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

QUADRATIC CONVERGENCE OF A NEWTON METHOD FOR NONLINEAR PROGRAMMING [1)]

by

O. L. Mangasarian [2)]

Technical Report #146

March 1972

1) This research was supported by NSF Grant GJ-362.

2) Computer Sciences Department, University of Wisconsin, Madison, Wisconsin 53706.

QUADRATIC CONVERGENCE OF A NEWTON METHOD FOR NONLINEAR PROGRAMMING

Abstract

A Newton algorithm for solving the problem minimize $f(x)$ subject to $g(x) \geq 0$, where $f: R^n \to R$ and $g: R^n \to R^m$ is given for the case when g is concave. At each step a convex quadratic program with linear constraints is solved by means of a finite algorithm to obtain the next point. Quadratic convergence is established.

1. INTRODUCTION

Levitin and Polyak [5] have proposed a Newton method for solving nonlinear programming problems of the form

1.1 $$\underset{x \in X}{\text{minimize}}\ f(x), \qquad X = \{x \mid x \in R^n,\ g(x) \leq 0\}$$

where $f: R^n \to R$ and $g: R^n \to R^m$. The method consists of taking a quadratic approximation f_i of f around a current point x_i, that is

1.2 $$f_i(x) := f(x_i) + \nabla f(x_i)(x - x_i) + \frac{1}{2}(x - x_i)\nabla^2 f(x_i)(x - x_i)$$

where ∇f denotes the n-dimensional gradient vector of f and $\nabla^2 f$ the $n \times n$ Hessian matrix of f, and solving the quadratic programming problem $\underset{x \in X}{\text{minimize}}\ f_i(x)$ to obtain x_{i+1}. Under suitable conditions they show that their algorithm has quadratic convergence (see definition 2.7 below). Unfortunately their method is not practical for nonlinear constraints, that is when g is nonlinear, because each subproblem, $\min_{x \in X} f_i(x)$, is, in general, as difficult as the original problem. In this work we show that for a restricted class of problems of type 1.1, the class of reverse convex problems [12,8,9] that is where g is concave, a practical Newton method is possible. In this method each subproblem consists of a quadratic approximation of f around x_i and a linear approximation of g around x_i. This then gives rise to the following quadratic programming subproblem with linear constraints

1.3 $$\underset{x \in X_i}{\text{minimize}} \; f_i(x), \quad X_i = \{x \mid x \in R^n, \; g(x_i) + \nabla g(x_i)(x - x_i) \leq 0\}$$

where f_i is defined by 1.2 and ∇g is the $m \times n$ Jacobian matrix of g. This subproblem can be efficiently solved by any of the finite and fast quadratic programming algorithms [2,3,13]. We will show that this algorithm also has a quadratic convergence rate.

In Section 2 of the paper we state the algorithm, the assumptions and define r-th order convergence. We also state in Section 2 the convergence theorem for the algorithm. Section 3 and the Appendix contain the proof of the convergence theorem.

2. ALGORITHM, ASSUMPTIONS AND CONVERGENCE RATE

2.1 Algorithm: Start with any x_0 in X. Having x_i we determine x_{i+1} by solving the quadratic program 1.3 by principal pivoting [2,3] or any other finite or fast quadratic programming algorithms [13].

To establish quadratic convergence we shall need the following assumptions:

2.2 $\nabla^2 f$, the Hessian of f, is Lipschitz continuous on X, that is $\|\nabla^2 f(y) - \nabla^2 f(x)\| \leq R\,\|y-x\|$, $\forall x,\ y \in X$, for some $R > 0$

2.3 $M_1 yy \leq y\nabla^2 f(x)y \leq M_2 yy$, $\forall x \in X$, $\forall y \in R^n$, for some $M_1, M_2 > 0$

2.4 $\alpha := \dfrac{2R}{M_1}\,\|x_1 - x_0\| < 1$

2.5 g is continuously differentiable and concave on some open set containing X

2.6 For each $x \in X$, there exists a $z \in R^n$ such that $\nabla g_i(x)z < 0$ for $i \in I(x) = \{i \mid g_i(x) = 0\}$.

