MATHEMATICS OF OPERATIONS RESEARCH
Vol. 11, No. 4, November 1986
Printed in U.S.A.

A GAUSS-NEWTON APPROACH TO SOLVING GENERALIZED INEQUALITIES*†

JIM BURKE‡ AND S.-P. HAN§

Generalized inequalities are systems of the form $g(x) \leqslant_K 0$, where g maps between normed linear spaces and " $\leqslant_K$ " denotes the partial order induced by the closed convex cone K (e.g. $K = \mathbf{R}_+^{m_1} \times \{0\}_{\mathbf{R}^{m_2}}$). In this paper a Gauss-Newton type algorithm is presented for minimizing the distance function

$$\rho(x) := \operatorname{dist}(g(x), -K) := \inf\{\|g(x) + k\| : k \in K\}.$$

The technique globalizes the well-known Newton methods for solving generalized inequalities, and overcomes the difficulties associated with subgradient methods for the global minimization of ρ.

1. Introduction. Generalized inequalities are systems of the form

$$g(x) \leqslant_K 0 \tag{1}$$

where g is a mapping between normed linear spaces X and Y and " $\leqslant_K$ " denotes the partial order induced by a closed convex cone K contained in Y (e.g. $K := \mathbf{R}_+^{m_1} \times \{0\}_{\mathbf{R}^{m_2}}$). Inequalities of this type are significant as they provide a unifying theoretical framework for investigating the structural characteristics of a wide variety of problems in applied mathematics (e.g. approximation, optimization, complementarity, variational inequalities). Moreover, these systems play a central role in the model formulation, design, and analysis of the numerical techniques employed in solving problems arising in mathematical programming, complementarity, and variational inequalities.

Most iterative methods for solving (1) depend upon the solvability of the linearized subproblems

$$g(\bar{x}) + g'(\bar{x})(x - \bar{x}) \leqslant_K 0. \tag{2}$$

However, since in general (2) may be inconsistent, more robust methods for solving (1) are required. Garcia-Palomares [6], and Garcia-Palomares and Restuccia [7, 8] develop a mini-max approach to this problem where $K := \mathbf{R}_+^{m_1} \times \{0\}_{\mathbf{R}^{m_2}}$. In fact their approach has provided a good deal of motivation for our viewpoint. In conjunction with this work many of the recently developed techniques for composite nondifferentiable optimization [19] also apply. But as in [6, 7, 8] these contributions depend upon the polyhedrality of both the cone and the norm employed. In the present paper we pursue a more geometric approach, thereby eliminating the dependency on polyhedrality and finite dimensionality.

*Received February 9, 1984; revised April 19, 1985.
AMS 1980 subject classification. Primary 65K05, Secondary 90C30.
IAOR 1973 subject classification. Main: Programming; Nonlinear.
OR/MS Index 1978 subject classification. Primary: 654 Programming/Nonlinear/Algorithms.
Key words. Generalized Inequalities, Gauss-Newton, Clarke Subgradient.
†Partially sponsored by the National Science Foundation under Grant No. DMS82-03603.
‡University of Washington.
§University of Illinois, Urbana.

0364-765X/86/1104/0632$01.25
Copyright © 1986, The Institute of Management Sciences/Operations Research Society of America

We begin in §2 by presenting the algorithm. In §3 the geometric underpinnings of the method are established, and in §4 certain stationarity criteria are developed. §5 considers regularity conditions designed to assure the existence and boundedness of our search directions, and finally, in §6, the convergence results are presented.

The notation that we employ is for the most part the same as that in Rockafellar [16]. A partial list is provided below for the reader's convenience.

—$g'(x; d) = \lim_{\lambda \downarrow 0}\{(g(x+\lambda d) - g(x))/\lambda\}$,

—$g'(x)$ is the Fréchet derivative of g at x,

—The space X^* is the space of continuous linear functionals on the normed linear space X, normed with the operator norm.

—Let C be a nonempty set in a normed linear space X, then

—cl C is the closure of C,

—co C is the convex hull of C,

—$\psi^*(x^*|C) := \sup\{\langle x^*, x\rangle;\ x \in C\}$ is the support functional of C,

—$\gamma(x|C) := \inf\{\gamma\colon\ x \in \gamma C\}$ is the gauge functional of C, and $C^0 := \{x^* \in X^*\colon \langle x^*, x\rangle \leqslant 1 \text{ for all } x \in C\}$.

If C is a convex cone in X, then $C^* := (-C)^0$.

—$\operatorname{argmin}\{f(x)\colon\ x \in S\} = \{\bar{x}\colon\ f(\bar{x}) = \min\{f(x)\colon\ x \in S\}\}$.

2. The algorithm. Iterative schemes that employ inequality (2) to generate updates for solving (1) are called Newton methods [4, 14, 15] as they are the natural generalization of Newton's method for solving equations. Such methods are locally quadratically convergent under the appropriate hypothesis and so constitute a powerful class of techniques for solving (1). But on a global scale these methods may not be well defined due to the possible infeasibility of (2), or for that matter, the infeasibility of (1). One way to overcome this difficulty is to develop methods for the global minimization of the functional

$$\rho(x) := \operatorname{dist}(g(x), -K) := \inf\{\|g(x) + k\|\colon\ k \in K\}.$$

In this connection two procedures immediately come to mind: (a) subgradient methods [12, 13], and (b) Gauss-Newton methods. The subgradient approach is not altogether satisfactory since in general convergence to stationary points of ρ cannot be guaranteed even if exact line searches are performed (e.g. see [5]). (One hope in this direction though is ϵ-subgradients [1, 11], and in fact this is the basis for the success of [6, 7, 8] since a natural and practical definition for the ϵ-subgradient of maximum functions has been provided in [5]. But, for now, a workable definition for more general non-convex functions does not exist.) On the other hand, we will show that the Gauss-Newton approach provides a natural vehicle for overcoming the difficulties of the subgradient approach. That is, there is a way to choose steplengths that is not encumbered by the discontinuities of the directional derivative $\rho'(x; d)$. The algorithm is as follows:

Step (0). Choose $c \in (0,1)$, $\gamma \in (0,1)$, $M \geqslant 0$, and $x_0 \in X$. Set $k = 0$.

