WIS-CS-185-73
The University of Wisconsin
Computer Sciences Department
1210 West Dayton Street
Madison, Wisconsin 53706

Received June 18, 1973

NONLINEAR PROGRAMMING THEORY AND COMPUTATION

by

O. L. Mangasarian

Computer Sciences Technical Report #185

July 1973

NONLINEAR PROGRAMMING THEORY AND COMPUTATION[1)2)]

by

O. L. Mangasarian[3)]

ABSTRACT

A survey of nonlinear programming theory and computational algorithms is given. Subjects covered are: optimality conditions, duality theory, unconstrained and constrained optimization algorithms.

[1] Supported by NSF Grant GJ35292.

[2] Chapter 6 of "Handbook of Operations Research" S. E. Elmaghraby and J. J. Moder, editors, VanNostrand Reinhold Company, New York.

[3] Computer Sciences Department, University of Wisconsin, 1210 West Dayton Street, Madison, Wisconsin 53706

CHAPTER 6 NONLINEAR PROGRAMMING THEORY AND COMPUTATION

6.1 INTRODUCTION

The basic nonlinear programming problem is to

$$\left.\begin{array}{ll} \text{minimize} & f(x_1, x_2, \ldots, x_n) \\ \text{subject to} & g_j(x_1, x_2, \ldots x_n) \leqq 0, \; j=1, \ldots, m \\ & h_j(x_1, x_2, \ldots, x_n) = 0, \; j=1, \ldots, k \end{array}\right\} \quad (6.1)$$

By using vector notation the above problem can be written more succinctly as

$$\left.\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & g(x) \leqq 0 \\ & h(x) = 0 \end{array}\right\} \quad (6.1)$$

where $f:R^n \rightarrow R$, $g:R^n \rightarrow R^m$ and $h:R^n \rightarrow R^k$, that is f, g and h are functions which map each point of the n dimensional real Euclidean space R^n into R, R^m and R^k respectively.

Problems such as (6.1) arise frequently in the decision [Dantzig & Veinott, 1968] and physical sciences [Fox, 1971, Stark & Nichols, 1972] and their systematic study beginning in the late 1940's [John, 1948] has grown into the discipline of nonlinear programming. In this survey we shall be concerned with the theory and computational algorithms of nonlinear programming. In section 6.2 we shall discuss various optimality conditions. Besides their intrinsic importance optimality conditions play an important role in the computational algorithms also. In section 6.3 we shall dwell briefly on one type of duality in nonlinear programming. Beginning with section 6.4 we shall be discussing various computational algorithms of nonlinear programming. In section 6.4 we shall discuss one dimensional minimization problems. In section 6.5 we shall discuss unconstrained minimization algorithms, that is the problem: $\underset{x \in R^n}{\text{minimize}}\ f(x)$. Finally in section 6.6 we shall discuss a variety of algorithms for solving problem (6.1).

We begin with a few paragraphs on notation and definitions.

We shall be interested in various types of "solutions" of problem (6.1). A point $\bar{x}$ in R^n satisfying $g(\bar{x}) \leqq 0$ and $h(\bar{x}) = 0$ is said to be: a <u>solution</u> or <u>global solution</u> of (6.1) iff $f(\bar{x}) \leqq f(x)$ for all x satisfying $g(x) \leqq 0$ and $h(x) = 0$; a <u>local solution</u> of (6.1) iff $f(\bar{x}) \leqq f(x)$ for all x satisfying $g(x) \leqq 0$, $h(x) = 0$ and $\|x-\bar{x}\| \leqq \delta$ for some $\delta > 0$ where $\|x-\bar{x}\| = (\sum_{j=1}^{n} (x_j-\bar{x}_j)^2)^{\frac{1}{2}}$; a unique local solution of (6.1) iff $f(\bar{x}) < f(x)$ for all x satisfying $g(x) \leqq 0$, $h(x) = 0$ and $\|x-\bar{x}\| \leqq \delta$ for some $\delta > 0$.

We shall also make use of the concepts of convex and concave functions. More general concepts can be found in [Mangasarian, 1969]. A function f defined on a set X in R^n is said to be convex at $\bar{x}$ (with respect to X) iff

$$\left.\begin{array}{l} x \in X,\ 0 < \lambda < 1,\ (1-\lambda)x+\lambda\bar{x} \in X,\ \text{imply:} \\ (1-\lambda)f(\bar{x}) + \lambda f(x) \geqq f((1-\lambda)\bar{x} + \lambda x) \end{array}\right\} \qquad (6.2)$$

If the last inequality of 6.2 is strict ($>$) for $x \neq \bar{x}$ then f is strictly convex at x, and if it is reversed ($\leqq$) then f is concave at $\bar{x}$. The function f is convex on X if it is convex at each $\bar{x}$ in X.

We shall use vector notation quite frequently. If x and y are in R^n, then x_i and y_i, $i=1,\ldots,n$, denote their respective components, $xy = \sum_{i=1}^{n} x_i y_i$ is the scalar product of x and y, and xy^T is the $n \times n$ matrix whose ij^{th} element is $x_i y_j$. Superscripts will denote specific vectors such as x^1 and x^2 in R^n say. Exponentiation will be distinguished by enclosing the quantity raised to a power by parentheses. If $f: R^n \to R$ is differentiable at $\bar{x}$ then $\nabla f(\bar{x}) = \left(\frac{\partial f(\bar{x})}{\partial x_1}, \ldots, \frac{\partial f(\bar{x})}{\partial x_n}\right)$ and if f is twice differentiable at $\bar{x}$ then $\nabla^2 f(\bar{x})$ is the $n \times n$ Hessian matrix whose ij element is $\frac{\partial^2 f(\bar{x})}{\partial x_i \partial x_j}$. We shall say that a function is differentiable (twice differentiable) <u>around</u> $\bar{x}$ if it is differentiable (twice differentiable) in a neighborhood of $\bar{x}$. The symbol $\blacksquare$ will mark the end of a statement of a theorem or of an algorithm.

6.2 OPTIMALITY CONDITIONS

Necessary optimality conditions are those conditions which some type of solution of (6.1) must satisfy. Sufficient optimality conditions are conditions which when satisfied guarantee that the point is some type of solution of (6.1). We give first some of the best known optimality conditions which involve first derivatives only and hence are called first order optimality conditions.

6.2.2 First Order Kuhn Tucker Conditions [Kuhn & Tucker 1951, Mangasarian 1969] (Necessity) Let $\bar{x}$ be a local or global solution of 6.1 and let f, g and h be differentiable at $\bar{x}$. Let g and h satisfy a first order constraint qualification at $\bar{x}$ [Mangasarian, 1969] such as this: h is continuously differentiable at $\bar{x}$, $\nabla h_i(\bar{x})$, $i=1,\ldots,k$, are linearly independent and there exists a z in R^n satisfying $\nabla g_i(\bar{x}) z > 0$ for $i=1,\ldots,m$ such that $g_i(\bar{x}) = 0$, and $\nabla h_j(\bar{x})z = 0$ for $j=1,\ldots,k$, then $\bar{x}$ and some $\bar{u}$ in R^m and $\bar{v}$ in R^k satisfy the Kuhn-Tucker conditions

$$\left.\begin{aligned} \nabla f(x) + \sum_{i=1}^{m} \bar{u}_i \nabla g_i(\bar{x}) + \sum_{i=1}^{k} \bar{v}_i \nabla h_i(\bar{x}) &= 0 \\ \bar{u}_i g_i(\bar{x}) &= 0, i=1,\ldots,m \\ g(\bar{x}) &\leqq 0 \\ h(\bar{x}) &= 0 \\ \bar{u} &\geqq 0 . \end{aligned}\right\} \quad (6.3)$$

(Sufficiency) Conversely if f and g are differentiable and convex at $\bar{x}$, h is linear and conditions (6.3) are satisfied, then $\bar{x}$ is a global solution of (6.1). ■

The Kuhn-Tucker conditions (6.3) are merely a statement that a linear combination of the gradients of the objective function with positive weights, the gradients of the active inequality constraints $(g_i(\bar{x}) = 0)$ with nonnegative weights, and the gradients of the equality constraints must vanish at $\bar{x}$. A constraint qualification is imposed to rule out singular solutions at which the Kuhn-Tucker conditions cannot hold. For example the origin in R^2 is the global solution of the problem minimize x_1 subject to $(x_1)^4 - x_2 \leqq 0$ and $-(x_1)^3 + x_2 \leqq 0$, but neither a constraint qualification nor the Kuhn-Tucker conditions (6.3) are satisfied there. In the absence of convexity, conditions (6.3) cannot rule out stationary points such as the origin for the function $(x)^3$. To rule out such stationary points, conditions involving second order derivatives have been developed [Fiacco & McCormick, 1968].

