\infty) \end{cases}$$

and determine ε by

(2.5.17) $$\varepsilon = \frac{1}{2} \sup \{\delta \ \varepsilon \ [0,\bar{\varepsilon}] \,|\, \hat{\mu}(\psi(\delta)) - \delta \geq 0\}.$$

Since C is regular and q changes only finitely many times as δ ranges over $[0,\infty)$, it follows from (2.5.10) that ψ is a well-defined antitone step-function. We use Lemma 2.4.7 in order to show that ε is well-defined and positive. Suppose that $\|P_q(x)f'(x)^T\| > 0$, $q = |J(x,0)|$. Clearly then, by the regularity of C, $\psi(\delta) = \psi(0) > 0$ for all $\delta \in [0,\delta_0]$ with some $\delta_0 > 0$ and $\psi(\delta) = \bar{\varepsilon}$ for all $\delta \in [\bar{\varepsilon},\infty)$. Therefore, since $\psi(0) > 0$ and ψ is antitone, we see that the definition (2.5.17) is of the same form as (2.4.7), and hence, by Lemma 2.4.7 that ε is well-defined and positive. Therefore, either $\varepsilon = 0$ and x is a conditional critical point of f on C or $\varepsilon > 0$ and $\psi(\varepsilon) > 0$.

We now turn to the question of choosing a suitable direction s. For notational simplicity and without loss of generality we assume that for the index set $J(x,\varepsilon)$, where ε is now fixed at x by the preceding algorithm, that the vector $r(x)$ given by (2.5.13) satisfies

(2.5.18) $$r_q(x) \geq r_i(x),\ i = 1,\dots,q-1,\ q = |J(x,\varepsilon)|.$$

In fact, this means only that at each point x the columns of $N_q(x)$ are suitably ordered. We now introduce an integer-valued function ℓ_1, with $q = |J(x,\varepsilon)|$, by

(2.5.19) $$\ell_1(x,\varepsilon) = \begin{cases} q-1, & \text{if } \|P_q(x)f'(x)^T\| < \frac{1}{2}\, r_q(x)\|P_{q-1}(x)n_q(x)\| \\ q, & \text{otherwise.} \end{cases}$$

At first thought one might consider choosing s, for example, by

(2.5.20) $$s = P_\ell(x)f'(x)^T/\|P_\ell(x)f'(x)^T\|,\ \ell = \ell_1(x,\varepsilon).$$

However, then we are only guaranteed that $s \in K(x,J(x,\varepsilon),0)$ which is not sufficient for feasibility. To see this, note first that if $\|P_q(x)f'(x)^T\| = 0$, $q = |J(x,\varepsilon)|$, then, by the Algorithm for ε, we have $\varepsilon = 0$ and x is a conditional critical point of f on C. Otherwise, (2.5.10) implies that

$$\|P_\ell(x)f'(x)^T\| \geq \|P_q(x)f'(x)^T\| > 0, \quad \ell = \ell_1(x,\varepsilon)$$

and hence s is well-defined. From

$$P_\ell(x)f'(x)^T \in \text{span}\,\{n_1(x),\ldots,n_\ell(x)\}^\perp$$

it follows that

$$(2.5.21) \qquad n_i(x)^T s = 0, \quad i = 1,\ldots,\ell = \ell_1(x,\varepsilon)$$

and for $\ell = q-1$, (2.5.19) shows that

$$\|P_q(x)f'(x)^T\| < \frac{1}{2}\, r_q(x)\|P_{q-1}(x)n_q(x)\|, \quad r_q(x) > 0, \quad q = |J(x,\varepsilon)|.$$

Now, by (2.5.8) and (2.5.13)

$$(2.5.22) \qquad P_{q-1}(x)f'(x)^T = r_q(x)P_{q-1}(x)n_q(x) + P_q(x)f'(x)^T$$

and hence, together with (2.5.5) and (2.5.6), we obtain

$$n_q(x)^T s = r_q(x)\|P_{q-1}(x)n_q(x)\|^2/\|P_{q-1}(x)f'(x)^T\| > 0.$$

Thus, in all cases

$$n_i(x)^T s \geq 0, \quad i = 1,\ldots,q$$

and, because the vectors $n_i(x)$, $i = 1,\ldots,q$ are the gradients $g_j'(x)^T$, $j \in J(x,\varepsilon)$ in a certain order, it follows that indeed $s \in K(x,J(x,\varepsilon),0)$. As noted before, we require a stronger condition on s in order to guarantee feasibility as well as the basic estimate (2.3.1) and, for s defined by (2.5.20), the latter is precluded by (2.5.21). We note that for the nonlinear case, Rosen [1961] developed an algorithm which allowed directions of the form (2.5.20). However, it was necessary for him to introduce a correction term whenever such directions led away from the constraint set. Our approach will be to modify the vector s by considering instead of (2.5.20) normalized vectors of the form $P_\ell(x)p+\bar{\alpha}\sum_{i=1}^{q}\beta_i n_i(x)$, $\ell = \ell_1(x,\varepsilon)$ where p is a suitable direction and $\bar{\alpha},\beta_i, i = 1,\ldots,q$ are certain real numbers. In other words, we add to $P_\ell(x)p$ some specified linear combination of the gradients $g_j'(x)^T$, $j \in J(x,\varepsilon)$. This procedure is similar to one considered by Kalfon, Ribiere, and Sogno [1968] for projected gradient directions, that is, for the case $p = f'(x)^T$. At first, we shall restrict ourselves as well to this gradient case and discuss some other cases later.

Let $f:D \subset R^n \to R^1$ and $g_j:D \subset R^n \to R^1$, $j \in J_0$ be continuously differentiable on the open set D. Let $C_0 \subset C$ be a compact subset of the regular constraint set C and consider the following procedure.

<u>Algorithm for s</u>

I_1. Let $x \in C_0$ be given and choose ε by the Algorithm for ε.

(i) If $\|P_q(x)f'(x)^T\| = 0$, $q = |J(x,\varepsilon)|$, then the process is terminated.

(ii) Otherwise, determine $\ell = \ell_1(x,\varepsilon)$ and let $p = f'(x)^T$.

II. With fixed isotone F-functions $\bar{\mu}_j:[0,\infty) \to [0,\infty)$, $j \in J_0^N$ for which $\bar{\mu}_j(t) \neq 0$ for all $t \in [0,\infty)$, define the vector $z(x) \in R^q$, $q = |J(x,\varepsilon)|$ by

$$(2.5.23)\qquad z_i(x) = \begin{cases} \bar{\mu}_j(f'(x)P_\ell(x)p), & \text{if } n_i(x) = g_j'(x)^T,\ j \in J^N(x,\varepsilon) \\ & i = 1,\dots,q \\ 0, & \text{if } n_i(x) = g_j'(x)^T,\ j \in J^L(x,\varepsilon) \end{cases}$$

and set

$$(2.5.24)\qquad \beta = [N_q(x)^T N_q(x)]^{-1} z(x),\quad \beta = (\beta_1,\dots,\beta_q)^T.$$

III. Choose $\bar{\alpha} \geq 0$ by

$$(2.5.25)\qquad \bar{\alpha} = \begin{cases} \frac{1}{2} f'(x)P_\ell(x)p/\left|\sum_{i=1}^q \beta_i f'(x)n_i(x)\right|, & \text{if } \left|\sum_{i=1}^q \beta_i f'(x)n_i(x)\right| \\ & \geq (f'(x)P_\ell(x)p)^{1/2} > 0 \\ \frac{1}{2}(f'(x)P_\ell(x)p)^{1/2}, & \text{otherwise} \end{cases}$$

and set

$$(2.5.26)\qquad s = \left[P_\ell(x)p + \bar{\alpha}\sum_{i=1}^q \beta_i n_i(x)\right] / \left\|P_\ell(x)p + \bar{\alpha}\sum_{i=1}^q \beta_i n_i(x)\right\|.$$

We now discuss the validity of the steps II and III: If $\|P_q(x)f'(x)^T\| = 0$, $q = |J(x,\varepsilon)|$, then the Algorithm for ε implies that $\varepsilon = 0$ and hence that x is a conditional critical point of f on C. Otherwise, $\varepsilon > 0$ and, by (2.5.6) and (2.5.10), $f'(x)P_\ell(x)p \geq f'(x)P_q(x)p > 0$, $\ell = \ell_1(x,\varepsilon)$, and the procedure continues to step II. For the validity of II we note that by the regularity of C together with $\varepsilon \leq \bar{\varepsilon}$, where $\bar{\varepsilon}$ is the constant of Definition 2.2.1, the vectors $n_i(x)$, $i = 1,\ldots,q$ and hence also $N_q(x)^T n_i(x)$, $i = 1,\ldots,q$ are linearly independent. Therefore, the system

$$\sum_{i=1}^{q} \beta_i N_q(x)^T n_i(x) = z(x) \tag{2.5.27}$$

has a unique solution $\beta = (\beta_1,\ldots,\beta_q)^T$, namely (2.5.24). Proceeding to III we observe that, by (2.5.11) and the continuity of all derivatives,

$$\left|\sum_{i=1}^{q} \beta_i f'(x) n_i(x)\right| \leq \kappa_1 \tag{2.5.28}$$

where κ_1 is independent of $x \in C_0$. Hence, from (2.5.25) and (2.5.28) it follows that $\bar{\alpha}$ is well-defined and positive and that

$$\bar{\alpha} \geq \bar{\mu}(f'(x)P_\ell(x)p) \tag{2.5.29}$$

where $\bar{\mu}(t) = \frac{1}{2} \min\{\frac{t}{\kappa_1}, t^{1/2}\}$ is clearly an F-function. If $J^N(x,\varepsilon) = \{\emptyset\}$, then $z(x) = 0$ and (2.5.26) reduces to (2.5.20); that is, s is well-defined. Otherwise, $z(x) \neq 0$ and hence also $\sum_{i=1}^{q} \beta_i n_i(x) \neq 0$. Moreover, $f'(x)P_\ell(x)p > 0$ together with (2.5.25) implies that $\|P_\ell(x)p+\bar{\alpha}\sum_{i=1}^{q} \beta_i n_i(x)\| \neq 0$; thus s is again well-defined.

It is important to note that steps II and III are not dependent on our particular choice of p in I_1. In other words, these steps remain well-defined as long as some direction p is chosen for which $f'(x)P_\ell(x)p > 0$ and some appropriate index ℓ is used which need not be equal to $\ell_1(x,\varepsilon)$. We shall present below several examples of such choices of p and ℓ.

We note further that the procedure defined here may stop at a conditional critical point of f on C which has at least one positive multiplier. This corresponds to the well-known possibility in the unconstrained case that a minimization procedure stops at a saddlepoint. We shall discuss later a procedure for moving away from such points so that ultimately we stop only at conditional critical points of f on C with nonpositive multipliers.

We return to the Algorithm for s and show next that there the directions s have the desired feasibility properties.

Theorem 2.5.2 Let $f:D \subset R^n \to R^1$ and $g_j:D \subset R^n \to R^1$, $j \in J_0$ be continuously differentiable on the open set D and assume that C is regular. Let $C_0 \subset C$ be a compact set and $x \in C_0$ a given point which is not a conditional critical point of f on C. Then the vector $s \in R^n$ given by the Algorithm for s is well-defined and satisfies, for $\ell = \ell_1(x,\varepsilon)$, $s \in K(x,J(x,\varepsilon)$, $f'(x)P_\ell(x)p)$ as well as $f'(x)s \geq \mu_0(f'(x)P_\ell(x)p)$ for some F-function μ_0.

Proof. Since $x \in C_0$ is not a conditional critical point of f on C, we find that ε is well-defined and positive, and that $f'(x)P_\ell(x)p > 0$, $\ell = \ell_1(x,\varepsilon)$. Thus, by our previous discussion of the steps II and III we see that also s is well-defined.

From (2.5.25) it follows that $\bar{\alpha} \leq \frac{1}{2}(f'(x)P_\ell(x)p)^{1/2}$. Thus, using (2.5.11), (2.5.24) and the uniform boundedness of f' on C_0 we obtain $\|p\| \leq \kappa_2$ and

$$\|P_\ell(x)p + \bar{\alpha} \cdot \sum_{i=1}^{q} \beta_i n_i(x)\| \leq \kappa_3$$

where $\kappa_2, \kappa_3 \in (0,\infty)$ are independent of $x \in C_0$. Together with (2.5.25) and $f'(x)P_\ell(x)p > 0$ this leads to

$$f'(x)s = [f'(x)P_\ell(x)p+\bar{\alpha}\sum_{i=1}^{q}\beta_i f'(x)n_i(x)]/\|P_\ell(x)p+\sum_{i=1}^{q}\beta_i n_i(x)\|$$

$$\geq [f'(x)P_\ell(x)p-\bar{\alpha}|\sum_{i=1}^{q}\beta_i f'(x)n_i(x)|]/\|P_\ell(x)p+\sum_{i=1}^{q}\beta_i n_i(x)\|$$

$$\geq \frac{1}{2} f'(x)P_\ell(x)p/\|P_\ell(x)p+\bar{\alpha}\sum_{i=1}^{q}\beta_i n_i(x)\|$$

$$\geq \frac{1}{2\kappa_3} f'(x)P_\ell(x)p,$$

that is, $f'(x)s \geq \mu_0(f'(x)P_\ell(x)p) > 0$ with the F-function $\mu_0(t) = \frac{1}{2\kappa_3} t$.

Now using (2.5.5) and (2.5.27) we have

$$n_i(x)^T s = [n_i(x)^T P_\ell(x)p+\bar{\alpha}\sum_{k=1}^{q}\beta_k n_i(x)^T n_k(x)]/\|P_\ell(x)p+\bar{\alpha}\sum_{k=1}^{q}\beta_k n_k(x)\|$$

$$= \bar{\alpha}z_i(x)/\|P_\ell(x)p+\bar{\alpha}\sum_{k=1}^{q}\beta_k n_k(x)\|$$

$$\geq \kappa_3^{-1}\bar{\alpha}z_i(x),\ i = 1,\ldots,\ell.$$

Suppose that $\ell = q-1$; then it follows from (2.5.19) that $r_q(x) > 0$ and hence from (2.5.5) and (2.5.22) that

$$n_q(x)^T s = [r_q(x)\|P_{q-1}(x)n_q(x)\|^2 + \bar{\alpha}z_q(x)]/\|P_{q-1}(x)p+\bar{\alpha}\sum_{k=1}^{q}\beta_k n_k(x)\|$$

(2.5.30) $$\geq \kappa_3^{-1}\bar{\alpha}z_q(x).$$

Therefore, by (2.5.23) and (2.5.29)

$$g_j'(x)s \geq \kappa_3^{-1}\bar{\mu}(f'(x)P_\ell(x)p)\bar{\mu}_j(f'(x)P_\ell(x)p),\ j \in J^N(x,\varepsilon)$$

$$= \mu_j(f'(x)P_\ell(x)p),\ j \in J^N(x,\varepsilon)$$

where $\mu_j(t) = \kappa_3^{-1}\bar{\mu}(t)\bar{\mu}_j(t)$, $j \in J^N(x,\varepsilon)$ are clearly isotone F-functions with $\mu_j(0) = \bar{\mu}(0) = 0$. In addition, we have

$$g_j'(x)s \geq 0, \; j \in J^L(x,\varepsilon)$$

and hence altogether $s \in K(x,J(x,\varepsilon),f'(x)P_\ell(x)p)$.

Note that with the help of (2.5.6) we can write the conclusions of this theorem also in the form

$$s \in K(x,J(x,\varepsilon),\|P_\ell(x)f'(x)^T\|^2), f'(x)s \geq \mu_0(\|P_\ell(x)f'(x)^T\|^2), \quad \ell = \ell_1(x,\varepsilon). \tag{2.5.31}$$

Note further that we used $p = f'(x)^T$ only to ensure that $f'(x)P_\ell(x)p > 0$ and that s is well-defined; in addition, we needed the estimate $\|p\| \leq \kappa_2$ with some fixed constant $\kappa_2 \in (0,\infty)$ independent of $x \in C_0$. Similarly, the properties of $\ell = \ell_1(x,\varepsilon)$ defined in (2.5.19) were needed only to obtain (2.5.30) in the case $\ell = q-1$. Consequently, Theorem 2.5.2 will remain valid for any uniformly bounded choice of p which satisfies $f'(x)P_\ell(x)p > 0$ and produces a well-defined s, and for any definition of ℓ which ensures that (2.5.30) holds if $\ell = q-1$.

In order to discuss a particular generalization of our choice of p we present first a lemma given by Ortega and Rheinboldt [1970; Theorem 14.4.1].

Lemma 2.5.3 Let $D_0 \subset D \subset R^n$ be any compact subset of D and $A: D_0 \to L(R^n)$ any continuous mapping such that $A(x)$ is positive definite for each $x \in D_0$. Then there exist constants $0 < \eta_1 \leq \eta_2$ such that

(2.5.32) $h^T A(x)^{-1} h \geq (\eta_1/\eta_2^2)\|h\|^2, \quad \forall\, x \in D_0,\ h \in R^n.$

<u>Proof</u>. Evidently, $A(x)^{-1}$ exists and is positive definite for all $x \in C_0$, and by the continuity of A on the compact set C_0 there clearly exist constants $0 < \eta_1 \leq \eta_2$ such that

$$\|A(x)\| \leq \eta_2,\ h^T A(x) h \geq \eta_1 \|h\|^2, \quad \forall x \in C_0,\ h \in R^n.$$

Hence, by the Cauchy-Schwarz inequality

$$\|A(x)h\| \geq \eta_1 \|h\|, \quad \forall x \in C_0,\ h \in R^n$$

and

(2.5.33) $\|h\| = \|A(x)A(x)^{-1}h\| \geq \eta_1 \|A(x)^{-1}h\|, \quad \forall\, x \in C_0,\ h \in R^n.$

Similarly,

$$\|h\| \leq \eta_2 \|A(x)^{-1}h\|, \quad \forall x \in C_0,\ h \in R^n$$

so that with $\hat{h} = A(x)^{-1}h$ we see that

$$h^T A(x)^{-1} h = \hat{h}^T A(x)^T \hat{h} = \hat{h}^T A(x) \hat{h}$$

$$\geq \eta_1 \|\hat{h}\|^2 \geq (\eta_1/\eta_2^2)\|h\|^2.$$

We return to our intended generalization of the choice of p. For this we replace step I_1 in the Algorithm for s by the following step I_2. As before, $C_0 \subset C$ is a compact subset of the regular set C.

I_2. Let $A:C_0 \to L(R^n)$ be a continuous mapping such that $A(x)$ is positive definite for each $x \in C_0$ and $0 < \eta_1 \leq \eta_2$ the constants of Lemma 2.5.3. Let $x \in C_0$ be given and choose ε by the Algorithm for ε. Define the integer-valued mapping ℓ_2, with $q = |J(x,\varepsilon)|$, by

$$\ell_2(x,\varepsilon) = \begin{cases} q-1, & \text{if } \|A(x)^{-1}P_q(x)f'(x)^T\| < \frac{1}{2}(\eta_1/\eta_2^2)r_q(x)\|P_{q-1}(x)n_q(x)\| \\ q, & \text{otherwise.} \end{cases} \tag{2.5.34}$$

(i) If $\|P_q(x)f'(x)^T\| = 0$, $q = |J(x,\varepsilon)|$, then the process is terminated.

(ii) Otherwise set $p = A(x)^{-1}P_\ell(x)f'(x)^T$, $\ell = \ell_2(x,\varepsilon)$.

The next theorem shows that the Algorithm for s with I_2 replacing I_1 produces a well-defined vector $s \in R^n$ and that s satisfies a relation similar to (2.5.31).

