SUPERLINEARLY CONVERGENT VARIABLE METRIC ALGORITHMS
FOR GENERAL NONLINEAR PROGRAMMING PROBLEMS†

Shih-Ping Han

TR 75-233

March 1975

Department of Computer Science
Cornell University
Ithaca, NY 14853

†This research was supported in part by the National Science Foundation under Grant ENG 75-10486.

SUPERLINEARLY CONVERGENT VARIABLE METRIC ALGORITHMS FOR GENERAL NONLINEAR PROGRAMMING PROBLEMS[†]

Shih-Ping Han

ABSTRACT:

In this paper variable metric algorithms are extended to solve general nonlinear programming problems. In the algorithm we iteratively solve a linearly constrained quadratic program which contains an estimate of the Hessian of the Lagrangian. We suggest the variable metric updates for the estimates of the Hessians and justify our suggestion by showing that, when some well known update such as the Davidon-Fletcher-Powell update are so employed, the algorithm converges locally with a superlinear rate. Our algorithm is in a sense a natural extension of the variable metric algorithm to the constrained optimization and this extension offers us not only a class of effective algorithms in nonlinear programming but also a unified treatment of constrained and unconstrained optimization in the variable metric approach.

† This research was supported in part by the National Science Foundation under Grant ENG 75-10486.

SUPERLINEARLY CONVERGENT VARIABLE METRIC ALGORITHMS FOR GENERAL NONLINEAR PROGRAMMING PROBLEMS

Shih-Ping Han

1. Introduction

The nonlinear programming problem to be considered in this paper is defined as

$$(P)\qquad \begin{array}{l} \text{minimize } f(x) \\ \text{subject to the constraints } g_i(x) \leq 0 \quad i=1,\ldots m, \end{array}$$

where $f, g_1,\ldots,g_m$ denote real-valued functions of a vector x in the n-dimensional Euclidean space R^n. For convenience we shall restrict ourselves to problems with inequality constraints only. The inclusion of equality constraints causes no difficulties and all the results go through with minor modifications.

In developing methods for solving constrained optimization problems much can be learned from the established methods for unconstrained optimization problems, which can be viewed as a special case of Problem P with m=0. We usually can obtain efficient methods by extending successful unconstrained optimization methods to the constrained case. In this paper some variable metric methods including the Davidon-Fletcher-Powell method are extended to solve Problem P and we establish a local superlinear convergence for these methods.

In Section 2 we state the algorithm and compare it with the related work in the nonlinear programming literature. We establish some general local convergence theorems for the algorithm in Section and discuss the superlinear rate of convergence in Section 4. In Section 5 some variable metric updates are incorporated into the algorithm for updating the estimates of the Hessian of the Lagrangian in the quadratic programming subproblems; by the results of Section 3 and 4, we show that the resulting algorithms converge locally with superlinear rates. Some comments and computational results are contained in Section 6.

We note here that all vectors are column vectors and a row vector will be indicated by superscript T. For convenience a column vector in R^{n+m} is sometimes written as (x,u) instead of $\begin{bmatrix} x \\ u \end{bmatrix}$. We use x_i to denote the i-th component of vector x. Superscripts are used to denote different vectors; i.e., x^1 and x^2. To avoid some cumbersome constants we restrict ourselves to the ℓ_2 vector and operator norm and denote it by $||\cdot||$. An ε-neighborhood $N(x,\varepsilon)$ of a point x in R^n is the set $N(x,\varepsilon) = \{y \in R^n : ||y-x|| < \varepsilon\}$, and $\bar{N}(x,\varepsilon)$ is its closure.

2. The algorithm

The class of algorithms to be considered are for finding a Kuhn-Tucker point of Problem P. The algorithm constructs a sequence of (n+m)-vectors $\{(x^k,u^k)\}$ which are estimates of a Kuhn-Tucker pair (x^*,u^*) of Problem P. This is done by solving a sequence of linearly constrained quadratic programming subproblems which can be effectively solved by the existing algorithms [1,5,6,28]. Each subproblem contains an estimate H_k of the Hessian of the Lagrangian of Problem P, and the matrix H_k can be updated by a variety of variable metric schemes which are well known in unconstrained optimization. Before the statement of the algorithm we first define the following quadratic programming problem $Q(x,H)$:

$$\min_{s} \quad \nabla f(x)^T s + \frac{1}{2} s^T H s$$

$$\text{s.t. } g_i(x) + \nabla g_i(x)^T s \leq 0, \quad i=1,\ldots m,$$

which can be associated with any x in R^n and any nxn matrix H.

Definition 2.1: A vector $\bar{z} = (x+\bar{s},\bar{u})$ in R^{n+m} is a z-solution of $Q(x,H)$ if $(\bar{s},\bar{u})$ is a Kuhn-Tucker pair of $Q(x,H)$. □

Now we can state the algorithm as follows:

Algorithm

Step 1. Start with an estimate $z^0 = (x^0,u^0)$ of a Kuhn-Tucker pair of problem P and an estimate H_0 of the Hessian of the Lagrangian evaluated at that Kuhn-Tucker pair.

Step 2. Set k = 0.

Step 3. Find a z-solution of $Q(x^k,H_k)$ and call this z-solution $z^{k+1} = (x^{k+1},u^{k+1})$. If there are more than one such z-solutions, choose one which is closest to z^k.

Step 4. If $z^{k+1} = (x^{k+1},u^{k+1})$ satisfies a prescribed convergence criterion, stop; otherwise, go to step 5.

Step 5. Update H_{k+1} by some updating scheme, then set k = k+1 and go to step 3. □

We propose to use variable metric schemes to update the matrix H_{k+1} from the matrix H_k and the vectors $s^k = x^{k+1} - x^k$ and $y^k = \nabla_x L(x^{k+1},u^{k+1}) - \nabla_x L(x^k,u^{k+1})$, where $L(x,u) = f(x) + u^T g(x)$ is the Lagrangian of Problem P; the Davidon-Fletcher-Powell update [7,11] and an update by Powell [23,24] will be so utilized in our algorithm and shown to possess local and superlinear convergence properties. It is noted here that the algorithm is just the variable metric algorithm in the degenerated case m=0 and therefore can be viewed as its natural extension to the constrained case.

The algorithm was first studied by Wilson [29] with $H_k = \nabla_{xx}L(x^k,u^k)$. Though converging quadratically, Wilson's algorithm, like Newton's method in unconstrained optimization, is expensive in computing second derivatives. Robinson [25] proposed a modified Wilson's algorithm which avoids the need for second derivatives and at the same time preserves the quadratic rate of convergence; however, in this algorithm we have to solve more difficult general linearly constrained minimization subproblems instead of quadratic programming subproblems. An approach adopted by Garcia and Mangasarian [12,13] is to update H_k as an estimate of $\nabla_{xx}L(x^k,u^k)$. The main difference between our algorithms and Garcia's is in the way we update matrices. In Garcia's algorithms an $(n+m)\times(n+m)$ matrix is updated in each iteration to approximate the matrix $\nabla_z^2 L(z^k)$ and the upper left $n\times n$ submatrix is used in the quadratic programming subproblems. This is wasteful especially if the number of constraints is very large. In our algorithms we update only $n\times n$ matrices which directly approximate the matrix $\nabla_{xx}L(z^*)$. To guarantee superlinear rates of convergence for Garcia's updates we need a very stringent condition called "uniform linear independence", i.e., each n consecutively generated vectors are, in some sense, uniformly linearly independent. Strictly speaking, this condition cannot be assumed beforehand and therefore superlinear rates of convergence have not really been guaranteed for his updates. However this condition is never needed in establishing superlinear rates for our algorithms. We would also like to note here that some other different extensions of the variable metric algorithm to constrained optimization have been studied by Goldfarb [15] and Gill and Murry [14] for the linearly constrained case.

