

AD-A077 104

WISCONSIN UNIV-MADISON MATHEMATICS RESEARCH CENTER

F/G 12/1

A GLOBALLY AND QUADRATICALLY CONVERGENT ALGORITHM FOR GENERAL N--ETC(U)

JUL 79 M J BEST , J BRAUNINGER , K RITTER

DAA629-75-C-0024

UNCLASSIFIED

MRC-TSR-1973

NL

| OF |
AD-
A077104



END
DATE
FILMED

12-79
DDC

LEVEL

9

AD A 077104

MRC Technical Summary Report #1973

A GLOBALLY AND QUADRATICALLY CONVERGENT
ALGORITHM FOR GENERAL NONLINEAR
PROGRAMMING PROBLEMS

Michael J. Best, Jürgen Bräuninger,
Klaus Ritter and Stephen M. Robinson

Mathematics Research Center
University of Wisconsin-Madison
610 Walnut Street
Madison, Wisconsin 53706

DDC
RECEIVED
NOV 26 1979
E

DDC FILE COPY

July 1979

(Received February 26, 1979)

Approved for public release
Distribution unlimited

Sponsored by

U.S. Army Research Office
P.O. Box 12211
Research Triangle Park
North Carolina 27709

National Research Council
Montreal Road
Ottawa, Ontario K1A 0R6
Canada

National Science Foundation
Washington, D.C. 20550

UNIVERSITY OF WISCONSIN - MADISON
MATHEMATICS RESEARCH CENTER

A GLOBALLY AND QUADRATICALLY CONVERGENT ALGORITHM
FOR GENERAL NONLINEAR PROGRAMMING PROBLEMS

Michael J. Best^{*}, Jürgen Bräuninger^{**}, Klaus Ritter^{**} and Stephen M. Robinson[†]

Technical Summary Report #1973
July 1979

ABSTRACT

↙ This paper presents an algorithm for the minimization of a nonlinear objective function subject to nonlinear inequality and equality constraints. The proposed method has the two distinguishing properties that, under weak assumptions, it converges to a Kuhn-Tucker point for the problem and under somewhat stronger assumptions, the rate of convergence is quadratic. The method is similar to a recent method proposed by Rosen in that it begins by using a penalty function approach to generate a point in a neighborhood of the optimum and then switches to Robinson's method. The new method has two new features not shared by Rosen's method. First, a correct choice of penalty function parameters is constructed automatically, thus guaranteeing global convergence to a stationary point. Second, the linearly constrained subproblems solved by the Robinson method normally contain linear inequality constraints while for the method presented here, only linear equality constraints are required. That is, in a certain sense, the new method "knows" which of the linear inequality constraints will be active in the subproblems. The subproblems may thus be solved in an especially efficient manner.

Preliminary computational results are presented.

AMS(MOS) Subject Classifications - 90C30, 65K05

Key Words - Nonlinear programming algorithms, penalty algorithms

Work Unit Number 5 - Mathematical Programming and Operations Research

* Department of Combinatorics and Optimization, University of Waterloo, Waterloo, Ontario, Canada

** Mathematisches Institut A, Universität Stuttgart, Pfaffenwaldring 57, 7000 Stuttgart 80, West Germany

† Mathematics Research Center, University of Wisconsin-Madison, Madison, Wisconsin

Sponsored by the United States Army under Contract No. DAAG29-75-C-0024, by the National Research Council of Canada under Research Grant A8189, and by the National Science Foundation under Grant No. MCS74-20584 A02.

SIGNIFICANCE AND EXPLANATION

This paper presents a computational method for solving the problem of minimizing a nonlinear function subject to nonlinear inequality and equality constraints. Such optimization problems need to be solved in many practical applications, and one of the difficulties to which most numerical algorithms are subject is that of finding a starting point close enough to the true solution. In this paper we show that by combining two different types of numerical methods, one can produce a method which can be started at any point and which, under fairly weak assumptions, will converge to a (local) solution of the optimization problem no matter how bad the starting point may have been.

Accession For	
NTIS GPO&I	
DDC TAB	
Unannounced	
Justification	
By	
Distribution/	
Availability Codes	
Dist	Avail and/or special
A	

The responsibility for the wording and views expressed in this descriptive summary lies with MRC, and not with the authors of this report.

A GLOBALLY AND QUADRATICALLY CONVERGENT ALGORITHM
FOR GENERAL NONLINEAR PROGRAMMING PROBLEMS

Michael J. Best^{*}, Jürgen Bräuninger^{**}, Klaus Ritter^{**} and Stephen M. Robinson[†]

1) INTRODUCTION, GENERAL DESCRIPTION OF THE METHOD

This paper presents an algorithm for the minimization of a nonlinear objective function subject to nonlinear inequality and equality constraints. The distinguishing features of the proposed method are that, under weak assumptions, it is guaranteed to converge to a Kuhn-Tucker point for the given problem (Theorem 1) and, under somewhat stronger assumptions, the rate of convergence is quadratic (Theorem 2).

The method is somewhat similar to a recently proposed method by Rosen [10]. Like the Rosen method, the new method uses a penalty function to find a point in a neighborhood of the optimum, and then switches to the method of Robinson [9] which, under appropriate assumptions, gives the quadratic rate of convergence.