Theorem 2.5.4 Let $f:D \subset R^n \to R^1$ and $g_j:D \subset R^n \to R^1$, $j \in J_0$ be continuously differentiable on the open set D and assume that $C_0 \subset C$ is a compact subset of the regular constraint set C. Suppose that $x \in C_0$ is not a conditional critical point of f on C and let $s \in R^n$ be chosen by the Algorithm for s with I_1 replaced by I_2. Then s is well-defined and, with $\ell = \ell_2(x,\varepsilon)$,

$$s \in K(x, J(x,\varepsilon), (\eta_1/\eta_2^2)\|P_\ell(x)f'(x)^T\|^2), \tag{2.5.35}$$

$$f'(x)s \geq \mu_0((\eta_1/\eta_2^2)\|P_\ell(x)f'(x)^T\|^2).$$

Proof. Since $x \varepsilon C_0$ is not a conditional critical point of f on C we see again that ε is well-defined and positive and that $\|P_q(x)f'(x)^T\| > 0$, $q = |J(x,\varepsilon)|$. Hence, by (2.5.10) and the positive definiteness of $A(x)^{-1}$ we see that

$$f'(x)P_\ell(x)p = (P_\ell(x)f'(x)^T)^T A(x)^{-1} P_\ell(x)f'(x)^T > 0, \quad \ell = \ell_2(x,\varepsilon)$$

and therefore that s is well-defined.

For the proof of (2.5.35) we first show that $s \varepsilon K(x,J(x,\varepsilon),f'(x)P_\ell(x)p)$. As in Theorem 2.5.2 it follows that

$$n_i(x)^T s \geqslant \bar{\kappa}_3^{-1}\bar{\alpha} z_i(x), \quad i = 1,\ldots,\ell = \ell_2(x,\varepsilon).$$

If $\ell = q-1$, then $r_q(x) > 0$ and, using (2.5.22) and (2.5.32) with $h = P_{q-1}(x)n_q(x)$ we have

$$\begin{aligned}
n_q(x)^T P_{q-1}(x)p &= n_q(x)^T P_{q-1}(x)A(x)^{-1}P_{q-1}(x)f'(x)^T \\
&= r_q(x)n_q(x)^T P_{q-1}(x)A(x)^{-1}P_{q-1}(x)n_q(x) \\
&\qquad + n_q(x)^T P_{q-1}(x)A(x)^{-1}P_q(x)f'(x)^T \\
&\geqslant r_q(x)(\eta_1/\eta_2^2)\|P_{q-1}(x)n_q(x)\|^2 \\
&\qquad - \|A(x)^{-1}P_q(x)f'(x)^T\|\|P_{q-1}(x)n_q(x)\| \\
&> \frac{1}{2} r_q(x)(\eta_1/\eta_2^2)\|P_{q-1}(x)n_q(x)\|^2 > 0.
\end{aligned}$$

Therefore,

$$n_q(x)^T s \geqslant \kappa_3^{-1}[\tfrac{1}{2} r_q(x)(\eta_1/\eta_2^2)\|P_{q-1}(x)n_q'(x)\|^2 + \bar{\alpha}z_q(x)]$$

$$\geqslant \kappa_3^{-1}\bar{\alpha}z_q(x)$$

and hence (2.5.30) holds. Thus, as in Theorem 2.5.2

$$s \in K(x,J(x,\varepsilon),f'(x)P_\ell(x)p),\ f'(x)s \geqslant \mu_0(f'(x)P_\ell(x)p),\ \ell = \ell_2(x,\varepsilon).$$

Now using (2.5.32) with $h = P_\ell(x)f'(x)^T$ we see that

$$f'(x)P_\ell(x)p \geqslant (\eta_1/\eta_2^2)\|P_\ell(x)f'(x)^T\|^2$$

and hence, by the isotonicity of μ_0, μ_j, $j \in J_0^N$ that

$$g_j'(x)s \geqslant \mu_j(f'(x)P_\ell(x)p)$$

$$\geqslant \mu_j((\eta_1/\eta_2^2)\|P_\ell(x)f'(x)^T\|^2),\ j \in J_0^N.$$

and

$$f'(x)s \geqslant \mu_0(f'(x)P_\ell(x)p) \geqslant \mu_0((\eta_1/\eta_2^2)\|P_\ell(x)f'(x)^T\|^2).$$

Another choice of the vector p is given by the well-known Gauss-Southwell algorithm. For this we replace step I_1 of the Algorithm for s with the following procedure.

I_3. Let $x \in C_0$ be given and choose ε by the Algorithm for ε. Define the integer-valued mapping ℓ_3 by

(2.5.36) $\qquad \ell_3(x,\varepsilon) = q = |J(x,\varepsilon)|.$

(i) If $\|P_q(x)f'(x)^T\| = 0$, $q = |J(x,\varepsilon)|$, then the procedure is terminated.

(ii) Otherwise, choose p as the coordinate direction $\pm e^i$, $1 \leqslant i \leqslant n$ such that

$$(2.5.37)\quad \begin{cases} |f'(x)P_\ell(x)e^i| \\ \qquad = \max\limits_{1\leqslant k\leqslant n} |f'(x)P_\ell(x)e^k|, \ \ell = \ell_3(x,\varepsilon) \\ p = \operatorname{sgn}\,(f'(x)P_\ell(x)e^i)e^i. \end{cases}$$

We show next that here too the Algorithm for s with I_3 replacing I_1 produces a well-defined vector $s \in R^n$ and that again s satisfies a relation similar to (2.5.31).

<u>Theorem 2.5.5</u> Let $f:D \subset R^n \to R^1$ and $g_j:D \subset R^n \to R^1$, $j \in J_0$ be continuously differentiable on the open set D and assume that $C_0 \subset C$ is a compact subset of the regular constraint set C. Suppose that $x \in C_0$ is not a conditional critical point of f on C and let $s \in R^n$ be chosen by the Algorithm for s with I_1 replaced by I_3. Then s is well-defined and, with $\ell = \ell_3(x,\varepsilon)$,

$$(2.5.38)\quad s \in K(x,J(x,\varepsilon),n^{-1/2}\|P_\ell(x)f'(x)^T\|_2),$$

$$f'(x)s \geqslant \mu_0(n^{-1/2}\|P_\ell(x)f'(x)^T\|_2).$$

<u>Proof</u>. As before, the fact that x is not a conditional critical point of f on C implies that $\varepsilon > 0$ and $\|P_q(x)f'(x)^T\| > 0$, $q = |J(x,\varepsilon)|$. Since (2.5.37) is equivalent to

$$f'(x)P_\ell(x)p = \|P_\ell(x)f'(x)^T\|_\infty, \quad \ell = \ell_3(x,\varepsilon) = q$$

we readily see that

$$f'(x)P_\ell(x)p \geq n^{-1/2}\|P_\ell(x)f'(x)^T\|_2 > 0, \qquad (2.5.39)$$
$$\ell = \ell_3(x,\varepsilon) = q$$

and hence that s is well-defined by (2.5.26).

The proof of (2.5.38) is evident since, as in Theorem 2.5.2,

$$n_i(x)^T s \geq \kappa_3^{-1}\bar{\alpha} z_i(x), \quad i = 1,\ldots,\ell = \ell_3(x,\varepsilon) = q$$

and hence

$$s \in K(x,J(x,\varepsilon),f'(x)P_\ell(x)p), f'(x)s \geq \mu_0(f'(x)P_\ell(x)p), \quad \ell = \ell_3(x,\varepsilon)$$

which, together with (2.5.39) and the isotonicity of μ_0, μ_j, $j \in J_0^N$, yields (2.5.38).

We close this section by indicating how one might proceed if any of the procedures discussed in this section terminates at a conditional critical point of f on C with at least one positive multiplier. Suppose that $x \in C$ satisfies

$$\|P_q(x)f'(x)^T\| = 0, \quad r_q(x) > 0, \quad q = |J(x,0)|.$$

Then it follows from Theorem 2.4.6 that $\sigma(x,J(x,0)) > 0$ where σ is defined by (2.4.3). Hence the Zoutendijk algorithm (2.4.8) is well-defined and we may choose a vector

$$s \in \hat{K}(x,J(x,\varepsilon),\sigma(x,J(x,\varepsilon)))$$

with $\varepsilon > 0$ given by (2.4.7). Clearly then, we obtain

$$f'(x)s \geq \mu_0(\sigma(x,J(x,\varepsilon))) > 0, \; g_j'(x)s \geq 0, \; j \in J^L(x,\varepsilon),$$

$$g_j'(x)s \geq \mu_j(\sigma(x,J(x,\varepsilon))) > 0, \; j \in J^N(x,\varepsilon).$$

This means that s is feasible and that one can take a step in the direction of $-s$ and obtain a reduced value of f on C. Hence, if there are only finitely many conditional critical points in a compact subset $C_0 \subset C$, then for minimization methods in which the iterates are of the form (2.2.6/7) and remain in C_0, we need only apply the above procedure at most finitely many times.

2.6 Convergence Theorems

In this section we combine the results of the preceding sections into complete convergence proofs for iterative methods of the form

$$(2.6.1) \qquad x^{k+1} = x^k - \omega_k \tau_k s^k, \; f(x^k) \geq f(x^{k+1}), \; k = 0,1,\ldots$$

for solving the constrained minimization problem. Several new methods can be generated in this way, and as corollaries of the basic theorems we also obtain results about variants of the Zoutendijk [1960] method of feasible directions as well as the Rosen [1960] gradient projection method for linear constraints.

Before presenting the convergence results we need a lemma which allows us to conclude that limit points of the sequence (2.6.1) are indeed conditional critical points of f on C.

Lemma 2.6.1 Let $f:D \subset R^n \to R^1$ and $g_j:D \subset R^n \to R^1$, $j \in J_0$, be continuously differentiable on the open set D and assume that the constraint set C is regular. Let $\{x^k\} \subset C_0$ be any convergent sequence with $\lim_{k\to\infty} x^k = x^*$ in a compact subset $C_0 \subset C$. Furthermore, let $\{\varepsilon_k\}$ be a given sequence of positive numbers such that $\lim_{k\to\infty} \varepsilon_k = 0$ and suppose that the sequence $\{\|P_\ell(x^k)f'(x^k)^T\|,\ \ell = \ell_2(x^k,\varepsilon_k)\}$ is positive and satisfies

$$\lim_{k\to\infty} \|P_\ell(x^k)f'(x^k)^T\| = 0.$$

Then,

$$\|P_q(x^*)f'(x^*)^T\| = 0,\ r(x^*) \leq 0,\ q = |J(x^*,0)|.$$

Proof. We first note that ℓ also depends directly on k but that we have repressed this dependence to keep the notation less cumbersome. The regularity of C ensures that there is an open neighborhood $V(x^*)$ of $x^* \in C_0$ such that

$$\text{rank } N_q(x) = q,\ q = |J(x,\varepsilon)|,\quad \forall x \in V(x^*) \cap C,\ \varepsilon \text{ in } [0,\bar{\varepsilon}]$$

where $\bar{\varepsilon}$ is the quantity of Definition 2.2.1; then $x^k \in V(x^*) \cap C$ for all sufficiently large $k \geq 0$. As in Lemma 2.4.9, $\lim_{k\to\infty} \varepsilon_k = 0$ implies that the sequence $\{J(x^k,\varepsilon_k)\}$ has a limit set $J^* \subset J(x^*,0)$ and we may choose a subsequence $\{x^{k_i}\} \subset \{x^k\}$ such that $x^{k_i} \in V(x^*) \cap C$ and $J(x^{k_i},\varepsilon_{k_i}) = J^*$ for all i. Furthermore, it follows from

(2.5.34) that at each x^{k_i} we have either $\ell = q^* = |J^*|$ or $\ell = q^*-1$. Hence at least one value ℓ^* arises infinitely often and the subsequence $\{x^{k_i}\}$ can be refined such that $\ell = \ell^*$ is fixed for all $i \geq 0$. Clearly $N_{\ell^*}(\cdot):V(x^*) \cap C \to L(R^{\ell^*},R^n)$ and the related operator $P_{\ell^*}(\cdot):V(x^*) \cap C \to L(R^n,R^n)$ are well-defined and continuous. Therefore, we have

$$(2.6.2) \qquad 0 = \lim_{i\to\infty} \|P_{\ell^*}(x^{k_i})f'(x^{k_i})^T\| = \|P_{\ell^*}(x^*)f'(x^*)^T\|.$$

Now $J^* \subset J(x^*,0)$ implies that $\ell^* \leq q^* \leq q = |J(x^*,0)|$ and hence, by (2.5.10), that

$$(2.6.3) \qquad 0 \leq \|P_q(x^*)f'(x^*)^T\| \leq \|P_{q^*}(x^*)f'(x^*)^T\| \leq \|P_{\ell^*}(x^*)f'(x^*)^T\| = 0.$$

If $\ell^* = q^*-1 \leq q-1$, then evidently also $\|P_{q-1}(x^*)f'(x^*)^T\| = 0$ and, using (2.5.5) and (2.5.22) with $x = x^*$ we see that

$$0 = n_q(x^*)^T P_{q-1}(x^*)f'(x^*)^T = r_q(x^*)\|P_{q-1}(x^*)\,\eta_q(x^*)\|^2.$$

Consequently, from (2.5.12) and (2.5.18) it follows indeed that

$$(2.6.4) \qquad 0 = r_q(x^*) \geq r_i(x^*),\ i = 1,\ldots,q-1,\ q = |J(x^*,0)|.$$

On the other hand, if $\ell^* = q^*$ then by (2.5.34) and (2.5.33) with $h = P_{q^*}(x^{k_i})f'(x^{k_i})^T$ we obtain

$$(2.6.5) \qquad r_{q^*}(x^{k_i})\|P_{q^*-1}(x^{k_i})\eta_{q^*}(x^{k_i})\| \leq 2(\eta_2^2/\eta_1^2)\|P_{q^*}(x^{k_i})f'(x^{k_i})^T\|$$

But $r_{q^*}(\cdot)$ is continuous on $V(x^*) \cap C_0$ by the same reasoning as used for $P_{q*}(\cdot)$, and hence (2.5.12) together with (2.6.2) implies that

$$0 \geq r_{q*}(x^*) \geq r_i(x^*),\ i = 1,\ldots,q^*-1.$$

Now, it follows from (2.6.3), (2.5.7) with $y = f'(x^*)^T$ as well as (2.5.13) that

$$f'(x^*)^T = \sum_{i=1}^{q} r_i(x^*)n_i(x^*) = \sum_{i=1}^{q^*} r_i(x^*)n_i(x^*)$$

and hence from the linear independence of $n_i(x^*)$, $i = 1,\ldots,q$ that

$$r_i(x^*) = 0,\ i = q^*+1,\ldots,q,\ q^* < q.$$

Consequently, (2.6.4) holds and the proof is complete.

We note that if $\ell = \ell_1(\cdot,\cdot)$, then (2.6.5) is satisfied for $\eta_1 = \eta_2 = 1$ and hence Lemma 2.6.1 is also valid for ℓ_1 in place of ℓ_2. However, in the case of $\ell = \ell_3(\cdot,\cdot)$ we can only show that (2.6.3) holds, since the proof of $r(x^*) \leq 0$ fails.

As a direct consequence of Lemma 2.6.1 and Theorem 2.5.1 we obtain a result which is equivalent to Lemma 2.2.6 and which we state as a separate corollary.

Corollary 2.6.2 Let $f: D \subset R^n \to R^1$ and $g_j: D \subset R^n \to R^1$, $j \in J_0$ be continuously differentiable on a compact subset C_0 of the regular constraint set C and suppose that $\{x^k\} \subset C_0$ is any sequence which satisfies

(2.6.6) $$\lim_{k\to\infty} \|P_{q_k}(x^k)f'(x^k)^T\| = 0$$

where $q_k = |J^k|$, $k \geq 0$ for some specific sequence of index sets $\{J^k\} \subset J_0$. Then the set

$$\Omega' = \{x \in C_0 \,\big|\, \|P_q(x)f'(x)^T\| = 0,\ q = |J(x,0)|\}$$

of conditional critical points of f on C is not empty and

$$\lim_{k\to\infty} [\inf_{x\in\Omega'} \|x^k - x\|] = 0.$$

In particular, if Ω' consists of a single point x^*, then

(2.6.7) $$\lim_{k\to\infty} x^k = x^*, \|P_q(x^*)f'(x^*)^T\| = 0,\ q = |J(x^*,0)|.$$

Note that if Ω' is finite and $\lim_{k\to\infty} \|x^k - x^{k+1}\| = 0$, then (2.6.7) holds.

We begin our discussion of convergence results under the following assumptions which will remain the same without further mention for the remainder of the section.

The functionals $f: D \subset R^n \to R^1$ and $g_j: D \subset R^n \to R^1$, $j \in J_0$, are continuously differentiable on the open convex set D and the constraint set C is regular and admissible. Moreover, x^0 is any point in C such that $L^0 \cap C$ is compact and that the set

$$\Omega = \{x \in L^0 \cap C \,\big|\, \|P_q(x)f'(x)^T\| = 0,\ q = |J(x,0)|\}$$

is finite.

We consider the following general class of directions $\{s^k\}$ for which

(2.6.8) $$\lim_{k\to\infty} f'(x^k)s^k = 0$$

implies (2.6.6).

Definition 2.6.3 Let $\{x^k\} \subset C$ be a given sequence and $\{\varepsilon_k\}$ the associated sequence determined by the Algorithm for ε (Section 2.5). Then a sequence $\{s^k\} \subset R^n$, with $s^k \in K(x^k, J(x^k, \varepsilon_k), \|P_{q_k}(x^k)f'(x^k)^T\|)$, $q_k = |J(x^k, \varepsilon_k)|$ for each $k \geq 0$ is projected-gradient-related to $\{x^k\}$ if there exists an F-function μ_0 such that

(2.6.9) $$|f'(x^k)s^k| \geq \mu_0(\|P_{q_k}(x^k)f'(x^k)^T\|),\ q_k = |J(x^k, \varepsilon_k)|,\ k = 0,1,\ldots .$$

We note that if no constraints are present then Definition 2.6.3 reduces to condition (2.1.14) which is precisely the Elkin [1968] definition of gradient-related sequences.

The results of Section 2.5 provide us with several examples of projected-gradient-related directions. Let $\{x^k\} \subset C_0$ be any sequence in a compact subset C_0 of the regular constraint set C and suppose that $\|P_{q_k}(x^k)f'(x^k)^T\| > 0$, $q_k = |J(x^k, \varepsilon_k)|$ for all $k \geq 0$ where ε_k is determined by the Algorithm for ε. Then the sequence $\{s^k\}$ generated by the Algorithm for s is projected-gradient-related to $\{x^k\}$ if

(i) $p^k = f'(x^k)^T$, $\ell = \ell_1(x^k, \varepsilon_k)$, $k = 0,1,\ldots$

(ii) $p^k = A(x^k)^{-1}P_\ell(x^k)f'(x^k)^T$, $\ell = \ell_2(x^k, \varepsilon_k)$, $k = 0,1,\ldots$

(iii) p^k is chosen by (2.5.37), $\ell = \ell_3(x^k, \varepsilon^k)$, $k = 0,1,\ldots$.

In (ii) A is assumed to satisfy the conditions of step I_2 of the Algorithm for s. In particular, if f is uniformly convex and twice continuously differentiable, then we may take $A = f''$. The validity of these examples follows directly from (2.5.31), (2.5.35), and (2.5.38), respectively, by noting that, because of (2.5.10) and the isotonicity of μ_0, μ_j, $j \in J_0^N$, these relations hold with q replacing ℓ.

We now turn to some complete convergence theorems. The first result provides for the convergence of projected-gradient-related methods in general.

Theorem 2.6.4 Consider the iteration (2.6.1) where at x^k, ε_k is chosen by the Algorithm for ε, $s^k \in R^n$ is any direction vector such that, with ω_k and τ_k obtained from any of the steplength algorithms described by Theorems 2.3.4 through 2.3.7, we have that $x^{k+1} \in L^0 \cap C$ and (2.3.1) holds with $\varepsilon = \varepsilon_k$ and $\sigma = \|P_{q_k}(x^k)f'(x^k)^T\|$, $q_k = |J(x^k, \varepsilon_k)|$. Then, if the sequence $\{s^k\}$ is projected-gradient-related to $\{x^k\}$ and Ω consists of only one point x^*, it follows that $\lim_{k\to\infty} x^k = x^*$.