Step (1). If $0 = \Delta(x_k) := \rho^*(x_k) - \rho(x_k)$, where

$$\rho^*(x_k) := \operatorname{dist}\big(g(x_k), -(K + \operatorname{Ran}[g'(x_k)])\big),$$
$$= \inf\{\operatorname{dist}(g(x_k) + g'(x_k)d, -K)\colon\ d \in X\}, \text{ stop.}$$

Otherwise go to (2).

Step (2). Choose $d_k \in D(x_k) := \operatorname{argmin}\{\operatorname{dist}(g(x_k) + g'(x_k)d, -K)\colon\ d \in X\}$ so that $\|d_k\| \leqslant \operatorname{dist}(0, D(x_k)) + M$.

Step (3). Set

$$x_{k+1} := x_k + \lambda_k d_k \quad \text{where}$$

$$\lambda_k := \max \gamma^s \quad \text{subject to} \quad s = 0, 1, 2, \ldots, \quad \rho(x_k + \gamma^s d_k) - \rho(x_k) \leqslant c\gamma^s \Delta(x_k).$$

Step (4). Set $k := k + 1$ and goto (1).

Clearly, the only difference between this algorithm and a more conventional one is the replacement of $\rho'(x; d)$ by $\Delta(x)$ in the Armijo type stepsize procedure, but, as we will see, this simple innovation is enough to guarantee global convergence properties under suitable hypothesis. (For the case when (2) is solvable, and so $\rho^*(x) = 0$, this stepsize strategy was first introduced by Pshenichnyi [14].) That the algorithm is well defined or that the stopping criteria makes sense is for the moment unclear. But as the analysis unfolds we will show that $\rho'(x; d) \leqslant \Delta(x) \leqslant 0$ for all x in X and d in $D(x)$, whenever $D(x) \neq 0$. Moreover, sufficient conditions for the nonemptiness of the sets $D(x)$ will also be derived. Finally, one should note that whenever $\rho^*(x) = 0$ and $D(x) \neq \emptyset$, the linearized inequality (2) is solvable.

3. The geometry.

(3.1) THEOREM. *Let K be a closed convex cone contained in the real normed linear space Y. Define the functional ϕ mapping Y into $\mathbf{R}$ by the relation $\phi(y) := \operatorname{dist}(y, -K) := \inf\{\|y + k\|: k \in K\}$. Then ϕ satisfies the equation*

$$\phi(y) = \operatorname{dist}(y, -K) = \gamma(y|B - K) = \psi^*(y|B^0 \cap K^*)$$

where $B := \{y: \|y\| \leqslant 1\}$.

PROOF. The equality of $\gamma(y|B - K)$ and $\psi^*(y|B^0 \cap K^*)$ follows from the standard results concerning the gauge functionals of convex sets that contain the origin (e.g. see [10, 16]) and the fact that $(B - K)^0 = B^0 \cap K^*$. The result now follows from the following derivation:

$$\begin{aligned} \inf\{\|y + k\|: k \in K\} &= \inf\{\gamma: z \in B, \gamma z \in y + K\} \\ &= \inf\{\gamma: y \in \gamma B - K\} \\ &= \inf\{\gamma: y \in \gamma(B - K)\} = \gamma(y|B - K). \quad \blacksquare \end{aligned}$$

The above theorem displays the very rich geometric structure of the functional ϕ. Moreover, since the functional ρ is simply the composition of ϕ and g, the structure of ϕ provides us with the necessary tools for analyzing ρ.

(3.2) THEOREM. *Let g be a continuously Fréchet differentiable map between the normed linear spaces X and Y, and let K be a closed convex cone contained in Y. Set*

$$D(x) := \operatorname{argmin}\{\operatorname{dist}(g(x) + g'(x)d, -K): d \in X\}.$$

(a) *If Y is a reflexive Banach space and the set* $\operatorname{Ran}(g'(x)) + K$ *is closed, then $D(x)$ is a nonempty closed convex set.*

(b) *If in addition to the hypothesis of* (a) *X is also assumed to be a reflexive Banach space, then $D(x)$ contains an element of least norm.*

PROOF. If $D(x)$ is nonempty then its convexity follows from the convexity of the norm and the linearity of $g'(x)$, and its closedness follows from the continuity of the distance function to convex sets (see Theorem (5.2) of §5). Hence by Vlasov [18, Proposition 2.3], (b) is valid. In order to obtain the nonemptiness of $D(x)$ we observe that

$$\begin{aligned}
&\inf\{\operatorname{dist}(g(x)+g'(x)d,-K)\colon d\in X\}\\
&\quad=\inf\{\|g(x)+g'(x)d+k\|\colon k\in K, d\in X\}\\
&\quad=\inf\{\|g(x)+z\|\colon z\in \operatorname{Ran}(g'(x))+K\}\\
&\quad=\operatorname{dist}(g(x),-[\operatorname{Ran}(g'(x))+K])
\end{aligned}$$

and again apply Vlasov [18, Proposition 2.3]. ∎

If the set $[\operatorname{Ran}[g'(x)]+K]$ is not closed it is possible, even in finite dimensions, that the set $D(x)$ is empty.

EXAMPLE. Define $g\colon \mathbf{R}\to\mathbf{R}^3$ by the relation $g(x):=(0,-1,-x)$, and define K to be the "ice cream cone" $K:=\{(x,y,z)\colon 2xz\geqslant y^2, x\geqslant 0, z\geqslant 0\}$. Then it is a simple matter to show that $D(0)=\emptyset$ and $\operatorname{Ran}[g'(0)]+K$ is not closed.