In both Rosen's method and the method proposed herein, the penalty function is used to locate a point in a neighborhood of the optimum. Robinson [9] has shown that if his method is initiated with a point sufficiently close to the optimum, then convergence will occur and, under appropriate assumptions, the rate of convergence will be R-quadratic. If the penalty function parameter, say μ , is required to approach $+\infty$, then its choice in the first phase of either algorithm is critical. If μ is chosen to be too small, then the point constructed by the penalty function may be too far from the optimum for Robinson's method to converge. Conversely, if μ is chosen to be too large, then although a point sufficiently close to the optimum will be constructed, the resulting penalty function will be numerically difficult and computationally expensive to minimize.

^{*} Department of Combinatorics and Optimization, University of Waterloo,
Waterloo, Ontario, Canada

^{**} Mathematisches Institut A, Universität Stuttgart, Pfaffenwaldring 57,
7000 Stuttgart 80, West Germany

Mathematics Research Center, University of Wisconsin-Madison, Madison, Wisconsin

Sponsored by the United States Army under Contract No. DAAG29-75-C-0024, by the National Research Council of Canada under Research Grant A8189, and by the National Science Foundation under Grant No. MCS74-20584 A02.

Rosen's approach is to use a certain fixed value of μ based on experimental results from a number of small test problems. The penalty function problem is then solved to produce a point which is then used to initiate Robinson's method. The inherent difficulty in this approach is that the choice of μ may be too small so that the resulting point may not be sufficiently close to the optimum for Robinson's method to converge. Rosen's method does not allow for this possibility.

The method proposed here explicitly allows for such a possibility in such a way that under weak assumptions, convergence to a stationary point is guaranteed. Each nonlinear constraint function is associated with its own penalty function parameter. These are initially set to a suitably small value. The penalty function is then minimized. The resulting point is then passed to Robinson's method. One or more iterations of the Robinson method are then performed. After each such iteration, certain tests are performed. Essentially these tests are performed to determine whether or not Robinson's method is converging, or, put another way, the tests determine whether or not the point found by the penalty function is in a suitably small neighborhood of the optimum. If the tests are successfully passed, then an additional iteration of Robinson's method is performed. If not, then certain of the penalty function parameters are increased and a return is made to the penalty function method. Under suitable assumptions, it is shown that after a finite number of steps, these tests are always successfully passed.

For a problem with nonlinear inequality constraints, the Robinson method subproblems contain linearizations of all such constraints and the resulting linear constraints are inequality constraints. In Rosen's approach, therefore, a significant portion of the computational effort required to solve a linearly constrained subproblem is devoted to determining which of these inequalities will be active at the optimum of the subproblem. In the new method, however, only linear equality constraints are passed to the Robinson subproblems as is proposed in [3]. Certain tests are used to decide prior to entering the Robinson subproblem, which of the linear inequalities should be active. It is shown that after a finite number of steps, the solutions to

the resulting subproblems are precisely those which would be determined from the original Robinson method.

Since the Robinson-type subproblems contain linear equality constraints only, an appropriate decrease in the computational effort required for their solution is anticipated.

As in the case for Rosen's approach, any efficient algorithm capable of minimizing a nonlinear objective function subject to linear equality constraints may be used to solve both the penalty function and the Robinson subproblems. Of course, if the given nonlinear problem contains linear inequality constraints as well as nonlinear constraints, the linear inequality constraints should be kept explicit in both the penalty problems and the Robinson subproblems. This can be done in an obvious fashion and for notational simplicity we do not give the details here. We do, however, consider linear equality constraints explicitly.

For the solution of the linearly constrained subproblems, Rosen recommends the algorithm of [7]. Because of their established rapid rates of convergence, we also suggest the algorithms in [2] and [8] as alternatives.

Preliminary computational results are given in Section 6.

Rosen suggests the use of a squared external penalty function. The difficulty with this is that it has discontinuous second derivatives and therefore, the rate of convergence of the algorithm used to solve it, e.g. [2] or [8], may not be as fast as possible. Here we use an external quartic penalty function which does have continuous second derivatives, so that the rapid convergence rates of [2] and [8] will apply. We point out that this particular type of penalty function may not be the most numerically efficient. However, to take advantage of the rapid rate of convergence of [2] or [8], a penalty function with continuous second derivatives should be used.

For any column vector x and matrix H , we use x' and H' to denote transpose, respectively.

2) FORMULATION OF THE PROBLEM

We consider the model problem

PR: $\min\{f(x) \mid x \in M\}, x \in E^n$, where

$$M = \{x \mid h_i(x) \leq 0, i = 1, \dots, m; h_i(x) = 0, i = m+1, \dots, m+q+p\},$$

$$h_i(x) = a_i'x - \beta_i, i = m+q+1, \dots, m+q+p, \text{ and}$$

$f(x), h_1(x), \dots, h_{m+q}(x)$ are continuously differentiable.

The model problem thus has m nonlinear inequality constraints, q nonlinear equality constraints and p linear equality constraints. We use the penalty function

$$P(x; u_1, \dots, u_{m+q}) = f(x) + \sum_{i=1}^m u_i (h_i(x)_+)^4 + \sum_{i=m+1}^{m+q} u_i (h_i(x))^4$$

$$\text{where } h_i(x)_+ = \begin{cases} h_i(x) & \text{if } h_i(x) \geq 0 \\ 0 & \text{if } h_i(x) < 0. \end{cases}$$