Proof. Since $\{s^k\}$ is projected-gradient-related to $\{x^k\}$, we see that $s^k \in K(x^k, J(x^k, \varepsilon_k), \|P_{q_k}(x^k)f'(x^k)^T\|)$, $q_k = |J(x^k, \varepsilon_k)|$ and that (2.6.9) holds, and hence, by (2.3.1) that

$$(2.6.10)\quad f(x^k) - f(x^{k+1}) \geq \Phi(f'(x^k)s^k, \|P_{q_k}(x^k)f'(x^k)^T\|, \varepsilon_k),\ q_k = |J(x^k, \varepsilon_k)|$$

with some given F-function Φ of three variables dependent only on C and the steplength algorithm. Therefore, by Lemma 2.1.2, either

(2.6.11) $\lim_{k\to\infty} f'(x^k)s^k = 0$ or $\lim_{k\to\infty} \|P_{q_k}(x^k)f'(x^k)^T\| = 0$ or $\lim_{k\to\infty} \varepsilon_k = 0$.

By construction we have $\lim_{k\to\infty} \varepsilon_k = 0$ if and only if $\lim_{k\to\infty} \|P_{q_k}(x^k)f'(x^k)^T\| = 0$ while (2.6.9) shows that $\lim_{k\to\infty} f'(x^k)s^k = 0$ implies that $\lim_{k\to\infty} \|P_{q_k}(x^k)f'(x^k)^T\| = 0$. Hence in any case (2.6.6) holds and, by Corollary 2.6.2, $\lim_{k\to\infty} x^k = x^*$.

Note that if L^0 is replaced by $L(f(x^0))$ and if $L(f(x^0)) \cap C$ is compact, and $\{x^k\} \subset L(f(x^0)) \cap C$, then the Goldstein and Goldstein-Armijo steplength algorithms (Theorems 2.3.9 and 2.3.10) also apply. On the basis of Theorem 2.6.4 we now give more concrete examples.

Theorem 2.6.5 Consider the iteration (2.6.1) where at x^k, ε_k is chosen by the Algorithm for ε and $s^k \in R^n$ is obtained by the Algorithm for s consisting of the steps I_1, II and III ($p^k = f'(x^k)^T$, $\ell = \ell_1(x^k,\varepsilon_k)$). Moreover, let $\omega_k \equiv 1$ and τ_k be chosen by (2.3.7) (minimization on $L^0(f(x^k)) \cap C$). Then the iterates (2.6.1) are well-defined, except at conditional critical points of f on C, remain in $L^0 \cap C$ and are strongly downward; and, if f is also hemivariate on $L^0 \cap C$, then $\lim_{k\to\infty} x^k = x^*$ with $x^* \in \Omega$.

Proof. Suppose that $x^0,\ldots,x^k$, $k \geq 0$ are already well-defined, lie in $L^0 \cap C$, and satisfy (2.3.11) with $\omega = 1$, as well as (2.3.1) with $\varepsilon = \varepsilon_{k-1}$ and $\sigma = \|P_{q_{k-1}}(x^{k-1})f'(x^{k-1})^T\|$, $q_{k-1} = |J(x^{k-1},\varepsilon_{k-1})|$. If $\|P_{q_k}(x^k)f'(x^k)^T\| = 0$, $q_k = |J(x^k,\varepsilon_k)|$, then the Algorithm for ε shows that $\varepsilon_k = 0$ and hence that $x^k \in \Omega$. Otherwise, by Theorem 2.5.2, s^k is well-defined at x^k, $\|P_\ell(x^k)f'(x^k)^T\| > 0$, $\ell = \ell_1(x^k,\varepsilon_k)$ and (2.5.31) holds. In particular, (2.5.31) implies that $f'(x^k)s^k > 0$ and, together with Theorem 2.3.5, that $x^{k+1} \in L^0 \cap C$, and satisfies (2.3.11) with

$\omega = 1$, as well as (2.3.1) with $\varepsilon = \varepsilon_k$ and $\sigma = \|P_{q_k}(x^k)f'(x^k)^T\|, q_k = |J(x^k,\varepsilon_k)|$. Therefore, by induction, either the entire sequence $\{x^k\}$ has these properties and, by (2.3.11), is strongly downward, or, for some $i \geq 0$, $\|P_{q_i}(x^i)f'(x^i)^T\| = 0$, $q_i = |J(x^i,0)|$ and $x^i \in \Omega$. In the latter case we are done; otherwise, (2.5.31) implies that the sequence $\{s^k\}$ is projected-gradient-related to $\{x^k\}$ and, by (2.3.1) that (2.6.10) holds. Therefore, as in Theorem 2.6.4 we see that (2.6.6) holds and, since f is hemivariate on $L^0 \cap C$, Lemma 2.1.5 shows that $\lim_{k\to\infty} \|x^k - x^{k+1}\| = 0$. Consequently, since the conditional critical points of f on C in $L^0 \cap C$ are isolated, Corollary 2.6.2 ensures that $\lim_{k\to\infty} x^k = x^*$ with $x^* \in \Omega$.

Note that with the help of the procedure described at the end of Section 2.5 we can always move away from conditional critical points of f on C in $L^0 \cap C$ which have at least one positive multiplier. Hence, since we used $\ell = \ell_1(x^k,\varepsilon_k)$ in Theorem 2.6.5 and there are only finitely many conditional critical points of f on C in $L^0 \cap C$, it follows from Lemma 2.6.1 that we can always assure the convergence of $\{x^k\}$ to a conditional critical point of f on C with nonpositive multipliers, that is, to a point which satisfies the necessary condition for a minimum of f on C. Again, any of the steplength algorithms of Section 2.3 can be used in conjunction with Theorem 2.6.5 provided the appropriate assumptions are made about f.

The proof of Rosen's [1960] gradient projection method for linear constraints is now a direct consequence of Theorem 2.6.5. Indeed, if $g_j : D \subset R^n \to R^1$, $j \in J_0$ are linear functionals, then the vector $z(x)$ of (2.5.23) is identically zero on C. Hence, it follows easily that for any x^k, which is not a conditional critical point of f on C, we have

$$s^k = P_\ell(x^k)f'(x^k)^T / \|P_\ell(x^k)f'(x^k)^T\|, \; \ell = \ell_1(x^k,\varepsilon_k),$$

and this is precisely the direction choice of Rosen [1960].

The next theorem illustrates some other results of Sections 2.3 and 2.5, and, more specifically, it combines the Gauss-Southwell directions with the Curry-Altman steplength algorithm.

Theorem 2.6.6 Consider the iteration (2.6.1) where at x^k, ε_k is chosen by the Algorithm for ε and $s^k \in R^n$ is obtained by the Algorithm for s consisting of the steps I_3, II, and III (p^k is chosen by (2.5.37) with $\ell = \ell_3(x^k,\varepsilon_k)$). Moreover, let ω_k and τ_k be given by the Curry-Altman steplength algorithm (Theorem 2.3.4). Then, the iterates (2.6.1) are well-defined except at conditional critical points of f on C, remain in $L^0 \cap C$, and are strongly downward; and, if f is also hemivariate on $L^0 \cap C$, then $\lim_{k\to\infty} x^k = x^*$ with $x^* \in \Omega$.

Proof. Suppose that $x^0,\ldots,x^k$, $k \geq 0$ are already well-defined, lie in $L^0 \cap C$, and satisfy (2.3.11) with $\omega = \omega_{k-1}$, as well as (2.3.1) with $\varepsilon = \varepsilon_{k-1}$ and $\sigma = \|P_{q_{k-1}}(x^{k-1})f'(x^{k-1})^T\|$, $q_{k-1} = |J(x^{k-1},\varepsilon_{k-1})|$. If $\|P_{q_k}(x^k)f'(x^k)^T\| = 0$, $q_k = |J(x^k,\varepsilon_k)|$, then the Algorithm for ε shows that $\varepsilon_k = 0$ and hence that $x^k \in \Omega$. Otherwise, by Theorem 2.5.5,

s^k is well-defined at x^k, $\|P_\ell(x^k)f'(x^k)^T\| > 0$, $\ell = \ell_3(x^k,\varepsilon_k)$, and (2.5.38) holds. In particular, (2.5.38) implies that $f'(x^k)s^k > 0$ and, together with the proof of Theorem 2.3.4, that $x^{k+1} \varepsilon L^0 \cap C$, and satisfies (2.3.11) with $\omega = \omega_k$, as well as (2.3.1) with $\varepsilon = \varepsilon_k$ and $\sigma = \|P_{q_k}(x^k)f'(x^k)^T\|$, $q_k = |J(x^k,\varepsilon_k)|$. Therefore, by induction, either the entire sequence $\{x^k\}$ has these properties and, by (2.3.11) is strongly downward, or for some $i \geq 0$, $\|P_{q_i}(x^i)f'(x^i)^T\| = 0$, $q_i = |J(x^i,0)|$ and $x^i \varepsilon \Omega$. In the latter case we are done; otherwise (2.5.38), together with $\ell_3(x^k,\varepsilon_k) = q_k = |J(x^k,\varepsilon_k)|$, ensures that $\{s^k\}$ is projected-gradient-related to $\{x^k\}$ and, by (2.3.1), that (2.6.10) holds. Therefore, the result follows as in Theorem 2.6.5.

We close this section with a result which generalizes the damped Newton-SOR theorem of Stepleman [1969] to the constrained case. Here we shall use the Curry-one-step-Newton method as our choice of steplength.

Theorem 2.6.7 Suppose that in addition to our basic assumptions, f is twice continuously differentiable and g_j, $j \varepsilon J_0$ are quasi-convex. Moreover, assume that $f''(x)$ is positive definite for all $x \varepsilon C$ and that Ω contains only one point x^*. Consider the iteration (2.6.1) where at x^k, ε_k is chosen by the Algorithm for ε, and, for given $\omega \varepsilon (0,2)$, let $p^k = A(x^k)^{-1}P_\ell(x^k)f'(x^k)^T$, $\ell = \ell_2(x^k,\varepsilon_k)$ with $A(x^k) = (\frac{1}{\omega})(D(x^k)-\omega L(x^k))$. Here $D(x^k)$ and

$-L(x^k)$ denote the diagonal and strictly lower triangular parts of $f''(x)$, respectively. Suppose that $s^k \in R^n$ is chosen by the Algorithm for s consisting of the steps I_2, II, and III and that τ_k and ω_k are given by (2.3.22) and (2.3.23), respectively. Then, either the iterates (2.6.1) are well-defined, remain in $L^0 \cap C$, and $\lim_{k\to\infty} x^k = x^*$ where x^* is the unique minimizer of f on C, or the iteration stops after a finite number of steps at x^*.

Proof. First observe that by the positive definiteness of $f''(x)$ for all $x \in C$ it follows from the convexity of C that the only conditional critical point x^* of f on C in $L^0 \cap C$ must be the unique minimizer of f on C. Now set $A(x) = (\frac{1}{\omega})[D(x)-\omega L(x)]$, $x \in L^0 \cap C$; then $A(x) + A(x)^T = (\frac{2}{\omega})D(x)-L(x)-L(x)^T = [(\frac{2}{\omega})-1]D(x) + f''(x)$. Therefore, since $\omega \in (0,2)$ and $h^T A(x)h = \frac{1}{2} h[A(x)+A(x)^T]h$, it follows that $A(x)$ is positive definite for all x in the compact set $L^0 \cap C$.

Suppose that $x^0,\ldots,x^k$, $k \geq 0$, are well-defined, satisfy (2.3.1) with $\varepsilon = \varepsilon_{k-1}$ and $\sigma = \|P_{q_{k-1}}(x^{k-1})f'(x^{k-1})^T\|$, $q_{k-1} = |J(x^{k-1},\varepsilon_{k-1})|$, and lie in $L^0 \cap C$. If $\|P_{q_k}(x^k)f'(x^k)^T\| = 0$, $q_k = |J(x^k,\varepsilon_k)|$, then, by the Algorithm for ε, $\varepsilon_k = 0$, and by our initial remark $x^k = x^*$. Otherwise, by Theorem 2.5.4, s^k is well-defined at x^k and (2.5.35) holds. In particular, (2.5.35) implies that $f'(x^k)s^k > 0$ and, together with the proof of Theorem 2.3.7, that $x^{k+1} \in L^0 \cap C$, and satisfies (2.3.1) with $\varepsilon = \varepsilon_k$ and $\sigma = \|P_{q_k}(x^k)f'(x^k)^T\|$, $q_k = |J(x^k,\varepsilon_k)|$. Therefore, by induction, either the entire sequence $\{x^k\}$ has these properties or, for some $i \geq 0$, $\|P_{q_i}(x^i)f'(x^i)^T\| = 0$, $q_i = |J(x^i,0)|$, $x^i \in L^0 \cap C$ which means that $x^i = x^*$. Otherwise, it follows from (2.5.35) that $\{s^k\}$ is projected-gradient-related to $\{x^k\}$ and, by (2.3.1), that (2.6.10) holds. Hence, by Theorem 2.6.4, $\lim_{k\to\infty} x^k = x^*$.

We conclude this chapter with the following observation. We have assumed throughout this chapter that C was regular. This guaranteed in essence that no degeneracies can occur, that is, that there are no points in C at which the normals to several constraining surfaces are linearly dependent. We only mention that a method of handling such degeneracies has been discussed in the literature (see, for example, Rosen [1960]).

CHAPTER III

Numerical Methods for a Class of Nonlinear Eigenvalue Problems

3.1 Constraint Set Construction

In this chapter we apply several nonlinear programming methods to solve an eigenvalue problem of the form

$$f'(x)^T = \lambda g'(x)^T, \ x \in D, \ \lambda \in R^1, \tag{3.1.1}$$

or its special case

$$f'(x)^T = \lambda Ax, \ x \in D, \ \lambda \in R^1, \tag{3.1.2}$$

with symmetric positive definite $A \in L(R^n)$.

As indicated in Section 2.1, both these problems correspond to a nonlinear programming problem of the form

$$\min \{f(x) | x \in C\} \tag{3.1.3}$$

with

$$C = \{x \in D | g(x) = 0\} = \{x \in D | g(x) \leqslant 0, -g(x) \leqslant 0\}. \tag{3.1.4}$$

Unfortunately this set has an empty interior and is not regular. Hence, before we can utilize the theory of Chapter II, it will be necessary to show that under suitable assumptions on g, the constraint set can be modified in such a way that it becomes both regular and admissible, and, moreover, such that the solution of

(3.1.3) with the modified set C still solves (3.1.1) or (3.1.2). Two such choices of the constraint set will be discussed in this section.

In the next section we analyze a minimization process due to Goldstein [1967] which differs from the methods discussed in Chapter II, but for which the constraint set is always a sphere of fixed radius under some elliptic norm. For (3.1.2) we have already seen that the constraint set (3.1.4) is exactly of this form. The remaining three sections of the chapter present a complete algorithm for the numerical computation of branches of solutions of (3.1.1) or (3.1.2), followed by a discussion of our numerical results for two specific applications. The first example concerns a problem of nonlinear heat generation in conducting solids, as analyzed by Joseph [1965], and the second one a problem of a heavy rotating string as described by Kolodner [1955].

We now turn to two modifications of the constraint set (3.1.4) which allow the application of the techniques of Chapter II to the problem (3.1.3). In the first approach the constraint set is changed to a set resembling an annulus, while in the second method a penalty function is added to the functional f so that solutions of (3.1.1) and (3.1.2) can be obtained from (3.1.3) using sets C of the form

$$C = \{x \in D \mid g(x) \leqslant 0\}. \tag{3.1.5}$$

To consider the first method, let $f, g: D \subset R^n \to R^1$ be continuously differentiable on the open convex set D and choose a number $\rho < 0$ in the range of g. The new constraint set C is then specified as

$$(3.1.6) \quad C = \{x \in D \mid g_1(x) \equiv g(x) \leq 0,\ g_2(x) \equiv \rho - g(x) \leq 0\}$$

where we assume that C satisfies (2.2.3). Then C is regular if, for example, $g'(x)^T \neq 0$ for all $x \in C$, since we may take any number in $(0, -\frac{\rho}{2})$ as the $\bar{\varepsilon}$ of Definition 2.2.1. Furthermore, C is admissible if it is compact or if (2.2.11) holds. Under these conditions, together with the relevant assumptions on f and on the initial point x^0 in C, all of the methods in Sections 2.4/2.6 and their variations apply to (3.1.3) with C defined by (3.1.6). Since the application of any of these methods produces a conditional critical point x^* of f on C, it follows that x^* solves (3.1.1) for some $\lambda \in R^1$.

In the special case (3.1.2) we have the constraint functional

$$(3.1.7) \qquad g: D \subset R^n \to R^1,\ g(x) = \frac{1}{2} x^T A x - \zeta,\ \zeta > 0$$

with symmetric, positive definite $A \in L(R^n)$. The next result concerns the constraint set C for this case.

<u>Theorem 3.1.1</u> Let $D \subset R^n$ be an open convex set and $g: D \subset R^n \to R^1$ the functional of (3.1.7) with a symmetric, positive definite matrix

$A \varepsilon L(R^n)$ and some $\zeta > 0$. If $\rho \varepsilon [-\frac{\zeta}{2},0)$, then the constraint set C of (3.1.6) is regular and admissible.

Proof. It follows from the positive definiteness of A that $0 \notin C$ and hence from the nonsingularity of A that $g'(x)^T = Ax \neq 0$ for all $x \varepsilon C$. Therefore, with $\bar{\varepsilon}$ in $(0,-\frac{1}{2}\rho)$ we see that C is regular. For the admissibility of C observe that because of the positive definiteness of A the set C is bounded. Hence, for any $x,y \varepsilon C$, we have

$$(3.1.8) \quad |g_i(x)-g_i(y)| \leqslant \sup_{x\varepsilon C} \|(Ax)^T\| \|x-y\| \leqslant \beta\|x-y\|, \; i = 1,2;$$

that is, (2.2.11) holds and Lemma 2.2.7 gives the desired result.

Note that it suffices to assume only the nonsingularity of $A \varepsilon L(R^n)$, provided that C is bounded, since the regularity of C is then trivial and the admissibility follows from (3.1.8).

The use of the constraint set (3.1.6) requires that the iterates generated by the minimization process remain in an annulus-shaped domain. This has the disadvantage that if a solution is on the "opposite side" from the initial point then the iterates cannot cross over directly but have to find their way to the other side inside the annulus-shaped set. A different approach to the solution of (3.1.1) or (3.1.2) which eliminates this disadvantage, but which may increase the occurrence of round-off errors, is to eliminate one of the constraints of (3.1.6) by introducing a penalty function.

Zangwill [1967] showed that such penalty functions allow the transformation of a convex programming problem into a single unconstrained minimization problem. Under different assumptions on the functionals involved we shall use a similar approach for solving (3.1.1) and (3.1.2). Let $f,g:D \subset R^n \to R^1$ be continuous on the open convex set D, and suppose that C is given by (3.1.5), and that for some $\alpha \in (0,\infty)$

(3.1.9) $$|f(x)| \leq \alpha, \quad \forall x \in C.$$

Now choose a number $\rho < 0$ and a point $x^0 \in C$ such that

(3.1.10) $$g(x^0) > \frac{\rho}{2}$$

and set

(3.1.11) $$\bar{\eta} = \left(\frac{4}{\rho^2}\right)(\alpha + f(x^0)).$$

Finally, define the functional

(3.1.12) $$h:D \subset R^n \to R^1, \; h(x) = f(x) + \bar{\eta}(\min\{g(x) - \frac{\rho}{2}, 0\})^2$$

and consider the new minimization problem

(3.1.13) $$\min\{h(x) | x \in C\}.$$

The next result relates solutions of (3.1.13) to those of (3.1.1) and (3.1.2).

