WIS-CS-191-73
COMPUTER SCIENCES DEPARTMENT
The University of Wisconsin
1210 West Dayton Street
Madison, Wisconsin 53706

Received September 14, 1973

ALGORITHMS FOR SOLVING SYSTEMS OF EQUATIONS AND INEQUALITIES WITH APPLICATIONS IN NONLINEAR PROGRAMMING

by

T. J. Huang

Computer Sciences Technical Report #191

September 1973

ALGORITHMS FOR SOLVING SYSTEMS OF EQUATIONS AND INEQUALITIES WITH APPLICATIONS IN NONLINEAR PROGRAMMING[1)]

by

T. J. Huang[2)]

ABSTRACT

Global algorithms are proposed for solving systems of nonlinear equations and inequalities. Global convergence is established under suitable assumptions. For some special cases, the rate of convergence is R-superlinear. A global algorithm for solving convex programming problems is also given. Computational results for one of the algorithms are given.

[1)] Supported by NSF Grant GJ-35292

[2)] Computer Sciences Department, University of Wisconsin, Madison, Wisconsin 53706. Now at the Department of Mathematics, National Taiwan University, Taipei, Taiwan, Republic of China.

ALGORITHMS FOR SOLVING SYSTEMS OF EQUATIONS AND INEQUALITIES WITH APPLICATIONS IN NONLINEAR PROGRAMMING

1. Introduction

In this work we will consider algorithms for solving systems of equations and inequalities. The problem can be stated as follows:

1.1 Find $x \in R^n$ such that $x \in X \subset R^n$

where $X = \{x \in R^n \mid g(x) \leqq 0 \text{ and } h(x) = 0\}$,

$$g : R^n \to R^m \text{ and } h : R^n \to R^\ell .$$

Recently, in [Robinson, 1972] and [Pshenichnyi, 1970] extended Newton's methods for solving 1.1 have been given. Local and R-quadratic convergence has been proved for their algorithms. Based on the global convergence of a generalized algorithm for nonlinear programming in [Mangasarian, 1972], we give globally convergent algorithms for solving 1.1. We rewrite the results of [Mangasarian, 1972] as a generalized algorithm for finding a zero of a non-negative lower semicontinuous function as follows:

1.2 Algorithm: (Finding a zero of a non-negative lower semicontinuous function) Let ϕ and f be functions defined on $D \subset R^n$. Assume that ϕ is a non-negative lower semicontinuous on D and f is differentiable on D. Start with any $x_o \in D$. Having x_i, determine $p_i \in R^n$ such that

1.3 $0 \leqq \sigma(\phi(x_i)) \leqq -\nabla f(x_i)p_i$, where σ is a forcing function from $[0,\infty)$ to $[0,\infty)$. (i.e., $\lim_{i\to\infty} \sigma(t_i) \to 0$ implies $\lim_{i\to\infty} t_i \equiv 0$). If $\nabla f(x_i)p_i = 0$ · stop, else find x_{i+1} by any one of the following stepsize methods.

1.4 Find λ_i such that $f(x_i+\lambda_i p_i) = \min_{\lambda\in[0,1] \text{ or } \lambda\in[0,\infty)} f(x_i+\lambda p_i)$

and let $x_{i+1} = x_i+\lambda_i p_i$.

1.5 Find $\lambda_i = \max\{1,\frac{1}{2},(\frac{1}{2})^2, \ldots\}$ such that

$$f(x_i) - f(x_i+\lambda_i p_2) \geqq c\lambda_i \nabla f(x_i)p_i, \text{ where}$$

$c\in(0,1)$ and let $x_{i+1} = x_i+\lambda_i p_i$.

1.6 <u>Convergence of Algorithm 1.2</u>: Suppose that $\{x_i\}$ can be generated by algorithm 1.2 that ∇f satisfies a Lipschitz condition then either $\{x_i\}$ terminates at x_j, for which $\phi(x_j)=0$, or for every accumulation point $(\bar{x},\bar{p})$ of $\{(x_i,p_i)\}$, $\phi(\bar{x})=0$.

<u>Proof</u>: If $\{x_i\}$ terminates at x_j, by recipe of the algorithm 1.2, we have $\phi(x_j)=0$. If $\{x_i\}$ is infinite, by Theorem 3.10 of [Mangasarian, 1972] we have $\nabla f(\bar{x})\bar{p}=0$. From 1.3 we have $\lim_{j\to\infty}\phi(x_{i_j})=0$ where $x_{i_j}\to\bar{x}$. But since ϕ is lower semicontinuous, $\lim_{j\to\infty} \phi(x_{i_j}) \geqq \phi(\bar{x}) \geqq 0$. This implies that $\phi(\bar{x})=0$. ∎

In Section 2, by adding a stepsize to the locally convergent Robinson algorithm we obtain a global algorithm for which we can show by Theorem 1.6 that every accumulation point of the generated sequence solves 1.1. However, we cannot show that the sequence is convergent. In Section 3, a modified algorithm is given such that the generated sequence is convergent. Moreover, an R-linear rate of convergence can be proved for the general case and an R-superlinear rate of convergence can be established if we further assume that $\nabla(h_j(x)^2)$, $j \in \{1,\ldots \ell\}$, is Lipschitz continuous with order $q > 1$, that is

$$\|\nabla(h_j(x)^2) - \nabla(h_j(y)^2)\| \leqq K\|x-y\|^q \quad \text{for some} \quad q > 1,$$

$x, y \in R^n$ and $j = 1,2\ldots\ell$.