6.2.3 <u>Second Order Kuhn-Tucker Conditions</u> (Necessity) Let $\bar{x}$ be a local or global solution of 6.1 and let f, g and h be twice continuously differentiable around $\bar{x}$. Let g and h satisfy a first order constraint qualification such as that of 6.2.2 and a second order constraint qualification such as:

$\nabla g_i(\bar{x})$, $i \in \{i \mid g_i(\bar{x}) = 0, i=1,\ldots,m\}$ and $\nabla h_j(\bar{x})$, $j=1,\ldots,k$, are linearly independent. Then $\bar{x}$ and some $\bar{u}$ in R^m and $\bar{v}$ in R^k satisfy (6.3), <u>and</u> for every y in R^n such that $y\nabla g_i(\bar{x}) = 0$, $i \in \{i \mid g_i(\bar{x}) = 0, i=1,\ldots,m\}$, and $y\nabla h_j(\bar{x}) = 0$ for $j=1,\ldots,k$, it follows that

$$y\nabla_{11}L(\bar{x},\bar{u},\bar{v})y \geqq 0 \tag{6.4}$$

where

$$L(x,u,v) = f(x) + \sum_{i=1}^{m} u_i g_i(x) + \sum_{i=1}^{k} v_i h_i(x) \tag{6.5}$$

and $\nabla_{11}L$ is the $n \times n$ Hessian matrix of second partial derivatives of L with respect to its first argument x. (Sufficiency) Conversely, let f, g and h be twice differentiable at $\bar{x}$, let $\bar{x}$ and some $\bar{u}$ in R^m and $\bar{v}$ in R^k satisfy (6.3), <u>and</u> let for every nonzero $y \in R^n$ such that

$$\left.\begin{array}{l} y\nabla g_i(\bar{x}) = 0 \text{ for } i=1,\ldots,m,\ g_i(\bar{x}) = 0 \text{ and } \bar{u}_i > 0 \\ y\nabla g_i(\bar{x}) \leqq 0 \text{ for } i=1,\ldots,m,\ g_i(\bar{x}) = 0 \text{ and } \bar{u}_i = 0 \\ y\nabla h_j(\bar{x}) = 0 \text{ for } i=1,\ldots,k, \end{array}\right\} \tag{6.6}$$

it follows that

$$y\nabla_{11}L(\bar{x},\bar{u},\bar{v})y > 0 . \tag{6.7}$$

Then $\bar{x}$ is a unique local solution of (6.1). ∎

The second order sufficient optimality condition plays an important role in some of the superlinearly convergent algorithms that we will discuss in section 6.6. It should be noted also that second order sufficient optimality conditions can also apply to linear programming problems in which case, since $\nabla_{11}L(\bar{x},\bar{u},\bar{v}) = 0$, the conditions (6.6) must imply that $y = 0$. For such a case $\bar{x}$ would be the unique solution to the linear program. Condition (6.4) is trivially satisfied by all solutions of linear programs.

More recently there have been efforts [Rockafellar, 1971, 1972a, 1972b, 1973, Buys, 1972, Mangasarian, 1973] directed towards relating solutions of nonlinear programming problems to solutions of nonlinear equations. The objective of such relation is to employ the vast machinery for solving nonlinear equations [Ortega & Rheinboldt, 1970, Ostrowski, 1966] in attacking nonlinear programming problems. One such relationship is embodied in the following optimality conditions which involve no inequalities whatsoever.

6.2.4 <u>First Order Modified Lagrangian Optimality Conditions</u>

[Mangasarian, 1973] (Necessity) Let $\bar{x}$ be a local or global solution of 6.1 and let f, g and h be differentiable at $\bar{x}$. Let g and h satisfy a first order constraint qualification at $\bar{x}$ such as that of 6.2.2. Then $\bar{x}$ and some $(\bar{y},\bar{z})$ in $R^m \times R^k$ constitute a stationary point of the modified Lagrangian function $M(x,y,z)$

$$\left.\begin{aligned} M(x,y,z) = f(x) + \frac{1}{4r}\sum_{i=1}^{m}((rg_i(x)+y_i)_+^4 - (y_i)^4) + \\ + \sum_{i=1}^{k}(\frac{r}{2}(h_i(x))^2 + z_i h_i(x)) \end{aligned}\right\} \quad (6.8)$$

where r is any fixed positive number and the notation $(\beta)_+^4$ denotes 0 if $\beta < 0$ and $(\beta)^4$ if $\beta \geqq 0$. In particular we have that

$$\nabla_1 M(\bar{x},\bar{y},\bar{z}) = 0\ ,\ \nabla_2 M(\bar{x},\bar{y},\bar{z}) = 0,\ \nabla_3 M(\bar{x},\bar{y},\bar{z}) = 0 \qquad (6.9)$$

where $\nabla_i M$ denotes the gradient of M with respect to its i^{th} argument, i = 1,2,3. (Sufficiency) Let f and g be differentiable and convex at $\bar{x}$, let h be linear and let conditions (6.9) hold. Then $\bar{x}$ is a global solution of (6.1). ∎

Note that the conditions (6.9) involve no inequalities of any sort, neither inequalities on any of the variables nor constraint inequalities. Thus, tools of nonlinear equations theory [Ortega & Rheinholdt, 1970] can be employed directly to solve (6.9). The relation between the optimal Kuhn Tucker multipliers $\bar{u}$, $\bar{v}$ of 6.3 and the optimal $\bar{y}$, $\bar{z}$ of 6.9 are given by

$$\left.\begin{aligned}\bar{u}_i &= (\bar{y}_i)^3, i=1,\ldots,m \\ \bar{v}_i &= \bar{z}_i, i=1,\ldots,k\end{aligned}\right\} \quad (6.10)$$

For more details concerning the modified Lagrangian see [Mangasarian, 1973, Arrow et al, 1971, Rockafellar 1971, 1972a, 1972b, 1973].

Another optimality condition that may be useful computationally is that associated with exact penalty functions. An exact penalty function from R^n into R associated with problem (6.1) is a function whose local or global minima for <u>finite</u> values of a parameter that it contains are associated with local or global minima of (6.1). In general such functions are continuous but only piecewise differentiable [Zangwill, 1967, Pietrzykowski, 1969, Howe, 1973, Evans et al, 1973]. However some exact penalty functions are differentiable locally [Fletcher, 1972]. We shall confine ourselves here to a simple piecewise differentiable exact penalty function.

6.2.5 <u>Exact Penalty Optimality Conditions</u> (Necessity)

Let $\bar{x}$ be a local or global solution of (6.1) and let either: (a) f, g and h be continuously differentiable around $\bar{x}$ and let one of the following constraint qualifications hold (1) the system

$\nabla f(\bar{x})y \leqq 0$, $\nabla g_i(\bar{x})y \leqq 0$, $i \in \{i \mid g_i(\bar{x}) = 0, i=1,\ldots,m\}$, $\nabla h(\bar{x})y = 0$ has no solution $y \neq 0$ in R^n, (2) $\nabla g_i(\bar{x})$, $i \in \{i \mid g_i(\bar{x}) = 0, i=1,\ldots,m\}$, $\nabla h_i(\bar{x}), i=1,\ldots,k$ are linearly independent, (b) or let f and g be convex on R^n, let h be linear and let the Slater constraint qualification hold, that is there exists an $\tilde{x}$ such that $g(\tilde{x}) < 0$ and $h(\tilde{x}) = 0$. Then there exists a real number $r_0 > 0$ such that for all $r \geqq r_0$: $P(\bar{x},r) \leqq P(x,r)$ for all x in some open neighborhood of $\bar{x}$, where

$$P(x,r) = f(x) + r\left[\sum_{i=1}^{m} g_i(x)_+ + \sum_{i=1}^{k} |h_i(x)|\right]$$

and $g_i(x)_+$ denotes $g_i(x)$ if $g_i(x) \geqq 0$ and 0 if $g_i(x) < 0$.

