WIS-CS-74-195
COMPUTER SCIENCES DEPARTMENT
The University of Wisconsin
1210 West Dayton Street
Madison, Wisconsin 53706

Received October 15, 1973

SUPERLINEARLY CONVERGENT QUASI-NEWTON ALGORITHMS FOR NONLINEARLY CONSTRAINED OPTIMIZATION PROBLEMS

by

U. M. Garcia Palomares
and
O. L. Mangasarian

Computer Sciences Technical Report #195'

March 1974

SUPERLINEARLY CONVERGENT QUASI-NEWTON ALGORITHMS FOR NONLINEARLY CONSTRAINED OPTIMIZATION PROBLEMS[1)]

by

U. M. Garcia Palomares[2)]

and

O. L. Mangasarian[3)]

ABSTRACT

A class of algorithms for nonlinearly constrained optimization problems is proposed. The subproblems of the algorithms are linearly constrained quadratic minimization problems which contain an updated estimate of the Hessian of the Lagrangian. Under suitable conditions and updating schemes local convergence and a superlinear rate of convergence are established. The convergence proofs require among other things twice differentiable objective and constraint functions, while the calculations use only first derivative data. Rapid convergence has been obtained in a number of test problems by using a program based on the algorithms proposed here.

[1)] Research supported by NSF Grant GJ-35292 at the University of Wisconsin.

[2)] Now at: Instituto Venezolano de Investigaciones Cientificas (IVIC), Departamento De Matematicas, Apartado 1827, Caracas, Venezuela

[3)] Computer Sciences Department, University of Wisconsin, Madison, Wisconsin 53706.

SUPERLINEARLY CONVERGENT QUASI-NEWTON ALGORITHMS FOR NONLINEARLY CONSTRAINED OPTIMIZATION PROBLEMS

1. INTRODUCTION

We develop in this paper a class of algorithms for finding Kuhn-Tucker points of the following nonlinear programming problem:

1.1 $\quad$ minimize $\{f(x) \mid g(x) \leqq 0\}$

where f and g are differentiable functions from R^n into R and R^m respectively. Our algorithms require calculations with gradients, hence f and g should at least be differentiable. However our convergence proofs require, among other things, twice differentiability of f and g. Starting with an initial and possibly infeasible guess x_0, of a solution, and an initial guess u_0 of the Kuhn-Tucker multipliers, the algorithms construct a sequence $\{(x_i,u_i)\}$ which under suitable conditions converges to a Kuhn-Tucker point $(\bar{x},\bar{u})$ of 1.1. The convergence we establish is a local one and the main condition required for it is closeness of the starting point to a point $(\bar{x},\bar{u})$ satisfying the second order sufficiency conditions, the linear independence of the gradients of the active constraints and positivity of the multipliers associated with them. The subproblems solved are linearly constained quadratic programming problems for which efficient and finite algorithms exist [3, 12]. These quadratic subproblems contain an $n \times n$ matrix which is an estimate of the Hessian of the Lagrangian of 1.1. If this estimate is close enough to the Hessian a linear convergence rate is obtained, and if it converges to the Hessian a superlinear rate of convergence is obtained. If the exact Hessian is used the algorithm becomes that of Wilson [14] for which Robinson [11] has established a quadratic convergence rate.

In order to achieve superlinear convergence it is sufficient but not necessary that the updated estimate of the Hessian of the Lagrangian converge to the Hessian. In fact any updating scheme which satisfies a less stringent condition, condition (b) of the algorithm below will achieve superlinear convergence. Thus a "finite difference" approximation of the Hessian similar to that employed by Goldstein and Price [6], in which only gradients are computed, can be used. Numerical experiments on all of Colville's test problems [2] were carried out with a specific updating scheme [5], in which under suitable conditions the updated matrices converge to the Hessian. The test results are quite encouraging and are discussed in more detail in section 4.

The subproblems generated here by matrix updating schemes can be considered either as quadratic approximations of the subproblems of either Robinson's [10] or Wilson's [14] algorithms, both of which converge quadratically [11]. However in Robinson's algorithm the subproblems to be solved have nonlinear objective functions which in general are not quadratic. In Wilson's algorithm explicit evaluation of second derivatives is required which need not be the case in our algorithms. In addition the Lagrangians of our subproblems here can also be considered as estimates of a quadratic approximation of the Lagrangians of the original problem 1.1 [5]. Topkis [13] has proposed a different quadratically convergent algorithm. However his subproblems are nonlinearly constrained.

Throughout this paper all vectors will be column vectors. The transpose will be denoted by the superscript T. The gradient or Jacobian with respect to x will be denoted by ∇, whereas the gradient or Jacobian with respect to any other variable will be denoted by the same symbol ∇ subscripted by that variable. Similarly ∇^2 will denote the Hessian with respect to x, whereas the Hessian with respect to another variable will be denoted by ∇^2 subscripted by that variable. Subscripts to vectors will denote iteration number, while superscripts will denote a vector component. The symbol $\|\cdot\|$ will denote an arbitrary but fixed norm.

2. ALGORITHM

The algorithm is described by the following steps. Let $z = (x,u)$, and let $\nabla^2 L(z)$ denote the $n \times n$ Hessian with respect to x of the Lagrangian $L(z) = f(x) + u^T g(x)$.