3. Local Convergence Theorems

In this section we shall present sufficient conditions for ensuring the local convergence of our algorithm; such conditions turn out to be satisfied by several updates. To begin with, we introduce some definitions and notation. For any $\tilde{z} = (\tilde{x},\tilde{u})$ in R^{n+m} and any $n \times n$ matrix H we define the function $F(\tilde{z},H,\cdot): R^{n+m} \to R^{n+m}$, as follows

$$F(\tilde{z},H,z) = \begin{bmatrix} \nabla f(\tilde{x}) + \nabla g(\tilde{x})u + H(x-\tilde{x}) \\ u_1(g_1(\tilde{x}) + \nabla g_1(\tilde{x})^T(x-\tilde{x})) \\ \cdot \\ \cdot \\ \cdot \\ u_m(g_m(\tilde{x}) + \nabla g_m(\tilde{x})^T(x-\tilde{x})) \end{bmatrix}. \tag{3.1}$$

where $z = (x,u)$. If H is symmetric then the equalities $F(\tilde{z},H,z) = 0$ are satisfied by a z-solution to the quadratic program $Q(\tilde{x},H)$. Let $z^* = (x^*,u^*)$ be a Kuhn-Tucker pair of Problem P; the Jacobian matrix of the function $F(z^*,H,z)$ with respect to z evaluated at z^* is denoted by

$$(3.2)\quad D_H = \nabla_z F(z^*,H,z)\Big|_{z^*} = \begin{bmatrix} H & \nabla g(x^*) \\ u_1^*\nabla g_1(x^*)^T & \cdot \\ \cdot & \cdot \\ \cdot & \operatorname{diag}(g_i(x^*)) \\ & i=1,\ldots m \\ u_m^*\nabla g_m(x^*)^T & \cdot \\ & \cdot \end{bmatrix}.$$

When $H = \nabla_{xx}L(z^*)$ the nonsingularity of D_H is essential for establishing our convergence theorems. To ensure this we use the following condition which was first studied by Fiacco and McCormick [10] and has been called "the Jacobian uniqueness condition" [19]. It should be noted that any condition which implies the nonsingularity of D_H with $H = \nabla_{xx}L(z^*)$ will work equally well.

Definition 3.2: A Kuhn-Tucker pair (x^*,u^*) of Problem P satisfies the Jacobian uniqueness condition if the following conditions are simultaneously satisfied

(i) $u_i^* > 0$ if $i \in I(x^*) = \{j : g_j(x^*) = 0\}$

(ii) $\{\nabla g_i(x^*) : i \in I(x^*)\}$ are linearly independent

(iii) For any $y \in R^n$, $y \neq 0$, such that $\nabla g_i(x^*)^T y = 0$ $i \in I(x^*)$, it follows that $y^T \nabla_{xx} L(x^*,u^*) y > 0$. □

In the following discussion $f \in LC^2[x]$ will mean that the function f has a second derivative which is Lipschitz continuous at x. Now we state a lemma which will be used later. This lemma can be established by using the mean value theorem [21, p. 78] and its proof can be found in [16].

Lemma 3.3: If f and g_i $(i=1,\ldots m) \in LC^2[x]$, then there exists a neighborhood $N(x^*,\varepsilon)$ and two positive numbers $\bar{K}$ and $\tilde{K}$ such that for any $\bar{x}$ and $\tilde{x}$ in $N(x^*,\varepsilon)$ and any $\hat{u}$ in R^m, we have

$$(3.3)\quad \begin{aligned} &\|\nabla_x L(\tilde{x},\hat{u}) - \nabla_x L(\bar{x},\hat{u}) - \nabla_{xx} L(x^*,u^*)(\tilde{x}-\bar{x})\| \\ &\leq (\bar{K}\max\{\|\tilde{x}-x^*\|,\|\bar{x}-x^*\|\} + \tilde{K}\|\hat{u}-u^*\|)\|\tilde{x}-\bar{x}\|. \square \end{aligned}$$

Next we provide an estimate of the distance between a z-solution of the quadratic program $Q(\tilde{x},H)$ and a Kuhn-Tucker pair $z^* = (x^*,u^*)$ of Problem P.

Lemma 3.4: Let $z^* = (x^*,u^*) \in R^{n+m}$ and f and $g_i (i=1,\ldots m)$ have continuous second derivatives. If H is an $n\times n$ matrix such that D_H is nonsingular and $F(z^*,H,z^*) = 0$, then there exists a neighborhood $N(z^*,\bar{\varepsilon})$ such that for any $\tilde{z}$ in $N(z^*,\bar{\varepsilon})$ the function

$$T_{\tilde{z},H}(z) = z - D_H^{-1} F(\tilde{z},H,z)$$

is a contraction in $N(z^*,\bar{\varepsilon})$. Furthermore if $\|D_H^{-1}\| \leq \tau$ and $\tilde{z}$ is close enough to z^* such that $\|F(\tilde{z},H,z^*)\| \leq \frac{\bar{\varepsilon}}{2\tau}$, then $T_{\tilde{z},H}$ has a unique fixed point $\bar{z}$ in $N(z^*,\bar{\varepsilon})$ and $\|\bar{z}-z^*\| \leq 2\tau\|F(\tilde{z},H,z^*)\|$.

Proof: Let $\|D_H^{-1}\| \leq \tau$; we choose δ such that $\tau\delta<\frac{1}{2}$ and $\bar{\varepsilon}$ such that for all z and $\tilde{z}$ in $N(z^*,\bar{\varepsilon})$ we have

$$\left\| \left[\begin{array}{cc} 0 & \nabla g(\tilde{x}) - \nabla g(x^*) \\ u_1 \nabla g_1(\tilde{x})^T - u_1 \nabla g_1(x^*)^T & \ddots \\ \vdots & \underset{i=1,\ldots m}{\mathrm{diag}(g_i(\tilde{x}) + \nabla g_i(\tilde{x})^T(x-\tilde{x}) - g_i(x^*))} \\ u_m \nabla g_m(\tilde{x})^T - u_m^* \nabla g_m(x^*)^T & \ddots \end{array} \right] \right\| \leq \delta$$

Hence

$$||\nabla_z T_{\tilde{z},H}(z)|| = ||I - D_H^{-1} \nabla_z F(\tilde{z},H,z)||$$

$$\leq ||D_H^{-1}||\ ||D_H - \nabla_z F(\tilde{z},H,z\ ||\ \leq \tau\delta < \frac{1}{2}\ .$$

Therefore $T_{\tilde{z},H}$ is a contraction in $N(z^*,\bar{\varepsilon})$. If z is close enough to z^* such that $||F(\tilde{z},H,z^*)|| \leq \frac{\bar{\varepsilon}}{2\tau}$, then

$$||T_{\tilde{z},H}(z^*) - z^*|| \leq ||D_H^{-1}||\ ||F(\tilde{z},H,z^*)|| \leq (1-\tfrac{1}{2})\ \bar{\varepsilon}.$$

By the contradiction mapping principle [18, p. 28] there exists a unique fixed point $\bar{z}$ of $T_{\tilde{z},H}$ in $N(z^*,\bar{\varepsilon})$ and

$$||\bar{z} - z^*|| \leq 2\tau\ ||F(\tilde{z},H,z^*)||. \ \square$$

Now we are ready to prove the following key theorem which establishes a general condition for the local convergence of our algorithm.