(Sufficiency) If for all $r \geqq r_0 > 0$, $P(\bar{x},r) \leqq P(x,r)$ for all x in some set S containing $\bar{x}$ and some feasible point of (6.1), then $\bar{x}$ solves (6.1) subject to the extra condition that $x \in S$. (Note that S may be taken as R^n or a sufficiently large neighborhood of $\bar{x}$.) ∎

Note also that in the sufficiency part, no differentiability, convexity or even continuity was assumed on f, g or h.

Finally we give one additional optimality condition which will be useful in connection with gradient projection algorithms.

6.2.6 <u>First Order Gradient Projection Optimality Conditions</u>

[Levitin & Polyak, 1966] (Necessity) Let $\bar{x}$ be a solution of (6.1), let g be continuous and convex on R^n, let h be linear and let f be differentiable at $\bar{x}$. Then for any $n \times n$ symmetric positive definite matrix H the gradient projection condition

$$Q(\bar{x} - \mu H \nabla f(\bar{x})) = \bar{x} \quad \text{for each } \mu \geqq 0$$

holds, where Q(z) is the projection of z on $X = \{x \mid g(x) \leqq 0, h(x) = 0\}$, that is

$$\mu \| H\nabla f(\bar{x}) \| = \underset{y \in X}{\text{minimum}} \| \bar{x} - \mu H \nabla f(\bar{x}) - y \| \quad \text{for each } \mu \geqq 0$$

(Sufficiency) If under the same assumptions, the gradient projection condition holds for <u>some</u> $\mu > 0$ and f is convex at $\bar{x}$, then $\bar{x}$ is a solution of (6.1). ■

6.3 DUALITY

Associated with the minimization problem (6.1) is a maximization problem which under suitable conditions gives the same extremum as (6.1). In particular we have the following <u>dual problem</u> [Wolfe, 1961] to the <u>primal problem</u> (6.1)

$$\left.\begin{aligned} &\text{maximize} \quad L(x,u,v) = f(x) + \sum_{i=1}^{m} u_i g_i(x) + \sum_{i=1}^{k} v_i h_i(x) \\ &\text{subject to} \quad \nabla_1 L(x,u,v) = \nabla f(x) + \sum_{i=1}^{m} u_i \nabla g_i(x) + \sum_{i=1}^{k} v_i \nabla h_i(x) = 0 \\ &\qquad\qquad u \geqq 0 \end{aligned}\right\} \quad (6.11)$$

There are also other dual formulations [Rockafellar, 1970, Mangasarian & Ponstein, 1965]. The dual problem can give rise to lower bounds to the minimum value of (6.1) and also to computational algorithms for solving (6.1) [Buys, 1972]. For more details concerning duality theory as presented here see [Mangasarian, 1969].

6.2.5 <u>Weak Duality Theorem</u> Let f and g be differentiable on R^n and let h be linear. Then

$$\left.\begin{aligned} &g(x^1) \leqq 0,\ h(x^1) = 0 \\ &\nabla_1 L(x^2,u^2,v^2) = 0,\ u^2 \geqq 0 \\ &f \text{ and } g \text{ convex at } x^2 \end{aligned}\right\rangle \quad \text{imply} \quad f(x^1) \geqq L(x^2,u^2,v^2)\ . \ \blacksquare$$

6.2.6 Duality Theorem Let f, g and h be differentiable on R^n, let $\bar{x}$ be a local or global solution of the primal problem (6.1) and let a first order constraint qualification such as that of 6.2.2 be satisfied at $\bar{x}$. If f, g and h are twice differentiable at $\bar{x}$, then $\bar{x}$ and some $\bar{u}$, $\bar{v}$ in $R^m \times R^k$ satisfy the first order Kuhn Tucker conditions for the dual problem (6.11) and $f(\bar{x}) = L(\bar{x},\bar{u},\bar{v})$. If f and g are convex on R^n and h is linear, then $\bar{x}$ and some $\bar{u}$, $\bar{v}$ in $R^m \times R^k$ constitute a global solution of the dual problem (6.11) and $f(\bar{x}) = L(\bar{x},\bar{u},\bar{v})$. ∎

6.2.7 Strict Converse Duality Theorem Let f, g and h be differentiable on R^n and let $(\bar{x},\bar{u},\bar{v})$ be a local or global solution of the dual problem (6.11). If the $n \times n$ Hessian matrix $\nabla_{11}L(\bar{x},\bar{u},\bar{v})$ is nonsingular, then $(\bar{x},\bar{u},\bar{v})$ satisfy the first order Kuhn Tucker conditions (6.3) of the primal problem (6.1). If in addition f and g are convex at $\bar{x}$ and h is linear, then $\bar{x}$ is also a global solution of the primal problem (6.1). ∎

6.4 ONE DIMENSIONAL MINIMIZATION ALGORITHMS

Because many algorithms of unconstrained and constrained optimization require the minimization of a function along a line we give in this section algorithms for solving the following problem

$$\left.\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & a \leqq x \leqq b \end{array}\right\} \qquad (6.12)$$

where f is a function defined on the real line segment [a,b]. Often in using one dimensional minimization as part of an n-dimensional optimization algorithm, the upper bound b will be missing from problem (6.12). However it will be known that the derivative $f'(a) < 0$. For such problems we first search for b such that $f'(b) > 0$ and then solve (6.12).

6.4.1 Golden Section and Fibonacci Search Algorithms [Wilde, 1964, Kowalik & Osborne, 1968]. Let $a^0 = a$, $b^0 = b$. Having a^i, b^i, determine a^{i+1}, b^{i+1} as follows:

(i) Define

$$\ell^i = b^i - \tau^i(b^i - a^i), \; r^i = a^i + \tau^i(b^i - a^i)$$

where for golden section

$$\tau^i = \tau = (\sqrt{5}-1)/2 \cong 0.618$$

and for Fibonacci search

$$\tau^i = \frac{F_{n-i-1}}{F_{n-i}}, \ i=0,1,\ldots,n-3, \ \tau^{n-2} = \frac{1+\varepsilon}{2} \ \text{or} \ \frac{1-\varepsilon}{2}$$

where $n(n \geqq 2)$ is a prescribed number of allowable function evaluations, ε is any small positive number less than 1, and F_j are the Fibonacci numbers: $F_0 = F_1 = 1$, $F_j = F_{j-1} + F_{j-2}$, $j \geqq 2$.

(ii) Set

$$a^{i+1} = \ell^i, b^{i+1} = b^i \quad \text{if} \quad f(\ell^i) > f(r^i)$$
$$a^{i+1} = a^i, b^{i+1} = r^i \quad \text{if} \quad f(\ell^i) \leqq f(r^i) \quad \blacksquare$$

6.4.2 Convergence of Golden Section and Fibonacci Search Algorithms

Assume that f has a minimum solution $\bar{x}$ in the interval $[a,b]$ and that f is unimodal on $[a,b]$, that is for ℓ, r in $[a,b]$ and $\ell < r$, if $f(\ell) > f(r)$ then $\bar{x} \in [\ell, b]$, if $f(\ell) < f(r)$ then $\bar{x} \in [a,r]$ and if $f(\ell) = f(r)$ then $\bar{x} \in [\ell, r]$. After $n(n \geqq 2)$ function evaluations the minimum solution $\bar{x}$ is in $[a_{n-1}, b_{n-1}]$ where

$$b_{n-1} - a_{n-1} = \begin{cases} (0.618)^{n-1}(b-a) & \text{for golden section} \\ \frac{1}{F_n}(b-a) \ \text{or} \ \frac{1+\varepsilon}{F_n}(b-a) & \text{for Fibonacci search} \ \blacksquare \end{cases}$$