Step 1: Set $i = 0$

Step 2: Having $z_i = (x_i, u_i)$ find a Kuhn-Tucker point $z_{i+1} = (x_{i+1}, u_{i+1})$ of the following linearly constrained quadratic program

$$Q(z_i): \begin{cases} \underset{x}{\text{minimize}} \quad \nabla f(x_i)^T (x-x_i) + \frac{1}{2}(x-x_i)G(z_i)(x-x_i) \\ \\ \text{subject to } g(x_i) + \nabla g(x_i)^T (x-x_i) \leqq 0 \end{cases}$$

If z_{i+1} is not unique, choose any Kuhn-Tucker point z_{i+1} which is closest to z_i in terms of the distance $\| z_{i+1} - z_i \|$. Here $G(z_i)$ is any $n \times n$ matrix which satisfies one or more of the following conditions.

a. $\| G(z_i) - \nabla^2 L(z_i) \| \leqq \frac{1}{10\beta}$ (linear convergence)

where β is a constant defined in 3.2 below.

b. $$\left\langle \begin{array}{l} \| G(z_i) - \nabla^2 L(z_i) \| \leqq \frac{1}{10\beta} \\ \\ \left\| \dfrac{(G(z_i) - \nabla^2 L(z_i))(x_{i+1} - x_i)}{\| z_{i+1} - z_i \|} \right\| \to 0 \end{array} \right\rangle$$ (superlinear convergence)

c. $G(z_i) = \nabla^2 L(z_i)$ (quadratic convergence)

Step 3: If some convergence criterion such as $\| x_{i+1} - x_i \| \leq \varepsilon$ is satisfied for some small preassigned positive number ε, stop. Otherwise set $i := i+1$ and go to step 2.

We make some remarks concerning the algorithm.

Remark 1: The key to the convergence of the algorithm is the matrix $G(z_i)$ which is an estimate of the Hessian $\nabla^2 L(z_i)$, and which need be defined on the sequence $\{z_i\}$ only. Condition (a) is a closeness requirement on $G(z_i)$ which is similar to a condition given by Ortega and Rheinboldt for an algorithm for unconstrained optimization [9, p. 401, theorem 12.3.3]. We shall establish in theorem 3.1 below a linear rate of convergence under condition (a), a superlinear rate of convergence under condition (b), and a quadratic rate of convergence under condition (c). In section 4 below we shall give updating schemes for the matrix $G(z_i)$ such that condition (b) is satisfied. When condition (c) is satisfied the above algorithm becomes Wilson's algorithm [14].

Remark 2: The rationale behind employing the quadratic subproblems $Q(z_i)$ is that the Lagrangian associated with it is (to within the constant term $f(x_i)$) equal to a quadratic approximation of the Lagrangian of the original problem (1.1) around z_i, provided that $G(z_i) = \nabla^2 L(z_i)$ [5]. Hence finding Kuhn-Tucker points of $Q(z_i)$ is equivalent to finding Kuhn-Tucker points for a quadratic approximation of $L(z)$ around z_i for case (c) above (Newton method) and to approximately doing that for case (b) above (quasi-Newton method).

Remark 3: The iterates x_i need not be feasible with respect to the original problem, that is $g(x_i) \leq 0$ need not be satisfied. If there are equalities $e^j(x) = 0$, $j=1,\ldots,k$, present in 1.1, they can be replaced by the $k+1$ inequalities $e^j(x) \leq 0$, $j=1,\ldots,k$, and

$$\sum_{j=1}^{k} -e^j(x) \leq 0 .$$

Alternatively the subproblem $Q(z_i)$ can be modified by adding the linear equality constraints $e(x_i) + \nabla e(x_i)^T(x-x_i) = 0$.

Remark 4: The quadratic subproblems $Q(z_i)$ are convex if the matrix $G(z_i)$ is positive semidefinite. Efficient and finite schemes for solving such problems exist [3, 12]. The updating scheme presented in section 4 below generates $G(z_i)$ which under suitable conditions [5] are positive definite matrices.

Remark 5: Condition (a) of the algorithm which also appears in condition (b) can be replaced by the weaker condition

$$\left\| \frac{(G(z_i)-\nabla^2 L(z_i))(x_{i+1}-x_i)}{\|z_{i+1}-z_i\|} \right\| \leqq \frac{1}{10\beta}$$

Remark 6: When $G(z_i)$ is positive definite problem $Q(z_i)$ has another interesting geometric interpretation as a gradient projection problem which is the following. It is equivalent to finding the closest point in the "linearized" set $\{x|g(x_i)+\nabla g(x_i)^T(x-x_i)\leqq 0\}$ to the point $x_i - G(z_i)^{-1}\nabla f(x_i)$ in terms of the norm $\|x\|_{G(x_i)} = (xG(x_i)x)^{1/2}$.

Remark 7: If $g(x)$ is convex and $X = \{x|g(x)\leqq 0\}$ is not empty then the feasible region $X_i = \{x|g(x_i) + \nabla g(x_i)^T(x-x_i) \leqq 0\}$ of $Q(z_i)$ is not empty and contains X, because for x in X

$$0 \geqq g(x) \geqq g(x_i) + \nabla g(x_i)^T(x-x_i)$$

If $g(x)$ is not convex, the theorem below ensures the nomemptiness of X_i under certain conditions including a closeness condition and second order sufficiency conditions. Note also that if X_i is nonempty then $Q(z_i)$ will always have a solution if the original problem 1.1 contains upper and lower bounds on x.