<u>Theorem 3.5</u>: Let f and $g_i (i=1,\ldots m) \in LC^2[x^*]$. If a Kuhn-Tucker pair $z^* = (x^*,u^*)$ of Problem P satisfies the Jacobian uniqueness condition then for any $r \in (0,1)$ there exist two positive numbers $\varepsilon(r)$ and $\delta(r)$ such that if $||\tilde{z} - z^*|| \leq \varepsilon(r)$ and $||H - \nabla_{xx}L(z^*)|| \leq \delta(r)$ then a closest z-solution $\hat{z} = (\hat{x},\hat{u})$ of $Q(\tilde{x},H)$ to $\tilde{z}$ exists and

$||\hat{z} - z^*|| \leq r||\tilde{z} - z^*||.$

<u>Proof</u>: Let $r \in (0,1)$ be given and D^* be the Jacobian of the equalities of the Kuhn-Tucker conditions evaluated at z^*; that is,

$$D^* = \begin{bmatrix} \nabla_{xx}L(z^*) & \nabla g(x^*) \\ u_1^*\nabla g_1(x^*)^T & \\ \vdots & \operatorname{diag}(g_i(x^*)) \\ & i=1,\ldots m \\ u_m^*\nabla g_m(x^*)^T & \end{bmatrix}$$

By the Jacobian uniqueness condition it follows that D^* is nonsingular [10, p. 80]. We set

$$\text{(3.4)} \qquad \lambda > ||D^{*-1}||, \quad \tau = \lambda/(1-r)$$

and choose $\bar{\varepsilon} > 0$ such that for all z and $\tilde{z}$ in $N(z^*,\bar{\varepsilon})$ the following conditions hold:

(3.5) (a)

$$\left\| \begin{bmatrix} 0 & \nabla g(\tilde{x}) - \nabla g(x^*) \\ u_1\nabla g_1(\tilde{x})^T - u_1^*\nabla g_1(x^*)^T & \\ \vdots & \operatorname{diag}(g_i(\tilde{x}) + \nabla g_i(\tilde{x})^T(x-\tilde{x}) - g_i(x^*)) \\ u_m\nabla g_m(x)^T - u_m^*\nabla g_m(x^*)^T & i=1,\ldots m \end{bmatrix} \right.$$

and for $i = 1,\ldots m$

(b) $\quad g_i(x^*) < 0$ implies $g_i(\tilde{x}) + \nabla g_i(\tilde{x})^T(x - \tilde{x}) < 0$

(c) $\quad u_i^* > 0$ implies $u_i > 0$.

Choose $\varepsilon(r)$ and $\delta(r)$ to satisfy the following conditions, where for simplicity we write henceforth ε and δ for $\varepsilon(r)$ and $\delta(r)$ respectively

(3.6) (a) $\max\{2\lambda\delta, 6\delta\tau\} \leqq r$

(b) $\varepsilon < \frac{\bar{\varepsilon}}{3}$

(c) For all $z \in N(z^*,\varepsilon)$ we have that

(i) $||\nabla f(x) + \nabla g(x)u^* - \nabla_{xx}L(z^*)(x-x^*)|| \leqq \delta||x - x^*||$

(ii) $$\left\|\begin{bmatrix} u_1^*(g_1(x) - g_1(x^*) + \nabla g_1(x)^T(x^* - x)) \\ \vdots \\ u_m^*(g_m(x) - g_m(x^*) + \nabla g_m(x)^T(x^* - x)) \end{bmatrix}\right\|$$

$\leqq \delta||x - x^*||$

(iii) for any $n\times n$ matrix H with $||H|| \leqq ||\nabla_{xx}L(z^*)|| + \frac{r}{\lambda}$ it follows $||F(z,H,z^*)|| < \frac{\bar{\varepsilon}}{2\tau}$

The existence of such δ and ε follows from the following considerations. We can choose δ first to satisfy (3.6.a), then choose ε small enough to satisfy (3.6.b) and (3.6.c). The existence of ε satisfying (3.6.c) follows from Lemma 3.3 by observing that $\nabla_{xx}L(x^*,u^*) = 0$. The last two conditions of (3.6.c) are easily satisfied.

Let $\bar{H} = \frac{1}{2}(H + H^T)$; if $||H - \nabla_{xx}L(z^*)|| \leqq \delta$ then it is obvious that $||\bar{H} - \nabla_{xx}L(z^*)|| \leqq \delta$. Thus we have

$$||D_{\bar{H}} - D^*|| \leqq ||\bar{H} - \nabla_{xx}L(z^*)|| \leqq \delta.$$

Hence by (3.4), (3.6.a) and the Banach perturbation lemma [21, p. 45] $D_{\bar{H}}$ is nonsingular and

$$||D_{\bar{H}}^{-1}|| \leqq \frac{\lambda}{1-r} = \tau .$$

From $||\bar{H} - \nabla_{xx}L(z^*)|| \leqq \delta$ and (3.6.a) we have

$$||\bar{H}|| \leqq ||\nabla_{xx}L(z^*)|| + ||\bar{H} - \nabla_{xx}L(z^*)||$$
$$\leqq ||\nabla_{xx}L(z^*)|| + \frac{r}{\lambda} .$$

Hence it follows from (3.6.c) and Lemma 3.4 that the function

$$T_{\tilde{z},\bar{H}}(z) = z - D_{\bar{H}}^{-1}F(\tilde{z},\bar{H},z)$$

is a contraction in $N(z^*,\bar{\varepsilon})$ and has a unique fixed point $\hat{z}$. Hence we have $F(\tilde{z},\bar{H},\hat{z}) = 0$ and

$$(3.7)\qquad ||\hat{z} - z^*|| \leqq 2\tau||F(\tilde{x},\bar{H},z^*)||.$$

Now since

$$(3.8)\quad ||F(\tilde{z},\bar{H},z^*)|| = \left\|\begin{bmatrix} \nabla f(\tilde{x}) + \nabla g(\tilde{x})u^* + \bar{H}(x^*-\tilde{x}) \\ u_1^*(g_1(\tilde{x}) + \nabla g_1(\tilde{x})^T(x^*-\tilde{x})) \\ \cdot \\ \cdot \\ \cdot \\ u_m^*(g_m(\tilde{x}) + \nabla g_m(\tilde{x})^T(x^*-\tilde{x})) \end{bmatrix} - \begin{bmatrix} 0 \\ u_1^*g_1(x^*) \\ \cdot \\ \cdot \\ \cdot \\ u_m^*g_m(x^*) \end{bmatrix}\right\|$$

$$\leqq ||\nabla f(\tilde{x}) + \nabla g(\tilde{x})u^* + \bar{H}(x^*-\tilde{x})|| + \left\|\begin{bmatrix} u_1^*(g_1(x) - g_1(x^*) + g_1(x)^T(x^*-x)) \\ \cdot \\ \cdot \\ \cdot \\ u_m^*(g_m(\tilde{x}) - g_m(x^*) + \nabla g_m(\tilde{x})^T(x^*-\tilde{x})) \end{bmatrix}\right\|$$

$$\leqq (2\delta + ||\bar{H} - \nabla_{xx}L(z^*)||)\ ||x^*-\tilde{x}|| \quad \text{(by 3.6.c)}$$
$$\leqq 3\delta\ ||x^*-\tilde{x}|| \leqq 3\delta||z^*-\tilde{z}|| ,$$

hence by (3.7), (3.8) and (3.6.a) we have

$$(3.9) \quad \|\hat{z} - z^*\| \leqq 6\delta\tau \, \|\tilde{z} - z^*\| \leqq r\|\tilde{z} - z^*\| .$$

Since $\hat{z}$ is the unique fixed point of $T_{\tilde{z},\bar{H}}$ in $N(z^*,\bar{\varepsilon})$, hence $\hat{z}$ is the unique zero of $F(\tilde{z},H,\cdot)$ in $N(z^*,\bar{\varepsilon})$. From $F(\tilde{z},\bar{H},\hat{z}) = 0$ it follows that

$$u_i(g_i(\tilde{x}) + \nabla g_i(\tilde{x})^T(\hat{x}-\tilde{x})) = 0 \qquad i=1,\ldots m.$$

By (3.5.b) and (3.5.c) we have $\hat{u} \geqq 0$ and

$$g(\tilde{x}) + \nabla g(\tilde{x})(\hat{x} - \tilde{x}) \leqq 0.$$

Hence $\hat{z}$ is a z-solution of $Q(\tilde{x},H)$.