Fibonacci search has the property of giving the smallest final interval length $b_{n-1} - a_{n-1}$ for a fixed number n of allowable function evaluations. However golden section is in general preferred because it is simpler to implement and it does not require a knowledge of n in advance. For other methods see [Danilin, 1971]. Since $b^i - a^i$ is the maximum error in the ith step, an estimate of the rate of convergence is given by the relation between $b^i - a^i$ and $b^{i-1} - a^{i-1}$. We have then $b^i - a^i = 0.618(b^{i-1} - a^{i-1})$ for golden section $b^i - a^i = \frac{F_{n-1}}{F_{n-i+1}}(b^{i-1} - a^{i-1})$ for Fibonacci search where it can be shown that F_{n-1}/F_{n-i+1} approaches 0.618 as i and n approach ∞. Because of the foregoing linear relation between $(b^{i-1} - a^{i-1})$ and $(b^i - a^i)$ we say that golden section has a linear convergence rate and Fibonacci search has an asymptotic linear convergence rate. To improve on these rates, we would like to get a ratio between the errors at the ith and (i-1)th steps which tends to zero rather than to a positive constant. Such rates of convergence a are termed superlinear because they are better than any linear convergence rate. One such method is the secant method (also called sometimes regula falsi). This is a discrete version of Newton's method for finding a zero of the derivative f' of f in $[a,b]$ if it exists. Newton's method itself consists of linearizing f' around the current point and taking the zero of the linearized function as the next point of the iteration. In particular we have the following secant algorithm.

6.4.3 <u>Secant Algorithm</u> [Ortega & Rheinboldt, 1970, Shampine & Allen, 1973] Start with two distinct points x^0, x^1 in $[a,b]$ such that $f'(x^0) \neq f'(x^1)$. Having x^i, x^{i-1} compute x^{i+1} as follows

$$x^{i+1} = x^i - \frac{f'(x^i)}{\dfrac{f'(x^i) - f'(x^{i-1})}{x^i - x^{i-1}}} \qquad \blacksquare$$

6.4.4 <u>Convergence of the Secant Algorithm</u> Let

$\left| \dfrac{f'''(x)}{2f''(y)} \right| \leqq M$ for some M and all x, y in some interval $[a',b']$ containing $[a,b]$ and let $M|\hat{x}-x^0| < 1$ and $M|\hat{x}-x^1| < 1$ where $\hat{x}$ is a zero of f' in $[a',b']$, that is $f'(\hat{x}) = 0$. Then the sequence $\{x^i\}$ converges to $\hat{x}$. If in addition $f''(x) > 0$ for all x in $[a',b']$, then the unique solution $\bar{x}$, of $\min_{x \in [a,b]} f(x)$, is given by

$\bar{x} = \hat{x}$ if $a \leqq \hat{x} \leqq b$, $\bar{x} = a$ if $\hat{x} < a$, and $\bar{x} = b$ if $\hat{x} > b$. $\blacksquare$

It can be shown that a bound on the error of the secant algorithm is given by

$$|x^i - \hat{x}| < \frac{\delta^{F_i}}{M}$$

where $\delta = \max\{M(x^0-\hat{x}), M(x^1-\hat{x})\} < 1$ and F_i is the ith Fibonacci number defined earlier under Fibonacci search. If we let e^i denote this error bound then

$$e^{i+1} = \delta^{F_{i-1}} e^i .$$

For large i , $\delta^{F_{i-1}}$ tends to $\delta^{0.724(1.618)^{i-1}}$ which approaches zero as i approaches ∞ . Hence the convergence rate is super-linear and is of order 1.618 per function evaluation. This is better than the order $\sqrt{2} = 1.414$ per function evaluation of Newton's method for one dimensional problems.

6.5 UNCONSTRAINED MINIMIZATION ALGORITHMS

We shall be concerned here with the problem

$$\underset{x \in R^n}{\text{minimize}} \quad f(x) \tag{6.13}$$

where $f: R^n \to R$. Because of space limitations we shall limit ourselves to two of the most effective algorithms for solving (6.13): the Davidon-Fletcher-Powell variable metric method and the method of conjugate gradients. Both of these methods have two important properties: (i) They have a superlinear rate of convergence and (ii) They find the minimum of a strictly convex quadratic function in n steps or less. The variable metric algorithm imitates the Newton method by computing $x^{i+1} - x^i = -\lambda^i H^i \nabla f(x^i)$ where H^i is an approximation in a certain sense to $\nabla^2 f(x^i)^{-1}$ and λ^i is a stepsize. The approximation is achieved by forcing H^i to satisfy the quasi-Newton condition $H^{i+1} (\nabla f(x^{i+1}) - \nabla f(x^i)) = x^{i+1} - x^i$ which must hold if f were a strictly convex quadratic function. The conjugate direction methods, on the other hand, picks directions p^i in R^n which satisfy the conjugacy relation $p^i \nabla^2 f(x^i) p^{i-1} = 0$. For a strictly convex quadratic function, the n first such directions are linearly independent and are orthogonal to $\nabla f(x^n)$, the gradient at the nth point, which must then vanish, so that x^n is optimal.

6.5.1 <u>Variable Metric Algorithm (Davidon-Fletcher-Powell)</u> [Fletcher & Powell, 1963] Start with any x^0 in R^n. Let $H^0 = I$, the $n \times n$ identity matrix. Having x^i and H^i determine x^{i+1} and H^{i+1} as follows:

(1) Let $p^i = -H^i \nabla f(x^i)$. Compute $x^{i+1} = x^i + \lambda^i p^i$ where λ^i is the first nonnegative root of $\nabla f(x^i + \lambda p^i) p^i = 0$, or, $\lambda^i \geqq 0$ and $f(x^i + \lambda^i p^i)$ is the first local minimum of $f(x^i + \lambda p^i)$ subject to $\lambda \geqq 0$.

(2) If $y^i = 0$ or $z^i = 0$ where $z^i = x^{i+1} - x^i$ and $y^i = \nabla f(x^{i+1}) - \nabla f(x^i)$ set $x^i = \bar{x}$ and stop. Else compute

$$H^{i+1} = H^i + \frac{z^i z^{i^T}}{z^i y^i} - \frac{(H^i y^i)(H^i y^i)^T}{y^i H^i y^i} \quad \blacksquare$$

6.5.2 <u>Convergence and Rate of Convergence of the Variable Metric Algorithm</u> [Powell, 1971, Broyden et al, 1972] If $f(x) = \frac{1}{2} xCx + ax$ and C is positive definite, then algorithm 6.5.1 arrives at the unique minimum solution $\bar{x}$ in n or less steps. More generally if f is twice continuously differentiable on R^n

and $\alpha\|y\|^2 \leqq y\nabla^2 f(x)y$ for some $\alpha > 0$ and all x,y in R^n, then the sequence $\{x^i\}$ generated by 6.5.1 terminates at or converges to $\bar{x}$, the unique solution of problem (6.13). If, in addition, $\|\nabla^2 f(y) - \nabla^2 f(x)\| \leqq R\|y-x\|$ for some $R > 0$ and all x,y in R^n, then $\|x^{i+1}-\bar{x}\| \leqq \delta^i\|x^i-\bar{x}\|$, where $\lim_{i\to\infty} \delta^i = 0$, that is $\{x^i\}$ converges superlinearly to $\bar{x}$. ∎

6.5.3 Conjugate Directions Algorithm [Flecher & Reeves, 1964, Polyak, 1969b, Polak & Ribière, 1969] Start with any x^0 in R^n and set $p^0 = -\nabla f(x^0)$. Having x^i, p^i determine x^{i+1}, p^{i+1} as follows

(1) $x^{i+1} = x^i + \lambda^i p^i$, where λ^i is the first nonnegative root of $\nabla f(x^i+\lambda p^i)p^i = 0$, or, $\lambda^i \geqq 0$ and $f(x^i+\lambda^i p^i)$ is the first local minimum of $f(x^i + \lambda p^i)$ subject to $\lambda \geqq 0$.