3. CONVERGENCE AND RATE OF CONVERGENCE OF THE ALGORITHM

In the theorem below we shall establish convergence of the algorithm iterates (x_i, u_i) to a Kuhn-Tucker point of our original problem and establish various convergence rates.

We shall use the concept of R-quadratic convergence rate as defined in [9]. We shall also use the second order sufficiency conditions for problem 1.1 [4]. A point $(\bar{x},\bar{u})$ in $R^n \times R^m$ is said to satisfy the second order sufficiency conditions, if it satisfies the first order Kuhn-Tucker conditions and if $y\nabla^2 L(\bar{x},\bar{u})y > 0$ for each nonzero $y \in R^n$ satisfying $\nabla g^j(\bar{x})y = 0$, $j \in \{j|\bar{u}^j>0\}$ and $\nabla g^j(\bar{x})y \leq 0$, $j \in \{j|\bar{u}^j=0,\ g^j(\bar{x})=0\}$. We shall also use the concept of strict complementarity at $(\bar{x},\bar{u})$, that is $\bar{u}_j > 0$ or $g_j(\bar{x}) < 0$ for $j=1,\ldots,m$.

3.1 Convergence and Rate of Convergence

Theorem: Let $(\bar{x},\bar{u})$ be a Kuhn-Tucker point satisfying the second order sufficiency conditions for problem 1.1 with strict complementarity and linear independence of the gradients of the active constraints $\nabla g^j(\bar{x})$, $j \in \{j|g^j(\bar{x})=0\}$. Suppose that f and g have second order derivatives which are Lipschitz continuous in an open neighborhood $N(\bar{x})$ around $\bar{x}$. Then there exist positive numbers δ_1 and δ_2 such that if the algorithm above is started at any point (x_0,u_0) with $\|(x_0,u_0) - (\bar{x},\bar{u})\| < \delta_1$ and $\|G(x_0,u_0) - \nabla^2 L(x_0,u_0)\| < \delta_2$, then the sequence $\{(x_i,u_i)\}$ exists and converges to $(\bar{x},\bar{u})$ with the following R-rates:

a. R-linear convergence rate, that is there exist $\delta \in (0,1)$, $\varepsilon > 0$, $\bar{i} \geq 0$ such that $\|z_i-\bar{z}\| \leq \varepsilon\delta^i$, for all $i \geq \bar{i}$.

b. R-superlinear convergence rate, that is for each $\delta \in (0,1)$, no matter how small, there exist $\varepsilon > 0$, $\bar{i} \geq 0$ such that $\|z_i-\bar{z}\| \leq \varepsilon\delta^i$, for all $i \geq \bar{i}$, provided that condition (b) of the algorithm is satisfied.

(c) R-quadratic convergence rate, that is there exist $\delta \in (0,1)$, $\varepsilon > 0$, $\bar{i} \geq 0$ such that $\| z_i - \bar{z} \| \leq \varepsilon \, \delta^{2^i}$, for all $i \geq \bar{i}$, provided that condition (c) of the algorithm is satisfied.

Proof: We first establish the convergence of the algorithm with an R-linear rate, then establish the other rates of convergence. Our proof for convergence with an R-linear rate is patterned after that of Robinson [10] in which he proves R-quadratic convergence of a different algorithm. Introduce the functions $h : R^{n+m} \to R^{n+m}$ and $d : R^{n+m} \times R^{n+m} \to R^{n+m}$ as follows

$$h(z) = \begin{bmatrix} \nabla f(x) + \nabla g(x)u \\ u^1 g^1(x) \\ \vdots \\ u^m g^m(x) \end{bmatrix} \qquad d(\hat{z},z) = \begin{bmatrix} \nabla f(\hat{x}) + G(\hat{z})(x-\hat{x}) + \nabla g(\hat{x})u \\ u^1(g^1(\hat{x}) + \nabla g^1(\hat{x})^T(x-\hat{x})) \\ \vdots \\ u^m(g^m(\hat{x}) + \nabla g^m(\hat{x})^T(x-\hat{x})) \end{bmatrix}$$

These functions are obtained from the equalities $h(z) = 0$ and $d(\hat{z},z) = 0$ which constitute the equalities of the first-order Kuhn-Tucker conditions for the original problem 1.1 and the quadratic subproblem $Q(\hat{z})$ respectively. McCormick [7] has shown that the second order sufficiency conditions, strict complementarity, and linear independence of the gradients of the active constraints at $(\bar{x},\bar{u})$ insure that $\nabla_z h(\bar{z})$ is nonsingular. Define

3.2 $$\beta = \frac{3}{2} \, \| \nabla_z h(\bar{z})^{-1} \|$$

Since $h(\bar{z}) = 0$ and $\nabla_z h(\bar{z})$ is nonsingular, there is some open neighborhood of $\bar{z}$ which contains no other zero of h, and hence no other Kuhn-Tucker point of 1.1. Combining this with the facts that $\nabla^2 f$ and $\nabla^2 g^j$, j=1,...,m , are Lipschitz continuous on $N(\bar{x})$ and that $\| G(z_i)-\nabla^2 L(z_i)\| \leq \frac{1}{10\beta}$ we conclude that there are positive constants μ and M such that $\bar{z}$ is the unique zero of h in $\bar{B}(\bar{z},\frac{1}{2}\mu)$, the closed ball of radius $\frac{1}{2}\mu$ around $\bar{z}$, and that for any z_{i_1} and z_{i_2} in the open ball $B(\bar{z},\mu)$ we have that