We now show that $\hat{z}$ is the closest z-solution of $Q(\tilde{x},H)$ to $\tilde{z}$. Since any z-solution of $Q(\tilde{x},H)$ is a zero of $F(\tilde{z},\bar{H},\cdot)$, it follows from the uniqueness of the zero of $F(\tilde{z},\bar{H},\cdot)$ in $N(z^*,\bar{\varepsilon})$ that $\hat{z}$ is the unique z-solution of $Q(\tilde{x},H)$ in $N(z^*,\bar{\varepsilon})$. If $\bar{z}$ is another z-solution of $Q(\tilde{x},H)$ then $\bar{z} \notin N(z^*,\bar{\varepsilon})$. Hence

$$\begin{aligned} \|\bar{z} - \tilde{z}\| &\geqq \|\bar{z} - z^*\| - \|\tilde{z} - z^*\| \\ &> \bar{\varepsilon} - \tfrac{1}{3}\bar{\varepsilon} = \tfrac{2}{3}\bar{\varepsilon} \qquad \text{(by 3.6.b)}. \end{aligned}$$

But
$$\begin{aligned} \|\hat{z} - \tilde{z}\| &\leqq \|\hat{z} - z^*\| + \|z^* - \tilde{z}\| \\ &< \tfrac{1}{3}\bar{\varepsilon} + \tfrac{1}{3}\bar{\varepsilon} = \tfrac{2}{3}\bar{\varepsilon} . \end{aligned}$$

Hence $\hat{z}$ is the closest z-solution of $Q(\tilde{x},H)$ to $\tilde{z}$, which in conjunction with (3.9) completes the proof. □

Before we give the following corollary recall that a sequence z^k converges to a point z^* Q-linearly if there exists $r \in (0,1)$ such that $||z^{k+1}-z^*|| \leq r||z^k-z^*||$, and Q-superlinearly if $||z^{k+1}-z^*|| = \theta_k||z^k-z^*||$ and $\lim_{k\to\infty} \theta_k = 0$.

<u>Corollary 3.6</u>: Let $z^* = (x^*,u^*)$ be a Kuhn-Tucker pair of Problem P which satisfies the Jacobian uniqueness condition and f and g_i $(i = 1,\ldots m) \in LC^2[x^*]$ and let $\{j_k\}$ be a nondecreasing sequence of nonnegative integers with $j_k \leq k$. If z^0 is close enough to z^* and $\{\alpha^k\}$ is a sequence of nonnegative numbers and bounded above by a sufficiently small positive number and if $||H_k - \nabla_{xx}L(z^{j_k})|| \leq \alpha_k$ then the sequence of points $\{z^k\}$ generated by the algorithm converges Q-linearly to z^*. Furthermore, if $k \to \infty$ implies that $j_k \to \infty$ and that $\alpha_k \to \infty$, then $\{z^k\}$ converges Q-superlinearly to z^*.

<u>Proof</u>: The first part follows directly from Theorem 3.5. We only need to prove the second part. Let $r \in (0,1)$. Since

$$||H_k - \nabla_{xx}L(z^*)|| \leq ||H_k - \nabla_{xx}L(z^{j_k})|| + ||\nabla_{xx}L(z^{j_k}) - \nabla_{xx}L(z^*)||$$

$$\leq \alpha_k + ||\nabla_{xx}L(z^{j_k}) - \nabla_{xx}L(z^*)||$$

and from the first part $\{z^k\}$ converges Q-linearly to z^* , thus we have that $H_k \to \nabla_{xx}L(z^*)$ when $\alpha_k \to 0$. Let $\varepsilon(r)$ and $\delta(r)$ be defined as in Theorem 3.5; then there exists $\bar{k} > 0$ such that for all $k \geq \bar{k}$ we have $||z^k - z^*|| < \varepsilon(r)$ and $||H_k - \nabla_{xx}L(z^*)|| < \delta(r)$. Therefore it follows from Theorem 3.5 that $||z^{k+1} - z^*|| \leq r\,||z^k - z^*||$ for all $k \geq \bar{k}$. Since r is arbitrary, it follows that

$$\lim_{k\to\infty} \frac{||z^{k+1} - z^*||}{||z^k - z^*||} = 0$$

which shows that $\{z^k\}$ converges Q-superlinearly to z^*. □

If we let $H_k = \nabla_{xx}L(z^{j_k}) + \alpha_k I$ then the resulting algorithm is an extension of the Levenberg-Marquardt algorithm to the constrained optimization; the local convergence properties of this algorithm follow immediately from the above corollary. When $j_k = k$ and $\alpha_k = 0$ the algorithm becomes Wilson's algorithm [29] for which Robinson has established a quadratic rate of convergence [26].

In the theorem below we give a sufficient condition on the update which guarantees that our algorithm generates points that converge to a Kuhn-Tucker pair of Problem P. The importance of this condition lies in the fact that some variable metric updates satisfy this condition and hence can be used in our algorithm to solve constrained optimization problems. This condition has been studied for unconstrained problems by Broyden, Dennis and Moré [3]. Throughout this work $||\cdot||'$ denotes any fixed matrix norm which may be different from $||\cdot||$.

<u>Theorem 3.7</u>: Let $z^* = (x^*,u^*)$ be a Kuhn-Tucker pair of Problem P and f and g_i $(i=1,\ldots m) \in LC^2[x^*]$. If the Jacobian uniqueness condition is satisfied at z^* and there are two nonnegative constants α_1 and α_2 such that

$$(3.10) \quad ||H_{k+1} - \nabla_{xx}L(z^*)||' \leqq (1 + \alpha_1||z^k-z^*||)||H_k - \nabla_{xx}L(z^*)||' + \alpha_2||z^k - z^*|| ,$$

then for any $r \in (0,1)$ we have two positive numbers $\varepsilon(r)$ and $\delta(r)$ such that if $||z^0 - z^*|| \leqq \varepsilon(r)$ and $||H_0 - \nabla_{xx}L(z^*)|| \leqq \delta(r)$, then the sequence $\{z^k\}$ generated by the algorithm is well defined and converges Q-linearly to z^*.

<u>Proof</u>: By the equivalence of matrix norms, there exist two positive numbers d and d' such that for any $n \times n$ matrix A we have

$$(3.11) \quad d||A||' \geqq ||A|| \ , \ d'||A|| \geqq ||A||' \ .$$

Let $r \in (0,1)$ be given. By Theorem 3.5 there exist two positive numbers $\bar{\varepsilon}$ and $\bar{\delta}$ such that if $||z - z^*|| \leqq \bar{\varepsilon}$ and $||H - \nabla_{xx}L(z^*)|| \leqq \bar{\delta}$ then the closest z-solution $\hat{z}$ of $Q(\bar{x},H)$ exists and $||\hat{z} - z^*|| \leqq r||\bar{z} - z^*||$.

We choose two positive numbers $\bar{\varepsilon}$ and $\bar{\delta}$ such that if $||\bar{z} - z^*|| \leqq \bar{\varepsilon}$ and $||H - \nabla_{xx}L(z^*)|| \leqq \bar{\delta}$ then the closest z-solution $\hat{z}$ of $Q(\bar{x},H)$ exists and $||z - z^*|| \leqq r||\bar{z} - z^*||$.

We choose two positive numbers ε and δ such that the following conditions are satisfied

$$(3.12) \quad \text{(a)} \quad \varepsilon \leqq \bar{\varepsilon}$$

$$\text{(b)} \quad 2dd'\delta \leqq \bar{\delta}$$

$$\text{(c)} \quad (2\alpha_1 d' + \alpha_2) \frac{\varepsilon}{1-r} \leqq d'\delta$$

If we can show that for each k we have

$$(3.13) \qquad ||z^k - z^*|| \leqq r^k \varepsilon$$

and

$$(3.14) \qquad ||H_k - \nabla_{xx}L(z^*)||' \leqq 2d'\delta \ ,$$

then by (3.12.a) and (3.12.b) we have $||z^k - z^*|| \leqq \bar{\varepsilon}$ and

$||H_k - \nabla_{xx}L(z^*)|| \leq 2dd'\delta \leq \bar{\delta}$; thus this theorem follows immediately from Theorem 3.5.