(2) If $\nabla f(x^i) = 0$, set $\bar{x} = x^i$ and stop. Otherwise compute

$$p^{i+1} = -\nabla f(x^{i+1}) + \alpha^{i+1} p^i$$

where

$$\alpha^{i+1} = \frac{\|\nabla f(x^{i+1})\|^2}{\|\nabla f(x^i)\|^2} \quad \text{(Fletcher-Reeves)}$$

or

$$\alpha^{i+1} = \frac{(\nabla f(x^{i+1}) - \nabla f(x^i))\nabla f(x^{i+1})}{\|\nabla f(x^i)\|^2} \quad \text{(Polyak-Polak-Ribière)}$$ ∎

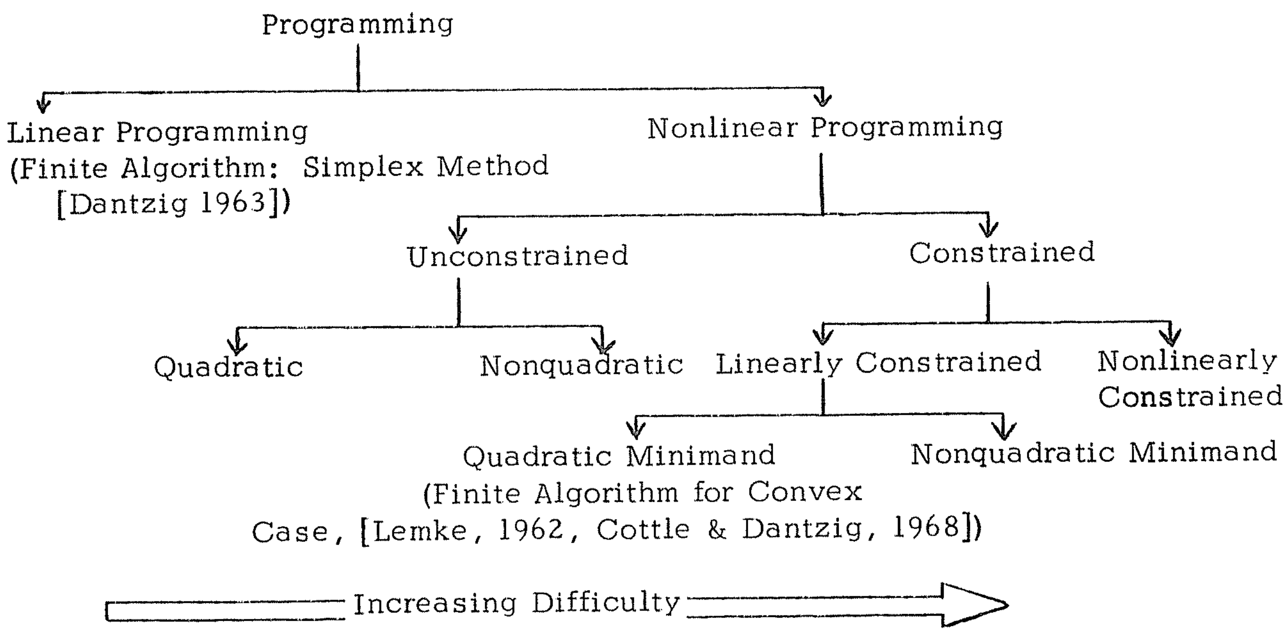
6.5.4 <u>Convergence and Rate of Convergence of the Conjugate Gradient Method</u> [Zoutendijk, 1970, McCormick & Ritter, 1970, 1972, Cohen, 1972] Let $\|\nabla f(y) - \nabla f(x)\| \leqq M\|y-x\|$ for some $M > 0$ and all y and x in R^n. If $f(x) = \frac{1}{2} xCx + ax$ and C is positive definite, then the conjugate direction algorithm arrives in n or less steps at $\bar{x}$, the unique minimum. More generally: (a) If the set $\{x | f(x) \leqq f(x^0)\}$ is bounded, then the Fletcher-Reeves algorithm either terminates at a stationary point $\bar{x}$, that is $\nabla f(\bar{x}) = 0$ or at least one accumulation point $\bar{x}$ of the sequence $\{x^i\}$ is stationary. (b) If f is twice continuously differentiable on R^n, then the Polyak-Polak-Ribière algorithm 6.5.3 either terminates at a stationary point $\bar{x}$ or each accumulation point $\bar{x}$ of the sequence $\{x^i\}$ is stationary. If in addition the set $\{x | f(x) \leqq f(x^0)\}$ is bounded and $\alpha\|y\|^2 \leqq y\nabla^2 f(x)y \leqq M\|y\|^2$ for some $M \geqq \alpha > 0$ and all x, y in R^n, then the Polyak-Polak-Ribière conjugate directions algorithm 6.5.3 is n-step superlinearly convergent, that is

$$\|x^{i+n} - \bar{x}\| \leqq \delta^i \|x^i - \bar{x}\|, \quad \lim_{i \to \infty} \delta^i = 0$$

provided that p^i is reset to $-\nabla f(x^i)$ after each n steps. ∎

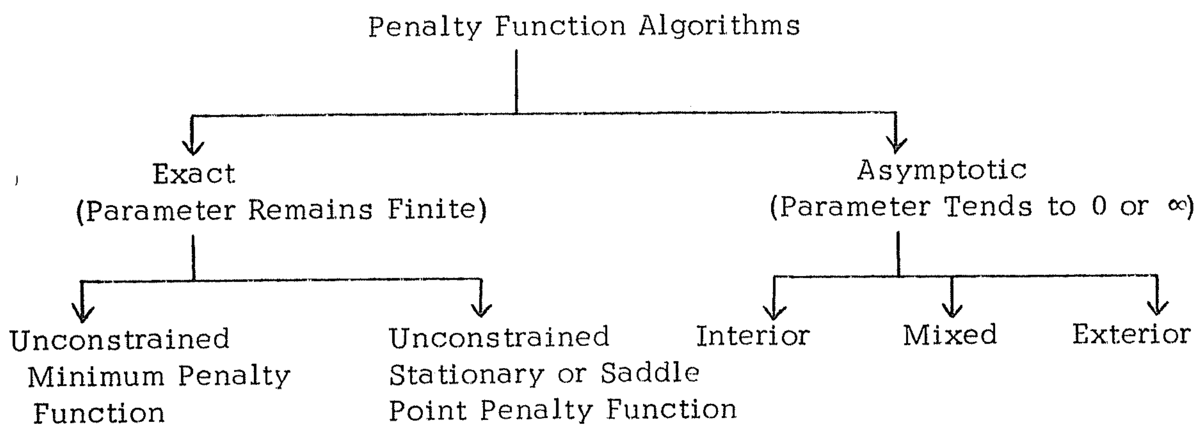
6.6 CONSTRAINED MINIMIZATION ALGORITHMS

Constrained optimization problems form the core of nonlinear programming and the typical programming problem is that given by (6.1). The difficulty of the problem depends on whether the functions f, g, h are linear, quadratic or neither. In particular we can distinguish the following classes of programming problems:



We shall concentrate here on algorithms for solving the rightmost problems given in the above table, that is, problems which are nonlinearly constrained, and, linearly constrained problems with a nonquadratic minimand. We shall give a number of algorithms for solving such problems.

We begin with penalty function algorithms. The basic idea here is to reduce the constrained problems to a sequence of "unconstrained" problems, the solutions of which approach a solution of the original problem. This is done by combining the constraints with the objective function in such a way that minimizing the combined "penalty function" penalizes constraint violation. Penalty function algorithms can be classified as follows:



Each of these methods has its advantages and disavantages which we shall mention after describing each algorithm.

6.6.1 Asymptotic Interior Penalty Algorithm [Fiacco & McCormick, 1968]
For any decreasing sequence of real positive numbers $\{r^i\}$ converging to zero, find solutions, x^i, of the problems minimize $P^i(x)$ subject to $g(x) < 0$ where

$$P^i(x) = f(x) - r^i \sum_{j=1}^{m} \frac{1}{g_j(x)}, \quad \text{for all } i$$

or

$$P^i(x) = f(x) - r^i \sum_{j=1}^{m} \log(-g_j(x)), \quad \text{for all } i$$

If the set $X^0 = \{x \mid g(x) < 0\}$ is not empty and $\inf_{x \in X^0} f(x) = \inf_{x \in X} f(x)$, where $X = \{x \mid g(x) \leqq 0\}$, and if f and g are lower semicontinuous on X, then every accumulation point of the sequence $\{x^i\}$ is a solution of minimize $f(x)$ subject to $g(x) \leqq 0$ (which is problem (6.1) but without equality constraints $h(x) = 0$). ∎

The main difficulties with the above algorithm are that equality constraints cannot be handled and the penalty parameter r^i must approach zero which creates conditioning problems [Lootsma, 1969, 1972, Murray, 1969]. To avoid the former difficulty we consider exterior penalty function algorithms which can handle both equalities and inequalities.