We use the following notation for the linearized problems:

$$Lh_i(x; x_j) := h_i(x_j) + \nabla h_i(x_j)'(x - x_j),$$

$$H(x; x_j, u_j) := f(x) - \sum_{i=1}^{m+q} (u_j)_i [h_i(x) - Lh_i(x; x_j)],$$

$$LPH(x_j, u_j) := \min\{H(x; x_j, u_j) \mid Lh_i(x; x_j) \leq 0, i = 1, \dots, m$$

$$Lh_i(x; x_j) = 0, i = m+1, \dots, m+q+p\},$$

$$LPH(x_j, u_j, I) := \min\{H(x; x_j, u_j) \mid Lh_i(x; x_j) = 0, i \in I\},$$

$$\text{where } I \subset \{1, \dots, m+q+p\}.$$

The algorithm requires a procedure to minimize a nonlinear objective function subject to linear equality constraints. As previously mentioned, the algorithms in [2], [7] and [8] are good candidates, but any efficient algorithm would suffice. We refer to such a procedure as procedure Ψ . Procedure Ψ is to solve a problem of the form

$$(*) \quad \min\{\zeta(x) \mid Lh_i(x; x_j) = 0, i \in I\},$$

where $\zeta(x)$ is any nonlinear function and I is a specified index set. If $\zeta(x)$ is equal to $H(x; x_j, u_j)$ and if the minimization problem has Kuhn-Tucker points, it is

assumed that procedure $\bar{\gamma}$ determines the Kuhn-Tucker point $(\bar{x}, \bar{u})$ which is closest to (x_j, u_j) . If $\zeta(x)$ is equal to $P(x; u_1, \dots, u_{m+q})$ and $I \subset \{m+q+1, \dots, m+q+p\}$ then it is assumed that procedure $\bar{\gamma}$ determines a Kuhn-Tucker point $(\bar{x}, \bar{u})$ such that $P(\bar{x}, u_1, \dots, u_{m+q}) \leq \gamma$ where the constant γ is independent of $u_1, \dots, u_{m+q}$; note that this property surely holds if $\bar{x}$ is an optimal solution of (*). In connection with (*) we use the notation

$$\begin{aligned}\bar{x} &= \Phi(\zeta(x), x_j, I), \\ (\bar{u})_i &= \theta_i(\zeta(x), x_j, I) \text{ for all } i \in I, \text{ and} \\ (\bar{u})_i &= 0 \text{ for all } i \notin I.\end{aligned}$$

Thus $(\bar{u})_i$ is the Kuhn-Tucker multiplier, or dual variable, associated with the linearization of constraint i . Note that these dual variables are the negatives of those in e.g., [9].

We remark that the given linear constraints are unchanged by the linearization.

3) THE ALGORITHM

Throughout the paper we require the following assumption to be satisfied.

Assumption 1.

- i) There is a number τ such that $f(x) \geq \tau$ for all x in the set
$$S = \{x | h_i(x) = a_i'x - \beta_i = 0, i = m+q+1, \dots, m+q+p\},$$
- ii) $\max\{h_i(x), i = 1, \dots, m, |h_i(x)|, i = m+1, \dots, m+q\} \rightarrow \infty$ as $\|x\| \rightarrow \infty, x \in S$,
- iii) $M \neq \emptyset$.

The first two parts of the above assumption assure that the subproblems

$$(**) \quad \min\{P(x, u_1, \dots, u_{m+q}) | x \in S\}$$

possess optimal solutions. The last part is a necessary assumption for the primary problem PR to have an optimal solution.

A precise statement of the algorithm for the solution of problem PR is as follows.

Step 0: Choose an initial point x_0 and constants $\alpha > 0, \beta > 0$. Set $I' = \{m+q+1, \dots, m+q+p\}$, $u_1 = \dots = u_{m+q} = 1$, $j = k = 0$ and go to Step 1.

Step 1: Use the procedure Ψ to compute

$$x_{j+1} = \Phi(P(x; u_1, \dots, u_{m+q}), x_k, I')$$

and $(u_{j+1})_i = \theta_i(P(x; u_1, \dots, u_{m+q}), x_k, I')$, for $i \in I'$.

Set $k = j + 1$, replace j by $j + 1$ and go to Step 2.

Step 2: If $(\max h_i(x_k), i = 1, \dots, m, |h_i(x_k)|, i = m+1, \dots, m+q) \leq \beta$, go to Step 3. Otherwise replace u_i by $10u_i$, if $i \leq m$ and $h_i(x_k) > \beta$ or $i > m$ and $|h_i(x_k)| > \beta$ and go to Step 1.

Step 3: Set

$$(u_k)_i = \begin{cases} -4u_i(h_i(x_k))^3 & \text{for } i = 1, \dots, m \\ -4u_i(h_i(x_k))^3 & \text{for } i = m+1, \dots, m+q, \end{cases}$$

$$z_k = \begin{pmatrix} x_k \\ u_k \end{pmatrix}.$$

Determine $\hat{I}_k$ such that $i \in \hat{I}_k$ if and only if $i \leq m$ and

$$|h_i(x_k)| < -(u_k)_i.$$

Set $I_k = \hat{I}_k \cup \{m+1, \dots, m+q+p\}$. If I_k has more than n elements replace u_i by $10u_i$, $i = 1, \dots, m+q$, and go to Step 1; otherwise go to Step 4.

Step 4: Use the procedure Ψ to solve the problem

$$\min \{H(x; x_j, u_j) \mid Lh_i(x; x_j) = 0, i \in I_k\}.$$

If no Kuhn-Tucker point exists replace u_i by $10u_i$, $i = 1, \dots, m+q$, and go to Step 1; otherwise set

$$x_{j+1} = \Phi(H(x; x_j, u_j), x_j, I_k)$$

and

$$(u_{j+1})_i = \begin{cases} \theta_i(H(x; x_j, u_j), x_j, I_k) & \text{if } i \in I_k \\ 0 & \text{if } i \notin I_k \end{cases}$$

Set

$$z_{j+1} = \begin{pmatrix} x_{j+1} \\ u_{j+1} \end{pmatrix},$$

replace j with $j+1$ and go to Step 5.

