Mathematical Programming Study 10 (1979) 128–141.
North-Holland Publishing Company

GENERALIZED EQUATIONS AND THEIR SOLUTIONS, PART I: BASIC THEORY

Stephen M. ROBINSON*
University of Wisconsin-Madison, Madison, Wisconsin, U.S.A.

Received 28 November 1977

We consider a class of "generalized equations," involving point-to-set mappings, which formulate the problems of linear and nonlinear programming and of complementarity, among others. Solution sets of such generalized equations are shown to be stable under certain hypotheses; in particular a general form of the implicit function theorem is proved for such problems. An application to linear generalized equations is given at the end of the paper; this covers linear and convex quadratic programming and the positive semidefinite linear complementarity problem. The general nonlinear programming problem is treated in Part II of the paper, using the methods developed here.

Key words: Variational Inequalities, Generalized Equations, Monotone Operators, Nonlinear Complementarity Problem, Nonlinear Programming, Economic Equilibria.

1. Introduction

In this paper we shall study the behavior of solutions of the *generalized equation*

$$0 \in f(x) + T(x), \tag{1.1}$$

where f is a continuously Fréchet differentiable function from an open set $\Omega \subset \mathbf{R}^n$ into $\mathbf{R}^n$ and T is a maximal monotone operator from $\mathbf{R}^n$ into itself (recall that an operator T is *monotone* if for each (x_1, w_1), (x_2, w_2) in graph T one has

$$\langle x_1 - x_2, w_1 - w_2 \rangle \geq 0,$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product, and *maximal monotone* if its graph is not properly contained in that of any other monotone operator). We use the term "generalized equation" because if T is identically zero, then (1.1) reduces to the equation $f(x) = 0$, and because systems like (1.1) retain some of the analytic properties of nonlinear equations, as we shall show in what follows.

We shall be particularly interested in conditions which, when imposed on f and T, will ensure that the set of solutions to (1.1) remains nonempty and is well behaved (in a sense to be defined) when f is subjected to small perturbations. To introduce these perturbations, we shall make use of a topological space P and a

* Sponsored by the United States Army under Contract No. DAAG29–75–C–0024 and by the National Science Foundation under Grant No. MCS74–20584 A02.

function $f: P \times \Omega \to \mathbf{R}^n$, so that we can replace (1.1) by

$$0 \in f(p, x) + T(x), \tag{1.2}$$

and study the set of x which solve (1.2) as p varies near a base value p_0.

A particular case of (1.2) of special interest for applications is that in which T is taken to be the operator $\partial\psi_C$, where for a closed convex set $C \subset \mathbf{R}^n$ one defines the indicator function ψ_C of C by

$$\psi_C(x) := \begin{cases} 0, & x \in C, \\ +\infty, & x \notin C, \end{cases}$$

and where ∂ denotes the subdifferential operator [13, Section 23]. This yields the special generalized equation

$$0 \in f(p, x) + \partial\psi_C(x), \tag{1.3}$$

which expresses analytically the geometric idea that $f(p, x)$ is an inward normal to C at x.

Many problems from mathematical programming, complementarity, mathematical economics and other fields can be represented in the form (1.3): for example, the nonlinear complementarity problem

$$F(x) \in K^*, \qquad x \in K, \qquad \langle x, F(x)\rangle = 0 \tag{1.4}$$

where $F: \mathbf{R}^n \to \mathbf{R}^n$, K is a nonempty polyhedral convex cone in $\mathbf{R}^n$, and $K^* := \{y \in \mathbf{R}^n \mid \langle y, k\rangle \geq 0 \text{ for each } k \in K\}$, can be written as

$$0 \in F(x) + \partial\psi_K(x).$$

Further information on nonlinear complementarity problems (often with $K = \mathbf{R}^n_+$, the non-negative orthant) may be found in, e.g., [2, 4, 7, 8]. The Kuhn–Tucker necessary conditions for mathematical programming [6] form a special case of (1.4); e.g., for the problem

$$\begin{aligned} &\text{minimize} \quad \theta(y), \\ &\text{subject to} \quad g(y) \leq 0, \qquad h(y) = 0. \end{aligned} \tag{1.5}$$

where θ, g and h are differentiable functions from $\mathbf{R}^m$ into $\mathbf{R}$, $\mathbf{R}^q$ and $\mathbf{R}^r$ respectively, one has the Kuhn–Tucker conditions