Theorem 3.1.2 Let $f,g:D \subset R^n \to R^1$ be continuously differentiable on the open convex set D and, with C defined by (3.1.5), suppose that f satisfies (3.1.9). For any $x^0 \in C$, such that (3.1.10) holds, let $\bar{\eta}$ be given by (3.1.11). Then the functional h of (3.1.12) is continuously differentiable on D and $L^0(h(x^0)) \subset L^0(f(x^0))$. Moreover, any solution x^* of (3.1.13) lies in the set $\{x \in D | \rho \leqslant g(x) \leqslant 0\}$ and solves (3.1.1) for some real number $\lambda \in R^1$.

Proof. By construction we have for any $x \in D$

$$(3.1.14) \quad h'(x)^T = \begin{cases} f'(x)^T, \text{ if } g(x) - \frac{\rho}{2} > 0 \\ f'(x)^T + 2\bar{\eta}(g(x) - \frac{\rho}{2})g'(x)^T, \text{ otherwise;} \end{cases}$$

and, in particular, h is continuously differentiable on D. Clearly (3.1.10) and (3.1.12) imply that $h(x^0) = f(x^0)$ as well as $f(x) \leqslant h(x)$ for all $x \in D$, and hence that

$$f(x) \leqslant h(x) \leqslant h(x^0) = f(x^0), \; x \in L^0(h(x^0))$$

or $x \in L(f(x^0))$. By definition $L^0(h(x^0))$ is the connected component of $L(h(x^0))$ containing x^0 and the curve in $L^0(h(x^0))$ connecting x^0 and x must also lie in $L(f(x^0))$. Therefore, we have shown that in fact $x \in L^0(f(x^0))$ and hence, that $L^0(h(x^0)) \subset L^0(f(x^0))$.

Suppose that x^* solves (3.1.13) and that w is any other point of C such that $g(w) < \rho$. Then it follows that

$$g(w) - \frac{\rho}{2} < \frac{\rho}{2} < 0,$$

and hence by (3.1.10) and (3.1.12) that

$$\begin{aligned} h(w) &= f(w) + \bar{\eta}(g(w) - \tfrac{\rho}{2})^2 > f(w) + \bar{\eta}\,\frac{\rho^2}{4} \\ &= f(w) - f(x^0) + \bar{\eta}\,\frac{\rho^2}{4} + f(x^0) \\ &\geq -\alpha - f(x^0) + \bar{\eta}\,\frac{\rho^2}{4} + f(x^0) \\ &= f(x^0) = h(x^0) \geq h(x^*), \end{aligned}$$

for any w for which $g(w) < \rho$. Therefore, if $g(x^*) < \rho$, then $h(x^*) > h(x^0) \geq h(x^*)$ which is a contradiction. Hence, since $x^* \varepsilon C$, we have $\rho \leq g(x^*) \leq 0$.

Finally, since x^* solves (3.1.13), it follows that x^* is a conditional critical point of h on C and hence that $h'(x^*)^T = \bar{\lambda} g'(x^*)^T$ for some $\bar{\lambda} \varepsilon R^1$. Thus, by (3.1.14) we see that $f'(x^*)^T = \lambda g'(x^*)^T$ where

$$\lambda = \begin{cases} \bar{\lambda}, \text{ if } g(x^*) - \frac{\rho}{2} > 0 \\ \\ \bar{\lambda} - 2\bar{\eta}(g(x^*) - \frac{\rho}{2}), \text{ otherwise.} \end{cases}$$

If C is both regular and admissible, we can now apply the results of Chapter II in order to solve (3.1.13) and thereby obtain solutions of (3.1.1) and (3.1.2). For the special case (3.1.2), Theorem 3.1.2 immediately applies with g defined by (3.1.7) and, clearly, the constraint set C of (3.1.5) is regular and admissible. In the general case, note that if $L^0(f(x^0)) \cap C$

is compact then so is $L^0(h(x^0)) \cap C$ and, since the conditional critical points of h on C are also conditional critical points of f on C, it follows that if those belonging to f are isolated, then so are those belonging to h. However, before applying any particular method of Chapter II to (3.1.13), we must still ascertain whether certain properties previously assumed only for f may now also be required of g due to its presence in (3.1.12).

3.2 The Goldstein Algorithm

In the previous section we considered two ways of choosing the constraint set so that the results of Chapter II are applicable. We now discuss a method of Goldstein [1967] which works directly with spheres as constraint sets and which overcomes some of the difficulties inherent in the techniques of Section 3.1. At the same time this method appears to have a considerably smaller range of applicability.

Let $A \in L(R^n)$ be symmetric and positive definite, and denote by $\|x\|_A = (x^T Ax)^{1/2}$ the corresponding elliptic norm on R^n. For ease of notation we will present the results of this section in terms of the unit ball $\bar{S}_A(0,1) = \{y \in R^n \mid \|y\|_A \leq 1\}$; the generalization to a ball $\bar{S}_A(0,\rho)$ with any radius $\rho > 0$ is evident.

Now consider the eigenvalue problem

$$(3.2.1) \qquad f'(x)^T = \lambda Ax, \; x \in \dot{S}_A(0,1), \; \lambda \in R^1$$

and the associated minimization problem

$$(3.2.2) \qquad \min \{f(x) \mid x \in \dot{S}_A(0,1)\}.$$

Then the Goldstein algorithm has the basic form

$$x^{k+1} = \frac{x^k - \tau_k A^{-1} f'(x^k)^T}{\|x^k - \tau_k A^{-1} f'(x^k)^T\|_A}, \quad k = 0,1,\ldots$$

with some suitable choice of parameter τ_k. We introduce the mapping

$$(3.2.3) \qquad h: R^n \to R^1, \quad h(x) = \|A^{-1} f'(x)^T\|_A^2 - (f'(x)x)^2.$$

Since, evidently

$$h(x) = (A^{-1/2} f'(x)^T)^T A^{-1/2} f'(x)^T - ((A^{-1/2} f'(x)^T)^T A^{1/2} x)^2,$$

it follows from the Cauchy-Schwarz inequality that $h(x) \geq 0$ whenever $x \in \dot{S}_A(0,1)$ and that any root of the equation

$$h(x) = 0, \quad x \in \dot{S}_A(0,1)$$

solves (3.2.1) with some $\lambda \in R^1$. With

$$(3.2.4) \qquad x(t) = \frac{x - tA^{-1} f'(x)^T}{\|x - tA^{-1} f'(x)^T\|_A}, \quad x \in \dot{S}_A(0,1), \ t \geq 0$$

this leads us to consider estimates of the form

$$(3.2.5) \qquad f(x) - f(x(\tau)) \geq \mu(h(x))$$

involving some F-function $\mu: [0,\infty) \to [0,\infty)$ as well as a suitable parameter τ.

Let $x^0 \in \mathring{S}_A(0,1)$ be an initial point and specify a general step of the iterative algorithm as follows:

At the iterate $x \in L(f(x^0)) \cap \mathring{S}_A(0,1)$, set $\alpha = 0$ if $h(x) = 0$, and otherwise, let $\alpha \in [0,\infty)$ be the smallest number which satisfies

$$(3.2.6) \qquad \zeta_1 f'(x)(x-x(\alpha)) \leq f(x) - f(x(\alpha)) \leq \zeta_2 f'(x)(x-x(\alpha)),$$

with fixed numbers $0 < \zeta_1 \leq \zeta_2 < 1$. Then the next iterate $x(\tau)$ is given by (3.2.4) with

$$(3.2.7) \qquad \tau = \min \{\alpha, (3\|A^{-1}f'(x)^T\|_A)^{-1}\}.$$

For the case $A = I$, a convergence theorem for this algorithm was stated by Goldstein [1967]. Unfortunately, his proof appears to be incomplete, and we provide next a corrected convergence result. As a first step, the following theorem establishes the validity of the basic inequality (3.2.5).

Theorem 3.2.1 Assume that $f:R^n \to R^1$ is continuously differentiable and for any $x \in L(f(x^0)) \cap \mathring{S}_A(0,1)$, let τ be given by (3.2.7). Then $x(\tau)$ given by (3.2.4) satisfies $x(\tau) \in L(f(x)) \cap \mathring{S}_A(0,1)$ and moreover (3.2.5) holds with some F-function μ independent of x and τ.

Proof. If $h(x) = 0$, then $\tau = 0$ and the result is trivial. Therefore assume that $h(x) > 0$. Then, with the abbreviations $\nu = f'(x)x$ and

$$(3.2.8) \qquad \beta(\tau) = \left(1 + \frac{h(x)\tau^2}{(1-\nu\tau)^2}\right)^{1/2},$$

it follows that

$$1 - \|x-\tau A^{-1}f'(x)^T\|_A = 1 - [1-2\nu\tau+\tau^2\|A^{-1}f'(x)^T\|_A^2]^{1/2}$$

$$= 1 - [(1-\nu\tau)^2+\tau^2 h(x)]^{1/2}$$

(3.2.9) $$= 1 - (1-\nu\tau)\beta(\tau) = \tau(\nu\beta(\tau) + \frac{1-\beta(\tau)}{\tau})$$

and hence that

$$f'(x)(x-x(\tau)) = f'(x)[\tau A^{-1}f'(x)^T - x(1-\|x-\tau A^{-1}f'(x)^T\|_A)]\|x-\tau A^{-1}f'(x)^T\|_A^{-1}$$

$$= [\tau\|A^{-1}f'(x)^T\|_A^2 - \nu\tau(\nu\beta(\tau) + \frac{1-\beta(\tau)}{\tau})]\|x-\tau A^{-1}f'(x)^T\|_A^{-1}$$

(3.2.10) $$= \tau[h(x)+\nu^2(1-\beta(\tau))-\nu(\frac{1-\beta(\tau)}{\tau})]\|x-\tau A^{-1}f'(x)^T\|_A^{-1}.$$

Now, by (3.2.7), we obtain that

$$\frac{2}{3} \leq 1 - \tau\|A^{-1}f'(x)^T\|_A \leq \|x-\tau A^{-1}f'(x)^T\|_A \leq 1 + \tau\|A^{-1}f'(x)^T\|_A \leq \frac{4}{3}$$

and, since $\nu\tau \leq \frac{1}{3}$, that

$$\frac{h(x)\tau^2}{(1-\nu\tau)^2} \leq \frac{\|A^{-1}f'(x)^T\|_A^2\tau^2}{(1-\nu\tau)^2} \leq \frac{1}{4} .$$

Therefore, expanding $\beta(\tau)$ as an alternating series we see that

(3.2.11) $$1 \leq \beta(\tau) \leq 1 + \frac{1}{2}\frac{h(x)\tau^2}{(1-\nu\tau)^2} \leq \frac{9}{8}$$

and, using (3.2.10), that

$$f'(x)(x-x(\tau)) \geq \tau[h(x)+\nu^2(1-\beta(\tau))+|\nu|\frac{(1-\beta(\tau))}{\tau}](\frac{3}{4})$$

$$\geq \frac{3}{4}\tau h(x)[1-\frac{\nu^2\tau^2}{2(1-\nu\tau)^2} - \frac{|\nu|\tau}{2(1-\nu\tau)^2}]$$

$$\geq \frac{3}{8}\tau h(x). \tag{3.2.12}$$

Consider the functional

$$\psi(\xi) = \begin{cases} 1 & \xi = 0 \\ (f(x)-f(x(\xi))/(f'(x)(x-x(\xi))), & \xi \in (0,(3\|A^{-1}f'(x)^T\|_A)^{-1}]. \end{cases} \tag{3.2.13}$$

Since $h(x) > 0$, it follows from the continuity of f' together with (3.2.12) that ψ is continuous on $(0,(3\|A^{-1}f'(x)^T\|_A)^{-1}]$. Thus, an application of L'Hospital's rule shows that $\lim_{\xi\to 0}\psi(\xi) = 1$ and hence that ψ is continuous on $[0,(3\|A^{-1}f'(x)^T\|_A)^{-1}]$. If $\tau = \alpha \leq (3\|A^{-1}f'(x)^T\|_A)^{-1} < \infty$, then it follows directly from (3.2.6) that $\psi(\tau) \geq \zeta_1$. Hence assume that $\psi(\tau) < \zeta_1$ and therefore necessarily that $\tau = (3\|A^{-1}f'(x)^T\|_A)^{-1} < \alpha$. By continuity ψ takes on all values between ζ_1 and 1 on $[0,\tau)$ and hence there exists an $\bar{\alpha} \in (0,\tau)$ such that $\zeta_1 \leq \psi(\bar{\alpha}) \leq \zeta_2$. Thus, $\bar{\alpha}$ satisfies (3.2.6) and $\bar{\alpha} < \tau < \alpha$ which contradicts the assumption that α is the smallest number in $[0,\infty)$ for which (3.2.6) holds. Therefore, in all cases $\psi(\tau) \geq \zeta_1$, and consequently, it follows from (3.2.12) that

$$f(x) - f(x(\tau)) = \psi(\tau)f'(x)(x-x(\tau))$$

$$\geq \frac{3}{8}\zeta_1 h(x)\tau. \tag{3.2.14}$$

To obtain the final estimate (3.2.5) we still require an estimate on $\|x-x(\tau)\|_A$. By (3.2.9) and (3.2.12) it is easily seen that

$$\|x-x(\tau)\|_A^2 = 2(1-x^TAx(\tau)) = 2(1 - \frac{1-\nu\tau}{\|x-\tau A^{-1}f'(x)^T\|_A})$$

$$= 2(1 - \frac{1}{\beta(\tau)}) = 2(\frac{\beta(\tau)-1}{\beta(\tau)})$$

$$\leq 2(\beta(\tau)-1) \leq \frac{h(x)\tau^2}{(1-\nu\tau)^2}$$

$$\leq \frac{9}{4} h(x)\tau^2$$

and hence that

(3.2.15) $$\|x-x(\tau)\|_A \leq \frac{3}{2} h(x)^{1/2}\tau.$$

Suppose that $\tau = \alpha$. If ω denotes the modulus of continuity of $A^{-1}f'^T$ on $\bar{S}_A(0,1)$ and η the corresponding isotone function of (2.1.15), then we see that

$$f(x) - f(x(\tau)) \geq f'(x)(x-x(\tau)) - \|x-x(\tau)\|_A\eta(\|x-x(\tau)\|_A).$$

Thus, by the isotonicity of η, (3.2.6), and (3.2.15), it follows that

$$\zeta_2 f'(x)(x-x(\tau)) \geq f'(x)(x-x(\tau)) - \frac{3}{2} h(x)^{1/2}\tau\eta(\frac{3}{2} h(x)^{1/2}\tau),$$

and, by (3.2.12), that

$$\frac{3}{2} h(x)^{1/2}\tau\hat{\eta}(\frac{3}{2} h(x)^{1/2}\tau) \geq \frac{3}{8} (1-\zeta_2)h(x)\tau.$$

Hence,

$$(3.2.16) \qquad \tau \geq \frac{2}{3} h(x)^{-1/2}\hat{\eta}^{-1}(\frac{1}{4}(1-\zeta_2)h(x)^{1/2})$$

with any strictly isotone F-function $\hat{\eta}:[0,\infty) \to [0,\infty)$ such that $\hat{\eta}(t) \geq \eta(t)$, for all $t \geq 0$, and therefore that $\hat{\eta}^{-1}$ exists and is a strictly isotone F-function.

On the other hand, if $\tau = (3\|A^{-1}f'(x)^T\|_A)^{-1} < \alpha$, then it follows from the continuity of f' and the compactness of $\dot{S}_A(0,1)$ that there is a constant $\kappa \in (0,\infty)$ such that $\tau \geq \kappa$.

Therefore, (3.2.14) and (3.2.16) together show that

$$(3.2.17) \qquad f(x) - f(x(\tau)) \geq \mu(h(x))$$

where $\mu:[0,\infty) \to [0,\infty)$, $\mu(t) = \frac{3}{8}\zeta_1 \min\{\kappa t, \frac{2}{3} t^{1/2}\hat{\eta}^{-1}(\frac{1}{4}(1-\zeta_2)t^{1/2})\}$ is evidently an F-function. Finally, (3.2.4) and (3.2.17) ensure that $x(\tau) \in L(f(x)) \cap \dot{S}_A(0,1)$.

Our algorithm so far does not specify how the parameter τ is obtained. One possibility for this is provided by the following procedure of Goldstein [1967] (see also Armijo [1966]).

Goldstein-Armijo Algorithm: Let $\bar{\mu}$ be a fixed F-function and $\zeta_1 \in (0,1)$ as well as $\gamma > 1$ given constants. Further, let $x \in L(f(x^0)) \cap \dot{S}_A(0,1)$.

I. If $h(x) = 0$, set $\tau = 0$, otherwise let $\tau' > 0$ be any real number such that

$$(3\|A^{-1}f'(x)^T\|_A)^{-1} \geq \tau' \geq \bar{\mu}(h(x)).$$

II. If

(3.2.18) $f(x) - f(x(\tau')) \geq \zeta_1 f'(x)(x-x(\tau'))$

let $\omega = 1$; otherwise determine ω as the largest number in the sequence $\{\gamma^{-j}\}_{j=1}^{\infty}$ such that

(3.2.19) $f(x) - f(x(\omega\tau')) \geq \zeta_1 f'(x)(x-x(\omega\tau'))$,

and then set $\tau = \omega\tau'$.

Theorem 3.2.2 Assume that $f:R^n \to R^1$ is continuously differentiable. For any $x \in L(f(x^0)) \cap \dot{S}_A(0,1)$ let τ be selected by the Goldstein-Armijo algorithm. Then $x(\tau) \in L(f(x)) \cap \dot{S}_A(0,1)$ and (3.2.5) holds with some F-function μ independent of x and τ.

Proof. If $h(x) = 0$, then $\tau = 0$ and the result is again trivial; hence assume that $h(x) > 0$. If (3.2.18) holds, then by (3.2.12),

(3.2.20) $f(x) - f(x(\tau')) \geq \frac{3}{8}\zeta_1 h(x)\tau' \geq \frac{3}{8}\zeta_1 h(x)\bar{\mu}(h(x))$.

On the other hand, if (3.2.19) applies, then by the definition of ω we have $\gamma\tau \leq \tau \leq (3\|A^{-1}f'(x)^T\|_A)^{-1}$ and

(3.2.21) $f(x) - f(x(\gamma\tau)) < \zeta_1 f'(x)(x-x(\gamma\tau))$.

Hence, the functional ψ of (3.2.13) is continuous on $[0,\gamma\tau]$ and takes on all values between ζ_1 and 1 on $[0,\gamma\tau)$. Consequently, there exists an $\hat{\alpha} \in (0,\gamma\tau)$ such that $\psi(\hat{\alpha}) = \zeta_1$ and hence, by (3.2.14) and (3.2.16), that

$$\gamma\tau > \hat{\alpha} \geq \frac{2}{3}\, h(x)^{-1/2}\hat{\eta}^{-1}(\tfrac{1}{4}(1-\zeta_1)h(x)^{1/2}),$$

and that

$$f(x) - f(x(\tau)) \geq \frac{3}{8}\,\zeta_1 h(x)\tau$$

$$\geq \frac{\zeta_1}{4\gamma}\, h(x)^{1/2}\hat{\eta}^{-1}(\tfrac{1}{4}(1-\zeta_1)h(x)^{1/2}). \tag{3.2.22}$$

Thus, (3.2.20) and (3.2.22) together show that (3.2.5) holds with $\mu(t) = \zeta_1 \min\{\frac{3}{8}\, t\bar{\mu}(t), \frac{1}{4\gamma}\, t^{1/2}\hat{\eta}^{-1}(\frac{1}{4}(1-\zeta_1)t^{1/2})\}$. Finally, (3.2.4) and (3.2.5) together show that $x(\tau) \in L(f(x)) \cap \mathring{S}_A(0,1)$.

In Theorems 3.2.1 and 3.2.2 the parameter τ cannot be interpreted as a steplength in the sense of Chapter II since $\|x-x(\tau)\|_A \neq \tau$. The next result shows that under a slightly stronger assumption on f we need not compute τ at each step but can choose it instead as a fixed constant.