Step 5: If

$$\|z_j - z_{j-1}\| \leq \alpha \left(\frac{1}{2}\right)^{j-k}$$

$$(u_j)_i \leq 0 \text{ for all } i \in \hat{I}_k, \text{ and}$$

$$lh_i(x_j; x_{j-1}) \leq 0 \text{ for all } i \notin I_k$$

then go to Step 4; otherwise replace μ_i by $10\mu_i$, $i = 1, \dots, m+q$, and go to Step 1.

4) CONVERGENCE

The following three lemmas establish properties of the subproblems (**). They are used to demonstrate the basic convergence theorem (Theorem 1). To simplify the notation we set

$$P_k(x) = P(x; \mu_{1k}, \dots, \mu_{m+q,k})$$

where we assume that the penalties $\mu_{ik} \geq 1$ for all i and k .

Lemma 1

For every $k = 0, 1, \dots$, the problem

$$\min\{P_k(x) \mid x \in S\}$$

has an optimal solution and, therefore, at least one Kuhn-Tucker point.

Proof: Let $\hat{x} \in M$. For every k set

$$S(\hat{x}) = \{x \in S \mid P_k(x) \leq P_k(\hat{x})\}.$$

Then

$$S(\bar{x}) = \{x \in S \mid P_k(x) \leq f(\bar{x})\} \subset \{x \in S \mid \sum_{i=1}^m (h_i(x)_+)^4 + \sum_{i=m+1}^{m+q} (h_i(x))^4 \leq f(\bar{x}) - \tau\},$$

and it follows from the continuity of $P_k(x)$ and part (ii) of Assumption 1 that $S(\bar{x})$ is compact. Thus the problem $\min\{P_k(x) \mid x \in S(\bar{x})\}$ has an optimal solution which is also an optimal solution to $\min\{P_k(x) \mid x \in S\}$.

For $k = 0, 1, \dots$, let $(\bar{x}_k, \bar{u}_k)$ be a Kuhn-Tucker point of the problem $\min\{P_k(x) \mid x \in S\}$ obtained by procedure $\forall$. Then the sequence $\{\bar{x}_k\}$ has the following properties.

Lemma 2

- i) The sequence $\{\bar{x}_k\}$ is bounded.
- ii) For every $\epsilon > 0$ there is $\rho(\epsilon) > 0$ such that, for all k , the following property holds.
 If $i \in \{1, \dots, m\}$ then $u_{ik} \geq \rho(\epsilon)$ implies $h_i(\bar{x}_k) \leq \epsilon$.
 If $i \in \{m+1, \dots, m+q\}$ then $u_{ik} \geq \rho(\epsilon)$ implies $|h_i(\bar{x}_k)| \leq \epsilon$.
- iii) Every cluster point of $\{\bar{x}_k\}$ is in M , provided
 $\min\{u_{ik} \mid i = 1, \dots, m+q\} \rightarrow \infty$ as $k \rightarrow \infty$.

Proof:

- i) For every k , $\bar{x}_k$ is in the set

$$\{x \in S \mid P_k(x) \leq \gamma\}$$

which by part (ii) of Assumption 1 is bounded.

- ii) Because $f(x) \geq \tau$ for all $x \in S$ we have for $i = 1, \dots, m$,

$$\begin{aligned} u_{ik} (h_i(\bar{x}_k)_+)^4 &\leq P_k(\bar{x}_k) - f(\bar{x}_k) \\ &\leq \gamma - \tau. \end{aligned}$$

Thus

$$h_i(\bar{x}_k) \leq \left(\frac{\gamma - \tau}{u_{ik}} \right)^{1/4}.$$

Similarly, for $i = m+1, \dots, m+q$,

$$|h_i(\bar{x}_k)| \leq \left(\frac{\gamma - \tau}{u_{ik}} \right)^{1/4}.$$

iii) Since the sequence $(\bar{x}_k)$ is bounded, it has at least one cluster point.

Suppose $\bar{x}$ is a cluster point and $\bar{x} \notin M$. Then $\bar{x}$ does not satisfy all constraints. Assume $h_1(\bar{x}) \geq 2\epsilon > 0$. It follows that $h_1(\bar{x}_k) > \epsilon$ for infinitely many k . Since $\min\{u_{ik} | i = 1, \dots, m+q\} \rightarrow \infty$ as $k \rightarrow \infty$ this is a contradiction to part (ii) of the lemma. Thus $\bar{x} \in M$.

Lemma 3

Let $\min\{u_{ik} | i = 1, \dots, m+q\} \rightarrow \infty$ as $k \rightarrow \infty$, let $\bar{x}$ be a cluster point of $(\bar{x}_k)$ and suppose $J \subset \{0, 1, \dots\}$ is an infinite set such that $\bar{x}_k \rightarrow \bar{x}$ as $k \rightarrow \infty$, $k \in J$. Define I in such a way that, for $i = 1, \dots, m+q+p$,

$$h_i(\bar{x}) = 0, i \in I, \text{ and } h_i(\bar{x}) < 0, i \notin I.$$

If the vectors $\nabla h_i(\bar{x})$, $i \in I$, are linearly independent then there are numbers u_i such that

$$\nabla f(\bar{x}) = \sum_{i \in I} u_i \nabla h_i(\bar{x}), u_i \leq 0 \text{ for } i \in I, i \leq m$$

and for every $i \in I$,

$$\left. \begin{aligned} -4u_{ik}(h_i(\bar{x}_k))^3 &\rightarrow u_i \text{ for } i \leq m \\ -4u_{ik}(h_i(\bar{x}_k))^3 &\rightarrow u_i \text{ for } i = m+1, \dots, m+q \end{aligned} \right\} \text{ as } k \rightarrow \infty, k \in J.$$

Furthermore, $h_i(\bar{x}_k) < 0$ for $i \notin I$ and $k \in J$ sufficiently large.