We note that the concavity assumption of 2.5 does not make the set X convex except for the degenerate case when g is linear. This case of concave g has been treated by Rosen [12] and Meyer [9,10] using other algorithms and is referred to as the <u>reverse</u> <u>convex</u> case. We also note that the existence of z satisfying $\nabla g_i(x)z < 0$ for $i \in I(x)$, which is a form of the Arrow-Hurwicz-Uzawa constraint qualification [1], is equivalent, by the Gordan theorem, [7, p. 31, Theorem 5] to the positive linear independence of $\nabla g_i(x)$, $i \in I(x)$, that is $\sum_{i \in I(x)} u_i \nabla g_i(x) = 0$, $u_i \geq 0$, $i \in I(x)$, implies that $u_i = 0$, $i \in I(x)$.

We define now r-th order convergence.

2.7 <u>Definition:</u> The sequence $\{x_i\}$ in R^n is said to converge to $\bar{x}$ with order $r \geq 1$ iff for $i = j, j+1, \ldots, j \geq 0$

$$\|x_i - \bar{x}\| \leq \mu \gamma^{r^i} \text{ for some } \mu > 0,\ 0 < \gamma < 1, \text{ if } r > 1$$

$$\|x_i - \bar{x}\| \leq \mu \gamma^{i} \text{ for some } \mu > 0,\ 0 < \gamma < 1, \text{ if } r = 1$$

It can be shown [4] that the number r of definition 2.7 is a lower bound to the root-order convergence factor O_R of Ortega and Rheinholdt [11].

We are ready now to state the main convergence result of this work.

2.8 Quadratic Convergence Theorem. Under assumptions 2.2 to 2.6, the sequence $\{x_i\}$ generated by algorithm 2.1 converges quadratically (that is with $r = 2$ in definition 2.7) to a Kuhn-Tucker point $\bar{x}$ [7, p. 94] of problem 1.1, that is

$$2.9 \quad \begin{cases} \nabla f(\bar{x}) + \bar{u}\,\nabla g(\bar{x}) = 0 \\ \bar{u}\,g(\bar{x}) = 0 \\ g(\bar{x}) \leqq 0 \\ \bar{u} \geqq 0 \end{cases}$$

for some $\bar{u} \in R^m$.

It is interesting to note that convergence of the above algorithm can also be established under different assumptions if we add a step-size selection procedure to the direction-finding quadratic problem 1.3. In fact the dual [7, Chapter 8] of problem 1.3 is the following quadratic program in $u \in R^m$

$$\underset{u \geqq 0}{\text{minimize}} \ \frac{1}{2}\,(\nabla f(x_i) + u\nabla g(x_i))\nabla^2 f(x_i)^{-1}(\nabla f(x_i) + u\nabla g(x_i)) - ug(x_i)$$

with $x - x_i = -\ \nabla^2 f(x_i)^{-1}(\nabla f(x_i) + u\nabla g(x_i))$. This is essentially problem 2.3a" of [8] for which convergence has been established under the fairly general procedure of dual, feasible direction algorithms. This connection may help establish convergence rates for other dual, feasible direction algorithms [8], and may also help in the devising of quadratically convergent algorithms without the concavity restriction on the constraint g.

3. PROOF OF QUADRATIC CONVERGENCE THEOREM

We begin by establishing a lemma which gives a sufficient condition for r-th order convergence.

3.1 Lemma (Sufficient condition for r-th order convergence) If the sequence $\{x_i\}$ in R^n satisfies

$$3.2 \qquad \|x_{i+1} - x_i\| \le \beta \|x_i - x_{i-1}\|^r, \quad i = 1, 2, \ldots, \text{ for some } \beta > 0 \text{ and } r \ge 1$$

and

$$3.3 \qquad \beta \|x_1 - x_0\|^{r-1} < 1$$

then $\{x_i\}$ converges to a limit $\bar{x}$ with order r in the sense of definition 2.7 such that for $i = 0, 1, \ldots$

$$3.4 \qquad \|x_i - \bar{x}\| \le \left(\beta^{\frac{1}{1-r}} \sum_{k=0}^{\infty} \gamma^{r^k - 1}\right) \gamma^{r^i}$$

$$3.5 \qquad \gamma = \beta^{\frac{1}{r-1}} \|x_1 - x_0\| < 1$$

if $r > 1$

$$3.6 \qquad \|x_i - \bar{x}\| \le \frac{\|x_1 - x_0\|}{1 - \gamma} \gamma^i, \quad \gamma = \beta < 1, \quad \text{if } r = 1$$