<u>Theorem 3.2.3</u> Assume that $f:R^n \to R^1$ is continuously differentiable and that

$$\|A^{-1}f'(x)^T - A^{-1}f'(y)^T\|_A \leq \xi\|x-y\|_A, \quad \forall\, x,y \in \mathring{S}_A(0,1). \tag{3.2.23}$$

Given the set $L(f(x^0)) \cap \mathring{S}_A(0,1)$ let τ be defined by

$$\tau = \min\{\frac{1}{3\kappa_1}, \frac{1}{4\xi}\} \tag{3.2.24}$$

where $\kappa_1 \in (0,\infty)$ is such that

$$\|A^{-1}f'(x)^T\|_A \leq \kappa_1, \quad \forall x \in \mathring{S}_A(0,1).$$

Then $x(\tau) \in L(f(x)) \cap \mathring{S}_A(0,1)$ and (3.2.5) holds for some F-function μ independent of x and τ.

Proof. If $h(x) = 0$, then the result is trivial; therefore, assume that $h(x) > 0$. Because of (3.2.23) we have

$$(3.2.25) \qquad f(x) - f(x(\tau)) \geq f'(x)(x-x(\tau)) - \frac{1}{2}\xi\|x-x(\tau)\|_A^2,$$

and, since $\tau \leq (3\kappa_1)^{-1} \leq (3\|A^{-1}f'(x)^T\|_A)^{-1}$, it follows from (3.2.12) and (3.2.15) that (3.2.25) can be continued to

$$f(x) - f(x(\tau)) \geq \frac{3}{8}\,\tau h(x) - \frac{9}{8}\,\xi\tau^2 h(x)$$

$$= \frac{3}{8}\,\tau h(x)(1-3\xi\tau).$$

Therefore, by (3.2.24) we see that (3.2.5) holds with $\mu(t) = \frac{3}{32}\min\{\frac{1}{3\kappa_1}, \frac{1}{4\xi}\}t$ and, again, (3.2.4) and (3.2.5) imply that $x(\tau) \in L(f(x)) \cap \mathring{S}_A(0,1)$.

Until now we have discussed three ways of choosing values of τ for which estimates of the form (3.2.5) are valid. As indicated before, this particular estimate is crucial in proving convergence of the iterative process. Following is a convergence theorem covering all three choices of τ.

Theorem 3.2.4 Assume that $f:R^n \to R^1$ is continuously differentiable and let $x^0 \in \mathring{S}_A(0,1)$ be any point such that the set

$$\Omega = \{y \in L(f(x^0)) \cap \mathring{S}_A(0,1) \,|\, h(y) = 0\}$$

is finite. Consider the iteration

(3.2.26) $$x^{k+1} = x^k(\tau_k),\ k = 0,1,\ldots$$

where at x^k, either

(i) τ_k is chosen by (3.2.7), or

(ii) τ_k is selected by the Goldstein-Armijo algorithm or, if f also satisfies (3.2.23),

(iii) τ_k is chosen by (3.2.24).

Then $\{x^k\} \subset L(f(x^0)) \cap \dot{S}_A(0,1)$ and $\lim_{k\to\infty} x^k = x^* \in \Omega$; that is, x^* solves (3.2.1).

Proof. Suppose that $x^0, x^1, \ldots, x^k$, $k \geq 0$ have already been obtained and lie in the set $L(f(x^0)) \cap \dot{S}_A(0,1)$. If $h(x^k) = 0$, then $x^k \in \Omega$ and the iteration stops. Otherwise, $h(x^k) > 0$ and Theorem 3.2.1, 3.2.2 or 3.2.3, respectively, shows that

(3.2.27) $$f(x^k) - f(x^{k+1}) \geq \mu(h(x^k))$$

with some F-function μ independent of x^k and τ_k. Moreover, by (3.2.15), we have

(3.2.28) $$\|x^k - x^{k+1}\|_A \leq \frac{3}{2}\, h(x^k)^{1/2}\tau_k .$$

Therefore, $x^{k+1} \in L(f(x^0)) \cap \dot{S}_A(0,1)$ and by induction either the entire sequence $\{x^k\}$ satisfies (3.2.27) and (3.2.28) and lies in $L(f(x^0)) \cap \dot{S}_A(0,1)$ or for some $i \geq 0$, $h(x^i) = 0$. In the latter case $x^i \in \Omega$ and otherwise, by Lemma 2.1.2, $\lim_{k\to\infty} h(x^k) = 0$. Since f' is continuous, the same holds for h and hence limit points of $\{x^k\}$ must be contained in Ω. Moreover, (3.2.28) shows that

$\lim_{k\to\infty} \|x^k - x^{k+1}\|_A = 0$ and, therefore, since Ω is finite and $L(f(x^0)) \cap \mathring{S}_A(0,1)$ is compact, that $\lim_{k\to\infty} x^k = x^* \in \Omega$.

3.3 A Complete Algorithm

In this section we incorporate the methods discussed so far into a complete algorithm for the numerical solution of nonlinear eigenvalue problems in R^n. This algorithm will then by applied to two particular problems considered by Joseph [1965] and Kolodner [1955], namely, an eigenvalue problem arising in the theory of nonlinear heat generation in conducting solids and another one concerning the movement of a heavy rotating string.

As before, we consider the eigenvalue problem

$$f'(x)^T = \lambda g'(x)^T, \; x \in D, \; \lambda \in R^1 \tag{3.3.1}$$

as well as its special case

$$f'(x)^T = \lambda Ax, \; x \in D, \; \lambda \in R^1 \tag{3.3.2}$$

where $f,g : D \subset R^n \to R^1$ are given functionals and $A \in L(R^n)$.

All of the methods discussed in the previous sections were guaranteed to converge to solutions of (3.3.1) or (3.3.2), respectively. However, as with most minimization processes, the rate of convergence is, in general, only linear. Thus, it is desirable to terminate the minimization algorithm once it has come sufficiently close to the solution and to use a faster, locally convergent process such as the quadratically convergent Newton method for the final phase.

In general, it is of little interest to obtain only a single solution of (3.3.1) or (3.3.2). However, once such a solution has been computed, the continuation idea of Chapter I can be used to obtain an entire branch of solutions through it. For this we solve numerically the initial value problem for the ordinary differential equation describing this branch. This can be done by means of any suitable discrete-variable method such as a Runge-Kutta method and a predictor-corrector method. In summary then our complete algorithm consists of the following major parts:

(i) Use of either a minimization process of the form described in Chapter II with a constraint set of the type given in Section 3.1 or of the Goldstein type process with fixed parameter τ of Section 3.2. This procedure is terminated if $\|f'(x)^T - \lambda g'(x)^T\|_\infty$ and the distance between successive iterates are less than some given tolerance.

(ii) Use of Newton's method to obtain a solution of (3.3.1) or (3.3.2), respectively, to the desired accuracy. The termination criteria is the same as under (i).

(iii) Use of the standard fourth-order Runge-Kutta method as a starter for a fourth-order predictor-corrector method to obtain complete branches of solutions of (3.3.1) or (3.3.2), respectively.

For the computations described in the following sections we have implemented part (i) by using a constraint set of the form (3.1.6) and a minimization algorithm of the form (2.6.1) consisting of the Algorithm for s with steps I_1,II, and III, and the Curry [1944] algorithm, as discussed in Sections 2.5 and 2.3, respectively. This particular choice was used only for ease of implementation; any other combination of methods developed in Chapter II could have been used as well. For comparison purposes we have also used a Goldstein-type procedure with fixed parameter τ.

In order to guarantee the applicability of Newton's method in part (ii) of the algorithm, we still need to discuss the local convergence of that process in the case of our particular problem. The following basic Newton Attraction Theorem can be found, for example, in Ortega and Rheinboldt [1970]; Theorem 10.2.2].

Theorem 3.3.1 Assume that $H:D \subset R^n \to R^n$ is G-differentiable on an open neighborhood $S_0 \subset D$ of a point $y^* \in D$ for which $Hy^* = 0$, and that H' is continuous at y^* and $H'(y^*)$ is nonsingular. Then there is a neighborhood $S \subset D$ such that for any $y^0 \in S$ the Newton iterates

$$y^{k+1} = y^k - H'(y^k)^{-1}Hy^k, \quad k = 0,1,\ldots$$

converges to y^*. If, in addition, there is a constant $\gamma < \infty$ such that

$$\|H'(y)-H'(y^*)\| \leq \gamma\|y-y^*\|, \quad \forall y \in S_0,$$

then the order of convergence is at least quadratic. In order to apply this theorem to the nonlinear eigenvalue problem (3.3.1), let $\xi > 0$ be a given number and define the operator

$$(3.3.3)\qquad H: D \times R^1 \subset R^{n+1} \to R^{n+1},\quad H(x,\lambda) = \begin{pmatrix} f'(x)^T - \lambda g'(x)^T \\ \|x\|_2^2 - \xi \end{pmatrix}.$$

Then we obtain the following convergence result.

Theorem 3.3.2 Let $f,g: D \subset R^n \to R^1$ be twice F-differentiable on the open neighborhood $S_0 \subset D$ of a point $x^* \in D$ for which there is a number $\lambda^* \in R^1$ such that $H(x^*,\lambda^*) = 0$ with the operator H of (3.3.3). Moreover, suppose that f'' and g'' are continuous at x^* and that $[f''(x^*)-\lambda^* g''(x^*)]^{-1}$ exists and satisfies

$$(3.3.4)\qquad \delta^* = 2x^{*T}[f''(x^*)-\lambda^* g''(x^*)]^{-1} g'(x^*)^T \neq 0.$$

Then there is a neighborhood $S \times \Lambda \subset D \times R^1$ of (x^*,λ^*) such that, for any $(x^0,\lambda^0) \in S \times \Lambda$, the Newton iterates

$$(3.3.5)\qquad (x^{k+1},\lambda^{k+1}) = (x^k,\lambda^k) - H'(x^k,\lambda^k)^{-1} H(x^k,\lambda^k),\quad k = 0,1,\ldots$$

converge to (x^*,λ^*). If, in addition, g'' is bounded on S_0 and the estimates

$$\left.\begin{aligned} \|f''(x)-f''(x^*)\|_1 &\leq \gamma_1 \|x-x^*\|_1 \\ \|g''(x)-g''(x^*)\|_1 &\leq \gamma_2 \|x-x^*\|_1, \end{aligned}\right\},\quad \forall x \in S_0$$

hold, then the order of convergence is at least quadratic.

Proof. On the basis of the assumptions about f and g it follows by direct computation that H is G-differentiable on $S_0 \times R^1$ and that

$$H'(x,\lambda) = \begin{bmatrix} f''(x) - \lambda g''(x) & -g'(x)^T \\ & \\ 2x^T & 0 \end{bmatrix}, \quad x \in S_0, \ \lambda \in R^1$$

Moreover, the F-differentiability of g' on S_0 implies that g' is continuous at x^* and hence, from the continuity of f'' and g'' at x^*, we see that H' is continuous at (x^*,λ^*).

Since $[f''(x^*)-\lambda^* g''(x^*)]^{-1}$ exists and (3.3.4) holds, it is easily verified that

$$(3.3.6) \qquad H'(x^*,\lambda^*)^{-1} = \frac{1}{\delta^*}\begin{bmatrix} B_1 & B_2 \\ & \\ B_3 & 1 \end{bmatrix}, \quad \begin{matrix} B_1 \in L(R^n) \\ B_2 \in L(R^1,R^n) \\ B_3 \in L(R^n,R^1) \end{matrix}$$

where

$$B_1 = \delta^*[f''(x^*)-\lambda^* g''(x^*)]^{-1}$$

$$- 2[f''(x^*)-\lambda^* g''(x^*)]^{-1} g'(x^*)^T x^{*T} [f''(x^*)-\lambda^* g''(x^*)]^{-1},$$

$$B_2 = [f''(x^*)-\lambda^* g''(x^*)]^{-1} g'(x^*)^T$$

and

$$B_3 = -2x^{*T}[f''(x^*)-\lambda^* g''(x^*)]^{-1}.$$

Now it follows directly from Theorem 3.3.1 that there is a neighborhood $S \times \Lambda \subset D \times R^1$ of (x^*,λ^*) such that for any $(x^0,\lambda^0) \in S \times \Lambda$, the iterates (3.3.5) converge to (x^*,λ^*).

For the last part of the theorem we see that, for any $(x,\lambda) \in S_0 \times \Lambda$

$$\|H'(x,\lambda)-H'(x^*,\lambda^*)\|_1$$

$$= \left\| \begin{matrix} f''(x)-f''(x^*)+\lambda^* g''(x^*)-\lambda g''(x) & g'(x^*)^T-g'(x)^T \\ 2(x-x^*)^T & 0 \end{matrix} \right\|_1$$

$$\leq \|f''(x)-f''(x^*)+\lambda^* g''(x^*)-\lambda g''(x)\|_1$$

$$+ 2\|x-x^*\|_1 + \|g'(x^*)^T-g'(x)^T\|_1$$

$$\leq \|f''(x)-f''(x^*)\|_1 + \lambda^*\|g''(x^*)-g''(x)\|_1$$

$$+ \|g''(x)\|_1 |\lambda-\lambda^*| + \|g'(x^*)^T-g'(x)^T\|_1 + 2\|x-x^*\|_1$$

$$\leq (\gamma_1+\gamma_2\lambda^*+\sup_{x\in S_0}\|g''(x)\|_1+2)\|x-x^*\|_1$$

$$+ \sup_{x\in S_0}\|g''(x)\|_1 |\lambda-\lambda^*|.$$

Hence, under any norm in R^{n+1}, there exists a $\bar{\gamma} \in (0,\infty)$ such that

$$\|H'(x,\lambda)-H'(x^*,\lambda^*)\| \leq \bar{\gamma}\|(x,\lambda)-(x^*,\lambda^*)\|, \quad \forall (x,\lambda) \in S_0 \times \Lambda,$$

and, by Theorem 3.3.1, the rate of convergence is at least quadratic.

The condition (3.3.4) corresponds to a requirement used by Anselone and Rall [1964] in the case of the linear eigenvalue problem.

As a direct corollary of Theorem 3.3.2, we can also phrase the following convergence result for (3.3.2).

Corollary 3.3.3 Let $A \in L(R^n)$ be symmetric and positive definite and $\xi > 0$ a given real number. Assume that $f: D \subset R^n \to R^1$ is twice differentiable on an open neighborhood $S_0 \subset D$ of a point $x^* \in D$ for which there is a number λ^* such that $H(x^*, \lambda^*) = 0$ where H is now the operator

$$(3.3.7) \qquad H: D \times R^1 \subset R^{n+1} \to R^{n+1}, \; H(x,\lambda) = \begin{pmatrix} f'(x)^T - \lambda A x \\ \\ x^T A x - \xi \end{pmatrix} .$$

Moreover, suppose that f'' is continuous at x^* and, that $[f''(x^*) - \lambda^* A]^{-1}$ exists and satisfies

$$\delta^* = 2(Ax^*)^T [f''(x^*) - \lambda^* A]^{-1} A x^* \neq 0.$$

Then there is a neighborhood $S \times \Lambda \subset D \times R^1$ of (x^*, λ^*) such that for any $(x^0, \lambda^0) \in S \times \Lambda$, the Newton iteration (3.3.5) with H given by (3.3.7) converges to (x^*, λ^*). If, in addition, for some $\gamma_1 \in (0, \infty)$

$$\|f''(x) - f''(x^*)\| \leq \gamma_1 \|x - x^*\|, \quad \forall x \in S_0,$$

then the convergence is at least quadratic.

Finally, we turn to part (iii) of the complete algorithm. Following a suggestion of Pimbley [1969] we solve the n-dimensional initial value problem

$$(3.3.8) \qquad \frac{dx}{d\lambda} = [f''(x)-\lambda g''(x)]^{-1}g'(x)^T, \ x(\lambda^*) = x^*,$$

$$\lambda^* - \alpha \leqslant \lambda \leqslant \lambda^* + \alpha.$$

The next result, which relates solutions of (3.3.8) to (3.3.1), is based on the well-known Peano Existence Theorem (see, for example, Hartman [1964]).

Theorem 3.3.4 For given $f,g:D \subset R^n \to R^1$, let $x^* \in \text{int}(D)$, $\lambda^* \in R^1$ be a solution of (3.3.1). Suppose that f and g are twice continuously differentiable in some neighborhood $S \subset D$ of x^* and that $[f''(x^*)-\lambda^* g''(x^*)]^{-1}$ exists. Then there is an interval $[\lambda^*-\alpha,\lambda^*+\alpha]$ with $\alpha > 0$ on which (3.3.8) possesses a unique continuous solution $x = x(\lambda)$ which satisfies

$$f'(x(\lambda))^T = \lambda g'(x(\lambda))^T, \ \lambda \in [\lambda^*-\alpha,\lambda^*+\alpha].$$

Proof. By the continuity of $f''(x) - \lambda g''(x)$ on $S \times R^1$ and the existence of $[f''(x^*)-\lambda^* g''(x^*)]^{-1}$ it follows that there is a compact neighborhood $V \subset S \times R^1$ of (x^*,λ^*) in which $[f''(x)-\lambda g''(x)]^{-1}g'(x)^T$ is well-defined, continuous and bounded. Hence, by the Peano Existence Theorem, there is an $\alpha_1 > 0$ such that (3.3.8) possesses at least one solution $x = x(\lambda)$ on the interval $[\lambda^*-\alpha_1,\lambda^*+\alpha_1]$.

From (3.3.8) we see that for this solution $x(\lambda)$,

$$f''(x(\lambda))\frac{dx(\lambda)}{d\lambda} = g'(x(\lambda))^T + \ g''(x(\lambda))\frac{dx(\lambda)}{d\lambda}$$

and hence that

$$(3.3.9)\quad \frac{d}{d\lambda}\,(f'(x(\lambda))) = \frac{d}{d\lambda}\,(\lambda g'(x(\lambda))),\ \lambda \in [\lambda^*-\alpha_1, \lambda^*+\alpha_1].$$

Since $x(\lambda^*) = x^*$ and (x^*,λ^*) solves (3.3.1), we may integrate (3.3.9) over $[\lambda^*,\lambda]$, $\lambda \leq \lambda^* + \alpha_1$ and obtain

$$f'(x(\lambda))^T = \lambda g'(x(\lambda))^T,\ \lambda \in [\lambda^*, \lambda^*+\alpha_1].$$

A similar argument shows that this relation also holds on $[\lambda^*-\alpha_1,\lambda^*]$ and hence that $(x(\lambda),\lambda)$ solves (3.3.1) for $\lambda \in [\lambda^*-\alpha_1,\lambda^*+\alpha_1]$. But by Theorem 1.2.21, there is a proper continuous branch of solutions $\bar{x}(\cdot)$ of (3.3.1) which is unique on some interval $[\lambda^*-\alpha_2,\lambda^*+\alpha_2]$, $\alpha_2 > 0$, and hence, $x(\cdot)$ must coincide with $\bar{x}(\cdot)$ on $[\lambda^*-\alpha, \lambda^*+\alpha]$ with $\alpha = \min\{\alpha_1,\alpha_2\}$.

Note that, if $A \in L(R^n)$ is symmetric and positive definite and if (x^*,λ^*) solves (3.3.2) with nonsingular $(f''(x^*)-\lambda^*A)$, then Theorem 3.3.4 can be applied to (3.3.8) with $g'(x)^T \equiv Ax$. Thus, in this case, the solution $x = x(\lambda)$, $|\lambda - \lambda^*| \leq \alpha$, forms a unique proper continuous branch of solutions of (3.3.2).

3.4 Nonlinear Heat Generation in Conducting Solids

Joseph [1965] considered a problem of nonlinear heat generation in a conducting plate, and we begin with a brief description of his results. An electrical current is flowing through an infinite conducting plate (see Fig. 1) of finite width.