6.6.2 <u>Exterior Penalty Algorithm</u> [Fiacco & McCormick, 1968] For any increasing sequence of real positive numbers $\{r^i\}$ which tends to $+\infty$, find solutions x^i of the unconstrained problems:

$$\underset{x \in R^n}{\text{minimize}} \quad P^i(x) \text{ , where}$$

$$P^i(x) = f(x) + r^i \left[\sum_{j=1}^{m} (g_j(x)_+)^2 + \sum_{j=1}^{k} (h_j(x))^2 \right]$$

where $g_j(x)_+$ denotes $g_j(x)$ if $g_j(x) \geqq 0$, and 0 if $g_j(x) < 0$. If f and g are continuous on R^n and problem (6.1) possesses a solution, then every accumulation point $\bar{x}$ of the sequence $\{x^i\}$ is a solution of (6.1). ■

Again the exterior penalty algorithm suffers from conditioning difficulties as r^i approaches ∞. To overcome these difficulties exact penalty function methods will be discussed shortly. However the parameter r^i need not approach $+\infty$ if we are willing to accept some feasibility tolerance $\varepsilon > 0$ for the accumulation point $\bar{x}$, that is $\sum_{j=1}^{m} (g_j(\bar{x})_+)^2 + \sum_{j=1}^{k} (h_j(\bar{x}))^2 \leqq \varepsilon$. Then the sequence $\{r^i\}$ need not tend to $+\infty$ but rather $\lim_{i \to \infty} r^i > \frac{2\,|f(\hat{x}) - f(x^0)|}{\varepsilon}$, where $\hat{x}$ is any feasible point, that is $g(\hat{x}) \leqq 0$, $h(\hat{x}) = 0$, and x^0 is a solution of the first iteration $\min_{x \in R^n} P^0(x)$. Mixed penalty methods [Fiacco & McCormick, 1968] are a combination of exterior and interior penalty method. A typical mixed penalty function for problem (6.1) is given by

$$P^i(x) = f(x) - r^i \sum_{j\in J_1} \log(-g_j(x)) + \frac{1}{r^i}\left[\sum_{j\in J_2} (g_j(x)_+)^2 + \sum_{j=1}^{k} (h_j(x))^2\right]$$

where J_1 is any subset of indices of the inequality constraints for which there exist an interior, that is there exists an $\hat{x}$ such that $g_j(x) < 0$, for $j \in J_1$. J_2 is the set of indices of the remaining inequality constraints. The subproblems to be solved here are: minimize $P^i(x)$ subject to $g_j(x) < 0$, $j \in J_1$ and the sequence $\{r^i\}$ is a decreasing sequence tending to 0. If f, g and h are continuous and $-\infty < \underset{x\in X}{\text{infimum}}\ f(x) = \underset{x\in X^1 \cap X^2}{\text{infimum}}\ f(x)$ where

$X = \{x | g(x) \leqq 0, h(x) = 0\}$, $X^1 = \{x | g_j(x) < 0,\ j \in J_1\}$ and

$X^2 = \{x | g_j(x) \leqq 0,\ j \in J_2, h(x) = 0\}$, then each accumulation point of the sequence $\{x^i\}$ solves (6.1).

More recently there have been attempts at inventing penalty functions in which the parameter remains finite. Such penalty functions are called exact. Unfortunately exact penalty functions are in general nondifferentiable [Zangwill 1967, Pietrzykowski, 1969, Howe, 1973] or only locally differentiable [Fletcher, 1972], or an unconstrained stationary or saddle point must be found rather than an unconstrained minimum [Rockafellar 1971, 1972a, 1972b, 1973, Arrow et al, 1971, Mangasarian, 1973].

6.6.3 Exact Minimum Penalty Algorithm [Zangwill, 1967, Pietrzykowski, 1969, Howe, 1973] For an increasing sequence of real numbers $\{r^i\}$ find unconstrained local (global) minima x^i of $P(x,r^i)$ where $P(x,r)$ is defined in 6.2.5. Stop when x^i is feasible, that is $g(x^i) \leqq 0$ and $h(x^i) = 0$. x^i is a local (global) solution of (6.1). ■

Note the advantage of this algorithm over asymptotic penalty methods is the finiteness of r^i .

The main disadvantage is that $P(x,r)$ is in general piecewise differentiable at most. Special methods must be used to minimize it [Polyak, 1969a].

We turn now to the final type of penalty function methods namely the exact stationary point penalty algorithm.

6.6.4 Exact Stationary Point Penalty Algorithm For some positive number $r > 0$ find a root $(\bar{x},\bar{y},\bar{z})$ of the m+n+k equalities (6.9). The vector $\bar{x}$ and the vectors $\bar{u}$ and $\bar{v}$ given by (6.10) satisfy the first order Kuhn Tucker conditions (6.3). If in addition f and g are convex at $\bar{x}$, and h is linear then $\bar{x}$ is a global solution of (6.1). ∎

Specific methods for solving the equalities (6.9) are given in [Mangasarian, 1973]. For a compendium of methods for solving nonlinear equations see [Ortega & Rheinboldt, 1970].

We next turn to the method of feasible directions. The basic idea here is to find a direction which simultaneously decreases the minimand and at the same time remains feasible. A tolerance ε^i to prevent jamming or zig-zagging must be introduced in the algorithm [Zoutendijk, 1960]. Note that in this method g can be nonlinear but h must be linear.

6.6.5 <u>Feasible Directions Algorithm</u> [Zoutendijk 1960, Zangwill, 1969]

Start with an $x^0 \in X = \{x \mid g(x) \leqq 0, h(x) = 0\}$, where h is linear, and some fixed positive number $\varepsilon^0 > 0$. Having x^i, ε^i determine $x^{i+1}, \varepsilon^{i+1}$ as follows. Solve the linear program

$$\underset{\delta,q}{\text{minimize}} \ \{\delta \mid \nabla f(x^i)q \leqq \delta, \nabla g_j(x^i)q \leqq \delta, \ j \in I_N(x^i, \varepsilon^i)$$

$$\nabla g_j(x^i)q \leqq 0, \ \ j \in I_L(x^i, \varepsilon^i), \nabla h(x^i)q = 0\}$$

where

$$I_N(x^i, \varepsilon^i) = \{j \mid -\varepsilon^i \leqq g_j(x^i) \leqq 0, \text{ and } g_j \text{ is nonlinear}\}$$

$$I_L(x^i, \varepsilon^i) = \{j \mid -\varepsilon^i \leqq g_j(x^i) \leqq 0, \text{ and } g_j \text{ is linear}\}$$

Call a solution of the linear program δ^i, q^i. If $\delta^i = 0$ stop, else set

$$\varepsilon^{i+1} = \begin{cases} \varepsilon^i & \text{if } \varepsilon^i \leqq -\delta^i \\ \frac{\varepsilon^i}{2} & \text{if } \varepsilon^i > -\delta^i \end{cases}$$

Set $p^i = \nu^i q^i$ where $\nu^i = \max\{1, \frac{1}{2}, \frac{1}{4}, \ldots\}$ such that $x^i + \mu p^i \in X$ for all $\mu \in [0,1]$. Determine $x^{i+1} = x^i + \lambda^i p^i$ such that $f(x^i + \lambda^i p^i) = \underset{0 \leqq \lambda \leqq 1}{\text{minimum}} f(x^i + \lambda p^i)$ or $\lambda^i = \max\{1, \frac{1}{2}, \frac{1}{4}, \ldots\}$ such that $f(x^i) - f(x^i + \lambda^i p^i) \geqq -\frac{1}{4} \nabla f(x^i) p^i$. If f has continuous first partial derivatives, g has Lipschitz continous partial derivatives and h is linear, then either the sequence $\{x^i\}$ terminates at a stationary point or every accumulation point $\bar{x}$ in stationary, that is there exist no feasible direction p at $\bar{x}$ satisfying $\nabla f(\bar{x})p < 0$, $\nabla g_j(\bar{x})p < 0$, $j \in \{j \mid g_j(\bar{x}) = 0\}$ and $\nabla h(\bar{x})p = 0$. If in addition f and g are convex on R^n and there exists some $\tilde{x}$ such that $g(\tilde{x}) < 0$ and $h(\tilde{x}) = 0$ then $\bar{x}$ solves (6.1). ∎

We consider next a gradient projection algorithm which is useful for linearly constrained problems. The algorithm we present here is a variation of the Levitin-Polyak algorithm [Levitin & Polyak, 1966] which in our version here includes provisions for considering only a subset of the constraints and also provisions for introducing a matrix which can play the role of the inverse of the Hessian of the objective function or an approximation thereof.