In Section 4, by using the same ideas as those in Sections 2 and 3, a global algorithm for solving convex programming problems is constructed. Finally, computational results for the algorithm of Section 3 are given in Section 5. We note here that $\|\cdot\|$ will denote Euclidean norm throughout this work.

2. A Globally Convergent Algorithm for Solving Systems of Equations and Inequalities

Let us restate problem 1.1 as follows:

2.1 Find $x \in R^n$ such that $x \in X$

where $X = \{x \in R^n | g(x) \leq 0$ and $h(x) = 0\}$, $g \in C^1 : R^n \to R^n$ and $h \in C^1 : R^n \to R^\ell$ (C^1 denotes the class of continuously differentiable functions on the domain of definition).

In [Robinson, 1972] and [Pshenichnyi, 1970], Newton's method was extended to solve 2.1. Applying the idea of Newton's method, both Robinson and Pshenichnyi linearized g and h around the current point at each step. Robinson determined the new point to be the projection of current point on the polyhedron generated by the linearization of h and g around current point. Pshenichnyi determines an intermediate feasible point to the linearized system and then determines the new point along the direction joining the feasible point to the linearized system and the current point by a stepsize method that we will discuss later. For both algorithms, local and R-quadratic convergence has been proved.

In this section a globally convergent algorithm is presented by adding a stepsize to Robinson's algorithm. As we show below, every accumulation point of the sequence $\{x_i\}$ generated by the algorithm solves 2.1. However we can improve this result under additional assumptions and some modification of the algorithm. In Section 3 the algorithm is modified by converting the linearized system to a system that contains only inequalities. This modified algorithm has the following advantages: 1) We have a larger feasible set at each step than the case where linearized equalities are present in which case the corresponding feasible set may be empty. 2) The sequence $\{x_i\}$ itself generated by the modified algorithm is convergent with an R-linear rate. Under additional assumptions on 2.1, we can show that the convergence rate is R-superlinear. We will discuss the details of the modified algorithm in Section 3. Now we state the algorithm.

2.2 <u>Algorithm for solving systems of equations and inequalities</u>

2.3 Set $i = 0$.

Having x_i, let $\tilde{x}_{i+1}$ be the solution of the following problem.

2.4 Minimize $\|x-x_i\|$

subject to

$$g_j(x_i) + \nabla g_j(x_i)(x-x_i) \leq 0, \quad j = 1, 2 \ldots m$$

$$h_j(x_i) + \nabla h_j(x_i)(x-x_i) = 0, \quad j = 1, 2 \ldots \ell$$

Let $p_i = \tilde{x}_{i+1} - x_i$. If $p_i = 0$ stop, otherwise

2.5 Find $x_{i+1} = x_i + \lambda_i p_i$ by any one of the following three stepsize methods:

2.5a (minimization along p_i). Find λ_i such that

$$F(x_i+\lambda_i p_i) = \min_{0\leq\lambda\leq 1} F(x_i+\lambda p_i)$$

2.5b (Armijo's stepsize). Find $\lambda_i = \max\{1, \frac{1}{2}, (\frac{1}{2})^2 \ldots\}$ such that

$$F(x_i) - F(x_i+\lambda_i p_i) \geqq -c\lambda_i \nabla \bar{F}(x_i) p_i,$$

where $c \in (0,1)$.

2.5c (Pshenichnyi's stepsize). Find $\lambda_i = \max\{1, \frac{1}{2}, (\frac{1}{2})^2 \ldots\}$ such that

$$F(x_i+\lambda_i p_i) \leqq (1-\varepsilon\lambda_i)\, F(x_i),$$

where $0<\varepsilon<1$ and

$$F(x) = \frac{1}{2k} \sum_{j=1}^{m} g_j(x)_+^{2k} + \frac{1}{2} \sum_{j=1}^{\ell} h_j(x)^2 ,$$

k = positive integer $\geqq 1$, and

$$g_j(x)_+^t = \begin{cases} g_j(x)^t & \text{if } g_j(x) \geqq 0 \\ 0 & \text{if } g_j(x) < 0 \end{cases}$$

2.6 <u>Properties of $F(x)$</u>: We state here some obvious facts about $F(x)$.

2.6a $F(x) \geqq 0$ for all $x \in R^n$ and $F(x) = 0$ if and only if x solves 2.1.

2.6b $F(x) \in C^1$ if $k \geqq 1$ and $F(x) \in C^2$ if $k \geqq 2$, $g \in C^2$ and $h \in C^2$.

Now we will establish convergence of algorithm 2.2.

2.7 Convergence theorem for algorithm 2.5: Assume that 2.4 is solvable at each step. Assume that one of the following two conditions is satisfied.

2.8 $k = 2$, $g \in C^2$ and $h \in C^2$.

2.9 $\|\nabla F(x) - \nabla F(y)\| \leqq K\|x-y\|$ for $x,y \in R^n$.

Then either the sequence $\{x_i\}$ terminates at $x_{\bar{i}}$ such that $F(x_{\bar{i}})=0$ or every accumulation point $\{\bar{x},\bar{p}\}$ of $\{x_i,p_i\}$ satisfies $F(\bar{x})=0$.