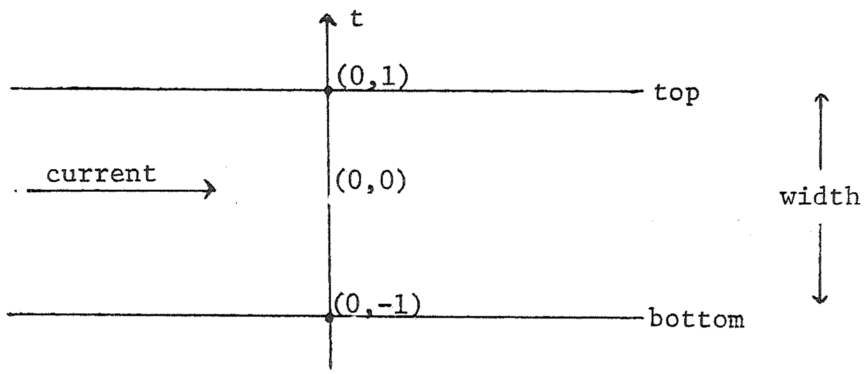


Fig. 1

In dimensionless coordinates, let the plate be specified by the set $\{(y,t) \in R^2 | -1 \leq t \leq 1\}$. It is assumed that the top and bottom surfaces of the plate are kept at equal and constant temperatures and that the thermal conductivity is constant. If the heat generation is assumed to be equal to the electrical power dissipation, then the governing differential equation is given by

$$(3.4.1) \qquad -u'' = \mu\bar{f}(u), \; u(1) = u(-1) = 0.$$

Here μ is proportional to the square of the current, $u = u(t)$ is the dimensionless temperature difference between t and the wall, and $\bar{f}$ is a temperature dependent function.

Consider the case when the nonlinear function $\bar{f}:[0,\infty) \to [0,\infty)$ is isotone on $[0,\infty)$. Then a result of Joseph [1965] is that there is some critical value μ_0 such that steady state solutions of (3.4.1) do not exist for $\mu > \mu_0$, while for $\mu \leq \mu_0$, (3.4.1) has both stable and unstable solutions. Furthermore, by symmetry, any solution assumes its maximum u_{max} along the center of the plate. In particular, for

(3.4.2) $$\bar{f}:[0,\infty) \to [0,\infty),\ \bar{f}(u) = e^u$$

and

$$\bar{f}:[0,\infty) \to [0,\infty),\ \bar{f}(u) = \bar{f}_m(u) \equiv 1 + u + \delta_2 u^2 + \ldots + \delta_m u^m,$$

$$\delta_i \geq 0,\ i = 2,\ldots,m,$$

solutions of (3.4.1) exist but are difficult to obtain analytically. On the other hand, for

(3.4.3) $$\bar{f}:[0,\infty) \to [0,\infty),\ \bar{f}(u) = f_2(u) \equiv 1 + u + \delta_2 u^2,\ \delta_2 = .195$$

Joseph [1965] displayed an analytical solution for u_{max} in terms of μ, and he noted that in this case $\mu_0 \simeq 1.28$. Moreover, he derived the bounds

(3.4.4) $$2u_{max}(1+ \frac{5}{6}\, u_{max} + \frac{11\delta_2}{15}\, u_{max}^2)^{-1} \leq \mu(u_{max})$$

$$\leq 2u_{max}(1+ \frac{2}{3}\, u_{max} + \frac{\delta_2}{2}\, u_{max}^2)^{-1}.$$

In order to compute approximate solutions of (3.4.1) we use the following standard discretization. With some given odd integer n let

(3.4.5) $$t_j = (-(\frac{n+1}{2})+j)h,\ h = \frac{2}{n+1},\ j = 0,1,\ldots,n+1,$$

and approximate $x''(t_j)$ by

$$\frac{1}{h^2}\,(x(t_{j+1}) - 2x(t_j) + x(t_{j-1})).$$

Then approximations x_i of the values $x(t_i)$, $i = 1,\ldots,n$ of the solution of (3.4.1) at the grid points t_i, $i = 1,\ldots,n$ are obtained as a solution of the system

$$(3.4.6)\qquad -x_{i+1} + 2x_i - x_{i-1} = h^2\mu\bar{f}(x_i),\ i = 1,\ldots,n,\ x_0 = x_{n+1} = 0.$$

In matrix notation this becomes

$$(3.4.7)\qquad Fx = \lambda Ax,\ \lambda = \frac{1}{\mu}$$

where

$$(3.4.8)\qquad F:R^n \to R^n,\ Fx = h^2(\bar{f}(x_1),\ldots,\bar{f}(x_n))$$

and $A \in L(R^n)$ is the symmetric, positive definite matrix

$$(3.4.9)\qquad A = \begin{bmatrix} 2 & -1 & & \bigcirc \\ -1 & \ddots & \ddots & \\ & \ddots & \ddots & -1 \\ \bigcirc & & -1 & 2 \end{bmatrix}$$

(see, for example, Ortega and Rheinboldt [1970]). For operators F of the form (3.4.8) it is easily seen that $F(x) = f'(x)^T$ with $f:R^n \to R^1$, $f(x) = h^2 \sum_{i=1}^{n} \int_0^{x_i} \bar{f}(t)dt$. Hence (3.4.7) can be written in the form

(3.4.10) $$f'(x)^T = \lambda Ax, \ \lambda = \frac{1}{\mu} .$$

Specifically, if $\bar{f}$ is defined by (3.4.2) or (3.4.3), then the corresponding functionals f are given by

(3.4.11) $$f:[0,\infty) \times \ldots \times [0,\infty) \subset R^n \to R^1, \ f(x) = h^2 \sum_{i=1}^{n} e^{x_i}$$

and

(3.4.12) $$f:[0,\infty) \times \ldots \times [0,\infty) \subset R^n \to R^1,$$

$$f(x) = h^2 \sum_{i=1}^{n} (x_i + \frac{1}{2} x_i^2 + \frac{\delta_2}{3} x_i^3), \ \delta_2 = .195,$$

respectively.

In order to obtain solutions of (3.4.10) which have a physical meaning in connection with (3.4.1), it is necessary to apply part (i) of the algorithm described in Section 3.3 to the problem

(3.4.13) $$\min \{-f(x) | \frac{1}{2} x^T Ax - \xi, \ \xi > 0\}$$

when f is, for example, given by (3.4.11) or (3.4.12).

We now indicate some of the computational aspects of solving (3.4.10). In implementing the complete algorithm of Section 3.3 we have used $n = 19$ in (3.4.5) and an initial vector

$$(1.,1.2,1.4,1.6,1.8,2.0,2.2,2.4,1.,\ldots,1.4)^T \in R^{19}$$

which is then normalized to lie in the constraint set. The choice of $\xi > 0$ in (3.4.13) simply determines the starting point for part (iii) of the algorithm and is otherwise unimportant. An error tolerance of $.1 \times 10^{-5}$ was used as indicated in part (i) of the algorithm, and several Newton steps were taken wherever this tolerance exceeded $.1 \times 10^{-5}$ to improve the accuracy of the results.

Tables I through VII and Graphs I and II display the numerical results which were obtained for (3.4.10) with f defined by (3.4.12). Table I compares the result obtained by implementing part (i) of the algorithm first with the Goldstein procedure of Section 3.2 using a fixed parameter τ and second with a minimization process of the type described in Chapter II (see Section 3.1). Tables II and III relate the quantities $\lambda = \frac{1}{\mu}$ and $\|x\|_2$ for the stable and unstable branches of solutions of (3.4.10), respectively. These results are then depicted in Graph I. In Tables IV and V are listed the data which corresponds directly to the results of Joseph [1965]. Here we have obtained values for $x_{max} = x_{10}$ and the corresponding values for μ and also the lower and upper bounds on μ given by (3.4.4). We note that Joseph [1965] only gives a graph of his results and that no accurate data was available for comparison. Our results indicate a critical value $\mu_0 \simeq 1.28411$ with corresponding $x_{max} \simeq 2.89567$. This differs somewhat from the results of Joseph [1965] who obtained $x_{max} \simeq 3.08$ when $\mu_0 \simeq 1.28$. However, on the unstable branch of solutions we obtained $\mu = 1.28110$ for $x_{max} = 3.08251$ (see Table V). These results are shown in Graph II. Finally, Tables VI and VII contain approximate solutions to (3.4.1) for selected values of μ. It should be noted that only $x_1,\ldots,x_{10}$ are given since the symmetry of the problem implies that

$x_{10+i} = x_{10-i}$, $i = 0,\ldots,10$.

We turn now to the problem (3.4.10) with f defined by (3.4.11). With precisely the same organization of the tables and the graphs as we have done for the preceding example, Tables VIII through XIV and Graphs III and IV summarize the computational results obtained for this problem. We note that in Tables XI and XII no lower or upper bounds were computed for μ. Further note that on the basis of these results, we obtain the estimate $\mu_0 \simeq .87676$ with corresponding $x_{max} \simeq 1.22790$.

3.5 The Heavy Rotating String Problem

In this section we present some numerical results for an eigenvalue problem related to a heavy rotating string with one free endpoint. This problem was studied by Kolodner [1955], and we first summarize his results. A heavy inelastic string of uniform cross-section ρ with one fixed endpoint is allowed to rotate with constant angular velocity ω, subject only to a tension $T(\cdot)$ and the force of gravity with acceleration constant g.

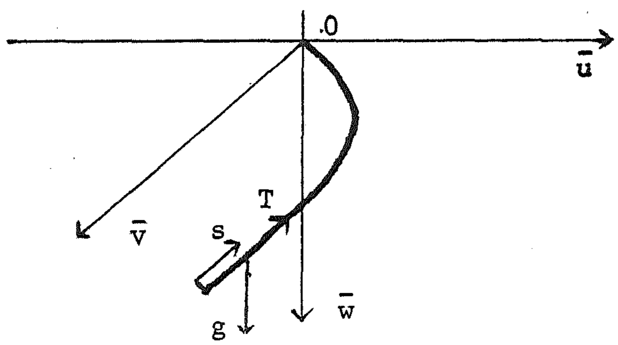


Fig. 2

We denote a point on the string in R^3 by $(\bar{u}(s,t),\bar{v}(s,t),\bar{w}(s,t))^T$, $(s,t) \in [0,1] \times [0,\infty)$, where s is the arclength measured from the free endpoint and t is the time. If $\bar{u}(s,t) = u(s)\cos \omega t$, $\bar{v}(s,t) = u(s)\sin \omega t$ and $\bar{w}(s,t) = w(s)$, $(s,t) \in [0,1] \times [0,\infty)$, with some vector function $(u(\cdot),v(\cdot),w(\cdot))^T$, then the equations of motion have the form

$$(Tu')' + \rho\omega^2 u = 0$$

(3.5.1)
$$(Tw')' + \rho g = 0 \quad , \quad 0 \leq s \leq 1$$

$$(u')^2 + (w')^2 = 1$$

$$T(0) = 0,\ u(1) = w(1) = 0.$$

With the variable change

$$y(s) = (\rho g)^{-1} T u'(s),$$

it is readily seen that

$$T(s) = \rho g (y^2+s^2)^{1/2}$$

$$u(s) = -\int_s^1 (y(\zeta)^2+\zeta^2)^{-1/2} y(\zeta)\, d\zeta$$

as well as

$$w(s) = \int_s^1 (y(\zeta)^2+\zeta^2)^{-1/2} \zeta\, d\zeta,$$

and that

(3.5.2)
$$-y'' = \omega^2 g^{-1} (y(s)^2+s^2)^{-1/2} y,\ y(0) = y'(1) = 0.$$

For this differential equation, Kolodner [1955] has shown that there is a critical value ω_0 such that for $\omega \leq \omega_0$ (3.5.2) has only the trivial solution while for all $\omega > \omega_0$, nontrivial solutions exist. ω_0 corresponds to a bifurcation point as discussed in Chapter I.

Since (3.5.2) is of the form

$$-y'' = \mu \bar{f}(s,y), \quad \mu = \frac{\omega^2}{g}, \quad y(0) = y'(1) = 0, \tag{3.5.3}$$

we use a discretization similar to that used for (3.4.1) in order to find approximate solutions of (3.5.2). Here, with some integer n, we set

$$t_j = jh, \quad h = \frac{1}{n+1}, \quad j = 0,\ldots,n+1 \tag{3.5.4}$$

and approximate the condition $y'(1) = 0$ by $x(t_n) = x(t_{n+1})$. Then, approximations x_i of the values $x(t_i)$, $i = 1,\ldots,n$ of a solution of (3.5.3) are obtained as solutions of the system

$$-x_{i+1} + 2x_i - x_{i-1} = h^2 \mu \bar{f}(t_i,x_i), \quad i = 1,\ldots,n-1, \quad x_0 = 0$$

$$x_n - x_{n-1} = h^2 \mu \bar{f}(t_n,x_n).$$

This is equivalent to the operator equation

$$Fx = \lambda Ax, \quad \lambda = \frac{1}{\mu} \tag{3.5.5}$$

where

$$F: R^n \to R^n, \quad Fx = h^2(\bar{f}(t_1,x_1),\ldots,\bar{f}(t_n,x_n))$$

and $A \in L(R^n)$ is now the symmetric positive definite matrix

$$A = \begin{bmatrix} 2 & -1 & & & \\ -1 & \ddots & \ddots & & \bigcirc \\ & \ddots & \ddots & \ddots & \\ \bigcirc & & \ddots & 2 & -1 \\ & & & -1 & 1 \end{bmatrix}$$

Therefore, as in Section 3.4, we can write (3.5.5) as

$$f'(x)^T = \lambda Ax, \; \lambda = \frac{1}{\mu} \tag{3.5.6}$$

with some functional $f: R^n \to R^1$. Specifically, for solving (3.5.2), we have

$$f: R^n \to R^1, \; f(x) = h^2 \sum_{i=1}^{n} (x_i^2 + i^2)^{1/2} \tag{3.5.7}$$

and, again, in order to obtain a physically meaningful solution we must apply part (i) of the algorithm of Section 3.3 to the problem

$$\min \{-f(x) \mid \tfrac{1}{2} x^T Ax - \xi, \; \xi > 0\}.$$

The complete algorithm of Section 3.3 was implemented in precisely the same manner as in Section 3.4; in particular, we used again $n = 19$ in (3.5.4) as well as the same initial vector and error criteria. Tables XV, XVI and the Graph V show the numerical results obtained for (3.5.6) with f defined by (3.5.7). In Table XV we again compare the results obtained by using, in part (i), the Goldstein procedure with fixed parameter τ versus those formed by implementing that part of the algorithm by means of a minimization algorithm of the type described in Chapter II. Table XVI lists the data obtained for $\|x\|_2$ as a function of μ, and the Graph V shows their pictorial representation. We note that Kolodner determined the estimate $\mu_0 = (\omega_0^2)/g \simeq 1.446$ while our estimate is $\mu_0 \simeq 1.48284$.

TABLE I

$$\bar{f}(u) = 1 + u + .195u^2$$

	Goldstein procedure with fixed τ	Minimization method as in Chapter II	Complete algorithm of Section 3.3
# iterations	60	60	
error	$.211 \times 10^{-4}$	$.146 \times 10^{-3}$	$.768 \times 10^{-7}$
μ	1.25439	1.25440	1.25438
x_1	.32990	.32978	.32989
x_2	.64284	.64261	.64283
x_3	.93417	.93387	.93415
x_4	1.19909	1.19871	1.19908
x_5	1.43291	1.43255	1.43290
x_6	1.63119	1.63080	1.63119
x_7	1.78996	1.78968	1.78996
x_8	1.90589	1.90563	1.90590
x_9	1.97647	1.97640	1.97650
x_{10}	2.00018	2.00016	2.00021

TABLE II

$\bar{f}(u) = 1 + u + .195u^2$

λ vs. $\|x\|_2$

STABLE BRANCH

λ	$\|x\|_2$	λ	$\|x\|_2$
.7789	8.2645	2.2683	.8855
.7795	8.0699	2.5584	.7642
.7801	7.9242	2.9337	.6494
.7807	7.8035	3.4380	.5405
.7931	6.6016	3.9862	.4573
.7994	6.2556	4.3316	.4168
.8059	5.9687	4.7424	.3772
.8191	5.5036	5.2343	.3383
.8397	4.9666	5.8526	.3001
.8765	4.2995	6.2164	.2813
.9167	3.7900	6.6285	.2626
.9607	3.3763	7.0990	.2441
1.0092	3.0264	7.6415	.2258
1.0629	2.7232	8.2738	.2076
1.1225	2.4557	9.0201	.1896
1.1893	2.2163	9.9143	.1718
1.2644	1.9998	11.0055	.1541
1.3498	1.8022	14.1115	.1191
1.4475	1.6206	16.4301	.1019
1.5604	1.4526	19.6030	.0848
1.6924	1.2965	24.7414	.0678
1.8490	1.1508	32.4002	.0510
2.0372	1.0141	47.9294	.0343
		92.0469	.0178

TABLE III

$$\bar{f}(u) = 1 + u + .195u^2$$

λ vs. $\|x\|_2$

UNSTABLE BRANCH

λ	$\|x\|_2$	λ	$\|x\|_2$
.7787	9.1749	1.0916	29.4889
.7794	9.4194	1.1160	30.7280
.7800	9.6021	1.1414	32.0142
.7806	9.7584	1.1681	33.3516
.7836	10.3445	1.1819	34.0408
.7867	10.7999	1.2561	37.7158
.7898	11.1938	1.3226	40.9706
.7930	11.5510	1.4162	45.5169
.8025	12.4968	1.4791	48.5470
.8431	15.6084	1.5478	51.8440
.8725	17.5120	1.6499	56.7254
.8880	18.4578	1.7358	60.8150
.9041	19.4088	1.8653	66.9584
.9207	20.3704	1.9759	72.1873
.9469	21.8416	2.1004	78.0615
.9652	22.8471	2.2417	84.7122
.9842	23.8762	2.3469	89.6587
1.0039	24.9322	2.4625	95.0863
1.0245	26.0183	2.7315	107.7012
1.0459	27.1375	2.9753	119.1154
1.0683	28.2933	3.2670	132.7513
		3.3773	137.9072

TABLE IV

$\bar{f}(u) = 1 + u + .195u^2$

μ vs. x_{max}

STABLE BRANCH

x_{max}	μ	lower bound on μ	upper bound on μ	x_{max}	μ	lower bound on μ	upper bound on μ
.05266	.10086	.10084	.10171	1.24200	1.11086	1.10127	1.25556
.10382	.19086	.19084	.19402	1.34294	1.14086	1.12994	1.29682
.15314	.27086	.27081	.27734	1.41961	1.16086	1.14893	1.32495
.20664	.35086	.35074	.36192	1.50621	1.18086	1.16779	1.35369
.24993	.41086	.41066	.42624	1.60604	1.20086	1.18649	1.38322
.30436	.48086	.48049	.50228	1.72464	1.22086	1.20497	1.41378
.40018	.59086	.59007	.62411	1.79387	1.23086	---	---
.50103	.69086	.68945	.73762	1.87244	1.24086	---	---
.60658	.78086	.77861	.84232	1.96394	1.25086	1.23193	1.46255
.70146	.85086	.84773	.92564	2.07390	1.26086	---	---
.81124	.92086	.91658	1.01089	2.21912	1.27086	---	---
.90164	.97086	.96555	1.07315	2.45725	1.28086	1.25653	1.52299
1.00531	1.02086	1.01430	1.13675	2.54232	1.28286	---	---
1.12687	1.07086	1.06276	1.20196	2.60450	1.28386	1.25808	1.53309

TABLE V

$$\bar{f}(u) = 1 + u + .195u^2$$

μ vs. x_{max}

UNSTABLE BRANCH

x_{max}	μ	lower bound on μ	upper bound on μ	x_{max}	μ	lower bound on μ	upper bound on μ
2.89567	1.28411	1.25568	1.54519	8.14453	.98610	.94305	1.26299
2.97393	1.28311	---	---	9.05680	.93610	.893308	1.20473
3.03244	1.28211	---	---	10.05027	.88610	.84387	1.14543
3.08144	1.28110	1.25117	1.54843	12.10561	.79610	.75556	1.03650
3.54271	1.26610	1.23289	1.54518	15.27725	.68610	.64863	.90023
4.05390	1.24110	1.20502	1.52835	18.26357	.60610	.571.46	.79932
4.56599	1.21110	1.11728	1.50279	21.58040	.53610	.50433	.70995
4.96194	1.18610	1.14651	1.47930	26.58669	.45610	.42801	.60671
5.57524	1.14610	1.10500	1.43925	30.70072	.40610	.38051	.54163
6.18691	1.10610	1.06400	1.39711	35.95709	.35610	---	---
7.13349	1.04610	1.00321	1.33123	42.91686	.30610	.28697	.41031
				44.58926	.29610	.27655	.39710