3.3a $$\| \nabla_z h(z_{i_1})-\nabla_z h(\bar{z})\| \leq \frac{1}{2\beta}$$

3.3b $$\| \nabla_z h(z_{i_1})-d'(z_{i_1},z_{i_2})\| \leq \frac{1}{2\beta}$$

where $d'(z_{i_1},z_{i_2})$ is the $(n+m) \times (n+m)$ matrix $\nabla_z d(z_{i_1},z)\big|_{z=z_{i_2}}$,

3.4 $$\| h(z_{i_2})-d(z_{i_1},z_{i_2})\| \leq M \| z_{i_2}-z_{i_1}\|^2 + \|(\nabla^2 L(z_{i_1})-G(z_{i_1}))(x_{i_2}-x_{i_1})\|$$

3.5 $$g^j(x_{i_1}) + \nabla g^j(x_{i_1})(x_{i_2}-x_{i_1}) < 0 \quad \text{for} \quad j \in \{j|g^j(\bar{x})<0\}$$

3.6 $$u_{i_1}^j > 0 \quad \text{for} \quad j \in \{j|\bar{u}^j>0\}$$

We shall need the following two lemmas, the first of which was also established by Robinson for his own algorithm [10].

3.7 Lemma

$$\left.\begin{array}{l}\hat{z} \in B(\bar{z},\frac{1}{2}\mu) \\ 4\beta \| h(\hat{z})\| \leq \mu\end{array}\right\rangle \Longrightarrow \left\langle\begin{array}{l}\exists \text{ a unique Kuhn-Tucker point } \tilde{z} \in \bar{B}(\hat{z},\frac{1}{2}\mu) \\ \text{of } Q(\hat{z}) \text{, and } \| \tilde{z}-\hat{z}\| \leq 2\beta \| h(\hat{z})\| \leq \frac{\mu}{2}\end{array}\right.$$

Proof: See lemma 1 of [10]

3.8 Lemma: The sequence $\{z_i\}$ generated by the algorithm under condition (a) exists, remains in $B(\bar{z}, \frac{\mu}{2})$ and converges R-linearly to $\bar{z}$.

Proof: Since $h(\bar{z}) = 0$ there exists a $\delta_1 \in (0, \frac{1}{4}\mu)$ and usch that

$$z \in B(\bar{z},\delta_1) \Longrightarrow \| h(z) \| \leq \frac{\eta^2}{8M\beta^2}$$

where

$$\eta = \min \{\frac{1}{2}, \frac{\mu M\beta}{2}, (2\mu M\beta)^{\frac{1}{2}}\}$$

Let $\delta_2 = \frac{1}{10\beta}$. Since $z_0 \in B(\bar{z},\delta_1)$ we have that

$$\| z_0 - \bar{z} \| < \delta_1 < \frac{1}{4}\mu \quad \text{and} \quad 4\beta \| h(z_0) \| \leq \frac{\eta^2}{2M\beta} \leq \mu$$

where the last inequality follows from the choice of η which satisfies $\eta^2 \leq 2\mu M\beta$. By lemma 3.7 it follows that there exist a unique Kuhn-Tucker point $\tilde{z}$ of $Q(z_0)$ in $\bar{B}(z_0, \frac{\mu}{2})$ with $\| \tilde{z} - z_0 \| \leq 2\beta \| h(z_0) \|$. Since z_1 is supposed to be a closest Kuhn-Tucker point of $Q(z_0)$ to z_0, we must have that $z_1 = \tilde{z}$ and hence

3.9 $$\| z_1 - z_0 \| \leq 2\beta \| h(z_0) \| \leq \frac{\eta^2}{4M\beta}$$

Since z_1 is a Kuhn-Tucker point of $Q(z_0)$, $d(z_0,z_1) = 0$.
Hence

$$\| h(z_1)\| = \| h(z_1)-d(z_0,z_1)\|$$

$$\leq M \|z_1-z_0\|^2 + \|(\nabla^2 L(z_0)-G(z_0))(x_1-x_0)\| \qquad \text{(by 3.4)}$$

$$\leq (M \|z_1-z_0\| + \|\nabla^2 L(z_0)-G(z_0)\|)\| z_1-z_0\|$$

$$\leq (\frac{\eta^2}{4\beta} + \frac{1}{10\beta}) \frac{\eta^2}{4M\beta} \qquad \text{(by 3.9)}$$

$$= (\frac{\eta^2}{2} + \frac{1}{5}) \frac{\eta^2}{8M\beta^2}$$

Suppose now that for some $i \geqq 1$ and all $k \in \{1,2,\ldots,i\}$ we have that

3.10 $$\| z_k-z_{k-1}\| \leqq 2\beta \| h(z_{k-1})\| , \quad z_{k-1} \in B(\bar{z}, \frac{\mu}{2})$$

and

3.11 $$\| h(z_k)\| \leqq \frac{\eta^2}{8M\beta^2} (\frac{\eta^2}{2} + \frac{1}{5})^k$$