Proof: If $\{x_i\}$ terminates at $x_{\bar{i}}$, $x_{\bar{i}}$ is a solution of 2.4. By 2.4 we have $g_j(x_{\bar{i}}) \leqq 0$ for $j=1,2,\ldots m$ and $h_j(x_{\bar{i}})=0$ for $j=1,2\ldots \ell$. So $x_{\bar{i}}$ solves 2.1. If $\{x_i\}$ is infinite by 2.4 we have

2.10 $g_j(x_i)_+^{2K} + g_j(x_i)_+^{2K-1} \nabla g_j(x_i)p_i \leqq 0$

for all i and $j=1,2\ldots m$,

2.11 $h_j(x_i)^2 + h_j(x_i) \nabla h_j(x_i)p_i = 0$

for all i and $j=1,2\ldots \ell$.

Summing 2.10 from $j=1$ to m and 2.11 from $j=1$ to ℓ, we have

2.12 $0 \leqq F(x_i) \leqq -\nabla F(x_i)p_i$.

Hence by Theorem 1.6, we have proven the theorem if 2.5a or 2.5b have been used. Now, we will show that the theorem still holds if we use Pshenichnyi's stepsize. We assume that $F(x_i)=\delta>0$ for $i \geqq M>0$.

Since $F(x) \in C^1$,

2.13 $F(x_i+\lambda p_i) = F(x_i) + \lambda \nabla F(x_i) p_i + \lambda(\nabla F(\xi_i) - \nabla F(x_i)) p_i$

holds. By 2.8 or 2.9, 2.12 and 2.13, we have

$$F(x_i+\lambda p_i) \leqq [1-\lambda(1-\frac{\lambda}{\delta}\|p_i\|^2 K)]F(x_i) \quad \text{for some} \quad K>0$$

and $i \geqq M$. Since $p_i \to \bar{p}$ as $i \in L$, the inequality of 2.5c holds for $\lambda \leq (1-\varepsilon)\delta/(\|p_i\|^2 K)$, as $i \in L$, where L is a subsequence. By the recipe of 2.5c we know that $2\lambda_i \geqq \min\{2, \frac{(1-\varepsilon)\delta}{\|p_i\|^2 K}\} \geqq \min\{2, \frac{1-\varepsilon}{\alpha K}\} = \delta'' > 0$, $i \in L$, where $\|p_i\|^2 \leqq \alpha$, $i \in L$, since $\{p_i\}_{i \in L} \to \bar{p}$. Hence by the inequality of 2.5c, we have $F(\bar{x})=0$. This is a contradiction. This completes the proof of the theorem. ∎

2.14 Remark: By 2.6a, we can reduce the problem 2.1 to an unconstrained minimization problem as follows.

$$\underset{x \in R^n}{\text{minimize}}\ F(x)$$

We note that $F(x)$ is not a convex function at a solution of 2.1 even if we assume g and h are convex. However efficient algorithms for finding a global solution of a non-convex unconstrained problem are not available.

2.19 <u>Remark</u>: Note that in 2.4 we need consider only violated constraints. We can change 2.4 into 2.4' in algorithm 2.2 as follows:

2.4' $$\min \|x-x_i\|$$

Subject to $$\begin{cases} g_j(x_i) + \nabla g_j(x_i)(x-x_i) \leqq 0 & j \in I_+\{x_i\} \\ h_j(x_i) + \nabla h_j(x_i)(x-x_i) = 0 & j \in E_o\{x_i\} \end{cases}$$

where $I_+\{x_i\} = \{j \mid g_j(x_i) > 0\}$

$E_o\{x_i\} = \{j \mid h_j(x_i) \neq 0\}$

The proof of theorem 2.7 goes through for algorithm 2.2 with 2.4' replacing 2.4.

3. The Modified Algorithm for Solving System of Equations and Inequalities

There are certain disadvantages associated with algorithm 2.2 which we shall try to overcome in a modification of the algorithm. 1) If the feasible region of 2.7 is empty, then algorithm 2.2 will not work. 2) The sequence $\{x_i\}$ generated by algorithm 2.2 may not be convergent itself. Here we will modify algorithm 2.2 so that the feasible region of 2.4 is enlarged and under suitable assumptions the sequence $\{x_i\}$ generated by the modified algorithm will converge R-linearly. For some special cases, as we mentioned in Section 1, we can show that the convergence is R-superlinear.

3.1 Modified algorithm for finding a feasible point of equalities and inequalities: Same as algorithm 2.5 except that subproblem 2.4 is replaced by the following subproblem.

3.2 Minimize $\|x-x_i\|$

subject to
$$g_j(x_i) + \nabla g_j(x_i)(x-x_i) \leq 0 \quad j=1,2\ldots m$$
$$h_j(x_i) + \nabla h_j(x_i)(x-x_i) \leq 0 \quad j\in E_+(x)=\{j\,|\,h_j(x)>0\}$$
$$h_j(x_i) + \nabla h_j(x_i)(x-x_i) \geq 0 \quad j\in E_-(x)=\{j\,|\,h_j(x)\leq 0\}$$

We will make some assumptions here which are required to prove the convergence theorem.

3.3 Assumptions

3.3a (Compactness) The set $\Omega = \{x\in R^n\,|\,F(x)\leq F(x_o)\}$ is compact for the starting point $x_o \in R^n$.