Now we establish (3.13) and (3.14) by induction. It is obvious that these inequalities are true for $k = 0$. Assume that they are true for j, $0 \leq j \leq k$; then it follows from (3.10) that

$$||H_{j+1} - \nabla_{xx}L(z^*)||' - ||H_j - \nabla_{xx}L(z^*)||' \leq 2\alpha_1 d'\varepsilon\delta r^j + \alpha_2 \varepsilon r^j.$$

By taking the sum from $j = 0$ to $j = k$, we get that

$$(3.15) \quad ||H_{k+1} - \nabla_{xx}L(z^*)||' \leq ||H_0 - \nabla_{xx}L(z^*)||' + (2\alpha_1 d'\delta + \alpha_2)\frac{\varepsilon}{1-r}$$

$$\leq d'\delta + d'\delta \quad \text{(by 3.12.c, 3.11 and initial choice of } H_0\text{)}$$

$$\leq 2d'\delta.$$

Therefore (3.14) is established. Next we show that (3.13) is true for $j = k+1$. From (3.15), (3.11) and (3.12.b) we have

$$(3.16) \qquad ||H_{k+1} - \nabla_{xx}L(z^*)|| \leq \bar{\delta}\ .$$

By the induction hypothesis and (3.12.a) we have

$$(3.17) \qquad ||z^k - z^*|| \leq r^k\varepsilon \leq \bar{\varepsilon}.$$

Thus it follows from (3.16), (3.17) and Theorem 3.5 that z^{k+1} exists and $||z^{k+1} - z^*|| \leq r||z^k - z^*|| \leq r^{k+1}$. Hence (3.13) is true for $j = k+1$ and the theorem is proven. □

We conclude this section with a corollary of the above theorem, which shows that if condition (3.10) is satisfied then the updated matrices will remain close to $\nabla_{xx}L(z^*)$.

Corollary 3.8: If all the assumptions of Theorem 3.7 hold, then for any $r \in (0,1)$ and $t > 0$ there exist two positive numbers $\varepsilon(r,t)$ and $\delta(r,t)$ such that if $||z^0 - z^*|| \leq \varepsilon(r,t)$ and $||H_0 - \nabla_{xx}L(z^*)|| \leq \delta(r,t)$, then the sequence $\{z^k\}$ generated by the algorithm is well defined and $||z^{k+1} - z^*|| \leq r||z^k - z^*||$, and furthermore the sequence of matrices $\{H_k\}$ satisfy

(3.18) $\quad ||H_k - \nabla_{xx}L(z^*)|| \leq t$

for each k.

Proof: Since in the proof of Theorem 3.7 we have established $||H_k - \nabla_{xx}L(z^*)|| \leq 2d'\delta$ for each k, hence (3.18) is obviously true if we choose $\varepsilon(r,t)$ and $\delta(r,t)$ to satisfy the additional condition $2d'\delta < t$. □

4. Q-Superlinear Rate of Convergence

This section is devoted to giving some sufficient conditions which guarantee that our algorithm has a Q-superlinear rate of convergence. In the next section we will discuss some specific updates which satisfy this condition. Actually Q-superlinear convergence conditions for a more general family of algorithms will be considered in this section; this family includes not only the algorithm considered in this paper but also the algorithms which iteratively solve optimization subproblems with linearized constraints. We define this family of algorithms in the following definition.

Definition 4.1: An algorithm for solving Problem P is called a linearized constraint algorithm if it generates (n+m)-vectors $z^k = (x^k, u^k)$ which satisfy

$$(4.1)\quad u_i^{k+1}\,(g_i(x^k) + \nabla g_i(x^k)(x^{k+1} - x^k)) = 0 \quad i=1,\ldots m \;\square$$

Besides our algorithm, some other examples of the linearized constraint algorithm are Wilson's algorithm [29], Robinson's algorithm [25] and dual variable metric algorithms [16].

The next result is due to Mangasarian [20] and is closely related to the work done by Dennis and Moré [9].

<u>Lemma 4.2</u>: Let z^* be a Kuhn-Tucker pair of Problem P satisfying the Jacobian uniqueness condition and let f and g_i $(i=1,\ldots m)$ have continuous second derivatives. A sequence $\{z^k\}$ converges Q-superlinearly if $\{z^k\}$ converges to z^* and

$$\lim_{k\to\infty} \frac{||E(z^{k+1})||}{||z^{k+1} - z^k||} = 0$$

where

$$E(z) = \begin{bmatrix} \nabla_x L(z) \\ u_1 g_1(x) \\ \cdot \\ \cdot \\ \cdot \\ u_m g_m(x) \end{bmatrix}.$$

<u>Proof</u>: By the Jacobian uniqueness condition it follows that $\nabla_z E(z^*)$ is nonsingular [10, p. 80]. Therefore there exists $c > 0$ such that for sufficiently large k we have

$$||E(z^{k+1})|| = ||E(z^{k+1}) - E(z^*)|| \geq c||z^{k+1} - z^*||.$$

Hence by the assumption of this lemma we obtain

$$\lim_{k\to\infty} \frac{||z^{k+1} - z^*||}{||z^{k+1} - z^k||} = 0$$

which implies

$$\lim_{k\to\infty} \frac{||z^{k+1} - z^*||}{||z^{k+1} - z^*|| + ||z^* - z^k||} = 0$$

and in turn implies

$$\lim_{k\to\infty} \frac{||z^{k+1} - z^*||}{||z^k - z^*||} = 0 . \quad \square$$

In the following theorem we establish a sufficient condition for Q-superlinear rate of convergence of a linearized constraint algorithm.

<u>Theorem 4.3</u>: Let z^* be a Kuhn-Tucker pair of Problem P satisfying the Jacobian uniqueness condition and let $f^\cdot$ and g_i $(i=1,\ldots m)$ have continuous second derivatives at x^*. If a sequence $\{z^k\}$ generated by a linearized constraint algorithm converges to z^* and

$$(4.2) \quad \lim_{k\to\infty} \frac{||\nabla_x L(z^{k+1})||}{||z^{k+1} - z^k||} = 0$$

then $\{z^k\}$ converges Q-superlinearly to z^*.

<u>Proof</u>: By Lemma 4.2 we only need to prove that

$$\lim_{k\to\infty} \frac{||E(z^{k+1})||}{||z^{k+1} - z^k||} = 0 .$$

Since $z^k \to z^*$ and $\{z^k\}$ is constructed by a linearized constraint algorithm, we have

$$\begin{aligned}
||E(z^{k+1})|| &\leq ||\nabla_x L(x^{k+1},u^{k+1})|| + \sum_{i=1}^{m} | u_i^{k+1} g_i(x^{k+1})| \\
&\leq ||\nabla_x L(x^{k+1},u^{k+1})|| + \\
&\quad \sum_{i=1}^{m} |u_i^{k+1}(g_i(x^{k+1}) - g_i(x^k) - \nabla g_i(x^k)(x^{k+1} - x^k))| \\
&\leq ||\nabla_x L(z^{k+1})|| + o\,(||x^{k+1} - x^k||).
\end{aligned}$$

Hence it follows from (4.2) that

$$\lim_{k\to\infty} \frac{||E(z^{k+1})||}{||z^{k+1} - z^k||} = 0 . \quad \square$$

Next we introduce a theorem which will be very useful for establishing Q-superlinear rates of convergence of our algorithms. This theorem uses the following result of Dennis and Moré [9].

<u>Lemma 4.4</u>: Let $\{a_k\}$ and $\{b_k\}$ be sequences of nonnegative numbers and $\alpha_1 \geq 0$, $\alpha_2 \geq 0$ such that

$$a_{k+1} \leq (1 + \alpha_1 b_k)a_k + \alpha_2 b_k$$

and

$$\sum_{i=1}^{\infty} b_k < \infty ,$$

then $\{a_k\}$ converges. $\square$

<u>Theorem 4.5</u>: Let z^* be a Kuhn-Tucker pair of Problem P satisfying the Jacobian uniqueness condition and f and g_i $(i=1,\ldots m)$ $\in LC^2[x^*]$. Assume further that $\{z^k\}$ is constructed by a linearized constraint algorithm. If there exist two nonnegative constants α_1 and α_2 and two sequences of nonnegative numbers $\{\rho_k\}$ and $\{\sigma_k\}$ such that the following conditions are satisfied

(i) $\sum_{k=1}^{\infty} ||z^k - z^*|| < \infty$

(ii) $\rho_k \to 0$ implies $\dfrac{||\nabla_x L(z^{k+1})||}{||z^{k+1} - z^k||} \to 0$

(iii) $\sigma_k \to 0$ implies $\dfrac{||\nabla_x L(z^{k+1})||}{||z^{k+1} - z^k||} \to 0$

(iv) $\rho_{k+1} \leqq (1 - \sigma_k + \alpha_1 ||z^k - z^*||)\rho_k + \alpha_2 ||z^k - z^*||$,

then $\{z^k\}$ converges Q-superlinearly to z^*.

