ON THE FINITE SOLUTION OF NONLINEAR INEQUALITIES

by

E. Polak and D.Q. Mayne

Memorandum No. UCB/ERL M78/80

6 September 1978

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

ON THE FINITE SOLUTION OF NONLINEAR INEQUALITIES*

E. Polak and D. Q. Mayne

Department of Electrical Engineering and Computer Sciences
and the Electronics Research Laboratory
University of California, Berkeley, California 94720

Abstract

We present an algorithm based on Newton's method and a systematic enlargement of a feasible region for solving finitely, systems of nonlinear inequalities. The method depends crucially on the superlinear rate of convergence of Newton's method.

*Research sponsored by National Science Foundation Grant ENG73-08214-A01, National Science Foundation (RANN) Grant ENV76-04264 and the Joint Services Electronics Program Contract F44620-76-C-0100.

1. Introduction

An examination of the engineering literature (see for example [1]) shows that not infrequently the designer is not so much interested in optimizing performance, as in meeting specifications. Generally, such specifications can be expressed as a system of differentiable inequalities

$$g^i(x) \leq 0, \quad j = 1,2,\ldots,m \tag{1.1}$$

which describe a set with a nonempty interior. An important special case in which a designer needs to solve a system of inequalities arrises in problems of design centering, tolerancing and tuning (see [2]) [3]). In such a problem, a designer is required to minimize some performance index, subject to constraints on the form

$$\max_{\omega \in \Omega} \min_{\tau \in T} \max_{j \in J} \zeta^j(x,\omega,\tau) \leq 0,$$

where x is the design vector (including tolerance and tuning range as components), ω is a tolerance parameter and τ is a tuning parameter. The optimization yields a <u>nominal</u> design $\hat{x}$ and the manufacturing process produces in a certain tolerance realization, $\hat{\omega} \in \Omega$. Should measurements show that $\max_{j \in J} \zeta^j(\hat{x},\hat{\omega},0) > 0$, it now becomes necessary to compute a value $\hat{\tau} \in T$, for the tuning parameter, such that $\max_{j \in J} \zeta^j(\hat{x},\hat{\omega},\hat{\tau}) \leq 0$. Normally, T has a simple description of the form $T = \{\tau | g^j(\tau) \leq 0,\ j = 1,2,\ldots,m_1\}$ and hence the required $\hat{\tau}$ can be computed by solving a system of inequalities of the form (1), with $g^{j+m_1}(\tau) \triangleq \zeta^j(\hat{x},\hat{\omega},\tau)$ for all $j \in J$.

Now, as it is well known, under certain conditions, it is possible to find a solution to such a system of inequalities in a finite number of iterations by means of any one of the existing feasible directions algorithms (see [7])

Unfortunately, feasible directions algorithms are rather slow and the question arises whether it is not possible to adapt a faster method, such as the Newton method described in [4,5] to find a solution to (1.1) in a finite number of iterations. In this paper we obtain an affirmative answer to this question. Our scheme is based on applying Newton's method for a controlled number of iterations ℓ_i to a progression of inequalities:

$$g^j(x) + \varepsilon_i \leq 0, \quad j = 1,2,\ldots,m \tag{1.2}$$

with $\varepsilon_i \searrow 0$ and $\ell_{i+1} > \ell_i$, and on the fact that under certain assumptions Newton's method converges quadratically.

2. The Algorithm

Consider the problem of finding a point $\hat{x}$ satisfying

$$g(x) \leqq 0 \tag{2.1}$$

where $g\colon \mathbb{R}^n \to \mathbb{R}^m$ is three times continuously differentiable. The first of the following assumptions is imposed by our desire to use Newton's method (see [ ]), while the second one is required to make finite solution of (2.1) possible. We shall use the notation $\underline{m} = \{1,2,\ldots,m\}$.