3.3b (Linear independence property) The set of vectors, $\{\nabla g_j(x),\ j=1,2 \ldots m;\ \ \nabla h_j(x),\ \ j=1,2, \ldots \ell\}$ satisfy the LI (linear independence) property, for all $x \in F$. That is for all $x \in F$

$$\sum_{j=1}^{m} \lambda_j \nabla g_j(x) + \sum_{j=1}^{\ell} \mu_j \nabla h_j(x) = 0 \quad \text{and} \quad \lambda_j \geqq 0,$$

$j=1,2 \ldots m$, imply that $\lambda_j=0$, $j=1,2 \ldots m$, $\mu_j=0$, $j=1,2 \ldots \ell$.

3.3c (Lipschitz condition on the gradients of g and h) There exists $K>0$, such that for all x and y in R^n we have

$$\|\nabla g_j(x) - \nabla g_j(y)\| = K\|x-y\| \qquad \text{for } j=1,2 \ldots m$$

$$\|\nabla h_j(x) - \nabla h_j(y)\| = K\|x-y\| \qquad \text{for } j=1,2 \ldots \ell$$

We will give three lemmas which are useful in proving the convergence theorem.

3.4 Lemma: Let $a_1, \ldots a_m$ be m vectors in R^n. Consider the linear system of inequalities:

3.5 $\qquad a_i x \leqq b_i \qquad i=1,2 \ldots m.$

Then 3.5 is solvable for every right hand side $(b_1 \ldots b_m) \in R^m$ if and only if $\{a_i\}_{i=1,2\ldots m}$ are positively linearly independent, i.e. $\sum_{i=1}^{m} \lambda_i a_i = 0$ and $\lambda_i \geqq 0$ for $i=1,2\ldots m$ imply $\lambda_i=0$ for $i=1,2\ldots m$.

Proof: See theorem 2 of [Robinson, 1971]. ∎

Let $a_1, a_2, \ldots, a_m$ be m vectors in R^n. Define $\delta(a_1, \ldots a_m)$ to be the following minimum:

3.6 $$\delta(a_1, \ldots, a_m) := \underset{(\lambda_1, \ldots, \lambda_m) \in S_m}{\text{minimum}} \left[\frac{1}{2} \left\| \sum_{i=1}^{m} \lambda_i a_i \right\|^2 \right],$$

where $S_m = \{(\lambda_1 \ldots \lambda_m) \mid \sum_{i=1}^{m} \lambda_i = 1 \text{ and } \lambda_i \geq 0 \text{ for } i=1,2\ldots m\}$.

We will relate 3.6 and 3.5 by the following lemma.

3.7 Lemma: The linear system 3.5 is solvable for every right hand side $b=(b_1, \ldots b_m) \in R^m$ if and only if $\delta(a_1, \ldots a_m) > 0$. Furthermore, if $\delta(a_1, \ldots a_m) > 0$ then the system 3.5 with right hand side $b_i = -c < 0$ for all i has a solution $\bar{x}$ defined by

$$\bar{x} = -\frac{c}{2\delta(a_1, \ldots a_m)} \sum_{i=1}^{m} \lambda_i a_i$$

where $\frac{1}{2} \left\| \sum_{i=1}^{m} \lambda_i a_i \right\|^2 = \delta(a_1, \ldots a_m)$.

Proof: The linear system 3.5 is solvable for every right hand side $b=(b_1 \ldots b_m) \in R^m$ if and only if $\{a_i\}$ $i=1,2,\ldots m$ are positively linearly independent by Lemma 3.4. Since S_m is a compact set, $\delta(a_1, \ldots a_m)$ attains its minimum in S_m. So $\delta(a_1, \ldots a_m) > 0$ if and only if $\{a_i\}$ $i=1,2 \ldots m$ are positively linearly independent. This proves the first part of the Lemma. Let $\frac{1}{2} \left\| \sum_{i=1}^{m} \lambda_i a_i \right\|^2 = \delta(a_1, \ldots a_m)$. By the

Kuhn-Tucker conditions there exists $\mu_i \geqq 0$ and $v \in R$ such that

$$3.8 \quad \begin{cases} \left(\sum_{i=1}^{m} \lambda_i a_i\right) a_j + v - \mu_j = 0 & \text{for } j=1,2,\ldots m, \\ \sum_{i=1}^{m} \lambda_i = 1 \quad , & \\ \lambda_i \geqq 0 & \text{for } i=1,2,\ldots m, \\ \mu_i \lambda_i = 0 & \text{for } i=1,2,\ldots m, \end{cases}$$

By multiplying the first m equalities of 3.8 by λ_j and summing with respect to j from 1 to m, we have

$$\left\|\sum_{i=1}^{m} \lambda_i a_i\right\|^2 + v = 0 \quad .$$

This implies that $v=-2\delta(a_1, \ldots a_m)$. From the fact that $\mu_j \geqq 0$ and the first m equalities of 3.8, we have

$$3.9 \qquad \left(\sum_{i=1}^{m} \lambda_i a_i\right) a_j \geqq 2\delta(a_1, \ldots a_m) \quad \text{for every } j=1,2,\ldots m.$$

From 3.9 we have $a_j \bar{x} \leqq -c$, for every j. This completes the proof of the Lemma. ∎

The next lemma provides bounds for $\|p_i\|$ in terms of $F(x_i)$.