Proof (Case 1: $r > 1$) We first prove by induction that

3.7 $$\|x_{i+1} - x_i\| \leq \beta^{\frac{1}{1-r}} \gamma^{r^i}, \quad i = 1, 2, \ldots.$$

By 3.2 and 3.5, inequality 3.7 holds for $i = 1$. Suppose 3.7 holds for $i - 1$. Then

$$\begin{aligned}
\|x_{i+1} - x_i\| &\leq \beta \|x_i - x_{i-1}\|^r && \text{(by 3.2)} \\
&\leq \beta \left(\beta^{\frac{1}{1-r}} \gamma^{r^{i-1}}\right)^r && \text{(by induction hypothesis)} \\
&= \beta^{\frac{1}{1-r}} \gamma^{r^i}
\end{aligned}$$

which completes the induction and hence 3.7 holds. Now for $j > i$ we have that

$$\begin{aligned}
\|x_j - x_i\| &\leq \|x_j - x_{j-1}\| + \|x_{j-1} - x_{j-2}\| + \ldots + \|x_{i+1} - x_i\| \\
&= \sum_{k=i}^{j-1} \|x_{k+1} - x_k\| \leq \beta^{\frac{1}{1-r}} \sum_{k=i}^{j-1} \gamma^{r^k} && \text{(by 3.7)}
\end{aligned}$$

Hence

3.8 $$\begin{aligned}
\|x_j - x_i\| &\leq \beta^{\frac{1}{1-r}} \sum_{k=i}^{j-1} \gamma^{r^k} \quad \text{for } j > i \\
&\leq \beta^{\frac{1}{1-r}} \sum_{k=i}^{\infty} \gamma^{r^k} \\
&= \beta^{\frac{1}{1-r}} \gamma^{r^i} \sum_{k=0}^{\infty} \gamma^{r^i(r^k-1)} \\
&\leq \beta^{\frac{1}{1-r}} \gamma^{r^i} \sum_{k=0}^{\infty} \gamma^{r^k-1} && \text{(since } \gamma^{r^i} < \gamma) \\
&= \nu \beta^{\frac{1}{1-r}} \gamma^{r^i}
\end{aligned}$$

where $\nu = \sum_{k=0}^{\infty} \gamma^{r^k - 1}$, which is a positive series for which

$$\frac{\gamma^{r^{k+1}} - 1}{\gamma^{r^k} - 1} = \gamma^{r^k (r-1)} \leqq \gamma^{r-1} < 1$$

and hence is convergent. Hence

3.9 $$\|x_j - x_i\| \leq \nu \beta^{\frac{1}{1-r}} \gamma^{r^i} \qquad \text{for } j > i$$

from which it follows that $\|x_j - x_i\| \to 0$ as $i, j \to \infty$ and hence $\{x_i\}$ is a Cauchy sequence which converges to some $\bar{x}$. By letting $j \to \infty$ in 3.9 we get that

3.10 $$\|x_i - \bar{x}\| \leqq \nu \beta^{\frac{1}{1-r}} \gamma^{r^i} = \beta^{\frac{1}{1-r}} \gamma^{r^i} \sum_{k=0}^{\infty} \gamma^{r^k - 1}$$

which establishes 3.4.