Then

$$\| z_i-\bar{z}\| \leqq \| z_0-\bar{z}\| + \sum_{k=1}^{i} \| z_k-z_{k-1}\|$$

But for any $k \in \{1,\ldots,i\}$

$$\| z_k - z_{k-1} \| \leqq 2\beta \, \| h(z_{k-1}) \| \leqq \frac{\eta}{4M\beta} \left(\frac{1}{2}\right)^{k-1} \quad (\text{since} \quad \eta \leqq \frac{1}{2})$$

Hence using $\| z_0 - \bar{z} \| < \frac{1}{4}\mu$ we get

$$\| z_i - \bar{z} \| < \frac{1}{4}\mu + \frac{\eta}{4M\beta} \sum_{k=1}^{i} \left(\frac{1}{2}\right)^{k-1} \leqq \frac{1}{4}\mu + \frac{\eta}{2M\beta} \leqq \frac{\mu}{2} \quad (\text{since} \quad \eta \leqq \frac{\mu M\beta}{2})$$

Hence $z_i \in B(\bar{z}, \frac{\mu}{2})$. Also from 3.11 we have

$$4\beta \, \| h(z_i) \| \leqq \frac{\eta}{2M\beta} \qquad (\text{since} \quad \eta \leqq \frac{1}{2})$$

$$\leqq \frac{\mu}{4} < \mu \qquad (\text{since} \quad \eta \leqq \frac{\mu M\beta}{2})$$

So we conclude from lemma 3.7 that z_{i+1} exists and is unique in $\bar{B}(z_i, \frac{1}{2}\mu)$ and

3.12 $$\| z_{i+1} - z_i \| \leqq 2\beta \, \| h(z_i) \| \leqq \frac{\mu}{2}$$

Since $d(z_i, z_{i+1}) = 0$ we have that

$$\| h(z_{i+1}) \| = \| h(z_{i+1})-d(z_i,z_{i+1}) \|$$

$$\leqq M \| z_{i+1}-z_i \|^2 + \| \nabla^2 L(z_i)-G(z_i) \| \ \| z_{i+1}-z_i \|$$

(by 3.4 and $\bar{B}(z_i,\frac{1}{2}\mu) \subset B(\bar{z},\mu)$)

$$\leqq 4M\beta^2 \| h(z_i) \|^2 + \frac{1}{10\beta} \ (2\beta \| h(z_i) \|$$

(by 3.12 and condition (a) of algorithm)

$$\leqq [\frac{\eta^2}{2}(\frac{\eta^2}{2}+\frac{1}{5})^i+\frac{1}{5}]\frac{\eta^2}{8M\beta^2}(\frac{\eta^2}{2}+\frac{1}{5})^i \qquad \text{(by 3.11)}$$

$$\leqq \frac{\eta^2}{8M\beta^2}(\frac{\eta^2}{2}+\frac{1}{5})^{i+1} \qquad (\text{since } \eta \leqq \frac{1}{2})$$

Hence by induction the sequence $\{z_i\}$ exists and both 3.10 and 3.11 hold for all $i \geqq 1$. Relations 3.10 and 3.11 imply that

$$\| z_k-z_{k-1} \| \leqq \frac{\eta^2}{4M\beta}(\frac{1}{3})^{k-1}$$

Hence for $i > k$

$$\| z_i-z_k \| \leqq \| z_i-z_{i-1} \| + \ldots . + \| z_{k+1}-z_k \|$$

$$\leqq \frac{\eta^2}{4M\beta}((\frac{1}{3})^{i-1} + \ldots . + (\frac{1}{3})^k) \leqq \frac{3\eta^2}{8M\beta}(\frac{1}{3})^k .$$

Hence $\{z_i\}$ is a Cauchy sequence which converges to some z' in $\bar{B}(\bar{z},\frac{1}{2}\mu)$ with the linear rate

$$\| z_i-z' \| \leqq \frac{3\eta^2}{8M\beta}(\frac{1}{3})^i .$$

Since $\{z_i\}$ converges and since $\nabla^2 f$ and $\nabla^2 g^j$, $j=1,\ldots,m$, are continuous, the sequence $\{\|\nabla^2 L(z_i)\|\}$ is bounded by σ, say. Hence

$$\frac{1}{10\beta} \geqq \| G(z_i)-\nabla^2 L(z_i)\| \geqq \| G(z_i)\| - \|\nabla^2 L(z_i)\| \geqq \| G(z_i)\| - \sigma$$

and the sequence $\{\| G(z_i)\|\}$ is bounded by $\sigma + \frac{1}{10\beta}$. It follows then that the Kuhn-Tucker conditions for $Q(z_i)$

$$\nabla f(x_i) + G(z_i)\,(x_{i+1}-x_i) + \nabla g(x_i)\,u_{i+1} = 0$$

$$u_{i+1}^T\,(g(x_i) + \nabla g(x_i)^T\,(x_{i+1}-x_i)) = 0$$

$$g(x_i) + \nabla g(x_i)^T\,(x_{i+1}-x_i) \leqq 0$$

$$u_{i+1} \geqq 0$$

become in the limit as $(x_i,u_i) \to (x',u')$ the Kuhn-Tucker conditions for problem 1.1. But since $\bar{z}$ is the unique Kuhn-Tucker point of 1.1 in $\bar{B}(\bar{z},\frac{1}{2}\mu)$ and since $z' \in \bar{B}(\bar{z},\frac{1}{2}\mu)$ it follows that $z' = \bar{z}$. This completes the proof of the lemma and establishes part (a) of theorem 3.1.