4. Stationary criteria for ρ. The following lemma provides the basis for the results of this section.

(4.1) LEMMA. *Let g be a continuously Fréchet differentiable map between the normed linear spaces X and Y, and suppose that the functional f on Y is positively homogeneous and sublinear, i.e. $f(\lambda x)=\lambda f(x)$ for all $\lambda\geqslant 0$ and $x\in X$, and $f(x+y)\leqslant f(x)+f(y)$ for all x and y in X. If we define $\mu\colon\equiv f\circ g$, then $\mu'(x;d)$ exists and satisfies the inequality*

$$\mu'(x;d)\leqslant f(g(x)+g'(x)d)-f(g(x))$$

for every x and d in X.

PROOF. Since f is positively homogeneous and sublinear, it is convex. Hence by Clarke [3] $\mu'(x;d)$ exists for all x and d in X. The result is thus a consequence of the following derivation:

$$\begin{aligned}
\mu'(x;d)&=\lim_{\lambda\downarrow 0}\frac{f(g(x+\lambda d))-f(g(x))}{\lambda}\\
&=\lim_{\lambda\downarrow 0}\frac{f(g(x)+\lambda g'(x)d+o(\lambda))-f(g(x))}{\lambda}\\
&\leqslant f(g(x)+g'(x)d)-f(g(x))+\lim_{\lambda\downarrow 0}f\left(\frac{o(\lambda)}{\lambda}\right). \quad ∎
\end{aligned}$$

Replacing f by the functional $\phi(y):=\operatorname{dist}(y,-K)$ of Theorem (3.1), we obtain, from the lemma, the relation

$$\rho'(x;d)\leqslant\Delta(x)\leqslant 0 \tag{4.1}$$

for all $x\in X$ and $d\in D(x)$, where Δ and D are defined in Steps 1 and 2 of the algorithm, respectively. Hence the Gauss-Newton directions, $D(x)$, are in fact descent directions for ρ. The key result of this section is now proved via inequality (4.1) and Theorem (3.1).

(4.2) THEOREM. *Let the assumptions of Theorem (3.2) concerning the spaces X and Y, the cone K, and the function g hold, let ρ, ρ^*, Δ, and D be as defined in the algorithm of*

§2, *and let* ϕ *be as defined in Theorem* (3.1). *Then the following conditions are equivalent*:

(i) $0 \in \partial\rho(x) := \partial\phi(g(x)) \circ g'(x) = \{z^* \circ g'(x) \colon z^* \in \partial\phi(g(x))\}$ *where* $\partial\phi$ *denotes the usual subgradient of a convex functional.*

(ii) $0 \leqslant \rho'(x; d)$ *for all* d *in* X.

(iii) $0 \leqslant \rho'(x; d)$ *for some or all* $d \in D(x)$.

(iv) $\Delta(x) = 0$.

(v) $0 \in D(x)$.

PROOF. We first note that the set $\partial\rho(x)$ is in fact the Clarke subdifferential of ρ at x. The equivalence $\partial\rho(x) = \partial\phi(g(x)) \circ g'(x)$ which we have presented as a definition for the sake of simplicity can be easily derived via Clarke [3, Proposition 13]. With this observation in mind the equivalence of (i) and (ii) can be found in Clarke [3]. Furthermore, the implication, (ii) ⇒ (iii), and the equivalence of (iv) and (v) are trivial. Also, the implication (iii) ⇒ (iv), is a simple consequence of Lemma (2.4). Thus it only remains to show that (iv) ⇒ (i).

By Theorem (3.1) and a standard result concerning the subgradient of a support functional (see e.g. Moreau [10]) we obtain the relation

$$\begin{aligned}\partial\rho(x) &= \partial\psi^*(g(x) \mid B^0 \cap K^*) \circ g'(x)\\ &= \left[\operatorname{argmax}\{\langle y^*, g(x)\rangle \colon y^* \in B^0 \cap K^*\}\right] \circ g'(x).\end{aligned}$$

Also, since $\Delta(x) = 0$, we have, as in Theorem (3.2), that

$$\begin{aligned}\operatorname{dist}(g(x), -K) &= \inf\{\operatorname{dist}(g(x) + g'(x)d, -K) \colon d \in X\}\\ &= \operatorname{dist}(g(x), -[\operatorname{Ran}(g'(0)) + K]).\end{aligned}$$

Hence, again by Theorem (3.1),

$$\psi^*(g(x) \mid B^0 \cap K^*) = \psi^*(g(x) \mid B^0 \cap K^* \cap \operatorname{Ran}[g'(x)]^{\perp}).$$

But then

$$\begin{aligned}&\operatorname{argmax}\{\langle y^*, g(x)\rangle \colon y^* \in B^0 \cap K^*\}\\ &\quad \supset \operatorname{argmax}\{\langle y^*, g(x)\rangle \colon y^* \in B^0 \cap K^* \cap \operatorname{Ran}[g'(x)]^{\perp}\},\end{aligned}$$

both of which are nonempty. Therefore, there is a $y_0^* \in \operatorname{Ran}(g'(x))^{\perp}$ such that

$$y_0^* \in \operatorname{argmax}\{\langle y^*, g(x)\rangle \colon y^* \in B^0 \cap K^*\}.$$

Hence $0 = y^* \circ g'(x) \in \partial\rho(x)$. ∎

Stronger results concerning the stationarity of ρ can be obtained by assuming that the norm possesses certain smoothness properties (see [2]). In particular, if Y is assumed to be a Hilbert space then an elegant generalization of the normal equations for systems of equations can be derived (see [2]). If Y is a Hilbert space and we let P_{K^*} denote the metric projection onto $K^* \subset Y$, then

$$0 \in \partial\rho(x) \quad \text{if and only if} \quad g'(x)^T P_{K^*}[g(x)] = 0.$$