<u>Proof</u>: From Lemma 4.4 and (iv) it follows that $\{\rho_k\}$ converges to a nonnegative number, say $\bar{\rho}$. If $\bar{\rho} = 0$ then by (ii) we have that

$$\lim_{k\to\infty} \frac{||\nabla_x L(z^{k+1})||}{||z^{k+1} - z^k||} = 0$$

and the desired conclusion follows directly from Theorem 4.3. Assume $\bar{\rho} \neq 0$. Then (iv) implies that

$$\sigma_k \rho_k \leqq \rho_k - \rho_{k+1} + ||z^k - z^*|| \; (\alpha_1 \alpha_k + \alpha_2) \; . \tag{4.3}$$

By taking sum of both sides of (4.3) over $k = 0,1,\ldots$ and taking (i) and the boundedness of $\{\rho_k\}$ into account it follows that

$$\sum_{k=1}^{\infty} \sigma_k \rho_k < \infty \; .$$

Since $\rho_k \to \bar{\rho}$ and $\bar{\rho} \neq 0$, it follows that $\sigma_k \to 0$. Hence by (iii) we also have that

$$\lim_{k\to\infty} \frac{||\nabla_x L(z^{k+1})||}{||z^{k+1} - z^k||} = 0 \; .$$

Theorem 4.3 now implies the desired result. □

We conclude this section with the following theorem which establishes a sufficient condition for the Q-superlinear rate of convergence of our algorithm. This theorem is closely related to a result due to Dennis and Moré [9]; a similar result was studied by Garcia and Mangasarian [13].

Theorem 4.6: Let z^* be a Kuhn-Tucker pair of Problem P satisfying the Jacobian uniqueness condition and f and g_i $(i=1,\ldots m) \in LC^2[x^*]$. Suppose $\{z^k\}$ is a sequence of points generated by the algorithm with respect to a sequence of matrices $\{H_k\}$. Then the sequence $\{z^k\}$ converges Q-superlinearly to z^* if $\{z^k\}$ converges to z^* and

$$(4.4) \qquad \lim_{k\to\infty} \frac{||(\frac{1}{2}(H_k + H_k^T) - \nabla_{xx}L(z^*))(x^{k+1} - x^k)||}{||z^{k+1} - z^k||} = 0$$

Proof: This theorem follows directly from Theorem 4.3 and the following lemma. □

Lemma 4.7: If all the assumptions of Theorem 4.6 hold then

$$(4.5) \qquad \lim_{k\to\infty} \frac{||\nabla_x L(z^{k+1})||}{||z^{k+1} - z^k||} = 0 \ .$$

Proof: Since $\{z^k\}$ is generated by the algorithm with respect to $\{H_k\}$, we have

$$\nabla_x L(x^k, u^{k+1}) + \frac{1}{2}(H_k + H_k^T)(x^{k+1} - x^k) = 0 \ .$$

Hence

$$||\nabla_x L(x^{k+1},u^{k+1})||$$

$$= ||\nabla_x L(x^{k+1},u^{k+1}) - \nabla_x L(x^k,u^{k+1}) - \tfrac{1}{2}(H_k + H_k^T)(x^{k+1} - x^k)||$$

$$\leqq ||\nabla_x L(x^{k+1},u^{k+1}) - \nabla_x L(x^k,u^{k+1}) - \nabla_{xx} L(z^*)(x^{k+1} - x^k)|| +$$

$$||(\tfrac{1}{2}(H_k + H_k^T) - \nabla_{xx}\ (z^*)(x^{k+1} - x^k)||.$$

Thus (4.5) follows from (4.4) and Lemma 3.3. □

5. Updates

We have developed some theorems for the local convergence and the superlinear rate of convergence of the algorithm. In this section we show that some concrete updates for the algorithm satisfy the hypothesis of these results and hence converge locally with superlinear rates. The updates to be discussed are extensions of some well known updates for unconstrained optimization problems; in this context the superlinear convergence of these updates has been established by Broyden, Dennis and Moré [3] and Dennis and Moré [9]. Some techniques of their proof are employed here.

For economy of notation let bared variables (such as $\bar{H}$) represent the (k+1)-st variables (H_{k+1}) and unadorned variables (H) represent the k-th variables (H_k). The updates to be considered can be written as

$$(5.1)\qquad \bar{H} = H + \frac{(y - Hs)c^T + c(y - Hs)^T}{c^T s} - \frac{s^T(y - Hs)cc^T}{(c^T s)^2}$$

where $s = \bar{x} - x$, $y = \nabla_x L(\bar{x},\bar{u}) - \nabla_x L(x,\bar{u})$ and c is any vector with $c^T s \neq 0$. A particular algorithm is determined once c is specified in formula (5.1). When formula (5.1) is used to update matrices in the algorithm the resulting algorithm will be called Algorithm A1 when $c = s$, Algorithm A2 when $c = y$ and Algorithm A3 when $c = D_0 s$ and D_0 is any fixed positive definite matrix. Thus we have the following updates:

$$(5.2) \qquad \bar{H} = H + \frac{(y-Hs)s^T + s(y-Hs)^T}{s^T s} - \frac{s^T(y-Hs)ss^T}{(s^T s)^2} \quad \text{(Algorithm A1)}$$

(Unconstrained case: Powell [23,24]),

$$(5.3) \qquad \bar{H} = H + \frac{(y-Hs)y^T + y(y-Hs)^T}{y^T s} - \frac{s^T(y-Hs)yy^T}{(y^T s)^2} \quad \text{(Algorithm A2)}$$

(Unconstrained case: Davidon-Fletcher-Powell [7,11]),

$$(5.4) \qquad \bar{H} = H + \frac{(y-Hs)s^T D_0^T + Ds(y-Hs)^T}{s^T D_0^T s} - \frac{s^T(y-Hs)Dss^T D_0^T}{(s^T D_0^T s)^2} \quad \text{(Algorithm A3}$$

In this section for any nonsingular $n \times n$ matrix matrix M we define the matrix norm $||\cdot||_M$ in such a way that for any $n \times n$ matrix A

$$(5.5) \qquad ||A||_M = \text{trace}[(MAM)^T(MAM)].$$

We next introduce the following lemma which is due to Broyden, Dennis and Moré [3].

<u>Lemma 5.1</u>: Let H be any $n \times n$ symmetric matrix and s, c and y be vectors in R^n with $c^T s \neq 0$ and define $\bar{H}$ by formula (5.2). If M is a nonsingular symmetric $n \times n$ matrix with

$$(5.6) \qquad ||Mc - M^{-1}s|| \leq \beta ||M^{-1}s||$$

for some $\beta \in [0, \frac{1}{3}]$, then for any symmetric $n \times n$ matrix A with $A \neq H$ we have

$$(5.7)\quad \|H - A\|_M \leq ((1 - \lambda\theta^2)^{\frac{1}{2}} + \lambda_1 \frac{\|Mc - M^{-1}s\|}{\|M^{-1}s\|})\|H - A\|_M + \lambda_2 \frac{\|y - As\|}{\|s\|}$$

where $\lambda \in (0,1)$, and λ_1 and λ_2 are constants which only depend on M and n, and

$$(5.8)\qquad \theta = \frac{\|M(H-A)s\|}{\|H-A\|_M \|M^{-1}s\|}$$

if $H \neq A$ and $\theta = 0$ otherwise. □

In the following theorem a sufficient condition is given to guarantee Q-superlinear convergence of the algorithm with an update of form (5.1).