(Case 2: $r = 1$) From 3.3 we have that $\beta < 1$, and from 3.2 we have that

$$\|x_{i+1} - x_i\| \leq \beta^i \|x_1 - x_0\|$$

Hence for $j > i$

$$\|x_j - x_i\| \leq \|x_j - x_{j-1}\| + \cdots\cdots + \|x_{i+1} - x_i\|$$

$$\leq (\beta^{j-1} + \cdots\cdots + \beta^i) \|x_1 - x_0\|$$

$$\leq \frac{\beta^i}{1 - \beta} \|x_1 - x_0\|$$

Hence $\{x_i\}$ is a Cauchy sequence which converges to some limit $\bar{x}$. By letting $j \to \infty$ we get that

$$\|x_i - \bar{x}\| \leq \frac{\beta^i}{1 - \beta} \|x_1 - x_0\|$$

which established 3.6. Q.E.D.

The above lemma 3.1 will help establish the rate of convergence of algorithm 2.1. However establish convergence to a stationary point, that is a point satisfying some necessary optimality criterion, we need the following definition and lemma.

3.11 Definition (Optimality function) An upper semicontinuous nonpositive function θ on X is an optimality function for problem 1.1 iff for each solution $\bar{x}$ of 1.1 $\theta(\bar{x}) = 0$.

If $X = R^n$, a typical optimality function for problem 1.1 is given by $\theta(x) = -\|\nabla f(x)\|^2$ if ∇f is continuous on R^n. If X is a compact convex set in R^n and ∇f is continuous on X, then an optimality function is given by $\theta(x) = \min_{y \in X} \nabla f(x)(y - x)$. We shall need a different optimality function here however, which is given by 3.15 below.

We give now a lemma that establishes convergence to a stationary point.

3.12 Lemma (Convergence to a stationary point) Let $\{x_i\}$ be a Cauchy sequence in the closed set X, and let θ be an optimality function defined by 3.11 for problem 1.1. If for some integers k, ℓ

3.13 $$-\theta(x_i) \leq \rho(x_{i-k}, x_{i-k+1}, \ldots, x_{i+\ell}), \quad i > k,$$

where ρ is some nonnegative function on $R^{k+\ell}$ such that $\lim_{i \to \infty} \rho(x_{i-k}, \ldots, x_{i+\ell}) = 0$, then the limit $\bar{x}$ of the sequence $\{x_i\}$ is stationary, that is $\theta(\bar{x}) = 0$.

Proof From 3.13, $0 \leq -\theta(x_i)$ and $\lim_{i \to \infty} \rho(x_{i-k}, \ldots, x_{i+\ell}) = 0$ we get that

$$\lim_{i \to \infty} -\theta(x_i) = 0$$

and hence by the lower semicontinuity of $-\theta$ we get that

$$-\theta(\bar{x}) \leq \lim_{i \to \infty} -\theta(x_i) = 0$$

which implies that $\theta(\bar{x}) = 0$, since θ is nonpositive on X, and $\bar{x} \in X$ because X is closed. Q.E.D.

We introduce now a specific optimality function associated with the Kuhn-Tucker optimality conditions 2.9 for problem 1.1.

3.14 Lemma (Optimality function associated with Kuhn-Tucker conditions) Let $\bar{x}$ be a solution of problem 1.1, let f be twice continuously differentiable and convex at $\bar{x}$, and let g be differentiable and concave at $\bar{x}$. Then $\theta(\bar{x}) = 0$ where

3.15 $$\theta(x) := \min_{y} \{ \nabla f(x)(y-x) + \frac{1}{2}(y-x)^T [illegible] f(x)(y-x) \mid y \in R^n, [illegible] \}$$

or equivalently the Kuhn-Tucker conditions 2.9 are satisfied at $\bar{x}$, if, in addition, conditions 2.3, 2.5 and 2.6 hold, then θ defined by 3.15 is an optimality function for problem 1.1 in the sense of definition 3.11.

3.16 Remark Under assumption 2.3 the minimum defined in 3.15 exists for any $x \in X$ because y is bounded by the inequality $\|y - x\| \leq \frac{2}{M_1} \|\nabla f(x)\|$, where M_1 is defined by 2.3.

<u>Proof</u> Since g is concave at $\bar{x}$, the reverse convex constraint qualification [7, p. 103] is satisfied and hence [7, p. 105, Theorem 7] the Kuhn-Tucker conditions 2.9 are satisfied at $\bar{x}$.