5. Regularity. The primary purpose of the regularity conditions required by the Newton methods for solving (1) is to guarantee the local solvability of the linearized problems (2). A rather fortuitous consequence of the imposition of such conditions is the local uniform boundedness of the Newton directions (i.e. $\operatorname{argmin}\{\|d\|: g(x) + g'(x)d \leqslant_K 0\}$). For our purposes the imposition of a condition that guarantees the solvability of (2) would be self-defeating as the possible insolvability of (2) is the primary motivation for our method. Yet we still require a condition that guarantees the boundedness of our search directions (i.e. the directions $\{d: \|d\| \leqslant \operatorname{dist}(0, D(x)) + M, d \in D(x)\}$). The question then arises as to whether the local uniform boundedness of our search directions forces the local solvability of (2). The following example answers this question in the negative.

EXAMPLE. Let $g: \mathbf{R} \to \mathbf{R}^2$ be defined by

$$g(x) := \begin{pmatrix} e - e^x \\ x \end{pmatrix}$$

and set $K := \mathbf{R}^2_+ := \{(x, y): x \geqslant 0, y \geqslant 0\}$. Then

$$\operatorname{argmin}\{\|d\|: d \in D(x)\} = \{(e^{x+1} - e^{2x} - x)(1 + e^{2x})^{-1}\}$$

where $\mathbf{R}^2$ is equipped with the Euclidean norm. Hence $D(x)$ is everywhere locally bounded and yet (2) is not solvable for any $x \in \mathbf{R}$.

Thus we are led to the following definition for regularity.

(5.1) DEFINITION. Let g be a continuously Fréchet differentiable map between the real normed linear spaces X and Y, and let K be a closed convex cone contained in Y. We say that g is K-regular at $x_o \in X$ if there is a neighborhood $N(x_o)$ of x_o such that

$$\sup\{\operatorname{dist}(0, D(x)): x \in N(x_o)\} < \infty$$

where $D(x) := \operatorname{argmin}\{\operatorname{dist}(g(x) + g'(x)d, -K): d \in X\}$. (Here we use the convention $\operatorname{dist}(0, \emptyset) = +\infty$.) The map g is said to be K-regular on set $S \subset X$ if it is K-regular at every point of S, and is said to be uniformly K-regular on S if $\sup\{\operatorname{dist}(0, D(x)): x \in S\} < +\infty$.

Although K-regularity is weaker than any of the well-known regularity conditions for generalized inequalities, and is the weakest condition under which our search directions can be guaranteed to be locally uniformly bounded, it may still fail to hold in what appear to be very well-behaved situations, as is illustrated by the next example.

EXAMPLE. Let $g: \mathbf{R} \to \mathbf{R}^2$ be defined by the relation

$$g(x) := \begin{pmatrix} x^2 + 1 \\ x \end{pmatrix},$$

and let $K := \mathbf{R}^2_+$. Then $x = 0$ is a global minimum for ρ and yet g is not K-regular at $x = 0$.

The above example is indicative of what can go wrong in the finite-dimensional setting. The following theorem clarifies this point and provides sufficient conditions for K-regularity in finite dimensions.

(5.2) THEOREM. *Let $g: \mathbf{R}^n \to \mathbf{R}^m$ be continuously Fréchet differentiable, let K be a closed convex cone contained in $\mathbf{R}^m$ and let ρ, ρ^*, and D be as defined in the algorithm of* §2. *If either*

(1) $\operatorname{Ran}(g'(\bar{x})) \cap \operatorname{int} K \neq \emptyset$, *or*

(2) (a) *g' is of locally constant rank at $\bar{x}$, and*

(b) $\operatorname{Ran}(g'(\bar{x})) \cap \operatorname{bdry} K = \{0\}$,

then g is K-regular at $\bar{x}$.

PROOF. (1) Since g is continuously Fréchet differentiable, there exists a neighborhood $N(\bar{x})$ of $\bar{x}$ on which $\operatorname{Ran}(g'(x)) \cap \operatorname{int} K \neq \emptyset$ for all $x \in N(\bar{x})$. Let $x \in N(\bar{x})$ and choose $y \in \operatorname{Ran}(g'(x)) \cap \operatorname{int} K$. Then there exists $\epsilon > 0$ such that $y + \epsilon B \subset \operatorname{int} K$, where B is the unit ball in $\mathbf{R}^m$. Hence $\lambda(y + \epsilon B) \subset \operatorname{int} K$ for all $\lambda \geqslant 0$. Therefore $\lambda\epsilon B = -\lambda y + \lambda(y + \epsilon B) \subset \operatorname{Ran}(g'(x)) + \operatorname{int} K$ for all $\lambda \geqslant 0$, and so $\operatorname{Ran}(g'(x)) + \operatorname{int} K = \mathbf{R}^m$, for every $x \in N(\bar{x})$. Now since $\rho(x) = \operatorname{dist}(g(x), -[\operatorname{Ran}(g'(x)) + K])$ and $\operatorname{Ran}(g'(\bar{x})) + \operatorname{int} K = \mathbf{R}^m$, there exists $\bar{d} \in D(\bar{x})$ such that $g(\bar{x}) + g'(\bar{x})\bar{d} \in -\operatorname{int}(K)$. Hence there is a neighborhood of $\bar{x}$, say $N_1(\bar{x})$, such that $g(x) + g'(x)\bar{d} \in -\operatorname{int}(K)$ for all $x \in N_1(\bar{x})$, since both g and g' are continuous, and so g is K-regular at $\bar{x}$.

(2) We begin by showing that if (a) and (b) hold, then there is a neighborhood $N(\bar{x})$ of $\bar{x}$ on which $\operatorname{Ran}(g'(x)) + K$ is closed, and so by Theorem (3.2), $D(x)$ is nonempty on this neighborhood.