Since each of conditions (b) and (c) of the algorithm imply condition (a), we have by the lemma 3.8 just proved that the sequence $\{z_i\}$ generated by the algorithm under condition (a), (b) or (c) is well defined and converges R-linearly to $\bar{z}$. To show R-superlinear convergence under condition (b) we have from 3.4, 3.10 and $d(z_i,z_{i+1}) = 0$ that for $i=1,2,\ldots,$

$$3.13 \qquad \frac{\| z_{i+1}-z_i \|}{\| z_i-z_{i-1} \|} \leq 2\beta[M\| z_i-z_{i-1} \| + \|(\nabla^2 L(z_{i-1})-G(z_{i-1}))\frac{(x_i-x_{i-1})}{\| z_i-z_{i-1} \|}\|]$$

If we let β_i denote the term in the square bracket in the inequality above, we have, from condition (b) and the fact that $\{z_i\}$ converges, that $\beta_i \to 0$. Hence

$$\| z_{i+1}-z_i \| \leq \beta_i \| z_i-z_{i-1} \| , \quad \beta_i \to 0$$

and by induction $\| z_{i+1}-z_i \| \leq \beta_i\beta_{i-1}\cdot\ldots\cdot\beta_1 \| z_1-z_0 \|$.

Hence for $k > i$

$$\| z_k-z_i \| \leq \| z_k-z_{k-1} \| + \ldots + \| z_{i+1}-z_i \|$$

$$\leq ((\beta_{k-1}\beta_{k-2}\cdot\ldots\cdot\beta_1) + \ldots + (\beta_i\beta_{i-1}\cdot\ldots\cdot\beta_1)) \| z_1-z_0 \|$$

For a given $\delta \in (0,1)$ pick $\bar{i}$ such that $\beta_i \leq \delta$ for $i \geq \bar{i}$. Then for $k > i \geq \bar{i}$ and $\gamma = \beta_1\beta_2\cdot\ldots\cdot\beta_{\bar{i}}\delta^{-\bar{i}}$

$$\| z_k-z_i \| \leq (\delta^{k-1}+\delta^{k-2}+ \ldots + \delta^i)\gamma \| z_1-z_0 \| \leq \frac{\| z_1-z_0 \|\gamma}{1-\delta}\delta^i$$

Since $z_k \to \bar{z}$ we have upon letting $k \to \infty$ in the last expression that

$$\| z_i-\bar{z} \| \leq \frac{\| z_1-z_0 \|\gamma}{1-\delta}\delta^i \quad \text{for } i \geq \bar{i} ,$$

which establishes the superlinear convergence of the sequence $\{z_i\}$ under condition (b).

We finally establish R-quadratic convergence under condition (c). We have from condition (c) and 3.13 that

$$\| z_{i+1}-z_i \| \leq 2\beta M \| z_i-z_{i-1} \|^2 , \quad i=1,2,\ldots .$$

Pick $\bar{i}$ such that $\gamma = 2\beta M \| z_{\bar{i}+1}-z_{\bar{i}} \| < 1$. We have then by induction that

$$\| z_{i+1}-z_i \| < (2\beta M)^{-1} \gamma^{2^{i-\bar{i}}} , \quad i=\bar{i},\bar{i}+1,\ldots .$$

Hence for $k > i \geq \bar{i}$

$$\| z_k-z_i \| \leq (2\beta M)^{-1} \sum_{\ell=i}^{k-1} \gamma^{2^{\ell-\bar{i}}} \leq (2\beta M)^{-1} \gamma^{2^{i-\bar{i}}} \sum_{\ell=0}^{\infty} \gamma^{2^{i-\bar{i}}(2^{\ell}-1)}$$

$$\leq (2\beta M)^{-1} \gamma^{2^{i-\bar{i}}} \sum_{\ell=0}^{\infty} \gamma^{2^{\ell}-1}$$

By defining $\varepsilon = (2\beta M)^{-1} \sum_{\ell=0}^{\infty} \gamma^{2^{\ell}-1}$ and $\delta = \gamma^{2^{-\bar{i}}}$ we obtain upon letting $k \to \infty$ in the last string of inequalities

$$\| z_i-\bar{z} \| \leq \varepsilon \, \delta^{2^i} , \quad i=\bar{i},\bar{i}+1,\ldots$$

which is the desired R-quadratic convergence rate. This completes the proof of theorem 3.1.

4. COMPUTATIONAL IMPLEMENTATION AND NUMERICAL EXPERIENCE

In order to insure superlinear convergence, condition (b) of the algorithm must be satisfied. For this purpose any updating scheme for an $(n+m) \times (n+m)$ matrix $H(z)$ which satisfies either of the following more stringent conditions will also satisfy condition (b):

4.1 The updated matrix $H(z)$ is an approximation of the Hessian $\nabla_z^2 L(z)$ of the Lagrangian with the property that

$$\| \nabla_z^2 L(z_i) - H(z_i) \| \to 0$$

if $z_i \to \bar{z}$

4.2 The updated matrix $H(z)$ is an approximation of the Jacobian $\nabla_z h(z)$ of the equalities of the Kuhn-Tucker conditions of problem 1.1, with the property that $\| \nabla_z h(z_i) - H(z_i) \| \to 0$ if $z_i \to \bar{z}$ $h(\bar{z}) = 0$ and $\nabla_z h(\bar{z})$ is nonsingular.