$\bar{f}(u) = 1 + u + .195u^2$

APPROXIMATE STABLE SOLUTION FOR SELECTED VALUES OF μ

μ	.10086	.20086	.30086	.40086	.50086	.60086	.70086
x_1	.00994	.09057	.03211	.04474	.05870	.07429	.09199
x_2	.01885	.03908	.06111	.08529	.11209	.14213	.17632
x_3	.02674	.05551	.08692	.12149	.15989	.20307	.25236
x_4	.03360	.06982	.10946	.15318	.20187	.25674	.31953
x_5	.03941	.08198	.12865	.18022	.23778	.30278	.37732
x_6	.04417	.09196	.14443	.20262	.26744	.34089	.42526
x_7	.04788	.09974	.15676	.21995	.28968	.37080	.46296
x_8	.05053	.10532	.16560	.23247	.30737	.39231	.49012
x_9	.05212	.10866	.17091	.23999	.31743	.40528	.50651
$x_{max} = x_{10}$	.05266	.10978	.17268	.24251	.32079	.40962	.51199

μ	.80086	.90086	1.00086	1.10086	1.20086	1.28386
x_1	.11247	.13684	.16704	.20718	.26924	.42106
x_2	.21602	.26340	.32234	.40097	.52306	.82344
x_3	.30975	.37845	.46420	.57900	.75796	1.20070
x_4	.39284	.48084	.59099	.73893	.97040	1.54611
x_5	.46454	.56948	.70117	.87854	1.15697	1.85284
x_6	.52417	.64342	.79337	.99581	1.31451	2.11435
x_7	.57116	.70182	.86639	1.08899	1.44020	2.32468
x_8	.60507	.74402	.91926	1.15662	1.53174	2.47880
x_9	.62555	.76955	.95128	1.19764	1.58738	2.57287
$x_{max} = x_{10}$	.63239	.77809	.96200	1.21138	1.60604	2.60450

TABLE VII

$\bar{f}(u) = 1 + u + .195u^2$

APPROXIMATE UNSTABLE SOLUTION FOR SELECTED VALUES OF μ

μ	1.28411	1.28311	1.28211	1.28110	1.20110	1.11610
x_1	.46442	.47602	.48468	.49208	.74325	.91956
x_2	.90949	.93253	.94973	.96444	1.46436	1.81586
x_3	1.32797	1.36206	1.38753	1.40931	2.15099	2.67355
x_4	1.71214	1.75665	1.78991	1.81835	2.78915	3.47468
x_5	2.05414	2.10815	2.14851	2.18303	3.36383	4.19960
x_6	2.34636	2.40864	2.45520	2.49502	3.85993	4.82810
x_7	2.58182	2.65089	2.70252	2.74669	4.26315	5.34082
x_8	2.75460	2.82870	2.88410	2.93151	4.56103	5.72060
x_9	2.86016	2.93737	2.99510	3.04450	4.74387	5.95433
$x_{max} = x_{10}$	2.89567	2.97393	3.03244	3.08251	4.80552	6.03319

μ	.90610	.50610
x_1	1.43220	3.34826
x_2	2.83874	6.66345
x_3	4.19625	9.89602
x_4	5.47557	12.97682
x_5	6.64324	15.82070
x_6	7.66368	18.33243
x_7	8.50185	20.41465
x_8	9.12621	21.97719
x_9	9.51165	22.94679
$x_{max} = x_{10}$	9.64199	23.27554

TABLE VIII

$\bar{f}(u) = e^u$

	Goldstein procedure with fixed τ	Minimization method as in Chapter II	Complete algorithm of Section 3.3
# Iterations	60	60	--
Error	$.149 \times 10^{-6}$	$.292 \times 10^{-3}$	$.237 \times 10^{-7}$
μ	.73176	.73173	.73176
x_1	.30875	.30835	.30875
x_2	.60754	.60672	.60754
x_3	.89289	.89173	.89289
x_4	1.16037	1.15886	1.16037
x_5	1.40450	1.40282	1.40450
x_6	1.61882	1.61700	1.61882
x_7	1.79621	1.79452	1.79621
x_8	1.92949	1.92803	1.92949
x_9	2.01239	2.01145	2.01239
$x_{max} = x_{10}$	2.04054	2.04017	2.04054

TABLE IX

$\bar{f}(u) = e^u$

λ vs. $\|x\|_2$

STABLE BRANCH

λ	$\|x\|_2$	λ	$\|x\|_2$
1.14073	3.62723	4.65910	.38585
1.14203	3.54752	4.88678	.36605
1.15656	3.17457	5.41613	.32707
1.17004	2.98838	6.07409	.28887
1.19813	2.72561	6.46689	.27006
1.22755	2.52538	6.91402	.25142
1.32515	2.09081	7.42756	.23296
1.43961	1.77439	8.02351	.21467
1.55127	1.55813	8.72343	.19656
1.68171	1.37006	9.55715	.17861
1.80294	1.23472	10.56706	.16082
2.02170	1.05112	11.81562	.14320
2.24904	.91232	13.39877	.12573
2.47137	.80888	15.47179	.10841
2.66927	.73514	18.30371	.09124
2.90163	.66434	22.40457	.07422
3.28263	.57408	28.87357	.05735
3.51329	.53057	40.59472	.04062
3.77881	.48806	68.33520	.02402
4.26196	.42607		

TABLE X

$f(u) = e^u$

λ vs. $\|x\|_2$

UNSTABLE BRANCH

λ	$\|x\|_2$	λ	$\|x\|_2$
1.14056	3.89880	3.67972	11.93360
1.14382	4.08124	4.13633	12.45985
1.14711	4.19430	4.72233	13.04077
1.15040	4.28501	5.21485	13.46649
1.15706	4.43401	5.82208	13.93086
1.16719	4.61598	6.18200	14.18038
1.18447	4.86522	6.58435	14.44340
1.20227	5.07857	7.05418	14.72173
1.22435	5.30735	8.21288	15.33368
1.24338	5.48309	8.94775	15.67345
1.26701	5.68153	9.82705	16.04123
1.29995	5.93105	10.89800	16.44272
1.40497	6.59240	12.23092	16.88554
1.51112	7.13419	13.93534	17.38027
1.66179	7.77292	16.19171	17.94219
1.81238	8.31024	19.31995	18.59457
2.07572	9.09033	23.94637	19.37563
2.42860	9.92806	31.48616	20.35477
2.84285	10.71978	45.95592	21.68016
3.20760	11.29717		

TABLE XI

$\bar{f}(u) = e^{u}$

μ vs. x_{max}

STABLE BRANCH

x_{max}	μ	x_{max}	μ
.01244	.02463	.53591	.68463
.10038	.18463	.56534	.70463
.14353	.25463	.59705	.72463
.18335	.31463	.63152	.74463
.23380	.38463	.66944	.76463
.26492	.42463	.71182	.78463
.28120	.44463	.76026	.80463
.31539	.48463	.81761	.82463
.35207	.52463	.88986	.84463
.38147	.55463	.99612	.86466
.42370	.59463	1.09614	.87463
.45808	.62463	1.11448	.87563
.49527	.65463	1.14017	.87663

TABLE XII

$\bar{f}(u) = e^u$

μ vs. x_{max}

UNSTABLE BRANCH

x_{max}	μ	x_{max}	μ
1.22790	.87676	3.29799	.40176
1.28703	.87426	3.48144	.36176
1.35323	.86926	3.62883	.33176
1.40176	.86426	3.84194	.29176
1.46115	.85676	4.07954	.25176
1.51192	.84926	4.27896	.22176
1.57178	.83926	4.42491	.20176
1.62588	.82926	4.66437	.16176
1.72333	.80926	4.95370	.14176
1.82240	.78676	5.17498	.12176
1.92459	.76176	5.43187	.10176
2.02159	.73676	5.57810	.09176
2.22509	.68176	5.73978	.08176
2.44137	.62176	5.93085	.07176
2.65963	.56176	6.36722	.05176
2.88601	.50176	7.51227	.02176
3.08515	.45176		

TABLE XIII

$\bar{f}(u) = e^u$

APPROXIMATE STABLE SOLUTION FOR SELECTED VALUES OF μ

μ	.05363	.10463	.20463	.30463	.40463	.50463
x_1	.00529	.01033	.02103	.03276	.04580	.06063
x_2	.01033	.01960	.03998	.06237	.08736	.11590
x_3	.01423	.02780	.05679	.08874	.12451	.16550
x_4	.01786	.03493	.07144	.11178	.15708	.20914
x_5	.02094	.04098	.08389	.13141	.18491	.24657
x_6	.02347	.04593	.09412	.14757	.20787	.27754
x_7	.02543	.04979	.10209	.16020	.22585	.30185
x_8	.02683	.05255	.10780	.16925	.23876	.31933
x_9	.02767	.05421	.11124	.17469	.24654	.32987
$x_{max} = x_{10}$	.02796	.05476	.11238	.17651	.24913	.33339

μ	.60463	.70463	.80463	.87463	.87663
x_1	.07807	.09987	.13119	.18205	.18845
x_2	.14961	.19195	.25321	.35361	.36633
x_3	.21413	.27549	.36486	.51271	.53155
x_4	.27115	.34976	.46492	.65721	.68186
x_5	.32025	.41402	.55218	.78483	.81483
x_6	.36102	.46763	.63546	.89328	.92801
x_7	.39311	.50999	.68369	.98036	1.01900
x_8	.41624	.54061	.72599	1.04413	1.08571
x_9	.43021	.55914	.75166	1.08305	1.12646
$x_{max} = x_{10}$	.43488	.56534	.76026	1.09614	1.14017

TABLE XIV

$\bar{f}(u) = e^u$

APPROXIMATE UNSTABLE SOLUTION FOR SELECTED VALUES OF μ

μ	.87676	.87176	.77176	.67176	.57176	.47176
x_1	.20105	.21456	.28915	.33577	.37849	.42185
x_2	.39137	.41831	.56799	.66215	.74863	.83650
x_3	.56873	.60881	.83321	.97550	1.10668	1.24027
x_4	.73060	.78329	1.08067	1.27103	1.44744	1.62773
x_5	.87427	.93869	1.30539	1.54261	1.76389	1.99116
x_6	.99693	1.07181	1.50164	1.78278	2.04698	2.32004
x_7	1.09582	1.17946	1.55326	1.98301	2.28578	2.60092
x_8	1.16848	1.25876	1.78414	2.13443	2.46837	2.81822
x_9	1.21294	1.30736	1.85907	2.22908	2.58346	2.95652
$x_{max} = x_{10}$	1.22790	1.32374	1.88448	2.26131	2.62284	3.00409

μ	.37176	.27176	.17176	.07176	.02176
x_1	.46917	.52500	.59896	.72580	.88436
x_2	.93240	1.04541	1.19479	1.45012	1.76820
x_3	1.38618	1.55808	1.78495	2.17138	2.65077
x_4	1.82509	2.05785	2.36487	2.88634	3.53025
x_5	2.24095	2.53634	2.92651	3.58844	4.40230
x_6	2.62184	2.98050	3.45610	4.26458	5.25659
x_7	2.95159	3.37113	3.93126	4.88968	6.06914
x_8	3.21019	3.68264	4.31886	5.41939	6.78761
x_9	3.37665	3.88613	4.57747	5.78712	7.31312
$x_{max} = x_{10}$	3.43429	3.95721	4.66901	5.92085	7.51227

TABLE XV

$$\bar{f}(s,y) = \omega^2 g^{-1} (y^2+s^2)^{-1/2} y$$

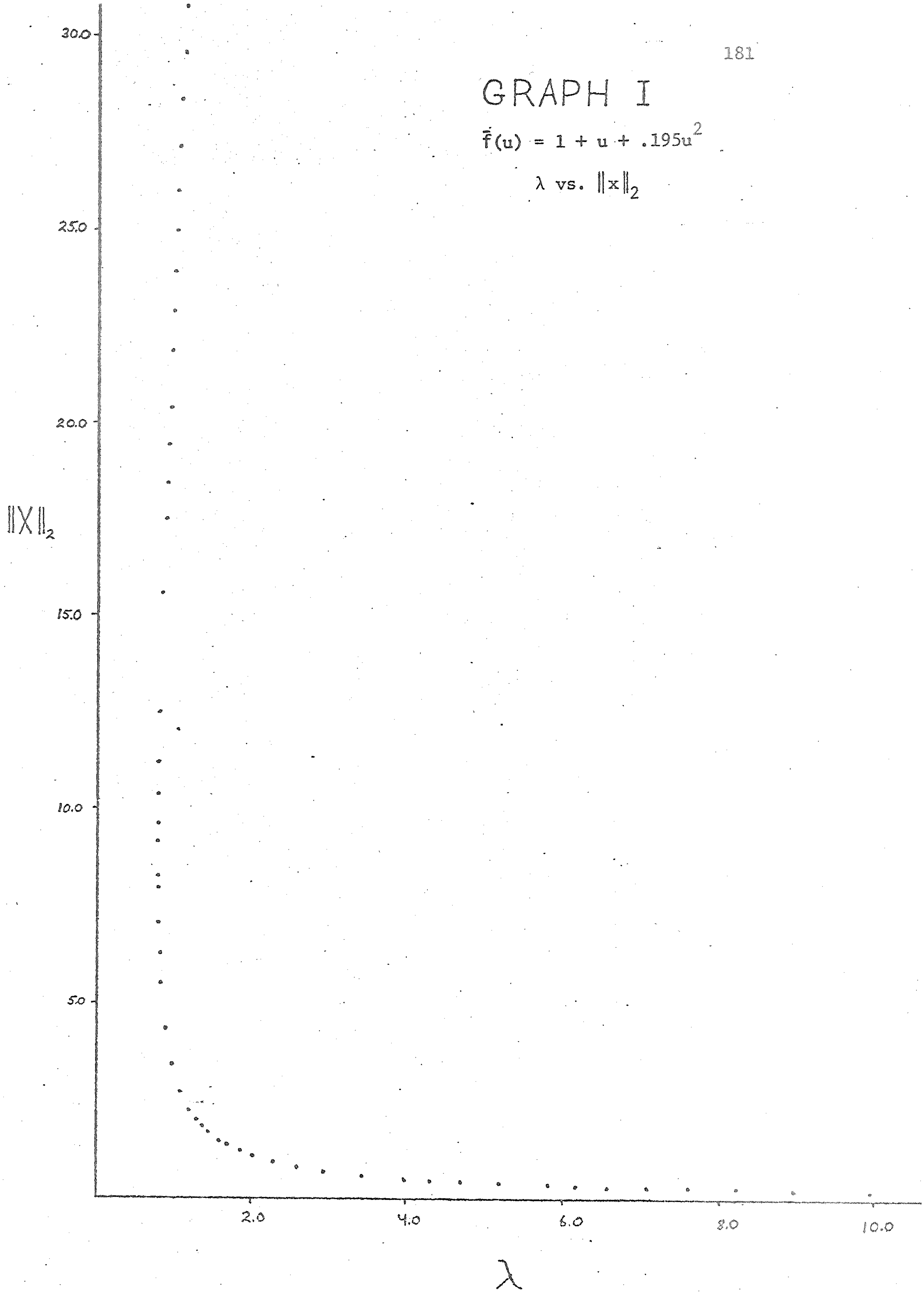
	Goldstein procedure with fixed τ	Minimization method as in Chapter II	Complete algorithm of Section 3.3
Execution time	6.194 sec.	3.0190 sec.	
# iterations	200	200	
Error	$.156 \times 10^{-2}$	$.355 \times 10^{-3}$	$.738 \times 10^{-7}$
λ	8.22367	8.20562	8.19808
x_1	.38391	.38348	.38318
x_2	.74676	.74655	.74604
x_3	1.08931	1.08937	1.08859
x_4	1.41156	1.41168	1.41083
x_5	1.71352	1.71383	1.71278
x_6	1.99521	1.99540	1.99445
x_7	2.25664	2.25689	2.25585
x_8	2.49783	2.49778	2.49700
x_9	2.71730	2.71868	2.71792
x_{10}	2.91805	2.91899	2.91861
x_{11}	3.09861	3.09938	3.09910
x_{12}	3.25900	3.25925	3.25941
x_{13}	3.39925	3.39926	3.39957
x_{14}	3.51938	3.51885	3.51959
x_{15}	3.61941	3.61862	3.61952
x_{16}	3.69939	3.69814	3.69937
x_{17}	3.75787	3.75785	3.75920
x_{18}	3.79786	3.79750	3.79903
x_{19}	3.81791	3.81734	3.82892

TABLE XVI

μ vs. $\|x\|_2$

$$\bar{f}(s,y) = \omega^2 g^{-1}(y^2+s^2)^{-1/2} y$$

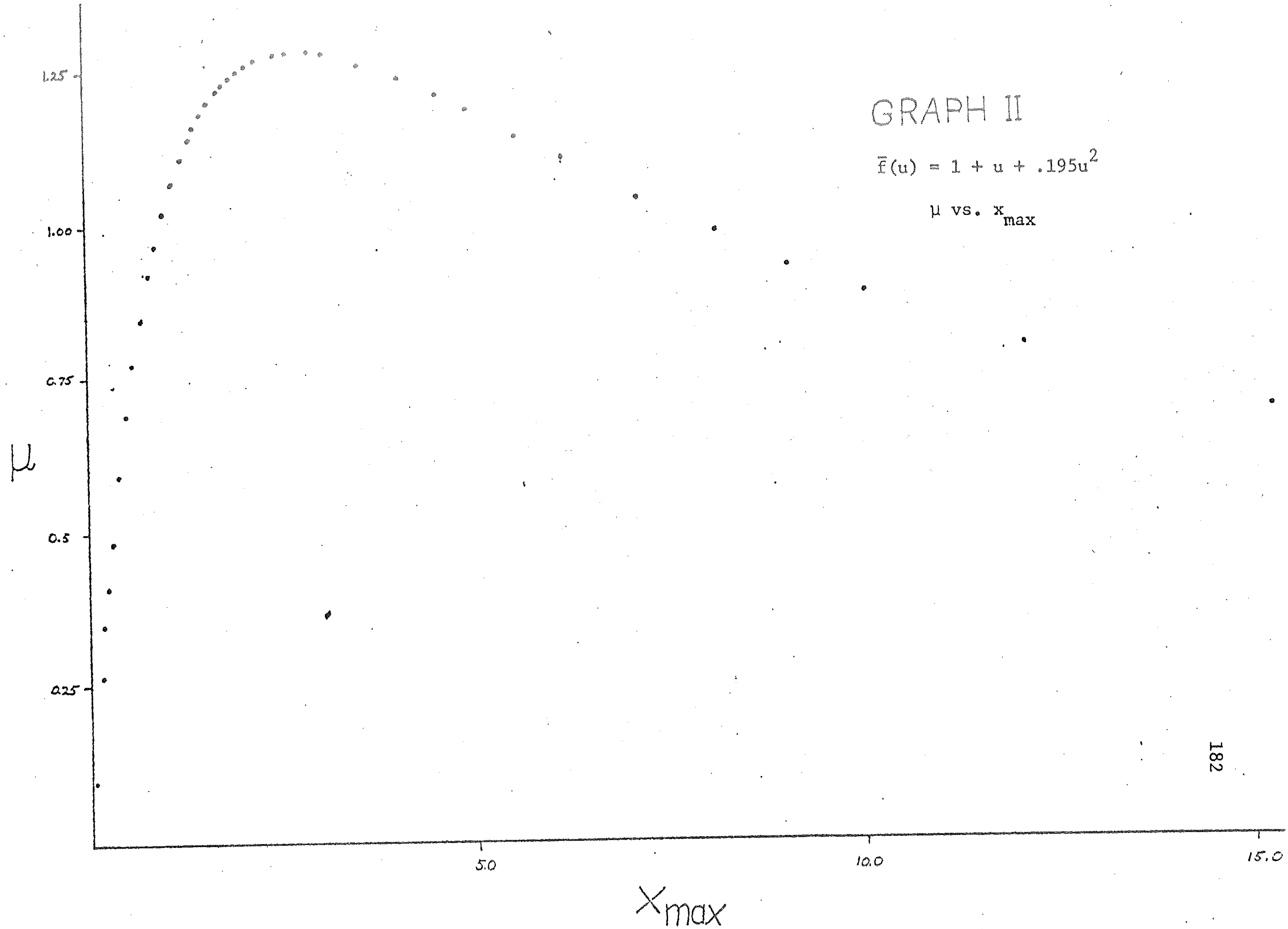
μ	$\|x\|_2$	μ	$\|x\|_2$
1.48284	$.1931 \times 10^{-9}$	3.28784	4.43921
1.48384	$.6246 \times 10^{-1}$	3.48784	4.78309
1.48584	.12540	3.68784	5.12244
1.48884	.18312	3.98784	5.62468
1.50084	.32469	4.28784	6.12047
1.53084	.53836	4.58784	6.61121
1.58084	.78195	4.88784	7.09793
1.58784	.81118	5.18784	7.58137
1.68784	1.16412	5.48784	8.06210
1.78784	1.45406	5.78784	8.54058
1.88784	1.71200	6.09894	9.01715
1.98784	1.95001	6.38784	9.49209
2.08784	2.17426	6.68784	9.96564
2.28784	2.59458	6.98784	10.43798
2.48784	2.98954	7.28784	10.90926
2.68784	3.36746	7.58784	11.37962
2.88784	3.73316	7.88784	11.84916
3.08784	4.08972	8.18784	12.31798