Since (a) holds we know that the projector, P_x, onto $\operatorname{Ran}(g'(x))$ is a continuous function of x on some neighborhood, say $N(\bar{x})$, of $\bar{x}$. Hence on $N(\bar{x})$, $\operatorname{Ran}(g'(x)) \cap \operatorname{bdry} K = \{0\}$, since otherwise there would exist sequences $\{x_i\}$ and $\{h_i\}$ with $h_i \in \operatorname{Ran}(g'(x_i)) \cap \operatorname{bdry} K$ and $\|h_i\| = 1$, for all i, $h_i \to h^*$, and $x_i \to \bar{x}$. But then $h_i = P_{x_i}h_i \to P_{\bar{x}}h^* = h^*$, a contradiction, since $\|h^*\| = 1$, $h^* \in \operatorname{bdry} K$, and $h^* \in \operatorname{Ran}(g'(\bar{x}))$.

We now proceed to show that if $\operatorname{Ran}(g'(x)) \cap \operatorname{bdry} K = \{0\}$, then $\operatorname{Ran}(g'(x)) + K$ is closed, yielding the first step of our proof.

Suppose to the contrary that $\operatorname{Ran}(g'(x)) + K$ is not closed, then there are sequences $\{z_i\} \subset \operatorname{Ran}(g'(x))$, and $\{k_i\} \subset K$ such that $z_i + k_i \to h \notin \operatorname{Ran}(g'(x)) + K$. Clearly, the sequences $\{z_i\}$ and $\{k_i\}$ must be unbounded since both $\operatorname{Ran}(g'(x))$ and K are closed sets. Thus with no loss of generality, we will suppose that $k_i/\|k_i\| \to k^*$, with $\|k^*\| = 1$, and since

$$\left\| \frac{k_i}{\|k_i\|} - \frac{h - z_i}{\|k_i\|} \right\| \to 0,$$

we can also assume that $z_i/\|k_i\| \to k^*$. Hence $k^* \in \operatorname{Ran}(g'(x)) \cap K$. Now if $k^* \in \operatorname{int} K$, then eventually so is $k_i/\|k_i\|$, so that eventually $(h - z_i)/\|k_i\|$ is in $\operatorname{int} K$. Hence $h \in \operatorname{Ran}(g'(x)) + \operatorname{int} K$, a contradiction, and so $k^* \in \operatorname{bdry} K$. But then $k^* = 0$, also a contradiction. Hence $\operatorname{Ran}(g'(x)) + K$ is closed.

We now have that $D(x) \neq \emptyset$ on $N(\bar{x})$. Define $\bar{P}_x$ to be the projector onto $\operatorname{Nul}(g'(x))^\perp$, and note that since g' is of locally constant rank, we can assume that $\bar{P}_x$ is a continuous function of x on $N(\bar{x})$. Now if $d \in D(x)$, then so is $\bar{P}_x d$ and so for every $x \in N(\bar{x})$, there is a $d_x \in D(x) \cap \operatorname{Nul}(g'(x))^\perp$ such that

$$\|d_x\| = \operatorname{dist}\left(0, D(x) \cap \operatorname{Nul}(g'(x))^\perp\right).$$

The proof is now concluded by assuming that g is not K-regular at $\bar{x}$ and deriving a contradiction.

Since g is not K-regular at $\bar{x}$, there exist sequences $\{x_i\}$ and $\{d_i\}$ with

$$\|d_i\| = \operatorname{dist}\left(0, D(x_i) \cap \operatorname{Nul}(g'(x_i))^\perp\right), \quad \|d_i\| \to \infty, \quad d_i/\|d_i\| \to d^*,$$

and $x_i \to \bar{x}$. Also, since $\rho(x)$ is upper-semicontinuous, we know that $\rho(x)$ is locally bounded at $\bar{x}$, hence $\rho(x_i)/\|d_i\| \to 0$. By the positive homogeneity of ϕ (see Theorem

(3.1), we find that

$$\frac{1}{\|d_i\|}\rho(x_i) = \frac{1}{\|d_i\|}\operatorname{dist}(g(x_i) + g'(x_i)d_i| - K)$$

$$= \operatorname{dist}\left(\frac{g(x_i)}{\|d_i\|} + g'(x_i)\frac{d_i}{\|d_i\|}| - K\right)$$

$$\to \operatorname{dist}(g'(\bar{x})d^*| - K),$$

and so $-g'(\bar{x})d^* \in \operatorname{Ran}(g'(\bar{x})) \cap K$. Now if $g'(\bar{x})d^* \in \operatorname{int}(-K)$, then eventually so is $(g(x_i) + g^*(x_i)d_i)/\|d_i\|$. Hence, eventually, $g(x_i) + g'(x_i)d_i$ is in $\operatorname{int}(-K)$. But then for each such i there is a $\lambda \in [0,1)$ for which $g(x_i) + g'(x_i)(\lambda\, d_i) \in \operatorname{int}(-K)$, contradicting the choice of the d_i's. Hence $-g'(\bar{x})d^* \in \operatorname{bdry}(K)$, and so by (b), $g'(\bar{x})d^* = 0$, yielding $d^* \in \operatorname{Nul}(g'(\bar{x}))$. But

$$\frac{d_i}{\|d_i\|} = \bar{P}_x\left(\frac{d_i}{\|d_i\|}\right) \to \bar{P}_{\bar{x}}d^* = d^*,$$

and so $d^* \in \operatorname{Nul}(g'(\bar{x}))^{\perp}$. Therefore $d^* = 0$, and $\|d^*\| = 1$, a contradiction. Hence g is K-regular on $N(\bar{x})$, and in particular at $\bar{x}$. ∎

We now have the following continuity results for ρ, ρ^*, and D:

(5.3) THEOREM. *Let g be a continuously Fréchet differentiable map between the real normed linear spaces X and Y, let K be a closed convex cone contained in Y, and let ρ, ρ^*, and D be as defined in the algorithm of* §2. *Then*

(i) *ρ is continuous on X,*

(ii) *ρ^* is upper semicontinuous on X,*

(iii) *if g is K-regular at x_o, then ρ^* is continuous at x_0, and*

(iv) *if Y is a reflexive Banach space and* $(\operatorname{Ran}(g'(x)) + K)$ *is closed for all $x \in X$, then the multi-valued map $D(x)$ is upper semicontinuous (i.e. if $x_i \to x^*$, and $d_i \to d^*$ with $d_i \in D(x_i)$, then $d^* \in D(x^*)$).*

PROOF. (i) This is a well-known and easily established result; for its proof, see for example [2].