It is clear that the upper left corner $n \times n$ submatrix of both $\nabla_z^2 L(z)$ and $\nabla_z h(z)$ is $\nabla^2 L(z) = \nabla^2 f(x) + \nabla^2 ug(x)$. Hence under either scheme 4.1 or 4.2 we obtain that $\| \nabla^2 L(z_i) - G(z_i) \| \to 0$, where $G(z_i)$ is the upper left corner $n \times n$ submatrix of $H(z_i)$, and condition (b) of the algorithm is satisfied.

Another possible choice for $G(z_i)$ is a finite difference approximation to the $n \times n$ matrix $\nabla^2 L(z_i)$, similar to that employed by Goldstein and Price [6]. We show now that this choice for $G(z_i)$ also satisfies the condition $\| \nabla^2 L(z_i) - G(z_i) \| \to 0$. In particular we define the j^{th} column of $G(z_i)$ as

$$G(z^i)^j = \frac{\nabla L(x_i + \theta_i I^j, u_i) - \nabla L(x_i, u_i)}{\theta_i} \qquad j=1,\ldots,n$$

where I^j is the j^{th} column of the $n \times n$ identity matrix, $\theta_i = r\|\nabla L(z_i)\|$ and r is some positive constant. Since f and g have Lipschitz continuous second derivatives in a neighborhood of $\bar{x}$, it follows that for (x,u) and $(x+y,u)$ in a neighborhood of $(\bar{x},\bar{u})$ and for some $K > 0$ that

$$\|\nabla L(x+y,u) - \nabla L(x,u) - \nabla^2 L(x,u)y\| \leqq K \|y\|^2$$

Hence if we let $\nabla^2 L(z_i)^j$ denote the j^{th} column of $\nabla^2 L(z_i)$ we have for $j=1,\ldots,n$ that

$$\|G(z_i)^j - \nabla^2 L(z_i)^j\| \leqq K \|I^j\|^2 \theta_i \leqq K_1 \|\nabla L(z_i)\|$$

where K_1 is a positive constant. Hence for $z_i \in B(\bar{z},\frac{\mu}{2})$

$$\|G(z_i) - \nabla^2 L(z_i)\| \leqq K_2 \|\nabla L(z_i)\| \leqq K_3 \|z_i - \bar{z}\| < K_3 \frac{\mu}{2}$$

where K_2 and K_3 are positive constants, and the last inequality holds by choosing μ small enough. By picking $z_0 \in B(\bar{z},\frac{\mu}{2})$ we have by the above inequality that condition a of the algorithm holds for z_0. By the proof of lemma 3.8, z_1 exists and is in $B(\bar{z},\frac{\mu}{2})$. Hence using the above inequality again we have that z_1 also satisfies condition a. By repeating this argument we have that all z_i satisfy condition a and hence by lemma 3.8 the sequence converges R-linearly to $\bar{z}$. But by the above inequality we also obtain that $\|G(z_i) - \nabla^2 L(z_i)\| \to 0$. Hence by theorem 3.1 the sequence $\{z_i\}$ converges R-superlinearly to $\bar{z}$.

The algorithm was implemented computationally by an updating scheme due to Garcia-Palomares [5] and which under suitable conditions satisfies condition 4.1 above. For explicitness we give the scheme here. Initially set both H_0 and C_0 equal to the $(n+m) \times (n+m)$ identity matrix. Subsequently we have that (using the simplifying convention that the index i is dropped and the index $i+1$ is replaced by 1)

$$s = z_1 - z$$

$$w = \nabla_z L_1 - \nabla_z L - Bs$$

$$t \in (o,1)$$

$$H_1 = H + \frac{t}{s^T Cs}(ws^T C + Csw^T) - \frac{t^2 s^T w}{(s^T Cs)^2} Css^T C$$

$$C_1 = \begin{cases} I & \text{if } i+1 = 0 \mod (n+m) \\ C - \dfrac{t}{s^T Cs} Css^T C & \end{cases}$$

In implementing the algorithm, only the $n \times n$ matrix $G(z_i)$ is updated during the initial iterations. When $\| x_{i+1} - x_i \| \leqq \varepsilon_1$ for some small $\varepsilon_1 > 0$, we begin to update the $(n+m) \times (n+m)$ matrix $H(z_i)$.