Assumption 2.1. For any $x \in \mathbb{R}^n$, $0 \notin \underset{j\in I(x)}{\text{co}} \nabla g^j(x)$ where

$$I(x) \triangleq \{j \in \underline{m} \mid g^j(x) \geq 0\} \tag{2.2}$$

(i.e. the gradients $\nabla g^j(x)$, $j \in I(x)$ satisfy the Robinson LI condition [6]). □

Assumption 2.2. There exists an $\hat{x}$ such that $g(\hat{x}) < 0$. □

Let

$$\nu \triangleq (1,1,\ldots,1)^T \in \mathbb{R}^n \tag{2.3}$$

and let $\varepsilon > 0$ be arbitrary. Let

$$g_\varepsilon(x) \triangleq g(x) + \varepsilon\nu \tag{2.4}$$

and let $g_\varepsilon(x)_+$ be defined by

$$[g_\varepsilon(x)_+]^j \triangleq \max\{g_\varepsilon^j(x),0\},\ j \in \underline{m} \tag{2.5}$$

Now, since by Assumption 2.2 there exists an $\hat{x}$ such that $g(\hat{x}) < 0$, it is clear that there exists an $\hat{\varepsilon} > 0$ such that for all $\varepsilon \in [0,\hat{\varepsilon}]$, there exists an $\hat{x}_\varepsilon$ such that $g_\varepsilon(\hat{x}_\varepsilon) \leqq 0$. If we <u>knew</u> such an $\varepsilon \in (0,\hat{\varepsilon}]$, we could apply the following version of Newton's method described in [5] to find $\hat{x}_\varepsilon$ (under the heading restoration iteration function a).

<u>Algorithm 2.1</u> (Newton Method - MP Version [5]).

<u>Parameters</u>: $\alpha \in (0,1/2)$, $\beta \in (0,1)$, $L >> 1$.

<u>Data</u>: x_0

<u>Step 0</u>: Set $i = 0$.

<u>Step 1</u>: Solve the QP for v_i

$$\min\{\|v\|^2 \mid g_\varepsilon(x_i) + \frac{\partial g_\varepsilon(x_i)}{\partial x} v \leqq 0\} \tag{2.6}$$

<u>Step 2</u>: If v_i exists and $\|v_i\| \leq L$, set $h_i = v_i$. Else set $h_i = -\frac{\partial g_\varepsilon(x_i)}{\partial x} g_\varepsilon(x_i)_+$ (i.e. set $h_i = -\frac{\partial}{\partial x}\frac{1}{2}\|g_\varepsilon(x_i)_+\|^2$).

<u>Step 3</u>: Compute the smallest integer $k \geq 0$ such that

$$\|g_\varepsilon(x_i+\beta^k h_i)_+\|^2 \leq (1-2\alpha\beta^k)\|g_\varepsilon(x_i)_+\|^2 \tag{2.7}$$

Step 4: Set $x_{i+1} = x_i + \beta^k h_i$, set $i = i+1$ and go to step 1. □

We now collect from [4,5], the relevant results of this method.

Theorem 2.1: Suppose that Assumption (2.1) is satisfied and that $\varepsilon \in (0,\hat{\varepsilon}]$ (i.e. there exists an $\hat{x}_\varepsilon$ such that $g_\varepsilon(\hat{x}_\varepsilon) \leqq 0$).

a) If Algorithm 2.1 constructs a bounded sequence $\{x_i\}$, then $x_i \to \hat{x}_\varepsilon$ as $i \to \infty$, satisfying $g_\varepsilon(\hat{x}_\varepsilon) \leqq 0$.

b) For any compact set U, there exists an $M \in (0,\infty)$, depending only on the values of the matrix $\frac{\partial g_\varepsilon(x)}{\partial x} = \frac{\partial g(x)}{\partial x}$ (and hence independent of ε) for $x \in U$, such that if for some i_0, $x_{i_0} \in U$ and $M\|g_\varepsilon(x_{i_0})_+\| < 1$, then for all $i \geq i_0$, $x_i \in U$ and hence $x_i \to \hat{x}_\varepsilon \in U$ as $i \to \infty$. Furthermore,

$$\|x_i - \hat{x}_\varepsilon\| \leq \frac{1}{M}\delta_\varepsilon^{2^{(i-i_0)}} \tag{2.8}$$

holds, with

$$\delta_\varepsilon \in (0, M\|g_\varepsilon(x_{i_0})_+\|) \tag{2.9}$$

c) If $\varepsilon \notin (0,\hat{\varepsilon}]$ (i.e., with $\psi_\varepsilon(\cdot)$ defined as in (2.10) below, $\min_{x \in \mathbb{R}^n} \psi_\varepsilon(x) > 0$), then $x_i \to \hat{x}_\varepsilon$ as $i \to \infty$, a minimizer of $\frac{1}{2}\|g_\varepsilon(x)_+\|^2$. □