Theorem 5.2: Let $z^* = (x^*,u^*)$ be a Kuhn-Tucker pair of Problem P satisfying the Jacobian uniqueness condition and f and g_i $(i=1,\ldots$ $LC^2[x^*]$. Suppose that in the algorithm the sequence of matrices $\{H_k\}$ are generated by the update (5.1) with any c^k such that $s^k \neq 0$,

$$(5.0)\quad \frac{\|Mc^k - M^{-1}s^k\|}{\|M^{-1}s^k\|} \leq \mu \max\{\|z^k - z^*\|, \|z^{k+1} - z^*\|\}$$

for a constant μ and an arbitrary but fixed nonsingular symmetric matrix M. If z^0 and H_0 are sufficiently close to z^* and $\nabla_{xx}L(z^*)$ respectively then the sequence $\{z^k\}$ generated by the algorithm is well defined and converges Q-superlinearly to z^*.

Proof: For any $r \in (0,1)$ let $\varepsilon(r)$ and $\delta(r)$ be defined as in Theorem 3.7 with matrix norm $\|\cdot\|'$ as $\|\cdot\|_M$, and let

$$(5.10)\qquad \alpha_1 = \lambda_1\mu, \qquad \alpha_2 = \lambda_2(\bar{K} + \tilde{K})$$

where $\bar{K}$ and $\tilde{K}$ are constants defined in Lemma 3.3 and λ_1 and λ_2 are defined in Lemma 5.1. We further require $\varepsilon(r)$ to satisfy

$$(5.11) \qquad \varepsilon(r) \leq \frac{1}{3\mu} .$$

First we show by induction that if $||z^0 - z^*|| \leq \varepsilon(r)$ and $||H^0 - \nabla_{xx}L(z^*)|| \leq \delta(r)$ then the generated sequence $\{z^j\}$ exists and converges to z^* Q-linearly; that is,

$$(5.12) \qquad ||z^{j+1} - z^*|| \leq r||z^j - z^*|| .$$

When $j = 0$ the existence of z^{j+1} and (5.12) follows directly from the choice of $\varepsilon(r)$ and $\delta(r)$ and Theorem 3.5. Assume that for all $j \leq k$, z^{j+1} exists and (5.12) holds; we now show that z^{k+2} exists and (5.12) is also true for $j = k+1$.

Assume that $s^k \neq 0$, since if $s^k = 0$ then z^{k+1} is a Kuhn-Tucker pair of Problem P and on the other hand by the Jacobian uniqueness condition z^* is the unique Kuhn-Tucker pair of Problem P in $N(z^*,\varepsilon(r))$. Hence in case $s^k = 0$ we have $z^{k+1} = z^*$ and the sequence $\{z^k\}$ converges to z^* in a finite number of steps. When $s^k \neq 0$ it follows from (5.9) and (5.11) that

$$\frac{||Mc^k - M^{-1}s^k||}{||M^{-1}s^k||} \leq \mu \max\{||z^k - z^*||, ||z^{k+1} - z^*||\}$$
$$< \mu\, \varepsilon(r) \leq \frac{1}{3} .$$

Thus by (5.1) of Lemma 5.1 we have that

(5.14)
$$||H_{k+1} - \nabla_{xx}L(z^*)||_M \leq ((1-\lambda\theta_k^2)^{\frac{1}{2}} + \lambda_1\mu||z^k - z^*||)||H_k - \nabla_{xx}L(z^*)||_M + \lambda_2 \frac{||y^k - \nabla_{xx}L(z^*)s^k||}{||s^k||}$$

where

$$(5.15)\quad \theta_k = \frac{||M(H_k - \nabla_{xx}L(z^*))s^k||}{||H_k - \nabla_{xx}L(z^*)||_M||M^{-1}s^k||}\,.$$

Lemma 3.3 shows that

$$||y^k - \nabla_{xx}L(z^*)s^k|| \leqq (\bar{K} + \tilde{K})||z^k - z^*||\ ||s^k||\,,$$

and therefore

$$(5.16)\quad \frac{||y^k - \nabla_{xx}L(z^*)s^k||}{||s||} \leqq (\bar{K} + \tilde{K})||z^k - z^*||.$$

By (5.14) and (5.16) in conjunction with (5.10) we obtain

(5.17)

$$||H_{k+1} - \nabla_{xx}L(z^*)||_M \leqq ((1-\lambda\theta_k^2)^{\frac{1}{2}} + \alpha_1||z^k - z^*||)||H_k - \nabla_{xx}L(z^*)||_M$$
$$+ \alpha_2||z^k - z^*||.$$

Hence the existence of z^{k+2} and the inequality $||z^{k+2} - z^*|| \leqq r||z^{k+1} - z^*||$ follows from (5.17) and Theorem 3.7 immediately.

So far we have proved that $\{z^k\}$ converges to z^* Q-linearly. Now we are going to use Theorem 4.5 to prove that the sequence $\{z^k\}$ converges to z^* with a Q-superlinear rate. Let $\rho_k = ||H_k - \nabla_{xx}L(z^*)||_M$ and $\sigma_k = \lambda\theta_k^2$ where θ_k is defined in (5.15). Condition (i) of Theorem 4.5 is satisfied because $\{z^k\}$ converges Q-linearly to z^*. It is obvious that $\rho_k \to 0$ implies

$$(5.18)\quad \lim_{k\to\infty} \frac{||(H_k - \nabla_{xx}L(z^*))s^k||}{||s^k||} = 0$$

and so does $\sigma_k \to 0$. Hence it follows from (5.18). Lemma 4.7 and the symmetry of H_k that conditions (ii) and (iii) of Theorem 4.5 are satisfied. By taking the inequality

$(1 - \lambda\theta^2)^{\frac{1}{2}} \leq 1 - \frac{\lambda}{2}\theta^2$ into account, it follows from (5.17) that

$$\rho_{k+1} \leq (1 - \sigma_k + \alpha_1||z^k - z^*||)\rho_k + \alpha_2||z^k - z^*||.$$

Therefore all conditions of Theorem 4.5 are satisfied and hence $\{z^k\}$ converges Q-superlinearly to z^*. □

In the above theorem we can see that inequality (5.9) is a key condition for updates of form (5.1) to possess local Q-superlinear convergence properties. We shall show in the following corollaries that inequality (5.9) can be established for Algorithms A1, A2 and A3 by a suitable choice of a matrix M and hence these algorithms generate Q-superlinearly convergent sequences when used to solve Problem P.

Corollary 5.3: Let $z^* = (x^*,u^*)$ be a Kuhn-Tucker pair of Problem P satisfying the Jacobian uniqueness condition and f and g_i $(i=1,\ldots m) \in LC^2[x^*]$. If the starting point z^0 and the starting matrix H_0 are sufficiently close to z^* and $\nabla_{xx}L(z^*)$ respectively, then Algorithm A1 and A3 generate sequences of points which converge Q-superlinearly to z^*.

Proof: Algorithm A1 is actually a special case of Algorithm A3 with $D_0 = I$. The corollary follows because inequality (5.9) is obviously true for Algorithm A3 by setting $M = (D_0^{-1})^{\frac{1}{2}}$. □

Corollary 5.4: Let the assumptions of Corollary 5.3 be satisfied. Assume further that $\nabla_{xx}L(z^*)$ is positive definite. Then the conclusion of Corollary 5.3 is also true for Algorithm A2.