(ii) Let x and z be elements of X and let $\epsilon > 0$. Choose $d \in X$ so that

$$\operatorname{dist}(g(x) + g^*(x)d, - k) \leqslant \rho^*(x) + \epsilon.$$

Then

$$\begin{aligned}\rho^*(z) &\leqslant \operatorname{dist}(g(z) + g'(z)d, - K)\\ &\leqslant \|(g(z) + g'(z)d) - (g(x) + g'(x)d)\| + \operatorname{dist}(g(x) + g'(x)d, - K)\\ &\leqslant \|g(z) - g(x)\| + \|g'(z) - g'(x)\|\,\|d\| + \rho^*(x) + \epsilon. \end{aligned} \tag{5.1}$$

Therefore,

$$\limsup_{z\to x} \rho^*(z) \leqslant \epsilon + \rho^*(x),$$

yielding the upper semicontinuity of ρ^*.

(iii) By K-regularity there is a neighborhood $N(x_o)$ of x_o and a constant $M \geqslant 0$ such that $\operatorname{dist}(0, D(x)) \leqslant M$ for all $x \in N(x_o)$. Note, in particular, that this implies that $D(x) \neq \emptyset$ for all $x \in N(x_o)$. Let $\epsilon > 0$ be given, choose x_1 and x_2 in $N(x_o)$, and select $d_i \in D(x_i)$ such that $\|d_i\| \leqslant \operatorname{dist}(0, D(x_i)) + \epsilon$, $i = 1,2$. Then employing a derivation similar to that of inequality (5.1) and using symmetry we get that

$$\begin{aligned}
&|\rho^*(x_1) - \rho^*(x_2)| \\
&\quad \leqslant \|g(x_1) - g(x_2)\| + \|g'(x_1) - g'(x_2) \\
&\qquad \times \|[\max\{\operatorname{dist}(0, D(x_1)), \operatorname{dist}(0, D(x_2))\} + \epsilon] \\
&\quad \leqslant \|g(x_1) - g(x_2)\| + \|g'(x_1) - g'(x_2)\|[M + \epsilon].
\end{aligned}$$

Therefore, ρ^* is continuous at x_0.

(iv) Theorem (3.2) tells us that $D(x) \neq \emptyset$ for all $x \in X$. Let $x_i \to x^*$ and $d_i \to d^*$ with $d_i \in D(x_i)$ for all $i = 1, 2, \ldots$. We need to show that $d^* \in D(x^*)$. First note that since the function $L(x, d) := g(x) + g'(x)d$ is continuous in both its variables x and d, we obtain via part (i) that the function $\sigma(x, d) := \operatorname{dist}(L(x, d)| - K)$ is continuous on $X \times X$. Combining this fact with the upper semicontinuity of ρ^* we get that

$$\begin{aligned}
\operatorname{dist}(g(x^*) + g'(x^*)d^*, -K) &= \sigma(x^*, d^*) \\
&= \lim_{i \to \infty} \sigma(x_i, d_i) \\
&= \lim_{i \to \infty} \rho^*(x_i) \\
&= \limsup_{i \to \infty} \rho^*(x^*) \\
&\leqslant \rho^*(x^*)
\end{aligned}$$

and so $d^* \in D(x^*)$.

6. Convergence. We begin this section by showing that the algorithm of §2 is indeed well defined. In Step (1), the algorithm is terminated if x_k is a stationary point of ρ, since, by Theorem (4.2), the condition $\Delta(x_k) = 0$ is equivalent to $0 \in \partial\rho(x_k)$. In Step (2) a search direction d_k is chosen from the set of Gauss-Newton directions $D(x_k)$ satisfying $\|d_k\| \leqslant \operatorname{dist}(0, D(x_k)) + M$. If $M > 0$ then the existence of such a d_k is guaranteed by assuming either that Y is a Banach space and the set $\operatorname{Ran}(g'(x_k)) + K$ is closed (Theorem (3.2)), or that g is K-regular at x_k (Definition (5.1)). If $M = 0$ then it must be further assumed that X is a Banach space (Theorem (3.2)). Step (3) of the algorithm produces a step-length via an Armijo type steplength procedure where $\rho'(x_k; d_k)$ is replaced by $\Delta(x_k)$. By Lemma (4.1) we know that $\rho'(x_k; d_k) \leqslant \Delta(x_k) \leqslant 0$, hence, since $\Delta(x_k) \neq 0$, the Armijo procedure is finitely terminating. Therefore the algorithm is well defined so long as $D(x_k) \neq 0$ for every iterate x_k.

(6.1) THEOREM. *Let g be a continuously Fréchet differentiable map between the normed linear spaces X and Y. (if the parameter M in the algorithm is zero, X must be taken to be a Banach space.) Suppose that K is a closed convex cone contained in Y, and that ρ, ρ^*, D and Δ are as defined in the algorithm of §2. Let $x_o \in X$ and set $S := \operatorname{cl}(\operatorname{co}\{x\colon 0 < \rho(x) \leqslant \rho(x_o)\})$. If*

(a) *g' is bounded on an open set containing S,*

(b) *g' is uniformly continuous on S, and*

(c) *g is uniformly K-regular on S,*

then if $\{x_i\}$ is the sequence generated by the algorithm of §2 with initial point x_o, then $\Delta(x_i) \to 0$, and $\{x_i\} \subset S$.