Now, let

$$\psi_\varepsilon(x) \triangleq \max_{j \in \underline{m}} g_\varepsilon^i(x) \tag{2.10}$$

and

$$\psi_\varepsilon(x)_+ \triangleq \max\{0, \psi_\varepsilon(x)\} \tag{2.11}$$

Then we have

$$\psi_\varepsilon(x) = \psi_0(x) + \varepsilon \tag{2.12}$$

and, by the relation between L_∞ and L_2 norms

$$\frac{1}{\sqrt{m}}\|g_\varepsilon(x)_+\| \leq \psi_\varepsilon(x)_+ \leq \|g_\varepsilon(x)_+\| \tag{2.13}$$

Now, suppose that the conditions of Theorem 2.1 apply, that $\varepsilon \in (0,\hat{\varepsilon})$, and that $\{x_i\}$ is a sequence constructed by Algorithm (2.1), converging to $\hat{x}_\varepsilon$. Then, since $\psi_\varepsilon(x)_+$ is locally Lipschitz continuous, with constant L, say, in any compact neighborhood of $\hat{x}_\varepsilon$, and $\psi_\varepsilon(\hat{x}_\varepsilon)_+ = 0$, we obtain

$$\psi_0(x_i) + \varepsilon = \psi_\varepsilon(x_i) \leq \psi_\varepsilon(x_i)_+ - \psi_\varepsilon(\hat{x}_\varepsilon)_+ \leq L\|x_i - \hat{x}_\varepsilon\| \text{ for all } i \tag{2.14}$$

and hence, since $g_\varepsilon(x_i)_+ \to 0$, there exists an i_0 such that by (2.8) and (2.14)

$$\psi_0(x_i) \leq -\varepsilon + \frac{L}{M}\delta_\varepsilon^{2^{i-i_0}} \quad \text{for all } i \geq i_0 \tag{2.15}$$

that is,

$$\max_{j\in \underline{m}} g^j(x_i) \leq 0 \tag{2.16}$$

for all $i \geq i_0$ such that

$$-\varepsilon + \frac{L}{M}\delta_\varepsilon^{i-i_0} \leq 0 \tag{2.17}$$

This shows that if we <u>knew</u> a correct value for ε, we would find a feasible point $\bar{x}$ satisfying $g(\bar{x}) \leqq 0$ very rapidly. Thus, our attention must be directed towards constructing a procedure for finding a satisfactory ε. We note in (2.17) that if we decrease ε suitably slowly, then because of the rapid decline of the term $\frac{L}{M}\delta_\varepsilon^{i-i_0}$, we should be able to find a satisfactory ε

and still achieve (2.16) in a finite number of iterations. Next, we note that Algorithm 2.1 minimizes $\frac{1}{2}\|g_\varepsilon(x)_+\|^2$. Since there is an $\hat{x}$ such that $g(\hat{x}) \leqq 0$, it follows that Algorithm 2.1 computes an $\hat{x}_\varepsilon$ such that, because of (2.12),

$$\frac{1}{2}[\psi_\varepsilon(\hat{x}_\varepsilon)_+]^2 \leq \frac{1}{2}\|g_\varepsilon(\hat{x}_\varepsilon)_+\|^2 \leq \frac{m}{2}\varepsilon^2 \tag{2.18}$$

i.e.

$$\psi_\varepsilon(\hat{x}_\varepsilon)_+ \leq \sqrt{m}\,\varepsilon \tag{2.19}$$

Hence,

$$\psi_0(\hat{x}_\varepsilon) + \varepsilon = \psi_\varepsilon(\hat{x}_\varepsilon) \leq \psi_\varepsilon(\hat{x}_\varepsilon)_+ \leq \sqrt{m}\,\varepsilon \tag{2.20}$$

As a result, if Algorithm 2.1 is initialized at x_0, then for any $\gamma \in (0,1)$, there exists a finite i such that

$$\psi_0(x_i) - (\sqrt{m} - 1)\varepsilon \leq \gamma[\psi_0(x_0) - (\sqrt{m}-1)\varepsilon] \tag{2.21}$$

i.e.

$$\psi_0(x_i) \leq \gamma\psi_0(x_0) + (1-\gamma)(\sqrt{m}-1)\varepsilon \tag{2.22}$$

The above observations form the basis for the algorithm below.