6.6.6 Gradient Projection Algorithm [Levitin & Polyak, 1966]
Start with any $x^0 \in X = \{x | g(x) \leqq 0, h(x) = 0\}$, where g is convex and h is linear, and some fixed number $\varepsilon^0 > 0$. Having x^i, ε^i determine $x^{i+1}, \varepsilon^{i+1}$ as follows. Solve the quadratic programming problem

$$\underset{y \in X^i}{\text{minimize}} \; \| x^i - H(x^i)^{-1} \nabla f(x^i) - y \|^2_{H(x^i)}$$

where $\|z\|^2_H = zHz$, $H(x^i)$ is any continuous symmetric, positive definite matrix and $X^i = \{x | g_j(x) \leqq 0, \; j \in I(x^i, \varepsilon^i), \; h(x) = 0\}$, $I(x^i, \varepsilon^i) = \{j \mid -\varepsilon^i \leqq g_j(x^i) \leqq 0\}$. Denote the solution of this quadratic program by y^i. If $y^i = x^i$ stop, else take $\varepsilon^{i+1} \geqq \hat{\varepsilon} > 0$ where $\hat{\varepsilon}$ is an arbitrary positive number. Set $p^i = \nu^i(y^i - x^i)$ where $\nu^i = \max\{1, \frac{1}{2}, \frac{1}{4}, \ldots\}$ such that $x^i + \mu p^i \in X$ for all $\mu \in [0,1]$. Determine $x^{i+1} = x^i + \lambda^i p^i$ such that $f(x^i + \lambda^i p^i) = \underset{0 \leqq \lambda \leqq 1}{\text{minimum}} \; f(x^i + \lambda p^i)$ or $\lambda^i = \max\{1, \frac{1}{2}, \frac{1}{4}, \ldots\}$ such that

$f(x^i) - f(x^i + \lambda^i p^i) \geqq -\frac{1}{4}\nabla f(x^i)p^i$. If f is convex and has continuous first partial derivatives on R^n , g has Lipschitz continuous gradients on R^n and there exists an $\tilde{x} \in X$ such that $h(\tilde{x}) = 0$ and $g(\tilde{x}) < 0$ and $M_1 \|z\|^2 \leqq zH(x^i)z \leqq M_2 \|z\|^2$ for all i , all $z \in R^n$ and some $M_2 \geqq M_1 > 0$, then either the sequence $\{x^i\}$ terminates at a solution of problem (6.1) or every accumulation point solves (6.1). ∎

We observe that most of the work in the above algorithm is contained in solving a quadratic program for determining y^i . For linearly constrained problems this can be quickly performed by principal pivoting or related methods [Lemke, 1962, Cottle & Dantzig, 1968] for which efficient computer codes are available [Teorey, 1972].

We terminate by giving a Newton algorithm for nonlinearly constrained problems for which local quadratic convergence can be established. The algorithm was originally proposed in [Wilson, 1963] and its quadratic convergence rate was established by Robinson [Robinson, 1973]. See [Robinson, 1972] for a related algorithm.

6.6.7 <u>Newton Algorithm for Nonlinearly Constrained Problems</u>

[Wilson, 1963, Robinson, 1973] Start with any $(x^0, u^0, v^0) \in R^n \times R^m \times R^k$.

Having (x^i, u^i, v^i) determine $(x^{i+1}, u^{i+1}, v^{i+1})$ to be the closest solution to (x^i, u^i, v^i) of the following linearly constrained quadratic programming problem

$$\min_p \quad \nabla f(x^i)p + \frac{1}{2} p \nabla_{11} L(x^i, u^i, v^i)p$$

$$\text{subject to} \quad g(x^i) + \nabla g(x^i)p \leqq 0$$

$$h(x^i) + \nabla h(x^i)p = 0$$

where L is the Lagrangian defined in (6.5). Call the solution p^i, the optimal Lagrange multipliers u^{i+1}, v^{i+1}, and set $x^{i+1} = x^i + p^i$. If (x^0, u^0, v^0) is sufficiently close to a point $(\bar{x}, \bar{u}, \bar{v})$ which satisfies the second order sufficiency condition of 6.2.3, and at which strict complementarity holds that is $\bar{u}_i > 0$ for $g_i(\bar{x}) = 0$, and $\nabla h_j(\bar{x}), j=1, \ldots, k$ and $\nabla g_i(\bar{x})$ for $i \in \{i \mid g_i(\bar{x}) = 0\}$ are linearly independent, and f, g and h are twice continuously differentiable around $\bar{x}$, then the sequence $\{x^i, u^i, v^i\}$ converges quadratically to $(\bar{x}, \bar{u}, \bar{v})$, that is

$$\|(x^i, u^i, v^i) - (\bar{x}, \bar{u}, \bar{v})\| \leqq \sigma\left(\frac{1}{2}\right)^{2^i}$$

where σ is some constant. ∎

REFERENCES

Arrow, K.J., F.J. Gould & S.M. Howe, "A General Saddle Point Result for Constrained Optimization," Department of Statistics, Chapel Hill, North Carolina, Mimeo Series No. 774, September 1971.

Broyden, C.G., J.E. Dennis, Jr. & J.J. Moré, "On the Local and Superlinear Convergence of Quasi-Newton Methods", Computer Science TR 72-137, Cornell University, August 1972.

Buys, J.D., "Dual Algorithms for Constrained Optimization Problems", Doctorate Dissertation, University of Leiden, June 1972.

Cohen, A., "Rate of Convergence of Several Conjugate Gradient Algorithms", SIAM J. Numer. Anal. Vol. 9, No. 2 (1972) pp. 248-259.

Cottle, R.W. & G. B. Dantzig, "Complementary Pivot Theory of Mathematical Programming", Linear Algebra & Appl., Vol. 1 (1968), pp. 103-125.

Danilin, Y.M., "Estimation of the Efficiency of an Absolute-Minimum-Finding Algorithm", USSR Computational Mathematics and Mathematical Physics, (English Translation) Vol. 11, No. 4 (1971), pp. 261-267.

Dantzig, G.B., "Linear Programming and Extensions", Princeton University Press. Princeton, N. J., 1963.

Dantzig, G.B. & A.F. Veinott (eds.), "Mathematics of the Decision Sciences", Parts I & II, Amer. Math. Soc., Providence, R.I., 1968.

Evans, J.P., F.J. Gould & J.W. Tolle, "Exact Penalty Functions in Nonlinear Programming", Mathematical Programming, Vol. 4, No. 1 (1973) pp. 72-97.

Fiacco, A.V. & G.P. McCormick, "Nonlinear Programming: Sequential Unconstrained Minimization Techniques", Wiley, New York, 1968.

Fletcher, R., "An Exact Penalty Function for Nonlinear Programming with Inequalities", T. P. 478 Atomic Energy Research Establishment. Harwell, February 1972.

Fletcher, R. & M.J.D. Powell, "A Rapidly Convergent Descent Method for Minimization", Computer J. Vol. 6 (1963) pp. 163-168.

Fletcher, R. & C.M. Reeves, "Function Minimization by Conjugate Gradients", Computer J., Vol. 7 (1964) pp. 149-154 .

Fox, R.L., "Optimization Methods for Engineering Design", Addison-Wesley, Reading, Mass., 1971.

Howe, S., "New Conditions for Exactness for a Simple Penalty Function", SIAM J. Control, Vol. 11, No. 2 (1973), pp. 378-381.