We show now that satisfying the Kuhn-Tucker conditions at $\bar{x}$ is equivalent to $\theta(\bar{x}) = 0$. By the Farkas theorem [7, p. 31, Theorem 6] the satisfaction of the Kuhn-Tucker conditions 2.9 is equivalent to

3.17 $$\nabla f(\bar{x})z < 0$$
$$\nabla g_i(\bar{x})z \leqq 0, \quad i \in I(\bar{x}) = \{i \mid g_i(\bar{x}) = 0\}$$

having no solution $z \in R^n$. This in turn is equivalent to

3.18 $$\begin{aligned} &\nabla f(\bar{x})z + \frac{1}{2} z\nabla^2 f(\bar{x})z < 0 \\ &g(\bar{x}) + \nabla g(\bar{x})z \leqq 0 \end{aligned}$$

having no solution $z \in R^n$. To see this last equivalence we note first that the forward implication is trivial because its equivalent contrapositive follows from the fact that if z solves 3.18, then z also solves 3.17 because $z\nabla^2 f(\bar{x})z \geqq 0$ [7, p. 89, Theorem 1]. To show the backward implication we prove its equivalent contrapositive, which follows from the fact that if $\bar{z}$ solves 3.17 then $\lambda\bar{z}$ solves 3.18 where

$$\lambda = \min \left\{1, \frac{-\nabla f(\bar{x})\bar{z}}{\bar{z}\nabla f(\bar{x})\bar{z}}, \frac{-g_i(\bar{x})}{|\nabla g_i(\bar{x})\bar{z}|}\right\}, \ i \notin I(\bar{x}) = \{i \mid g_i(\bar{x}) = 0\}$$

Hence 3.18 has no solution $z \in R^n$ which is equivalent to $\theta(\bar{x}) = 0$, upon making the change of variable $z = x - \bar{x}$.

Finally we show that θ as defined by 3.15 is an optimality function in the sense of definition 3.11. We first observe that for any $x \in X$, $g(x) \leq 0$ and hence

$$\theta(x) = \min_y \{\nabla f(x)(y-x) + \frac{1}{2}(y-x)\nabla^2 f(x)(y-x) \mid y \in R^n,$$
$$g(x) + \nabla g(x)(y-x) \leq 0\} \leq 0$$

where the last inequality follows from taking $y = x$. In the Appendix we show that θ is an upper semicontinuous function on X and hence satisfies definition 3.11.

Q.E.D.

We are now ready to prove the main theorem of the paper.

<u>Proof of Theorem 2.8</u> We will show that the algorithm 2.1 generates a sequence $\{x_i\}$ satisfying the assumptions of lemmas 3.1 and 3.12 and hence we have a sequence that converges quadratically to a stationary point, and by lemma 3.14 this is equivalent to a Kuhn-Tucker point.

Since x_{i+1} is a solution of 1.3, then [7, p. 141, Theorem 3i]

$$\nabla f_i(x_{i+1})(x_{i+1} - x_i) \leq 0$$

where f_i is defined by 1.2. This is equivalent to

$$(\nabla f(x_i) + (x_{i+1} - x_i)\nabla^2 f(x_i))(x_{i+1} - x_i) \leq 0$$

and so

$$\begin{aligned} f_i(x_{i+1}) - f(x_i) &= \nabla f(x_i)(x_{i+1} - x_i) + \frac{1}{2}(x_{i+1} - x_i)\nabla^2 f(x_i)(x_{i+1} - x_i) \\ &\leq -\frac{1}{2}(x_{i+1} - x_i)\nabla^2 f(x_i)(x_{i+1} - x_i) \\ &\leq -\frac{M_1}{2}\|x_{i+1} - x_i\|^2 \qquad \text{(by 2.3)} \end{aligned}$$