3.10 Lemma: Suppose that assumption 3.3a and 3.3b hold. Then the sequence $\{x_i\}$ generated by algorithm 3.1 is well defined. Furthermore there exists a constant $C>0$ such that

$$\|p_i\|^2 = \|\tilde{x}_{i+1} - x_i\|^2 \leqq C\, F(x_i) \quad \text{for every } i,$$

where $\tilde{x}_{i+1}$ is a solution vector of 3.2 and $F(x_i)$ is defined in 2.8 with $k = 1$.

Proof: The first part of Lemma is easily proved by 3.3b and Lemma 3.4. Define $\delta_1(x) := \delta(\nabla g_1(x), \ldots, \nabla g_m(x), \varepsilon_1(x)\nabla h_1(x), \ldots \varepsilon_\ell(x)\nabla h_\ell(x))$

where $$\varepsilon_j(x) = \begin{cases} +1 & \text{if } h_j(x) > 0 \\ -1 & \text{if } h_j(x) \leqq 0 \end{cases}$$ and

$$\delta_2(x) := \underset{(\lambda_1 \ldots \lambda_{m+\ell}) \in \bar{S}_{m+\ell}}{\text{minimum}} \left[\| \sum_{j=1}^{m} \lambda_j \nabla g_j(x) + \sum_{j=1}^{\ell} \lambda_{m+j} \nabla h_j(x) \|^2 \right]$$

where $\bar{S}_{m+\ell} = \{(\lambda_1, \ldots \lambda_{m+\ell}) \mid \sum_{j=1}^{m} \lambda_j + \sum_{j=1}^{\ell} |\lambda_{m+j}| = 1, \quad \lambda_j \geqq 0 \quad \text{for}$

$j=1,2 \ldots m\}$.

Since $\bar{S}_{m+\ell}$ is compact, $\delta_2(x)$ attains its minimum in $\bar{S}_{m+\ell}$. So $\{\nabla g_j(x),\ j=1,2 \ldots m;\ \nabla h_j(x),\ j=1,2 \ldots \ell\}$ satisfy LI property if and only if $\delta_2(x) > 0$. Since $\bar{S}_{m+\ell}$ is compact, $\delta_2(x)$ is continuous on F. But F is compact, so $\delta_2(x) \geqq \delta > 0$ for all $x \in F$. By the definition of $\delta_1(x)$,

$$\delta_1(x) = \underset{(\lambda_1, \ldots \lambda_{m+\ell}) \in S_{m+\ell}}{\text{minimum}} \| \sum_{j=1}^{m} \lambda_j \nabla g_j(x) + \sum_{j=1}^{\ell} \lambda_{m+j} \varepsilon_j(x) \nabla h_j(x) \|^2$$

$$= \underset{(\lambda_1 \ldots \lambda_m, \lambda'_{m+1} \ldots \lambda'_{m+\ell}) \in S'_{m+\ell}}{\text{minimum}} \| \sum_{j=1}^{m} \lambda_j \nabla g_j(x)$$

$$+ \sum_{j=1}^{\ell} \lambda'_{m+j} \nabla h_j(x) \|^2$$

where $S'_{m+\ell}=\{\lambda_1 \ldots \lambda_m, \lambda'_{m+1}, \ldots \lambda'_{m+\ell}) \mid \sum_{j=1}^{m} \lambda_j + \sum_{j=1}^{\ell} |\lambda'_{m+j}| = 1,$
$\lambda_j \geqq 0, \quad j=1,2\ldots m$ and $\lambda'_{m+j}\ \varepsilon_j(x) \geqq 0, \quad j=1,2,\ldots \ell\}.$
So $\delta_1(x) \geqq \delta_2(x)$ by $\bar{S}_{m+\ell} \supset S'_{m+\ell}$. This implies
$\delta_1(x) \geqq \delta > 0$ for all $x \in F$.

Let $\{x_i\}$ be the sequence generated by algorithm 3.1. If $\{x_i\}$ is finite, then the lemma follows trivially. If $\{x_i\}$ is infinite, then $F(x_i)>0$ for every i. Define $K_i := F(x_i)$. By taking $-c=-\sqrt{K_i} < 0$ in Lemma 3.7, there exists $\|p'_i\| < \frac{\sqrt{K_i}}{2\delta_1(x_i)^{\frac{1}{2}}}$ such that p'_i is solution of the following linear system:

$$\sqrt{K_i} + \nabla g_j(x_i)p \leqq 0 \qquad j=1,2\ldots m$$

$$\sqrt{K_i} + \varepsilon_j(x_i)\ \nabla h_j(x)p \leqq 0 \qquad j=1,2,\ldots \ell$$

Since $\sqrt{K_i} \geqq g_j(x_i)$ for $j=1,2\ldots m$ and $\sqrt{K_i} \geqq |h_j(x)|$ for $j=1,2\ldots\ell$, so p'_i is also a feasible point of problem 3.2. This implies that

$$\|p_i\| \leqq \frac{\sqrt{K_i}}{2\delta_1(x_i)^{\frac{1}{2}}} \leqq \frac{\sqrt{K_i}}{2\delta^{\frac{1}{2}}} \qquad \text{for} \quad i=0,1,2\ldots$$

So $\|p_i\|^2 \leqq \frac{1}{4\delta} F(x_i)$. The theorem is proved by taking $C = \frac{1}{4\delta}$. ∎

Now we will prove the convergence theorem for algorithm 3.1.