Proof: Since $\nabla_{xx}L(z^*)$ is positive definite, we can define

(5.19) $$M = (\nabla_{xx}L(z^*)^{-1})^{\frac{1}{2}} .$$

In Algorithm A2 we have $c^k = y^k$. Hence it follows from (5.19) and Lemma 3.3 that

$$\begin{aligned} \|Mc^k - M^{-1}s^k\| &\leq \|M\| \ \|y^k - \nabla_{xx}L(z^*)s^k\| \\ &\leq \|M\|(\bar{K} + \tilde{K}) \max\{\|z^k - z^*\|, \|z^{k+1} - z^*\|\}\|s^k\|, \end{aligned}$$

and therefore

$$\frac{\|Mc^k - M^{-1}s^k\|}{\|s^k\|} \leq \|M\|^2(\bar{K} + \tilde{K}) \max\{\|z^k - z^*\|, \|z^{k+1} - z^*\|\}.$$

Thus inequality (5.9) is satisfied with $\mu = \|M\|^2 \ (K + K)$, and hence the desired result follows. □

By a similar argument we also can establish the local Q-superlinear convergence for the following updates [16];

(5.20) $$\bar{H} = H + \frac{(y-Hs)s^T + s(y-Hs)^T}{s^Ts} ;$$

(5.21) $$\bar{H} = H + \frac{(y-Hs)y^T + y(y-Hs)^T}{y^Ts} .$$

We also note here that the local Q-linear convergence can be established for the nonsymmetric Broyden's update [2] and Pearson's update. [22] which are respectively

(5.22) $$\bar{H} = H + \frac{(y-Hs)s^T}{s^Ts}$$

and

(5.23) $$H = H + \frac{(y-Hs)y^T}{y^Ts} .$$

However, we have not succeeded in establishing the Q-superlinear rate for them because of the nonsymmetry. When we apply the Powell's symmetrization procedure to updates (5.22) and (5.23) we obtain updates (5.2) and (5.3) respectively [8,23]; updates (5.20) and (5.21) are the resulting updates by taking only one step of such a procedure.

6. Comments and Computational Experiences

Some comments are stated below:

(1) Our algorithms are in a sense a natural extension of variable metric algorithms to general nonlinear programming; this extension offers us a fruitful field of future research. A lot of results in the extensive literature of variable metric algorithms need to be investigated and developed for nonlinear programming and the whole theory can be treated in a unified way in both constrained and unconstrained optimization.

(2) Our convergence theorems still hold if the updated matrices are perturbed by suitably small amounts. This flexibility suggests that the use of difference approximations to derivatives could be used in the updates. We note here that variable metric algorithms have been so modified by Stewart [27] in unconstrained optimization and his modification appears to work well in practice.

(3) All the results in this paper are local. One approach studied by this author to achieving the global convergence is to determine a stepsize in each iteration to maintain the monotonic decrease of an exact penalty function or an augmented Lagrangian; some global convergence theorems have already been established [16].

Computational tests of the algorithm on this paper have been performed and are still going on. A report on the test results is expected to be published in the near future. However, it would be unfair to finish without at least giving some idea of the power of these methods in practice. We state in the table below the test results of Algorithm A2 (DFP) for Colville's test Problems 1 and 2. The computations were done on the UNIVAC 1110 system at the University of Wisconsin, Madison. The principal pivoting method [5.6] was used in solving the quadratic programming subproblems.

Table 1

Prob.	Obj. Fct Value	Standard Time Ratio
1	-32.3487	.00541 1)
2 2)	-32.3488	.577
3 3)	-32.3488	.039

1) This result is better than any one reported in the Colville's report [4].

2. Ifeasible starting point.

3. Feasible starting point.

Acknowledgement

This work constitutes a portion of the author's doctoral thesis under the supervision of Professor O. L. Mangasarian at the Department of Computer Sciences, University of Wisconsin-Madison. The author is also indebted to Professor J.E. Dennis, R.R. Meyer, J.J. Moré and S.M. Robinson for reading the manuscript and making many valuable suggestions.

References

1. BARTELS, R. H. and GOLUB, G. H. and SAUNDERS, M. A. : Numerical techniques in mathematical programming, in: Nonlinear programming, ed. J. B. Rosen, O. L. Mangasarian and K. Ritter, Academic Press, 123-176 (1970).

2. BROYDEN, C. G. : A class of methods for solving nonlinear simultaneous equations, Math. Comp. 19, 577-593 (1965).

3. BROYDEN, C. G. and DENNIS, J. E. and MORÉ, J. J. : On the local and superlinear convergence of quasi-Newton methods, J. Inst. Maths. Applics. 12, 223-245 (1973).

4. COLVILLE, A. R. : A comparative study on nonlinear programming codes, IBM New York Scientific Center, Techn. Rep. 320-2949, 1968.

5. COTTLE, R. W. : The principal pivoting method of quadratic programming, in : Mathematics of the decision sciences, part 1, ed. G. B. Dantzig and A. F. Veinott, Amer. Math. Soc., Providence, R. I., 144-162 (1968).

6. COTTLE, R. W. and DANTZIG, G. B. : Complementary pivot theory of mathematical programming, Linear Algebra and Appl. 1, 103-125 (1968).

7. DAVIDON, W. C. : Variable metric method for minimization, A.E.C. Res. and Dev. Report # ANL-5990, (1959).

8. DENNIS, J. E. : On some methods based on Broyden's secant approximation to the Hessian, in : Numerical methods for nonlinear optimization, ed. F. A. Lootsma, Academic Press, 19-34 (1972).

9. DENNIS, J. E. and MORÉ, J. J. : A characterization of superlinear convergence and its application to quasi-Newton methods, Mathematics of Computation, Vol. 28, No. 126 (1974).

10. FIACCO, A. V. and McCORMICK, G. P. : Nonlinear programming : Sequential unconstrained minimization techniques, John Wiley & Sons, Inc., New York, (1968).

11. FLETCHER, R. and POWELL, M. : A rapidly convergent descent method for minimization, Comput. J., Vol. 6, 163-168 (1963).

12. GARCIA-PALOMARES, U. M. : Superlinearly convergent quasi-Newton method for nonlinear programming, Ph.D. dissertation, University of Wisconsin, Madison 1973.

13. GARCIA-PALOMARES, U. M. and MANGASARIAN, O. L. : Superlinearly convergent quasi-Newton algorithms for nonlinearly constrained optimization problems, Computer Sciences Technical Report # 195, University of Wisconsin, Madison, (1974).

14. GILL, P. E. and MURRAY, W. : Quasi-Newton methods for linearly constrained optimization, Nat. Phys. Lab. Report NAC 32.

15. GOLDFARB, D. : Extension of Davidon's variable metric method to maximization under linear and inequality constraints, SIAM J. Appl. Math., Vol. 17, 739-764 (1969).

16. HAN, S. P. : Superlinearly convergent variable metric methods for general nonlinear programming problems, Ph. D. dissertation, University of Wisconsin, Madison, (1974).

17. ISAACSON, E. and KELLER, H. B. : Analysis of numerical methods, John Wiley & Sons, Inc., New York, (1966).

18. LIUSTERNIK, L. A. and SOBOLEV, V. J. : Elements of functional analysis, Frederick Ungan Publishing Co., New York, N.Y., (1961).

19. LOOTSMA, F. A. : A survey of methods for solving constrained minimization problems via unconstrained minimization, in : Numerical methods for nonlinear optimization, ed. F. A. Lootsma, Academic Press, 313-347 (1972).

20. MANGASARIAN, O. L. : Private communication.

21. ORTEGA, J. M. and RHEINBOLDT, W. C. : Iterative solution of nonlinear equations in several variables, Academic Press, New York (1970).

22. PEARSON, J. D. : Variable metric methods of minimization, Comput. J. 12, 171-178 (1969).

23. POWELL, M. J. D. : A new algorithm for unconstrained optimization, in : Nonlinear programming, ed. J. B. Rosen, O. L. Mangasarian and K. Ritter, Academic Press, New York (1970).

24. POWELL, M. J. D. : A fortran subroutine for unconstrained minimization, requiring first derivatives of the objective functions, A.E.R.E. Harwell report R64-69 (1970).

25. ROBINSON, S. M. : A quadratically convergent algorithm for general nonlinear programming problems, Mathematical Programming 3 , 145-156 (1972).

26. ROBINSON, S. M. : Perturbed Kuhn-Tucker points and rates of convergence for a class of nonlinear-programming algorithms, MRC Tech. Rept. #1298, Madison, Wisconsin, (1973).

27. STEWART, G. W. : A modification of Davidon's minimization method to accept difference approximations of derviatives, J. ACM 14, 72-83 (1967).

28. STOER, J. : On the numerical solution of constrained least-squares problems, SIAM J. Num. Anal. 8, 382-411 (1971).

29. WILSON, R. B. : A simplicial method for concave programming, Ph. D. dissertation, Harvard University, Cambridge, Mass., (1963).