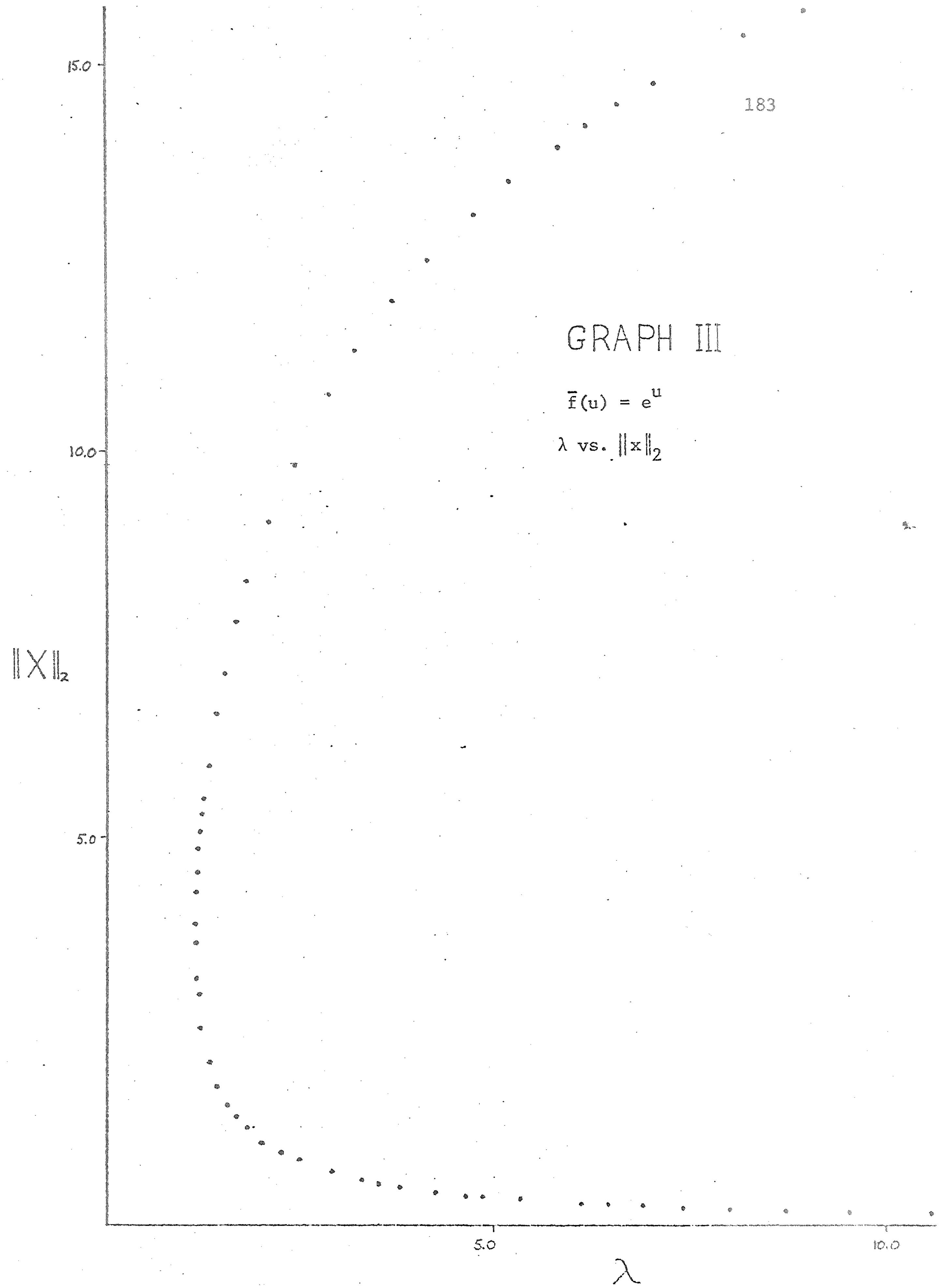

GRAPH I
f(u) = 1 + u + .195u^2
λ vs. ‖x‖₂
30.0
25.0
20.0
15.0
10.0
5.0
‖X‖₂
2.0
4.0
6.0
8.0
10.0
λ

GRAPH II

$\bar{f}(u) = 1 + u + .195u^2$

μ vs. x_{max}

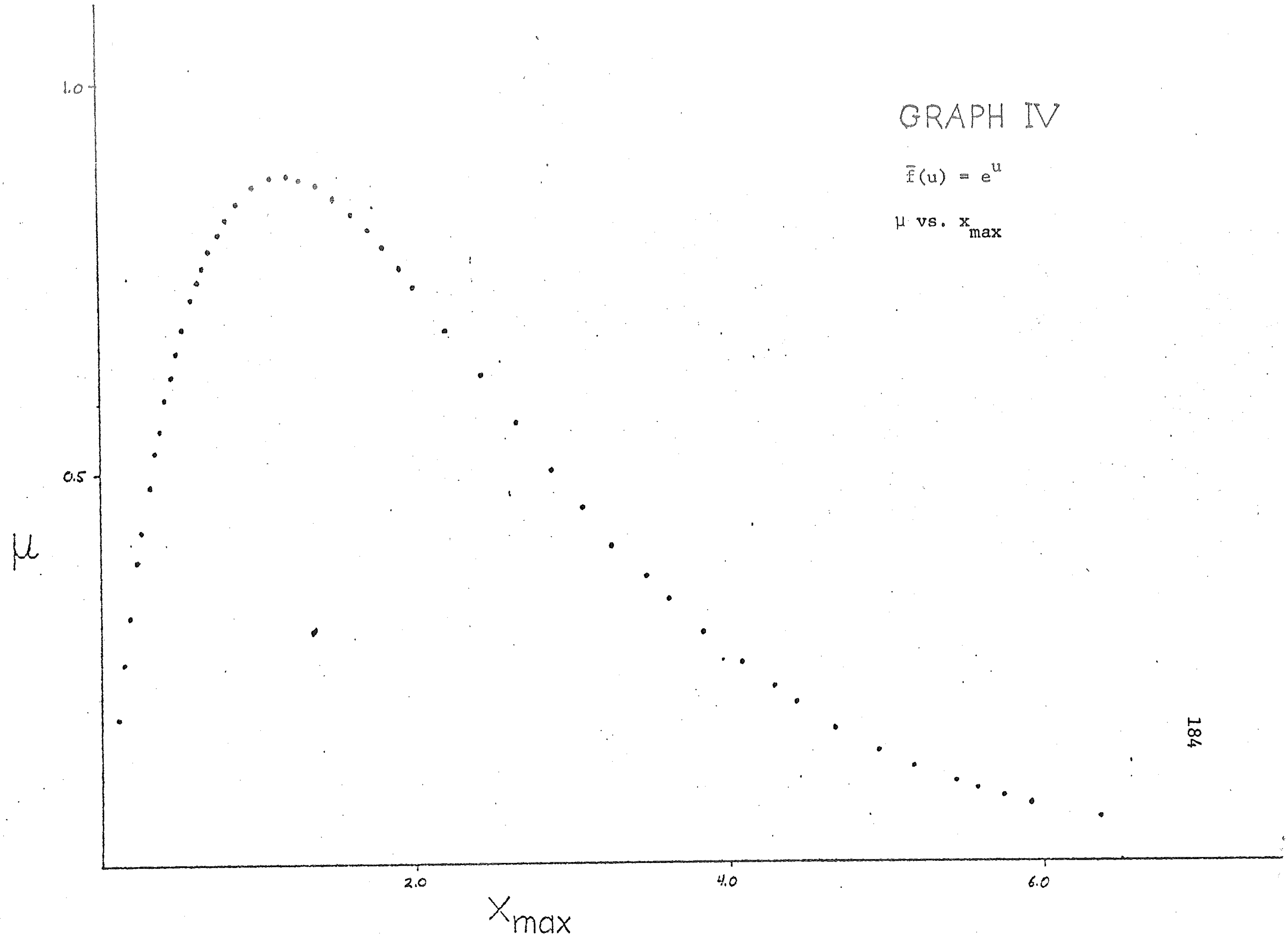


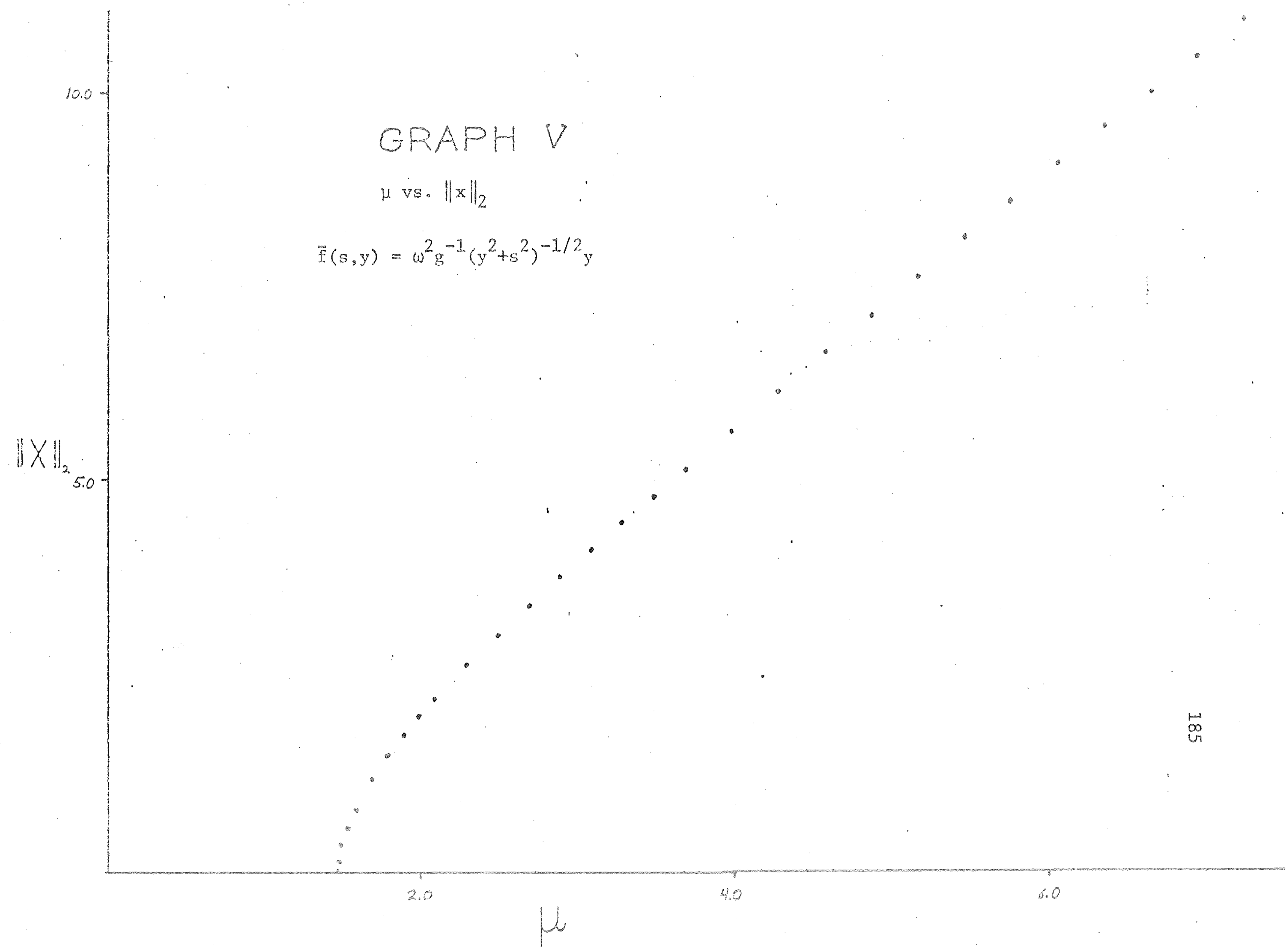

GRAPH III
$\bar{f}(u) = e^u$
λ vs. $\|x\|_2$
15.0
10.0
5.0
$\|X\|_2$
5.0
10.0
λ

GRAPH IV

$\bar{f}(u) = e^u$

μ vs. x_{max}




GRAPH V
μ vs. ‖x‖₂
f̄(s,y) = ω²g⁻¹(y²+s²)^(-1/2) y
‖X‖₂
10.0
5.0
μ
2.0
4.0
6.0

REFERENCES

Altman, M. [1964], Stationary points in nonlinear programming, Bull. Acad. Polon. des Sci., Sér. des sci. math., astr. et phys. 12, 29-35.

Altman, M. [1966], Generalized gradient methods of minimizing a functional, Bull. Acad. Polon. des Sci., Sér. des sci. math., astr. et phys. 14, 313-318.

Anselone, P., and Rall, L. [1968], The solution of characteristic value-vector problems by Newton's method, Numer. Math. 11, 38-45.

Armijo, L. [1966], Minimization of functions having Lipschitz continuous first partial derivatives, Pacific J. Math. 16, 1-3.

Berger, M., and Berger, M. [1968], Perspectives in Nonlinearity: An Introduction to Nonlinear Analysis, W. Benjamin, Inc., New York, N.Y.

Canon, M., Cullum, C., and Polak, E., [1970], Theory of Optimal Control and Mathematical Programming, McGraw Hill, Inc., New York, N.Y.

Curry, H. [1944], The method of steepest descent for nonlinear minimization problems, Quart. Appl. Math. 2, 258-263.

Dieudonné, J. [1960], Foundations of Modern Analysis, Academic Press, Inc., New York, N.Y.

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

Goldstein, A. [1962], Cauchy's method of minimization, Numer. Math. 4, 146-150.

Goldstein, A. [1964], Minimizing functionals on Hilbert space, in Computing Methods in Optimization Problems, A. Balakrishnan and L. Neustadt (Eds.), Academic Press, Inc. New York, 159-165.

Goldstein, A. [1965], On steepest descent, SIAM J. Control 3, 147-151.

Goldstein, A. [1966], Minimizing functionals on normed linear spaces, SIAM J. Control 4, 81-89.

Goldstein, A. [1967], Constructive Real Analysis, Harper and Row, Inc., New York, N.Y.

Golomb, M. [1934], Zur Theorie der nichtlinearen Integralgleichungen, Integralgleichungenssysteme und allgemeinen Funktionalgleichungen, Math. Z. 39, 45-75.

Gordan, P. [1873], Über die Auflösung linearer Gleichungen mit reellen Coefficienten, Math. Ann. 6, 23-28.

Granas, A. [1961], Introduction to Topology of Functional Spaces, Mathematics Lecture Notes, University of Chicago.

Hartman, P. [1964], Ordinary Differential Equations, J. Wiley and Sons, Inc., New York, N.Y.

Householder, A. [1964], The Theory of Matrices in Numerical Analysis, Blaisdell, Inc., New York, N.Y.

Joseph, D. [1965], Nonlinear heat generation and stability of the temperature distribution in conducting solids, Int. J. Heat and Mass Transfer 8, 281-288.

Kalfon, P., Ribiere, G., and Sogno, J. [1968], A method of feasible directions using projection operators, Proc. IFIP Congress 68, North Holland Publishing Co., Amsterdam, The Netherlands.

Kolodner, I. [1955], Heavy rotating string: A nonlinear eigenvalue problem, Comm. Pure and Appl. Math. 8, 395-408.

Krasnosel'skii, M. [1964], Topological Methods in the Theory of Nonlinear Integral Equations, translated by A. Armstrong, Pergamon Press, Oxford, England.

Kuhn, H., and Tucker, A. [1951], Nonlinear programming, in Proc. Second Berkeley Symp. Math., Stat., and Prob., J. Neyman (Ed.), University of California Press, Berkeley, Calif., 481-492.

Künzi, H., and Oettli, W. [1969], Nichtlineare Optimierung: Neuere Verfahren Bibliographie, Lecture Notes in Operations Research and Mathematical Systems 16, Springer-Verlag, Berlin, Germany,

Ljusternik, L. [1934], On conditional extrema of functionals, Mat. Sb. 41, 390-401.

Ljusternik, L., and Sololev, W. [1955], Elemente der Funktionanalysis, Akademie-Verlag, Berlin, Germany.

Mangasarian, O. [1963], Equivalence in nonlinear programming, Naval Research Logist. Quart. 10, 299-306.

Mangasarian, O. [1969], Nonlinear Programming, McGraw Hill, Inc., New York, N.Y.

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

Ostrowski, A. [1966], Solution of Equations and Systems of Equations, 2nd Edition, Academic Press, Inc., New York, N.Y.

Pimbley, G. [1969], Eigenfunction Branches of Nonlinear Operators, and Their Bifurcations, Lecture Notes in Mathematics 104, Springer-Verlag, Berlin, Germany.

Riedrich, T. [1968], Über positive Eigenwerte finiter Abbildungen, Math. Nachr. 37, 119-134.

Rosen, J. [1960], The gradient projection method for nonlinear programming. Part I: Linear constraints, J. SIAM 8, 181-217.

Rosen, J. [1961], The gradient projection method for nonlinear programming. Part II: Nonlinear constraints, J. SIAM 9, 514-532.

Stepleman, R. [1969], Finite dimensional analogues of variational and quasi-linear elliptic Dirichlet problems, Ph.D. Dissertation, University of Maryland.

Topkis, D., and Veinott, A. [1967], On the convergence of some feasible direction algorithms for nonlinear programming, SIAM J. Control 5, 268-279.

Vainberg, M. [1964], Variational Methods for the Study of Nonlinear Operators, translated by A. Feinstein, Holden-Day, San Francisco, Calif.

Vainberg, M., and Aizengendler, P. [1968], The theory and methods of investigation of branch points of solutions, in Progress in Mathematics, R. Gamkrelidze (Ed.), Plenum Press, Inc. New York, N.Y.

Zangwill, W. [1967], Nonlinear programming via penalty functions, Manag. Sci., Series A 13, 344-358.

Zoutendijk, G. [1960], Methods of Feasible Directions: A Study in Linear and Nonlinear Programming, Elsevier Pub. Co., Inc., Amsterdam, The Netherlands.