3.11 Convergence and R-linear rate of convergence of algorithm 3.1

Let assumptions 3.3 be satisfied. Then the sequence $\{x_i\}$ generated by algorithm 3.1 is well defined and if $k=1$ in 2.8 and $\{x_i\}$ is infinite, then $\{x_i\}$ converges R-linearly to an $\bar{x}$ which solves 2.1.

Proof: That the sequence $\{x_i\}$ is well defined follows from the assumption 3.3b. By Lemma 3.10, there exists $C>0$ such that

3.12 $\|p_i\|^2 \leqq CF(x_i)$ for all i. By 3.2 we have

3.13 $\nabla F(x_i)p_i \leqq -F(x_i) < 0$.

Since $F \in C^1$ from 3.3c, 3.12 and 3.13 we have

3.14
$$\begin{aligned} F(x_i+\lambda p_i) &= F(x_i) + \lambda\nabla F(x_i)p_i + \lambda[\nabla F(\xi_i)-\nabla F(x_i)]p_i \\ &\leq F(x_i) + (-\lambda F(x_i)) + \lambda^2 CMF(x_i) \\ &= F(x_i) - \lambda F(x_i)(1-\lambda CM) \quad \text{for some } M>0 \\ &\qquad \text{and every } i \end{aligned}$$

where $\|\nabla F(x) - \nabla F(y)\| = M\|x-y\|$ (This is true because of 3.3c) If Pshenichnyi's stepsize has been used, we have

3.15 $\lambda_i \geqq \min\{1, \frac{1-\varepsilon}{2CM}\} = \delta > 0$

By the inequality of 2.8c and 3.15,

3.16 $F(x_{i+1}) \leqq (1-\delta\varepsilon)\ F(x_i)$ for every i.

This implies, since $0 < 1 - \delta\varepsilon < 1$, that $F(x_i)$ converges to 0 as i tends to ∞.

Since 3.12 holds for every i,

3.17 $\qquad \|x_{i+1} - x_i\| \leqq \sqrt{CF(x_i)}$ for every i.

But $\sqrt{F(x_i)}$ converges to 0 with $Q_1(\{\sqrt{F(x_i)}\}) < 1$. We conclude that $\{x_i\}$ converges to $\bar{x}$ R-linearly and $F(\bar{x})=0$ by 3.12 and Lemma 3.1 of [Huang, 1973a]. This proves the theorem for the case when the Pshenichnyi's stepsize has been used. If the Armijo's stepsize has been used, λ_i can be obtained after finite division of unity by 2 because $\nabla F(x_i)p_i < 0$. From 3.2 we have $\nabla F(x_i)p_i \leqq -F(x_i)$. By the inequality of 2.5b, we have $F(x_{i+1}) \leqq (1-c\lambda_i)F(x_i)$ for every i, where λ_i is obtained by using Armijo's stepsize. Similarly to the proof for Pshenichnyi's stepsize we can show that the theorem also holds for Armijo's stepsize. Finally, if 2.5a has been used, 3.16 also holds. The same proof for Pshenichnyi's stepsize applies to 2.5a. This completes the proof of the theorem. ∎

The next theorem will show that algorithm 3.1 converges R-superlinearly under an additional assumption on h.

3.18 Theorem (R-superlinear convergence of algorithm 3.1). Let the hypotheses of theorem 3.11 hold and let h satisfy

3.19 $\qquad \|\nabla(h_j(x)^2) - \nabla(h_j(y)^2)\| \leqq K\|x-y\|^q$ for some $q > 1$,

$j=1, \ldots \ell$ and all x and y in R^n. Then the sequence $\{x_i\}$ generated by algorithm 3.1 is well defined and if $k=1$, $\{x_i\}$ is infinite and 2.5a and 2.5c have been used to find x_{i+1}, $\{x_i\}$ converges to $\bar{x}$ R-superlinearly, where $\bar{x}$ solves 2.1.

Proof: That the sequence is well defined follows from 3.3b. By Lemma 3.10, there exists $C > 0$ such that 3.12 holds for every i. Since $g \in C^1$, from 3.12, 3.3c and 3.2 we have

$$3.20 \qquad g_j(x_i+\lambda p_i) = g_j(x_i) + \lambda \nabla g_j(x_i)p_i + \lambda(\nabla g_j(\xi_i) - \nabla g_j(x_i))p_i$$
$$\leq (1-\lambda)g_j(x_i) + \lambda^2 KCF(x_i) \quad \text{for some } \xi_i \in R^n \text{ and for } j=1,2 \ldots m.$$

So, $0 \leq g_j(x_i+\lambda p_i)_+ \leq (1-\lambda)g_j(x_i)_+ + \lambda^2 KCF(x_i)$. Square both sides of above inequality, we get

$$3.21 \qquad g_j(x_i+\lambda p_i)_+^2 \leq (1-\lambda)^2 g_j(x_i)_+^2 + 2\lambda^2(1-\lambda)KCg_j(x_i)_+F(x_i) + \lambda^4K^2C^2F(x_i)^2$$
$$\leq g_j(x_i)_+^2 - \lambda g_j(x_i)_+^2 + 2\lambda^2(1-\lambda)KCg_j(x_i)_+F(x_i) + \lambda^4K^2C^2F(x_i)^2.$$