Proof: Because of the continuity of $h_i(x)$ we have $h_i(\bar{x}_k) < 0$ for $i \notin I$ and $k \in J$ sufficiently large. Thus

$$\nabla P_k(\bar{x}_k) = \nabla f(\bar{x}_k) + \sum_{\substack{i \in I \\ i \leq m}} 4u_{ik}(h_i(\bar{x}_k))^3 \nabla h_i(\bar{x}_k) + \sum_{i=m+1}^{m+q} 4u_{ik}(h_i(\bar{x}_k))^3 \nabla h_i(\bar{x}_k).$$

Set

$$\begin{aligned} u_{ik} &= -4u_{ik}(h_i(\bar{x}_k))^3 \text{ for } i \in I, i \leq m \\ u_{ik} &= -4u_{ik}(h_i(\bar{x}_k))^3 \text{ for } i = m+1, \dots, m+q. \end{aligned}$$

Since $\bar{x}_k$ is a Kuhn-Tucker point of the problem $\min\{P_k(x) | h_i(x) = 0,$

$i = m+q+1, \dots, m+q+p\}$ there are u_{ik} , $i = m+q+1, \dots, m+q+p$ such that

$$VP_k(\bar{x}_k) = \sum_{i=m+q+1}^{m+q+p} u_{ik} \nabla h_i(\bar{x}_k).$$

Therefore, we have for $k \in J$ sufficiently large,

$$\nabla f(\bar{x}_k) - \sum_{i \in I} u_{ik} \nabla h_i(\bar{x}_k) = 0$$

and

$$\|\nabla f(\bar{x}) - \sum_{i \in I} u_{ik} \nabla h_i(\bar{x})\| \leq \|\nabla f(\bar{x}_k) - \nabla f(\bar{x})\| + \left\| \sum_{i \in I} u_{ik} (\nabla h_i(\bar{x}_k) - \nabla h_i(\bar{x})) \right\|.$$

Because the vectors $\nabla h_i(\bar{x})$, $i \in I$, are linearly independent the sequence $(u_{ik}, i \in I, k \in J)$ is bounded. Let u_i , $i \in I$, be a cluster point of this sequence. Then

$$\left\| \nabla f(\bar{x}) - \sum_{i \in I} u_{ik} \nabla h_i(\bar{x}) \right\| \rightarrow 0 \text{ as } k \rightarrow \infty, k \in J$$

implies

$$\nabla f(\bar{x}) = \sum_{i \in I} u_i \nabla h_i(\bar{x}) \text{ and } u_i \leq 0 \text{ for } i \leq m.$$

Since the representation of $\nabla f(\bar{x})$ in terms of $\nabla h_i(\bar{x})$ is unique, $u_{ik} \rightarrow u_i$, $i \in I$ as $k \rightarrow \infty$, $k \in J$.

We are now ready to present the basic convergence property of the algorithm as follows.

Theorem 1

Let the sequence $\{z_j\} = \{(x_j, u_j)\}$ be generated by the algorithm. If at most finitely many x_j are computed by Step 1 then $\{z_j\}$ converges at least linearly to a Kuhn-Tucker point $\bar{z} = (\bar{x}, \bar{u})$ of PR. If infinitely many x_j are determined by Step 1 then every cluster point $\bar{x}$ of this subsequence is in M . If the gradients of the constraints active at $\bar{x}$ are linearly independent then there is $\bar{u}$ such that $(\bar{x}, \bar{u})$ is a Kuhn-Tucker point.

Proof: Suppose only finitely many x_j are computed by Step 1. Then it follows from Step 5 that $\{z_j\}$ is a Cauchy sequence which converges at least linearly to some z . Furthermore, since $lh_i(x_{j+1}; x_j) \leq 0$ for all $i \in I_k$, z_{j+1} is a Kuhn-Tucker point for $LPH(x_j; u_j)$. By Corollary 1 in [9], z is a Kuhn-Tucker point of PR. The second

part of the theorem follows from Lemma 2 and Lemma 3 because in this case $\min(u_{ij} | i = 1, \dots, m+q) \rightarrow \infty$.

5) QUADRATIC RATE OF CONVERGENCE.

In order to establish the rate of convergence of the method we require the following additional assumption to be satisfied throughout the remainder of the paper.

Assumption 2

The sequence $\{z_j\} = \{(x_j, u_j)\}$ generated by the algorithm converges to a limit point $\bar{z} = (\bar{x}, \bar{u})$ with the following properties.

- i) The gradients of the constraints active at $\bar{x}$ are linearly independent, i.e., $\bar{z}$ is a Kuhn-Tucker point (Theorem 1).
- ii) The strict complementary slackness condition is satisfied at $(\bar{x}, \bar{u})$.
- iii) $f(x)$, $h_i(x)$, $i = 1, \dots, m+q$ are twice continuously differentiable in an open neighborhood of $\bar{x}$.
- iv) The second order sufficiency conditions are satisfied at $(\bar{x}, \bar{u})$.