John, F., "Extremum Problems with Inequalities as Subsidiary Conditions", in K.O. Friedrichs, O.E. Neugebauer & J.J. Stoker (eds.), "Studies and Essays: Courant Anniversary Volume", pp. 187-204. Interscience, New York, 1948.

Kowalik, J. & M.R. Osborne, "Methods for Unconstrained Optimization Problems", Elsevier, New York, 1968.

Kuhn, H.W. and A.W. Tucker, "Nonlinear Programming" in J. Neyman (ed.), "Proceedings of the Second Berkeley Symposium in Mathematical Statistics and Probability", pp. 481-492, University of California Press, Berkeley, Calif., 1951.

Künzi, H.P. & W. Oettli, "Nichtlineare Optiemierung: Neuere Verfahren Bibliographie", Lecture Notes in Operations Research and Mathematical Systems No. 16, Springer, Berlin, 1969. (Most comprehensive bibliography on nonlinear programming up to 1969)

Lemke, C.E., "A Method of Solution for Quadratic Programs", Management Science, Vol. 8 (1972), pp. 442-455.

Levitin, E.S. & B.T. Polyak, "Constrained Minimization Methods", USSR Computational Mathematics and Mathematical Physics (English Translation) Vol. 6, No. 5 (1966) pp. 1-50.

Lootsma, F.A., "Hessian Matrices of Penalty Functions for Solving Constrained Optimization Problems", Philips Res. Rept. 24 (1969) pp. 322-331.

Lootsma, F.A. (ed.), "Numerical Methods for Non-linear Optimization", Academic Press, London, 1972.

Luenberger, D.G., "Introduction to Linear and Nonlinear Programming", Addison-Wesley, Reading, Mass., 1973.

McCormick, G.P. & K. Ritter, "On the Convergence and Rate of Convergence of the Conjugate Gradient Method", Math. Res. Center Report #1118, University of Wisconsin, Madison, December 1970. To appear in J. Optimization Theory & Applications under the title of "Alternate Proofs of the Convergence Properties of the Conjugate Gradient Method."

McCormick, G.P. & K. Ritter, "Methods of Conjugate Directions Versus Quasi-Newton Methods", Math. Prog. Vol. 3, No. 1 (1972), pp. 101-116.

Mangasarian, O.L., "Nonlinear Programming", McGraw-Hill, New York, 1969.

Mangasarian, O.L., "Techniques of Optimization", J. of Eng'g for Industry, Vol. 94, Series B, No. 2 (1972), pp. 365-372.

Mangasarian, O.L., "Unconstrained Lagrangians in Nonlinear Programming", Computer Sciences Technical Report #174, University of Wisconsin, Madison, March 1973.

Mangasarian, O.L. & J. Ponstein, "Minmax and Duality in Nonlinear Programming", J. Math. Anal. Appl., Vol. 11 (1965) pp. 504-518.

Murray, W., "Hessian Matrices of Barrier and Penalty Functions Arising in Optimization", Nat. Phys. Lab. Rept. Ma 77, 1969.

Ortega, J.M. & W.C. Rheinboldt, "Iterative Solution of Nonlinear Equations in Several Variables", Academic Press, New York, 1970.

Ostrowski, A., "Solution of Equations and Systems of Equations", Academic Press, New York, 1966 (Second Edition)

Pietrzykowski, T., "An Exact Potential Method for Constrained Maxima", SIAM J. Num. Anal. Vol. 6, No. 2 (1969), pp. 299-304.

Polak, E. & G. Ribière, "Note sur la Convergence de Méthodes de Directions Conjuguées", Rev. Fr. Inform. Rech. Opér., Vol. 16-R1 (1969) pp. 35-43.

Polak, E., "Computational Methods in Optimization: A Unified Approach", Academic Press, New York, 1971.

Polyak, B.T., "Minimization of Non-Smooth Functionals", USSR Computational Mathematics and Mathematical Physics (English Translation) Vol. 9, No. 3 (1969a) pp. 14-29.

Polyak, B.T., "The Method of Conjugate Gradients in Extremum Problems", USSR Computational Mathematics and Mathematical Physics (English Translation) Vol. 9, No. 4 (1969b) pp. 94-112.

Powell, M.J.D., "On the Convergence of the Variable Metric Algorithm", Inst. of Math. & Its Appl., Vol. 7 (1971), pp. 21-36.

Robinson, S.M., "A Quadratically-Convergent Algorithm for General Nonlinear Programming Problems", Mathematical Programming, Vol. 3, No. 2 (1972), pp. 145-156.

Robinson, S. M., "Perturbed Kuhn-Tucker Points and Rates of Convergence for a Class of Nonlinear Programming Algorithms", University of Wisconsin, Madison, Mathematics Research Center Report #1298, January 1973.

Rockafellar, R.T., "Convex Analysis", Princeton University Press, Princeton, New Jersey, 1970.

Rockafellar, R.T., "New Applications of Duality in Convex Programming", Proceedings of the Fourth Conference on Probability, Brasov, Romania, 1971.

Rockafellar, R.T., "A Dual Approach to Solving Nonlinear Programming Problems by Unconstrained Optimization", Mathematics Department, University of Washington, Seattle, December 1972a.

Rockafellar, R.T., "The Multiplier Method of Hestenes and Powell Applied to Convex Programming", Mathematics Department, University of Washington, Seattle, December 1972b.

Rockafellar, R.T., "Augmented Lagrange Multiplier Functions and Duality in Nonconvex Programming", Mathematics Department, University of Washington, Seattle, January 1973.

Shampine, L.F. & R.C. Allen, Jr., "Numerical Computing: An Introduction", Saunders, Philadelphia, 1973.

Stark, R.M. & R.L. Nichols, "Mathematical Foundations for Design-Civil Engineering Systems", McGraw-Hill, New York, 1972.

Teorey, T., "QUADPR, Quadratic Programming Subroutine", University of Wisconsin, Madison Academic Computing Center, June 1972.

Wilde, D.J., "Optimum Seeking Methods", Prentice-Hall, Englewood Cliffs, N. J., 1964.

Wilson, R.B., "A Simplicial Method for Convex Programming", Ph.D. Dissertation, Harvard University, 1963.

Wolfe, P., "A Duality Theorem for Nonlinear Programming", Quarterly of Applied Mathematics, Vol. 19 (1961) pp. 239-244.

Zangwill, W.I., "Nonlinear Programming via Penalty Functions", Management Science, Vol. 13, No. 5 (1967), pp. 344-358.

Zangwill, W.I., "Nonlinear Programming: A Unified Approach", Prentice-Hall, Englewood Cliffs, N.J., 1969.

Zoutendijk, G., "Methods of Feasible Directions", Elsevier, Amsterdam, 1960.

Zoutendijk, G., "Nonlinear Programming, Computational Methods", in J. Abadie (ed.), "Integer and Nonlinear Programming", North Holland, Amsterdam (1970), pp. 37-86.

BIBLIOGRAPHIC DATA SHEET	1. Report No. WIS-CS-185-73	2.	3. Recipient's Accession No.
4. Title and Subtitle			5. Report Date July 1973
			6.
7. Author(s) O. L. Mangasarian			8. Performing Organization Rept. No. 185
9. Performing Organization Name and Address The University of Wisconsin Computer Sciences Department, 1210 W. Dayton Street Madison, Wisconsin 53706			10. Project/Task/Work Unit No.
			11. Contract/Grant No. NSF GJ-35292
12. Sponsoring Organization Name and Address National Science Foundation Washington, D.C. 20550			13. Type of Report & Period Covered
			14.

15. Supplementary Notes

16. Abstracts

A survey of nonlinear programming theory and computational algorithms as given. Subjects covered are: optimality conditions, duality theory, unconstrained and constrained optimization algorithms.

17. Key Words and Document Analysis. 17a. Descriptors

Optimization, Nonlinear Programming, Mathematical Programming, Programming.

17b. Identifiers/Open-Ended Terms

17c. COSATI Field/Group

18. Availability Statement	19. Security Class (This Report) UNCLASSIFIED	21. No. of Pages 41
	20. Security Class (This Page) UNCLASSIFIED	22. Price