Again since $h^2 \in C^1$, from 3.12, 3.19 and 3.2 we have

$$3.22 \qquad h_j(x_i+\lambda p_i)^2 \leq (1-\lambda)h_j(x_i)^2 + 2\lambda^{1+q}K[CF(x_i)]^{\frac{1+q}{2}} \quad \text{for } j=1, 2, \ldots \ell.$$

Summing 3.21 with respect to j from 1 to m and 3.22 with j from 1 to ℓ, we have

$$3.23 \qquad F(x_i+\lambda p_i) \leqq F(x_i) - \lambda F(x_i)\ [1-M\lambda^r F(x_i)^r]$$

for some $M > 0$, $r = \min\{\frac{1}{2}, \frac{q-1}{2}\}$ and i sufficiently large. Since $F(x_i) \to 0$ by Theorem 3.11, we have that $\lambda_i = 1$ for $i \geqq N > 0$ if Pshenichnyi's stepsize has been used. This implies that

$$3.24 \qquad F(x_{i+1}) \leqq MF(x_i)^{1+r} \quad \text{for} \quad i \geqq N > 0$$

By Lemma 3.10 and Lemma 3.1 of [Huang, 1973a], we can conclude that $\{x_i\}$ converges to $\bar{x}$ R-superlinearly and $F(\bar{x}) = 0$. In fact we have $O_R(\{x_i\}) \geqq 1+r$. By the above proof we can show 3.24 holds for stepsize 2.5a. ∎

4. Applications in Nonlinear Programming

In this section we will apply the same idea as that of sections 2 and 3 to a nonlinear programming problem. We will construct a global algorithm for solving the following convex programming problem.

4.1 Minimize $\{f(x) | g(x) \leqq 0$ and $Ax = b\}$, where $f \in C^0$ and $g \in C^1$. f is quasiconvex and g_j, $j=1, \ldots m$ are convex functions from R^n to R and R^m respectively, and A is a $\ell x n$ matrix and $b \in R^\ell$. (C^0 denotes the class of continuous functions and C^1 denotes the class of continuous differentiable functions.)

Consider an algorithm for solving 4.1 as follows:

4.2 Algorithm: (For convex programming problems)

4.3[1] Start with $x_o \in R^n$ such that $x_o \notin \{x \in R^n | g(x) \leqq 0\}$ and $f(x_o) \leqq f(\bar{x})$, where $f(\bar{x}) = \text{minimum } \{f(x) | g(x) \leqq 0$ and $Ax = b\}$.

4.4 Set $i: = 0$

4.5 At x_i, find $\tilde{x}_{i+1}$ by solving

$$\min\{f(x) | g(x_i) + \nabla g(x_i)(x - x_i) \leqq 0 \text{ and } Ax = b\}$$

and define $p_i = \tilde{x}_{i+1} - x_i$.

---

(1) Conditions 4.3 can be easily satisfied as indicated in remark 4.10 below.

4.6 If $\tilde{x}_{i+1}$ is feasible, that is $g(\tilde{x}_{i+1}) \leqq 0$ and $A\tilde{x}_{i+1} = b$, let $x_{i+1} = \tilde{x}_{i+1}$ and stop, otherwise go to 4.7.

4.7 Find x_{i+1} by any one of three stepsize methods stated in algorithm 2.2 with $F(x) = \frac{1}{2} \sum_{j=1}^{m} g_j(x)_+^2$,

4.8 If x_{i+1} is feasible stop, otherwise set $i: = i+1$ and go to 4.5.

4.9 <u>Convergence Theorem</u>: Let x_i be a sequence generated by the above algorithm, then either $\{x_i\}$ terminates at some point $x_{\bar{i}}$ such that $x_{\bar{i}}$ solves 4.1, or for every accumulation point $(\bar{x},\bar{p})$ of $\{x_i,p_i\}$, we have that $\bar{x}$ solves 4.1.

<u>Proof</u>: If $\{x_i\}$ terminates at $x_{\bar{i}}$ for some $\bar{i} \geqq 1$, then there are two possibilities: either $\{x_i\}$ terminates because of 4.6 or because of 4.8. If $\{x_i\}$ terminates by 4.6 it is clear that $x_{\bar{i}}$ solves 4.1 because of g is convex and the feasible region of 4.5 contains that of 4.1. If $\{x_i\}$ terminates at $x_{\bar{i}}$ because of 4.8, then $f(x_{\bar{i}}) \leqq \max\{f(x_{\bar{i}-1}), f(\tilde{x}_{\bar{i}})\}$ because f is quasiconvex and $x_{\bar{i}} \in [x_{\bar{i}-1}, \tilde{x}_{\bar{i}}]$. But $f(x_{\bar{i}-1}) \leqq f(\bar{x})$ and $f(\tilde{x}_{\bar{i}}) \leqq f(\bar{x})$, so $f(x_{\bar{i}}) \leqq f(\bar{x})$. Since $x_{\bar{i}}$ is feasible, we have that $x_{\bar{i}}$ solves 4.1. Assume the sequence $\{x_i\}$ is infinite, we have as before that $f(x_i) \leqq f(\bar{x})$ for every i. Using the same

proof as that in theorem 2.10, we have that for every accumulation $(\bar{x},\bar{p})$ of $\{(x_i,p_i)\}$ $\bar{x}$ is feasible. This implies that $\bar{x}$ is a solution of 4.1. ∎

4.10 Remark: To get a starting point $x_o \in R^n$ such that the conditions of 4.3 are satisfied, we can start with any $x' \in R^n$. Then the solution x'' of 4.5 with $x_i = x'$ is either the desired starting point x_o or x'' solves 4.1.