For $z = (x, u)$ define

$$F(z) = (Vf(x)' - \sum_{i=1}^{m+q+p} (u_i Vh_i(x)', (u_i)_1 h_1(x), \dots, (u_i)_m h_m(x), h_{m+1}(x), \dots, h_{m+p+q}(x))' .$$

It was shown in [6] that $VF(\bar{z})$ is nonsingular. Set $\omega = \|VF(\bar{z})^{-1}\|$.

Prior to demonstrating the main convergence rate result we require the following two lemmas and theorem.

Lemma 4.

Let $\bar{z} = (\bar{x}, \bar{u})$ be as in Assumption 2. Then there is $\mu = \mu(\bar{z}) > 0$ such that the following properties hold

- i) If $i \leq m$ and $h_i(\bar{x}) < 0$ then $Lh_i(x_1; x_2) < 0$ for every $x_1, x_2 \in B(\bar{z}, \mu) = \{z | \|z - \bar{z}\| \leq \mu\}$.
- ii) If $i \leq m$ and $(\bar{u})_i < 0$, then $(u)_i < 0$ for every $z \in B(\bar{z}, \mu)$.
- iii) If $\bar{z} \in B(\bar{z}, \frac{1}{2}\mu)$ and $4\omega\|F(\bar{z})\| \leq \mu$ then there is $\tilde{z} \in B(\bar{z}, \frac{1}{2}\mu)$ such that $\tilde{z}$ is the unique Kuhn-Tucker point of $LPH(\bar{z})$ in $B(\bar{z}, \frac{1}{2}\mu)$.

Proof: The lemma follows from Lemma 1 in [9] and the proof of Theorem 2 in [9].

The following is a restatement of Robinson's result [9].

Theorem 2

There exist numbers $\delta = \delta(\bar{z}) > 0$ and $\bar{\delta} > 0$ such that for any starting point $z_0 \in B(\bar{z}, \delta)$ the sequence $\{z_j\}$ in Robinson's algorithm exists and converges R-quadratically to $\bar{z}$. In particular, there is some constant Q such that

$$\|z_j - \bar{z}\| \leq Q\left(\frac{1}{2}\right)^{(2^j)}$$

Furthermore, for all $j \geq 0$,

$$\|z_{j+1} - z_j\| \leq \bar{\delta}\left(\frac{1}{2}\right)^{(2^j)}$$

$$4\omega\|F(z_j)\| \leq \mu$$

$$\|z_j - \bar{z}\| \leq \frac{\mu}{2}.$$

Proof: This theorem is equivalent to Theorem 2 in [9]. The last three inequalities are not stated in that theorem but are derived in its proof.

The next result justifies our use of linear equality constraints rather than the usual inequality constraints in $LPH(x_j, u_j, I)$.

Lemma 5

- i) Let $I(\bar{z}) \subset \{1, \dots, m + p + q\}$ be defined in such a way that $i \in I(\bar{z})$ if and only if $h_i(\bar{z}) = 0$. If $k \rightarrow \infty$ then for all sufficiently large k the set I_k determined in Step 3 of the algorithm is equal to $I(\bar{z})$.
- ii) There is j_0 such that, for $j \geq j_0$, $LPH(z_j)$ and $LPH(z_j, I(\bar{z}))$ have Kuhn-Tucker points and $z_{j+1}^* = \bar{z}_{j+1}$, where z_{j+1}^* and $\bar{z}_{j+1}$ are the Kuhn-Tucker points obtained by applying procedure Ψ to $LPH(z_j, I(\bar{z}))$ and $LPH(z_j)$, respectively.

Proof:

- i) Suppose $h_i(\bar{z}) < 0$. For k sufficiently large we have $h_i(x_k) < 0$ and $(u_k)_i = 0$. Thus $|h_i(x_k)| > -(u_k)_i$ and $i \notin I_k$. Now suppose $h_i(\bar{z}) = 0$ and $i \leq m$. By the strict complementary slackness condition $(\bar{u})_i < 0$.

Because $h_1(x_k) \rightarrow h_1(\bar{x})$ and, by Lemma 3, $(u_k)_i \rightarrow (\bar{u})_i$ as $k \rightarrow \infty$, we have $|h_1(x_k)| < - (u_k)_i$ for k sufficiently large, i.e. $i \in I_k$.

ii) Choose j_0 such that $j \geq j_0$ implies $z_j \in B(\bar{z}, \frac{1}{2}\nu)$ and $4\omega \|F(z_j)\| \leq \nu$. Then it follows from part (iii) of Lemma 4 that $LPH(z_j)$ has a Kuhn-Tucker point and that $\bar{z}_{j+1}$ is the unique Kuhn-Tucker point of $LPH(z_j)$ in $B(z_j, \frac{1}{2}\nu)$. We claim first that $\bar{z}_{j+1}$ is also a Kuhn-Tucker point of $LPH(z_j, I(\bar{z}))$. To prove this claim we have to show that

$$(\bar{u}_{j+1})_i = 0, i \notin I(\bar{z}), \text{ and } Lh_1(\bar{x}_{j+1}, x_j) = 0, i \in I(\bar{z}).$$