PROOF. Let us suppose that $\Delta(x_i) \nrightarrow 0$, then the sequence $\{x_i\}$ is infinite and there exist a subsequence $\{x_{ij}\}$ and a $\beta < 0$ such that $\Delta(x_{ij}) \leqslant \beta < 0$ for all $j = 1, 2, \ldots$. Since $\{\rho(x_i)\}$ is a strictly decreasing sequence that is bounded below, we know that $(\rho(x_{i+1}) - \rho(x_i)) \to 0$ and so $\lambda_i \Delta(x_i) \to 0$. But then $\lambda_{ij} \to 0$, since $\Delta(x_{ij}) \leqslant \beta < 0$ for all $j = 1, 2, \ldots$. With no loss of generality we further assume that $\lambda_{ij} < 1$ for all $j = 1, 2, \ldots$. The Armijo inequality of Step (3) of the algorithm, and the Lebourg mean value theorem [9], now tell us that

$$c\gamma^{-1}\lambda_{ij}\Delta(x_{ij}) < \rho\big(x_{ij} + \gamma^{-1}\lambda_{ij}d_{ij}\big) - \rho(x_{ij})$$
$$\leqslant \gamma^{-1}\lambda_{ij}\rho'(z_j; d_{ij})$$

for some z_j on the open line segment joining x_{ij} to $x_{ij} + \gamma^{-1}\lambda_{ij}d_{ij}$. Dividing this inequality through by $\gamma^{-1}\lambda_{ij}$, we obtain the inequality $c\Delta(x_{ij}) < \rho'(z_j; d_{ij})$. By Lemma (4.1) we can replace the right-hand side of this inequality by

$$\operatorname{dist}\big(g(z_j) + g'(z_j)d_{ij}, -K\big) - \rho(z_j)$$

to obtain

$$c\Delta(x_{ij}) < \operatorname{dist}\big(g(z_j) + g'(z_j)d_{ij}, -K\big) - \rho(z_j)$$
$$\leqslant 2\|g(z_j) - g(x_{ij})\| + \|g'(z_j) - g'(x_{ij})\|\,\|d_{ij}\| + \Delta(x_{ij}).$$

But $\lambda_{ij}d_{ij} \to 0$ as $j \to \infty$, since $\lambda_{ij} \to 0$ and g is uniformly K-regular on S. Hence

$$0 \leqslant (1-c)\limsup_j \Delta(x_{ij}),$$

where $-\mu(x_0) \leqslant \limsup_j \Delta(x_{ij}) \leqslant \beta < 0$, a contradiction.

If Y is taken to be a Hilbert space then a stronger result can be obtained.

(6.2) THEOREM. *Let the hypothesis of Theorem* (6.1) *hold. If* $\{x_i\}$ *is the sequence generated by the algorithm with initial point* x_0, *then either* $\rho(x_i) \to 0$ *or* $\operatorname{dist}(0, \partial\rho(x_{ij})) \to 0$ *for every subsequence* $\{x_{ij}\}$ *of the* x_i*'s for which* $\{\rho(x_{ij})\}$ *is bounded away from zero.*

(For the proof of this theorem, see [2].)

The above theorems only provide information concerning the limiting properties of certain stationarity criterion for ρ, making no claims about convergence of the x_i's. In such a general setting this is of course the best one could hope for without placing further restrictions upon S, g, g', and ρ (see [2]). But such a discussion lies far beyond the scope and purpose of this paper.

In finite dimensions the existence of cluster points of the sequence can be guaranteed by suitable hypothesis either on S or g. In this case Theorem (6.1) guarantees these points to be stationary points for ρ. Furthermore, if $\rho^*(x_k) = 0$ and $M = 0$, then our search directions are identical to those found in Robinson's Newton method [15]. In fact if the sequence has a cluster point $\bar{x}$ that satisfies Robinson's PL I condition [15], then it is possible to show, using techniques similar to those found in Robinson [15] or Garcia-Palomares and Restuccia [7], that the entire sequence converges to $\bar{x}$ at a quadratic rate. This result is not presented here as it requires a great deal of extra effort, but it may be found in [2]. We conclude this section with a corollary describing the convergence behavior for finite dimensions when it is assumed that the sequence remains bounded.

(6.3) COROLLARY. *Let g: $\mathbf{R}^n \to \mathbf{R}^m$ be continuously Fréchet differentiable on $\mathbf{R}^n$. Suppose that K is a closed convex cone contained in $\mathbf{R}^m$, and that ρ, ρ^*, D, and Δ are as defined in the algorithm of §2. Let $x_0 \in \mathbf{R}^n$. If the sequence $\{x_i\}$ generated by the algorithm with initial point x_0 is bounded, then either*

(i) *$\Delta(x_i) \to 0$ and every cluster point $\bar{x}$ of the sequence $\{x_i\}$ satisfies $0 \in \partial\rho(\bar{x})$, or*

(ii) *there is a cluster point $\bar{x}$ of $\{x_i\}$ such that g is not K-regular at $\bar{x}$.*

PROOF. We will assume that (ii) does not hold and prove that (i) does hold.

First note that if indeed $\Delta(x_i) \to 0$ and $\bar{x}$ is a cluster point of $\{x_i\}$, then g is by assumption K-regular at $\bar{x}$ and so by Theorem (5.3), $\Delta(x) := \rho^*(x) - \rho(x)$ is continuous at $\bar{x}$. Hence $\Delta(\bar{x}) = 0$ and $0 \in \partial\rho(\bar{x})$ by Theorem (4.2). Thus we would be done if we knew that $\Delta(x_i) \to 0$. Let us assume to the contrary that $\Delta(x_i) \nrightarrow 0$ and derive a contradiction. The proof proceeds in much the same way as that of Theorem (6.1).