Hence

$$\|x_{i+1} - x_i\|^2 \leq -\frac{2}{M_1}(f_i(x_{i+1}) - f(x_i))$$

or by 3.15

3.19 $$\|x_{i+1} - x_i\|^2 \leq -\frac{2}{M_1}\theta(x_i)$$

Let

3.20 $$s = -\nabla f(x_i) + \nabla f(x_{i-1}) + \nabla^2 f(x_{i-1})(x_i - x_{i-1})$$

By McLeod's vector mean value theorem [6],

$$s = -\sum_{j=1}^{n} \sigma_j \left[\nabla^2 f(x_j) - \nabla^2 f(x_{i-1})\right](x_i - x_{i-1})$$

for some $\sigma_j \geq 0$, $\sum_{j=1}^{n} \sigma_j = 1$, $x_j \in (x_i, x_{i-1})$. So by 2.2

$$\|s\| \leq \sum_{j=1}^{n} \sigma_j R \|x_j - x_{i-1}\| \; \|x_i - x_{i-1}\|$$

$$\leq \sum_{j=1}^{n} \sigma_j R \|x_i - x_{i-1}\| \; \|x_i - x_{i-1}\| = R \|x_i - x_{i-1}\|^2$$

Hence

3.21 $$\|s\| \leq R \|x_i - x_{i-1}\|^2$$

Now

$$f_i(x_{i+1}) - f(x_i) = \nabla f(x_i)(x_{i+1} - x_i) + \frac{1}{2}(x_{i+1} - x_i)\nabla^2 f(x_i)(x_{i+1} - x_i)$$

$$\geq \nabla f(x_i)(x_{i+1} - x_i)$$

$$= (\nabla f(x_{i-1}) + (x_i - x_{i-1})\nabla^2 f(x_{i-1}) - s)(x_{i+1} - x_i) \quad \text{(by 3.20)}$$

$$= \nabla f_{i-1}(x_i)(x_{i+1} - x_i) - s(x_{i+1} - x_i)$$

$$\geq -s(x_{i+1} - x_i) \qquad \text{(by Theorem 3i, p. 141 [7])}$$

$$\geq -\|s\| \; \|x_{i+1} - x_i\|$$

$$\geq -R\|x_i - x_{i-1}\|^2 \; \|x_{i+1} - x_i\| \quad \text{(by 3.21)}$$

Hence

3.22 $$-\theta(x_i) = -f_i(x_{i+1}) + f(x_i) \leq R \|x_i - x_{i-1}\|^2 \quad \|x_{i+1} - x_i\|$$

Combining 3.19 and 3.22 we get that

3.23 $$\|x_{i+1} - x_i\| \leq \frac{2R}{M_1} \|x_i - x_{i-1}\|^2$$

Conditions 3.23, 2.4 and lemma 3.1 imply that the sequence $\{x_i\}$ generated by the algorithm 2.1 converge quadratically to a limit $\bar{x}$, which must be in X because X is closed. Condition 3.22 and Lemma 3.12 imply that $\theta(\bar{x}) = 0$, and by Lemma 3.14, $\bar{x}$ satisfies the Kuhn-Tucker conditions 2.9.

Q.E.D.

APPENDIX

The upper semicontinuity of θ, defined by 3.15, follows from the following results of Meyer: Lemma 1.3 of [10] and Theorem 4 and Lemmas 3 and 5 of [9, section 2]. For the sake of completeness and because the last reference is an unpublished dissertation we give below the proof of the upper semicontinuity of θ.

A.1 Meyer's Theorem [9,10] Let H be a subset of R^n, let $\varphi: R^n \times H \to R$ be continuous on $R^n \times H$, let $g: H \to R^m$ have continuous first partial derivatives on H, let

A.2 $$\theta(x) = \min_y \{\varphi(y,x) \mid y \in R^n,\ g(x) + \nabla g(x)(y-x) \leq 0\}$$

be well defined for each $x \in H$, and let for each $x \in H$

A.3 $$\nabla g_i(x) z < 0, \qquad i \in I(x) := \{i \mid g_i(x) = 0\}$$

have a solution $z \in R^n$. Then θ is upper semicontinuous on H.