If $i \notin I(\bar{z})$, then $h_1(\bar{x}) < 0$ and, by part (i) of Lemma 4, $Lh_1(\bar{x}_{j+1}, x_j) < 0$. Thus $(\bar{u}_{j+1})_i = 0$ if $i \notin I(\bar{z})$. If $i \in I(\bar{z})$ and $i \leq m$, then $h_1(\bar{x}) = 0$ and, by the strict complementary slackness condition, $(\bar{u})_i < 0$. Therefore it follows from part (ii) of Lemma 4 that $(\bar{u}_{j+1})_i < 0$ which implies $Lh_1(\bar{x}_{j+1}, x_j) = 0$. Now let $\hat{z} = (\hat{x}, \hat{u})$ be any Kuhn-Tucker point of $LPH(z_j, I(\bar{z}))$ in $B(z_j, \frac{1}{2}\nu)$. For every $i \notin I(\bar{z})$ it follows again from part (i) of Lemma 4 that $Lh_1(\hat{x}, x_j) < 0$, whereas for $i \in I(\bar{z}), i \leq m$, the strict complementary slackness condition and part (ii) of Lemma 4 imply $(\hat{u})_i < 0$. Therefore, $\hat{z}$ is a Kuhn-Tucker point of $LPH(z_j)$ and $\hat{z} = z_{j+1}^* = \bar{z}_{j+1}$.

We are now ready to demonstrate the rate of convergence of the algorithm.

Theorem 3

- i) Step 1 of the algorithm occurs at most finitely often.
- ii) The sequence $\{z_j\}$ generated by the algorithm converges R-quadratically to a Kuhn-Tucker point $\bar{z}$ of PR.

Proof:

We first assume that Step 1 occurs at most finitely often and show that then the sequence $\{z_j\}$ converges R-quadratically. Let j_1 be such that $j_1 \geq j_0$ as defined in Lemma 5 and $j \geq j_1$ implies $z_j \in B(\bar{z}, \delta)$. We first prove that, for $j \geq j_1$, the set I_k used in Step 4 of the algorithm is equal to $I(\bar{z})$. Indeed let $i \in I(\bar{z}), i \leq m$.

Then $h_1(\bar{x}) = 0$ and the strict complementary slackness condition and part (ii) of Lemma 4 imply $(u_{j+1})_i < 0$. Thus $i \in I_k$. Conversely, if $i \notin I(\bar{z})$, then $h_1(\bar{x}) < 0$ which by part (i) of Lemma 4 implies $Lh_1(x_{j+1}; x_j) < 0$, i.e., $i \notin I_k$. Therefore, we conclude from part (ii) of Lemma 5 that, for $j \geq j_1$, the sequence generated by the algorithm is identical with the sequence in Robinson's algorithm which, by Theorem 2, converges R-quadratically.

Now suppose Step 1 occurs infinitely often. Let j be such that x_{j+1} is determined by Step 1 of the algorithm. For j sufficiently large, x_{j+1} satisfies the conditions of Step 2 and, by Lemma 5, the set $I_k = I_{j+1}$ defined in Step 3 is equal to $I(\bar{z})$. Therefore, it follows from part (ii) of Lemma 5 that z_{j+2} is determined by Step 4 and is equal to the Kuhn-Tucker point obtained by applying procedure Ψ to $LPH(z_{j+1})$. Hence, $(u_{j+2})_i \leq 0$, $i \in \hat{I}_k$, and $Lh_1(x_{j+2}; x_{j+1}) \leq 0$, $i \notin I_k$. Furthermore, for j sufficiently large, $\|z_{j+2} - z_{j+1}\| \leq \frac{1}{2}\alpha$. This implies that for j sufficiently large the algorithm will remain in the cycle Step 4-Step 5. This contradiction shows that our assumption that Step 1 occurs infinitely often is wrong. This completes the proof of the theorem.

COMPUTATIONAL RESULTS

In this section we present the results of numerical tests on a variety of problems. The problems considered are taken from Himmelblau [5] and Asaadi [1].

Table 1 summarizes the results. It is interesting to note that in all cases, only a single penalty function iteration was required. The final accuracy achieved (maximum of primal and dual infeasibilities) was quite good.

All tests were performed on a CDC CYBER - 174 computer using the FTN compiler with OPT = 2. Colville's standard timing program [4] executed in an average time of 6.144 seconds, and the standardized times given in Table 1 were computed using this time.

We point out that the results for Problem 2 of Colville are not as good as those given in [9] (at least for the starting point for which convergence was obtained in [9]).

The reason for this is that the method presented here always begins with a penalty function iteration, and in some cases this may not be necessary. However, for real problems with possibly bad starting data, it is likely to be safer to begin this way.

For the original Robinson method, convergence did not occur for a particular initial point for Colville Test Problem 2 [9, p. 154]. However, convergence was obtained for the method proposed herein, as indeed is predicted by the theory (Theorem 1).

For the smaller test problems, the execution times should not be taken too literally but simply regarded as "small". This is because their execution times are so small that they are the same order of magnitude as the accuracy of the computer's timing routine.

Problem Source ^a No.	Variables	Constraints			Final Objective Value	Max. Primal Infeasibility	Max. Dual Infeasibility	Penalty Iterations	Subson Iterations	Function Evaluations	Gradient Evaluations	Exec. Time (sec's)	Standardized Exec. Time
		Linear Equality	Linear Inequality	Nonlinear Equality									
1	A	2	0	0	2.64666666666667	0.	0.	1	1	11	13	.019	.0215
2	A	4	0	0	-44.0000000000000	.1E-9	.6E-8	1	3	248	137	.043	.0103
3	A	3	0	0	.111111111111099	0.	.3E-13	1	1	19	21	.015	.0024
4	A	5	0	0	-2.9197004360530	.3E-7	.9E-11	1	2	238	204	.098	.0113
5	A	7	0	0	680.63005737294	.6E-9	.3E-9	1	1	465	274	.165	.0272
6	A	10	0	0	100.01660646158	.8E-10	.1E-8	1	3	974	387	.432	.0723
7	A	20	0	0	133.77827510914	.9E-5	.2E-8	1	2	7260	2224	3.543	.5767
11(3)	M(C)	5	0	0	-200455.530671783	.3E-12	.2E-11	1	3	292	86	.081	.0132
11(3)∞	M(C)	5	0	0	-200455.530671783	.3E-12	.3E-12	1	3	342	109	.094	.0153
18(2)	M(C)	15	0	0	32.34867894387	0.	.5E-5	1	3	3553	1085	2.895	.4875
18(2)∞	M(C)	15	0	0	32.34867894387	0.	.7E-5	1	3	4104	1142	3.381	.5503