Algorithm 2.2.

Parameters: $\alpha \in (0,\frac{1}{2})$, $\beta \in (0,1)$, $L >> 1$, $\gamma_1,\gamma_2 \in (0,1)$, $\delta \in (0,1)$, a sequence of integers $\{\ell_k\}_{k=0}^{\infty}$ such that $\ell_{k+1} > \ell_k$ for all k.

Data: $z_0 \in \mathbb{R}^n$, $\varepsilon_0 > 0$.

Step 0: Set $k = 0$.

Step 1: If $\psi_0(z_k) \leq 0$, stop. Else set $i = 0$, $x_0 = z_k$, $\varepsilon = \varepsilon_k$.

Step 2: Solve QP (2.6) for v_i.

Step 3: If v_i exists and $\|v_i\| \leq L$, set $h_i = v_i$. Else set $h_i = -\frac{\partial g(x_i)}{\partial x} g_\varepsilon(x_i)_+$.

Step 4: Compute the smallest integer $j \geq 0$ such that

$$\|g_\varepsilon(x_i+\beta^j h_i)_+\| \leq (1-2\alpha\beta^j)\|g_\varepsilon(x_i)_+\|^2 \tag{2.23}$$

Step 5: Set $x_{i+1} = x_i + \beta^j h_i$.

Step 6: If $i \geq \ell_k$ and

$$\psi_0(x_{i+1}) \leq \gamma_1\psi_0(x_0) + (1-\gamma_1)(\sqrt{m}-1)\varepsilon_k \tag{2.24}$$

Set $z_{k+1} = x_{i+1}$, $\varepsilon_{k+1} = \gamma_2\varepsilon_k$, set $k = k+1$ and go to step 1. Else, set $i = i+1$ and go to step 2. □

Lemma 2.1: Suppose that Algorithm 2.2 constructs an infinite sequence $\{z_k\}$. Then any accumulation point $\hat{z}$ of $\{z_k\}$ satisfies $\psi_0(\hat{z}) \leq 0$ (i.e. $g(\hat{z}) \leqq 0$).

Proof: By construction, the sequence $\{z_k\}$ satisfies (see (2.24))

$$\psi_0(z_{k+1}) \leq \gamma_1\psi_0(z_k) + (1-\gamma_1)(\sqrt{m}-1)\varepsilon_0\gamma_2^k, \quad k = 0,1,2,\ldots \tag{2.25}$$

Since $\gamma_1,\gamma_2 \in (0,1)$, it follows from (2.25) that

$$\overline{\lim}\ \psi_0(z_k) \leq 0 \tag{2.26}$$

and hence, if $\hat{z}$ is an accumulation point of $\{z_k\}$, then $\psi_0(\hat{z}) \leq 0$. □

Theorem 2.2: Suppose that Assumptions 2.1 and 2.2 are satisfied. If Algorithm 2.2 constructs a bounded sequence $\{z_k\}$ then there is a finite index $s \geq 0$ such that $g(z_s) \leqq 0$.

Proof: Suppose, for the sake of contradiction that Algorithm 2.2 constructs an infinite, bounded sequence $\{z_k\}$. Then, by Lemma 2.1, there exists a subsequence, indexed by $K \subset \{0,1,2,\ldots\}$ such that $z_k \to \hat{z}$, with $\psi_0(\hat{z}) \leq 0$. Hence, since $\varepsilon_k \to 0$ as $k \to \infty$, there exists a $k_0 \in K$ such that, for all $k \in K$, $k \geq k_0$, the set $\{x | \psi_{\varepsilon_k}(x) \leq 0\} \neq \phi$ and, with M as in Theorem 2.1(b),

$$\|g_{\varepsilon_k}(z_k)_+\| \leq \sqrt{m}\psi_{\varepsilon_k}(z_k)_+ \leq \sqrt{m}(\psi_0(z_k)_+ + \varepsilon_k) < \frac{1}{M} \tag{2.27}$$

Thus if $\hat{x}_{\varepsilon_k}$ is the limit of the infinite sequence $\{x_i\}$ generated by algorithm 2.1 when it has been initialised at $x_0 = z_k$, with $k \geq k_0$, then

$$\|z_{k+1} - \hat{x}_{\varepsilon_k}\| \leq \frac{1}{M} \delta_{\varepsilon_k}^{2^{\ell_k}} \tag{2.28}$$

for all $k \in K$, $k \geq k_0$, where $\delta_{\varepsilon_k} \in (0, M\|g_{\varepsilon_k}(z_k)_+\|)$.