Proof [9, Lemma 3, Section 2]

a) We first show that if $\lim_{i \to \infty} z_i = z$ and for each i $\lim_{j \to \infty} z_{ij} = z_i$ then there exists n_j, $j = 1,2,\ldots$, such that $\lim_{j \to \infty} z_{n_j j} = z$. Let $N(1)$ be chosen such that $\|z_i - z\| < 1$ for $i \geq N(1)$ and let $N'(1)$ be chosen such that $\|z_{N(1)j} - z_{N(1)}\| < 1$ for $j \geq N'(1)$. Suppose we have chosen $N(1), N(2), \ldots, N(k)$ and

$N'(1), N'(2), \ldots, N'(k)$. Choose $N(k+1)$ and $N'(k+1)$ so that $N'(k+1) > N'(k)$, $\|z_i - z\| < 1/(k+1)$ for $i \geq N(k+1)$ and $\|z_{N(k+1)j} - z_{N(k+1)}\| < 1/(k+1)$ for $j \geq N'(k+1)$. Let $N(0) = 1$ and define $n_j = N(\ell)$ when $N'(\ell) \leq j < N'(\ell+1)$. It is easily verified that $z_{n_j j} \to z$ as $j \to \infty$.

b) [9, Theorem 4 and Lemma 5, Section 2] We next show that the point-to-set mapping

$$\Gamma(x) = \{z \mid z \in R^n,\ g(x) + \nabla g(x)(z - x) \leq 0\}$$

is lower semicontinuous at x, that is if $z \in \Gamma(x)$ and $x_i \to x$ then there exist $z_i \in \Gamma(x_i)$ for $i \geq k$, for some k, and $z_i \to z$.

Let $z^* = x + \gamma z$ where z is a solution of A.3 and

$$\gamma = \min \{1, -g_i(x)/2|\nabla g_i(x)z|\},\quad i \notin I(x).$$

Then $g(x) + \nabla g(x)(z^* - x) < 0$. Let z be an arbitrary point in $\Gamma(x)$. It is clear that $\bar{z} = \lambda z^* + (1 - \lambda)z$, $\lambda \in (0, 1]$ also satisfies $g(x) + \nabla g(x)(\bar{z} - x) < 0$. Hence we can construct a sequence $\{z_i\}$ such that $g(x) + \nabla g(x)(z_i - x) < 0$ and $z_i \to z$. If $x_j \to x$, then by the continuity of g and ∇g, $g(x_j) + \nabla g(x_j)(z_i - x_j) < 0$ for sufficiently large j and hence $z_i \in \Gamma(x_j)$ for sufficiently large j. Hence, for every i there exists a sequence $\{z_{ij}\}$ such that z_{ij} belongs to $\Gamma(x_j)$ for every j and $\lim_{j \to \infty} z_{ij} \to z_i$. Hence by part (a) above, there exists a sequence $\{z_{n_j j}\}$ such that $z_{n_j j} \to z$. But $z_{n_j j} \in \Gamma(x_j)$, so we have that Γ is lower semicontinuous at x.

c) We finally prove that θ is upper simicontinuous at x [10, Lemma 1.3]. Let $z \in \Gamma(x)$ be such that $\theta(x) = \varphi(z,x)$ and let $\{x_i\}$ be an arbitrary sequence in H converging to x. Choose $\{x_{n_i}\}$ and $\{z_{n_i}\}$ such that $\theta(x_{n_i}) \to \overline{\lim}\ \theta(x_i)$ and $z_{n_i} \to z$ with $z_{n_i} \in \Gamma(x_{n_i})$. We then have $\theta(x) = \varphi(z,x) = \lim \varphi(z_{n_i}, x_{n_i}) \geq \underline{\lim}\ \theta(x_{n_i}) = \overline{\lim}\ \theta(x_i)$, and hence θ is upper semicontinuous at x. Since x is arbitrary point for which $\theta(x)$ of A.2 is defined, θ is upper semicontinuous at all such points which constitute the set H.

Q.E.D.

The upper semicontinuity of θ as defined in 3.15 follows immediately upon identifying $\varphi(y,x) = \nabla f(x)(y-x) + \frac{1}{2}(y-x)\nabla^2 f(x)(y-x)$, and $H = X$.