Since $\Delta(x_i) \nrightarrow 0$, the sequence $\{x_i\}$ is infinite and there are a subsequence $\{x_{i_j}\}$ and a $\beta < 0$ such that $\Delta(x_{i_j}) \leqslant \beta < 0$ for all $j = 1, 2, \ldots$. Since $\{\rho(x_i)\}$ is a strictly decreasing sequence that is bounded below, we know that $(\rho(x_{i+1}) - \rho(x_i)) \to 0$ and so $\lambda_i\Delta(x_i) \to 0$. But then $\lambda_{i_j} \to 0$, since $\Delta(x_{i_j}) \leqslant \beta < 0$ for all $j = 1, 2, \ldots$. With no loss of generality we further assume that $\lambda_{i_j} < 1$ for all $j = 1, 2, \ldots$ and $x_{i_j} \to \bar{x} \in \mathbf{R}^n$. Moreover, since by assumption g is K-regular at $\bar{x}$, we shall assume that $\{x_{i_j}\} \subset N(\bar{x})$ where $N(\bar{x})$ is some compact neighborhood of $\bar{x}$ on which g is uniformly K-regular. Now, exactly as in Theorem (6.1), we can employ the Armijo inequality and the Lebourg mean value theorem to obtain the inequality

$$0 < (1 - c)\Delta(x_{i_j}) + 2\|g(z_j) - g(x_{i_j})\| + \|g'(z_j) - g'(x_{i_j})\| \|d_{i_j}\|,$$

where z_j is some point on the line segment joining x_{i_j} to $x_{i_j} + \gamma^{-1}\lambda_{i_j}d_{i_j}$. Now taking limits and remembering that by K-regularity, Δ is continuous at $\bar{x}$ and the d_{i_j}'s are bounded, we get that $0 < (1 - c)\Delta(\bar{x}) \leqslant (1 - c)\beta < 0$, a contradiction, thereby establishing the result. ∎

References

[1] Auslenden, A. (1977). Convex Programming with Errors: Methods of ϵ-Subgradients. *C. R. Acad. Sci. Paris Sér. A-13* **284** A109–A112.

[2] Burke, J. V. (1983). Methods for Solving Generalized Inequalities with Applications to Nonlinear Programming. Ph.D. Thesis, Department of Mathematics, University of Illinois at Urbana-Champaign.

[3] Clarke, F. H. (1981). Generalized Gradients of Lipschitz Functionals. *Adv. in Math.* **40** 52–67.

[4] Daniel, J. W. (1973). Newton's Method for Nonlinear Inequalities. *Numer. Math.* **21** 381–387.

[5] Demyanov, V. F. and Malozemov, V. N. (1974). *Introduction to Minimax*. John Wiley & Sons, Inc., New York.

[6] Garcia-Palomares, U. M. (1983). On the Minimax Solution of a Nonlinear System of Mixed Equalities and Inequalities. Preprint, Departmento de Procesos y sistemas, Universidad Simon Bolivar, Caracas, Venezuela.

[7] ______ and Restuccia, A. (1981). Application of the Armijo Stepsize Rule in the Solution of a Nonlinear System of Equalities and Inequalities. *J. Optimi. Theory Appl.* **41** 405–415.

[8] ______ and ______ (1981). A Global Quadratic Algorithm for Solving a System of Mixed Equalities and Inequalities. *Math. Programming* **21** 290–300.

[9] Lebourg, G. (1975). Valeur Moyenne pour Gradient Généralisé. *C. R. Acad. Sci. Paris* **281** 795–797.

[10] Moreau, J.-J. (1966). Fonctionelles Convexes. Lecture Notes, Séminaire "Equations aux Derivées Partielles, Collège de France.

[11] Nurminski, E. (1978). Nondifferentiable Optimization with ϵ-Subgradient Methods. Working Paper 78-55, International Institute for Applied Systems Analysis.

[12] Polak, E., Mayne, D. Q. and Wardi, Y. (1983). On the Extension of Constrained Optimization Algorithms from Differentiable to Nondifferentiable Problems. *SIAM J. Control Optimi.* **21** 179–203.

[13] Polyak, B. T. (1964). Gradient Methods for Solving Equations and Inequalities. *USSR Computational Math.* **4** 17–32.

[14] Pshenichnyi, B. N. (1970). Newton's Method for the Solution of Systems of Equalities and Inequalities. *Math. Notes Acad. Sci. USSR* **8** 827–830.

[15] Robinson, S. M. (1972). Extension of Newton's Method to Nonlinear Functions with Values in a Cone. *Numer. Math.* **19** 341–347.

[16] Rockafellar, R. T. (1970). *Convex Analysis*. Princeton University Press, Princeton, N.J.

[17] Schittkowski, K. (1975). Numerical Solution of Systems of Nonlinear Inequalities. *Optimization and Operations Research*; *Lecture Notes in Economics and Mathematical Systems* **117**. Springer-Verlag, Berlin, 259–272.

[18] Vlasov, L. P. (1975). Approximative Properties of Sets in Normed Linear Spaces. *Russian Math. Surveys* **28** 1–61.

[19] Fletcher, R. (1981). *Practical Methods for Optimization. Vol. II, Constrained Optimization*. Wiley, New York.

BURKE: DEPARTMENT OF MATHEMATICS, UNIVERSITY OF WASHINGTON, SEATTLE WASHINGTON 98195

HAN: DEPARTMENT OF MATHEMATICS, UNIVERSITY OF ILLINOIS, URBANA, ILLINOIS 61801

Copyright 1986, by INFORMS, all rights reserved. Copyright of Mathematics of Operations Research is the property of INFORMS: Institute for Operations Research and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.