4.11 Remark: Note that if $\{x \in R^n | f(x) \leq f(\bar{x})\}$ is compact, then 4.5 has a solution for every i.

4.12 Remark: If $f \in C^1$ and convex, then by a well known technique problem 4.1 can be reduced to one with a linear objective function. Problem 4.5 then becomes a linear programming problem.

5. Computational Results

In this section, computational results for testing algorithm 3.1 are given. Two problems are tested, one of them is given by Robinson [Robinson, 1972].

Test problem #1 (Robinson):

Find x_1 and x_2 such that

$$x_1^2 + x_2^2 - 1 \leq 0$$

$$x_1^2 + (x_2-1)^2 - 1 \leq 0$$

$$(x_1-1)^2 + (x_2-1)^2 - 1 = 0$$

Test problem #2

Find x_1 and x_2 such that

$$100\, x_1^2 + x_2^2 - 1 \leqq 0$$

$$50\, x_1^2 + (x_2-1)^2 - 1 \leqq 0$$

$$(x_1-1)^2 + 50(x_2-1)^2 - 1 = 0$$

Both problems are tested by using different starting points.

<u>Test problem #1</u>

1) Starting point = (.55, .1)

k	$F(x_k)$
0	1.28×10^{-2}
1	5.52×10^{-5}
2	1.30×10^{-5}
3	$\leqq 1.00 \times 10^{-8}$

2) Starting point = (0., -1.)

k	$F(x_k)$
0	2.50×10^{1}
1	1.32×10^{0}
2	2.23×10^{-2}
3	2.13×10^{-5}
4	$\leqq 1.00 \times 10^{-8}$

3) Starting point = (100., 100.)

k	$F(x_k)$
0	1.18×10^{9}
1	6.64×10^{7}
2	4.33×10^{6}
3	2.50×10^{5}
4	1.34×10^{5}
5	6.32×10^{2}
6	2.58×10^{1}
7	1.87×10^{0}
8	8.34×10^{-1}
9	1.19×10^{-2}
10	7.16×10^{-6}
11	$\leqq 1.00 \times 10^{-8}$

We note here that this problem has been tested by Robinson using Robinson's algorithm [Robinson, 1972] with starting points (.55, .1) and (0., -1.). Comparing Robinson's results with those given above, we observe that for the first two starting points we have almost the same efficiency for these two algorithms. However, using Robinson's algorithm with starting point (100., 100.), we observe that no feasible point can be found in the first step.

Test problem #2

1) Starting point = (0.1, 1.1)

k	$F(x_k)$
0	1.56×10^{0}
1	4.55×10^{-2}
2	5.66×10^{-3}
3	6.60×10^{-4}
4	5.03×10^{-5}
5	1.90×10^{-7}
6	$\leqq 1.00 \times 10^{-8}$

2) Starting point = (10., 10.)

k	$F(x_k)$
0	1.49×10^{8}
1	8.64×10^{6}
2	5.10×10^{5}
3	3.08×10^{4}
4	1.93×10^{3}
5	1.14×10^{2}
6	6.15×10^{0}
7	3.64×10^{-1}
8	2.01×10^{-2}
9	8.40×10^{-4}
10	1.43×10^{-5}
11	2.00×10^{-8}

References

1. T.J. Huang [1973a]: Convergence and convergence rate - general theory. Also CH 1.

2. O.L. Mangasarian [1972]: Dual, feasible direction algorithms, in Techniques of Optimization. Ed. by Balakrishnan, A.V., Academic Press, pp. 67-88.

3. B.N. Pshenichnyi [1970]: Newton's method for the solution of systems of equalities and inequalities. Mathematics Notes of Academy of Sciences of U.S.S.R. 8 (1970) pp. 827-830.

4. Stephen M. Robinson [1971]: On rate of change in the solution set of a perturbed linear problem. Technical summary report No. 1134, Mathematics Research Center, University of Wisconsin.

5. Stephen M. Robinson [1972]: Extension of Newton's method to nonlinear functions with values in a cone. Numer. Math 19, pp. 341-347.

BIBLIOGRAPHIC DATA SHEET	1. Report No. WIS-CS-191-73	2.	3. Recipient's Accession No.

4. Title and Subtitle ALGORITHMS FOR SOLVING SYSTEMS OF EQUATIONS AND INEQUALITIES WITH APPLICATIONS IN NONLINEAR PROGRAMMING	5. Report Date September 1973 6.
7. Author(s) T. J. Huang	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. GJ-35292
12. Sponsoring Organization Name and Address	13. Type of Report & Period Covered 14.

15. Supplementary Notes

16. Abstracts

Global algorithms are proposed for solving systems of nonlinear equations and inequalities. Global convergence is established under suitable assumptions. For some special cases, the rate of convergence is R-superlinear. A global algorithm for solving convex programming problems is also given. Computational results for one of the algorithms are given.

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

17b. Identifiers/Open-Ended Terms

17c. COSATI Field/Group

18. Availability Statement Available to the public	19. Security Class (This Report) UNCLASSIFIED	21. No. of Pages 29
	20. Security Class (This Page UNCLASSIFIED	22. Price