ACKNOWLEDGEMENTS

I am grateful to Dr. Robert R. Meyer for pointing out to me the appropriate results in his publications from which the upper semicontinuity of θ follows and to Dr. Hilbert K. Schultz for pointing out a flaw in an earlier proof of Lemma 3.1. I am also grateful to my students T. J. Huang for the proof of the case $r = 1$ of Lemma 3.1, and U. Garcia Palomares for the observation following Theorem 2.8.

REFERENCES

1. K. J. Arrow, L. Hurwicz and H. Uzawa: Constraint qualifications in maximization problems, Naval Research Logistics Quarterly 8, 1961, 175-191.

2. R. W. Cottle: The principal pivoting method of quadratic programming, in Mathematics of the Decision Sciences, Part 1, ed. G. B. Dartzig and A. F. Veinott, Amer. Math. Soc., Providence, R. I., 1968, 144-162.

3. R. W. Cottle and G. B. Dartzig: Complementary pivot theory of mathematical programming, Linear Algebra and Appl. 1, 1968, 103-125.

4. T. J. Huang and O. L. Mangasarian: On superlinear convergence, forthcoming.

5. E. S. Levitin and B. T. Polyak: Constrained minimization methods, USSR Computational Mathematics and Mathematical Physics (English tran.) 6, 1966, 1-50.

6. R. M. McLeod: Mean value theorems for vector valued functions, Proc. Edinburgh Math. Soc. 14, 1965, 197-209.

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

8. O. L. Mangasarian: Dual, feasible direction algorithms, University of Wisconsin, Mathematics Research Center Report #1173, August 1971, IFIPS Colloquium on Optimization, Los Angeles, October 1971, forthcoming Academic Press, New York.

9. R. R. Meyer: The solution of non-convex optimization problems by iterative convex programming, Ph.D. Dissertation, Computer Sciences Department, University of Wisconsin, Madison, 1968.

10. R. R. Meyer: The validity of a family of optimization methods, SIAM J. Control 8, 1970, 41-54.

11. J. Ortega and W. Rheinboldt: Iterative Solution of Nonlinear Equations in Several Variables. Academic Press, New York, 1970.

12. J. B. Rosen: Iterative solution of nonlinear optimal control problems, SIAM J. Control 4, 1966, 223-244.

13. J. Stoer: On the numerical solution of constrained least-squares problems, SIAM J. Num. Anal. 8, 1971, 382-411.

BIBLIOGRAPHIC DATA SHEET	1. Report No. WIS-CS-72-146	2.	3. Recipient's Accession No.

4. Title and Subtitle
QUADRATIC CONVERGENCE OF A NEWTON METHOD FOR NONLINEAR PROGRAMMING

5. Report Date
March 1972

6.

7. Author(s)
O. L. Mangasarian

8. Performing Organization Rept. No.

9. Performing Organization Name and Address
Computer Sciences Department
The University of Wisconsin
1210 West Dayton Street
Madison, Wisconsin 53706

10. Project/Task/Work Unit No.

11. Contract/Grant No.
NSF GJ-362

12. Sponsoring Organization Name and Address

National Science Foundation
Washington, D. C.

13. Type of Report & Period Covered

14.

15. Supplementary Notes

16. Abstracts

A Newton algorithm for solving the problem minimize $f(x)$ subject to $g(x) \leq 0$, where $f: R^n \to R$ and $g: R^n \to R^m$ is given for the case when g is concave. At each step a convex quadratic program with linear constraints is solved by means of a finite algorithm to obtain the next point. Quadratic convergence is established.

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

Nonlinear Programming, Optimization, Newton algorithm, Minimization.

17b. Identifiers/Open-Ended Terms

17c. COSATI Field/Group

18. Availability Statement	19. Security Class (This Report)	21. No. of Pages
Available to the public.	UNCLASSIFIED	23
	20. Security Class (This Page) UNCLASSIFIED	22. Price

FORM NTIS-35 (10-70) USCOMM-DC 40329-P71