Another feature of the implementation of the algorithm prevents the iterates x_i from becoming more infeasible with respect to feasible region $\{x | x \in R^n, g(x) \leqq 0\}$ of problem 1.1. This is done by adding a heuristic stepsize to the algorithm which is determined by decreasing the infeasibility along the direction $x_{i+1} - x_i$ generated by solving problem $Q(z_i)$. In particular we let $r(x_i) = \text{maximum } \{r, g^1(x_i), \ldots, g^m(x_i)\}$ for some small positive number r, and move a sufficiently small amount $\lambda_i > 0$. along $x_i + \lambda(x_{i+1} - x_i)$ such that $r(x_i + \lambda_i(x_{i+1} - x_i)) \leqq r(x_i)$. This was achieved by cutting λ in half until the infeasibility decrease condition $r(x_i+\lambda_i(x_{i+1}-x_i)) \leqq r(x_i)$ is satisfied. This stepsize procedure was an important factor in achieving convergence from starting points that were not close to the solution. That the procedure is well defined can be seen from the following. If for some $j \in \{1,\ldots,m\}$, $g^j(x_i) > 0$, then since $g^j(x_i) + \nabla g^j(x_i)^T(x_{i+1}-x_i) \leqq 0$, we have that $\nabla g^j(x_i)^T(x_{i+1}-x_i) \leqq -g^j(x_i) < 0$. Hence we can always move a sufficiently small amount $\lambda_i > 0$, because the direction $(x_{i+1}-x_i)$ is a downhill direction. If convergence to a solution is obtained with this stepsize, and $\lambda_i = 1$ for sufficiently large i, then the convergence rates given by theorem 3.1 hold.

The algorithm together with the heuristic stepsize was coded in FORTRAN V for the UNIVAC 1108. The principal pivoting method of Cottle and Dantzig [3] was used in solving the subproblems $Q(z_i)$. It was tested on all eight of Colville's test problems [2]. The results given in Table 1 are very encouraging; in more than half the test problems the proposed algorithm was faster than all the algorithms reported on by Colville. We also list in Table 1 the timings obtained by Best [1] and Robinson [10] using their algorithms.

TABLE 1

SUMMARY OF COMPUTATIONAL
PERFORMANCE OF THE QUASI-NEWTON ALGORITHM
ON THE COLVILLE TEST PROBLEMS

		QUASI-NEWTON ALGORITHM				BEST STD. TIME IN COLVILLE REPORT	ROBINSON [11] STANDARD TIME	BEST [1] STANDARD TIME
COLVILLE PROBLEM	STARTING POINT	MINIMUM FUNCTION POINT	EXECUTION TIME (SECS.) UNIVAC 1108	STANDARD TIME[d]	RANK			
I	feasible	-32.3487	.1300	.0048		.0061		
II	feasible	32.3486[a]	10.710	.3967	10	.1228	.1113	
II	infeasible	32.3486[a]	4.7738	.1768	2	.1933	.0877	
III	feasible	-30665.5	.1044	.0039	2	.0052	.0038	
III	infeasible	-30665.5	.1654	.0061	2	.0069	.0064	
IV	feasible	.10E-10[b]	.2632	.0098	4	.0020		
V	feasible	126318.[c]	.2814	.0104	1	.0206		
V	infeasible	125220.[c]	1.8662	.0691	4	.0156		
VI	infeasible	8884.90	.4034	.0149	1	.0264		
VII	infeasible	246.7[a]	1.746	.0647		.0290		
VIII	feasible	-1162.0[a]	4.554	.1687	8	.0224		

a. The answer to the original maximization problem is the negative of the one given in this table, because we minimize -f(x) instead of maximizing f(x)

b. $E\text{-}10 = 10^{-10}$

c. This is the best minimum that has been found for this problem

d. The standard time is determined by dividing the execution time in seconds by 27, a factor which is determined for each computer by a timing package provided by Colville.

Acknowledgement

We are grateful to Ms. Ileana Teodoru for pointing out a flow in an earlier version in theinduction of lemma 3.8.

REFERENCES

1. M. J. Best, "A feasible conjugate direction method to solve linearly constrained minimization problems," Journal of Optimization Theory and Applications, to appear.

2. A. R. Colville, "A comparative study on nonlinear programming codes", IBM New York Scientific Center, Report 320-2949, 1968.

3. R. W. Cottle and G. B. Dantzig, "The principal pivoting method of quadratic programming", in: Mathematics of the Decision Sciences, Ed. G. B. Dantzig and A. F. Veinott, Jr., American Mathematical Society, Providence, R. I., 1968, 144-162.

4. A. V. Fiacco and G. P. McCormick, "Nonlinear programming: SUMT", John Wiley and Sons, New York, 1968.

5. U. M. Garcia Palomares, "Superlinearly convergent quasi-Newton methods for nonlinear programming", Ph.D. dissertation, University of Wisconsin, Madison, 1973.

6. A. A. Goldstein and J. F. Price, "An effective algorithm for minimization", Numer. Math. 10, 1967, 184-189.

7. G. P. McCormick, "Penalty function versus nonpenalty function methods for constrained nonlinear programming problems", Mathematical Programming, 1971, 217-238.

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

9. J. M. Ortega and W. C. Rheinboldt, "Iterative solution of nonlinear equations of several variables", Academic Press, New York, 1970.

10. S. M. Robinson, "A quadratically-convergent algorithm for general nonlinear programming problems", Mathematical programming 3, 1972, 145-156.

11. S. M. Robinson, "Perturbed Kuhn-Tucker points and rates of convergence for a class of nonlinear programming algorithms", MRC Technical Report #1298, University of Wisconsin, Madison, 1973.

12. J. Stoer, "On the numerical solution of constrained least-squares problems", SIAM J. on Numerical Analysis 8, 1971, 382-411.

13. D. M. Topkis, "Superlinear convergence of a second order algorithm for constrained optimization," ORC Technical Report 71-33, University of California, Berkeley, 1971.

14. R. B. Wilson, "A simplicial method for convex programming", Ph.D. dissertation, Harvard University, Cambridge, 1963.