∞ infeasible starting point

Table 1
NUMERICAL RESULTS OF TEST PROBLEMS

^a - Asaad [1], G - Galvillo [6], H - Himmelslaw [5].

REFERENCES

- [1] Asaadi, J., "A computational comparison of some non-linear programs", Mathematical Programming 4, 1973, pp. 144-154.
- [2] Best, M.J. and Ritter, K., "A class of accelerated conjugate direction methods for linearly constrained minimization problems", Mathematics of Computation, 30, 1976, pp. 478-504.
- [3] Bräuninger, J., "A modification of Robinson's algorithm for general nonlinear programming problems requiring only approximate solutions of subproblems with linear equality constraints", in: Optimization Techniques (ed. J. Stoer), Lecture Notes in Control and Information Sciences 6, 7, Springer-Verlag, Berlin 1977.
- [4] Colville, A.R., "A comparative study on nonlinear programming codes", IBM, New York Scientific Center, Report No. 320-2949, 1968.
- [5] Himmelblau, D.M., "Applied nonlinear programming", McGraw-Hill, New York, 1972.
- [6] McCormick, G.P., "Penalty function versus nonpenalty function methods for constrained nonlinear programming problems", Mathematical Programming, 1, 1971, pp. 217-238.
- [7] Murtagh, B.A. and Saunders, M.A., "Nonlinear programming for large, sparse systems", Technical Report SOL 76-15, Systems Optimization Lab., Stanford, August, 1976.
- [8] Ritter, K., "A superlinearly convergent method for minimization problems with linear inequality constraints", Mathematical Programming, 4, 1973, pp. 44-71.
- [9] Robinson, S.M., "A quadratically-convergent algorithm for general nonlinear programming problems", Mathematical Programming, 3, 1972, pp. 145-156.
- [10] Rosen, J.B., "Two-phase algorithm for nonlinear constraint problems", Technical Report 77-8, Computer Science Department, University of Minnesota, 1977.

MJB/JB/KR/SMR:ed

(14) SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

MAC-TSA-REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS
BEFORE COMPLETING FORM

1. REPORT NUMBER

1973

2. GOVT ACCESSION NO.

3. RECIPIENT'S CATALOG NUMBER

(9) Technical

4. TITLE (and Subtitle)

A GLOBALLY AND QUADRATICALLY CONVERGENT
ALGORITHM FOR GENERAL NONLINEAR PROGRAMMING
PROBLEMS.

5. TYPE OF REPORT & PERIOD COVERED
Summary Report, no specific
reporting period

6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s)

Michael J. Best, Jurgen Brauningner,
Klaus Ritter and Stephen M. Robinson

8. CONTRACT OR GRANT NUMBER(s)

(15) A8189

DAAG29-75-C-00242

VNSF-MCS74-20584 A02

9. PERFORMING ORGANIZATION NAME AND ADDRESS

Mathematics Research Center, University of
610 Walnut Street Wisconsin
Madison, Wisconsin 53706

10. PROGRAM ELEMENT, PROJECT, TASK
AREA & WORK UNIT NUMBERS

Work Unit Number 5 -
Mathematical Programming
and Operations Research

11. CONTROLLING OFFICE NAME AND ADDRESS

(See Item 18 Below)

12. REPORT DATE

(11) July 1979

13. NUMBER OF PAGES

17

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)

(12) 22

15. SECURITY CLASS. (of this report)

UNCLASSIFIED

15a. DECLASSIFICATION/DOWNGRADING
SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

U.S. Army Research Office National Research Council National Science
P.O. Box 12211 Montreal Road Foundation
Research Triangle Park Ottawa, Ontario K1A 0R6 Washington, D.C. 20550
North Carolina 27709 Canada

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Nonlinear programming algorithms, penalty algorithms

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This paper presents an algorithm for the minimization of a nonlinear
objective function subject to nonlinear inequality and equality constraints.
The proposed method has the two distinguishing properties that, under weak
assumptions, it converges to a Kuhn-Tucker point for the problem and under
somewhat stronger assumptions, the rate of convergence is quadratic. The
method is similar to a recent method proposed by Rosen in that it begins by

(continued)

DD FORM 1473

1 JAN 73

EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

221 200

using a penalty function approach to generate a point in a neighborhood of the optimum and then switches to Robinson's method. The new method has two new features not shared by Rosen's method. First, a correct choice of penalty function parameters is constructed automatically, thus guaranteeing global convergence to a stationary point. Second, the linearly constrained subproblems solved by the Robinson method normally contain linear inequality constraints while for the method presented here, only linear equality constraints are required. That is, in a certain sense, the new method "knows" which of the linear inequality constraints will be active in the subproblems. The subproblems may thus be solved in an especially efficient manner.

Preliminary computational results are presented.