Since $\{x | g_{\varepsilon_k}(x) \leq 0\} \neq \phi$, it follows that $\psi_{\varepsilon_k}(\hat{x}_{\varepsilon_k})_+ = 0$ so that

$$\psi_0(z_{k+1}) + \varepsilon_k \leq \psi_{\varepsilon_k}(z_{k+1})_+ - \psi_{\varepsilon_k}(\hat{x}_{\varepsilon_k})_+$$

$$\leq L_k \|z_{k+1} - \hat{x}_{\varepsilon_k}\| \tag{2.29}$$

for $k \geq k_0$, where L_k is the Lipschitz constant associated with a compact set containing the bounded sequence $\{x_i\}$ initiated at $x_0 = z_k$. Since for each $k \in K$, $k \geq k_0$, the sequence $\{x_i\}$ is contained in a sphere of radius at most $1/M$ centered on $\hat{x}_{\varepsilon_K}$, and since the sequence $\{z_k\}$ is bounded, it follows that the collection $\{L_k\}$ of Lipschitz constants can be bounded from above by an overall constant L. Thus from (2.28) and (2.29), it follows that

$$\psi_0(z_{k+1}) \leq \varepsilon_0 \gamma_1^k + \frac{L}{M} \delta_{\varepsilon_k}^{2^{\ell_k}}. \tag{2.30}$$

for all $k \in K$. $k \geq k_0$, where $\delta_{\varepsilon_k} \in (0, M\|g_{\varepsilon_k}(z_k)_+\|)$. Now, by Lemma (2.1), $\psi_0(Z_k)_+ \to 0$ as $k \to \infty$, $k \in K$ and by (2.13)

$$\|g_{\varepsilon_k}(z_k)_+\| \leq \sqrt{m}\,(\psi_0(z_k)_+ + \varepsilon_k) \tag{2.31}$$

Consequently, $\|g_{\varepsilon_k}(z_k)_+\| \to 0$ as $k \to \infty$, $k \in K$, which shows that $\delta_{\varepsilon_k} \to 0$ as $k \to \infty$, with $k \in K$. Hence, since $\ell_k \to \infty$, there exists a $k_1 \geq k_0$, $k_1 \in K$ such that

$$\psi_0(z_{k+1}) \leq -\varepsilon_0 \gamma_1^{k_1} + \frac{L}{M}\delta^2_{\varepsilon_{k_1}}{}^{\ell_{k_1}} \leq 0 \tag{2.32}$$

But then the algorithm must have stopped in Step 1 for $k = k_1+1$ and hence $\{z_k\}$ cannot be infinite. This completes our proof. □

References

[1] N. Zakian and U. Al-Naib, "Design of Dynamical and Control Systems by the Method of Inequalities," Proc. IEE, Vol. 120, No. 11, 1973, pp. 1421-1427.

[2] J. W. Bandler and H. L. Abdul-Malek, "Advances in the Mathematical Programming Approach to Design Centering, Tolerancing and Tuning," Proc. 1978 JACC, Philadelphia, Oct. 14-21, 1978, pp. 329-344.

[3] E. Polak and A. Sangiovanni-Vincentelli, "On Optimization Algorithms for Engineering Design Problems in the Distributed Constraints, Tolerances and Tuning," Proc. 1978 JACC, Philadelphia, Oct. 19-21, 1978, pp. 344-353.

[4] S. M. Robinson, "Extension of Newton's Method to Mixed Systems of Nonlinear Equations and Inequalities," Tech. Summ. Rep. 1161, Math. Research Center, Univ. of Wisconsin, Madison, Wisconson, 1971.

[5] H. Mukai and E. Polak, "On the Use of Approximations in Algorithms for Optimization Problems with Equality and Inequality Constraints," SIAM. J. Numer. Anal., Vol. 15, No. 4, 1978, pp. 674-693.

[6] S. M. Robinson, "Bounds for Error in the Solution Set of a Perturbed Linear Program," Linear Algebra and Its Applications, Vol. 6, 1973, pp. 69-81.

[7] E. Polak, Computational Methods in Optimization, Academic Press, New York, 1971.