$$\begin{aligned} &\theta'(y) + ug'(y) + vh'(y) = 0, \qquad g(y) \leq 0, \\ &h(y) = 0, \qquad u \geq 0, \qquad \langle u, g(y)\rangle = 0, \end{aligned}$$

and these can be written in the form (1.4) by taking $n = m + q + r$, $K = \mathbf{R}^m \times \mathbf{R}^q_+ \times \mathbf{R}^r$, $x = (y, u, v)$ and

$$F(x) = \begin{bmatrix} [\theta'(y) + ug'(y) + vh'(y)]^{\mathrm{T}} \\ -g(y) \\ -h(y) \end{bmatrix}.$$

There are also important applications of (1.3) to economic equilibrium problems [15], among others. It is of interest to note that in most of the applications mentioned one finds that C is a polyhedral convex set, and we shall see that particularly strong results can be obtained for such problems.

It is also worth pointing out that problems of linear or quadratic programming lead to linear generalized equations: for example, if $P \subset \mathbf{R}^m$ and $Q \subset \mathbf{R}^l$ are two polyhedral convex cones, H and A are matrices of dimensions $m \times m$ and $l \times m$ respectively, $c \in \mathbf{R}^m$ and $a \in \mathbf{R}^l$, then we can consider the quadratic programming problem

$$\begin{aligned} &\text{minimize} \quad \tfrac{1}{2}\langle x, Hx\rangle + \langle c, x\rangle, \\ &\text{subject to} \quad a - Ax \in Q^*, \quad x \in P, \end{aligned} \tag{1.6}$$

where Q^* is the dual cone of Q. The necessary optimality conditions for (1.6) are (assuming without loss of generality that H is symmetric):

$$\begin{aligned} &x^{\mathrm{T}}H + c + uA \in P^*, && a - xAx \in Q^*, \\ &x \in P, && u \in Q, \\ &\langle x^{\mathrm{T}}H + c + uA, x\rangle = 0, && \langle u, a - Ax\rangle = 0. \end{aligned}$$

These can be formulated in a somewhat more transparent manner by writing them as

$$0 \in \begin{bmatrix} H & A^{\mathrm{T}} \\ -A & 0 \end{bmatrix}\begin{pmatrix} x \\ u \end{pmatrix} + \begin{bmatrix} c \\ a \end{bmatrix} + \partial\psi_{P\times Q}\begin{pmatrix} x \\ u \end{pmatrix},$$

a linear generalized equation which, if P and Q are taken to be $\mathbf{R}^m$ and $\mathbf{R}^l$ respectively (i.e., in the case of quadratic programming with equality constraints and unconstrained variables) reduces to an ordinary linear equation. We shall see that linear generalized equations are basic to the analysis done here in much the same way as linear equations are basic to the analysis of nonlinear equations.

The organization of this paper is as follows: in the next section we state and prove the main result (Theorem 1) after defining a property used in the statement. We also discuss a way of simplifying (by restricting) one of the hypotheses. In Section 3 we examine a class of multivalued functions frequently found in applications, and show that they have one of the key properties needed in Theorem 1. Finally, in Section 4 we apply the results of Sections 2 and 3 to linear generalized equations. Applications to nonlinear problems will be the subject of Part II of this paper.

2. Main results

Before stating the main theorem, we require a preliminary definition dealing with a certain continuity property of multivalued functions (or *multifunctions*, as we shall call them).

Definition 1. Let X and Y be normed linear spaces. A multifunction $F : X \to Y$ is *upper Lipschitzian with modulus* λ, or U.L.(λ), at a point $x_0 \in X$ with respect to a set $V \subset X$, if for each $v \in V$ one has

$$F(v) \subset F(x_0) + \lambda \|v - x_0\| B_Y,$$

where B_Y is the unit ball in Y. We say F is *locally* U.L.(λ) at x_0 if it is U.L.(λ) at x_0 with respect to some neighborhood of x_0.

This property is close to the Lipschitz continuity for multifunctions defined by Rockafellar [14, Section 3], except that we do not require $F(x_0)$ to be a singleton; in the problems we shall consider $F(x_0)$ will often be multivalued. Note that the distance from any point of $F(v)$ to the set $F(x_0)$ is bounded above by $\lambda\|v - x_0\|$, although the distance from a point of $F(x_0)$ to $F(v)$ may be large.

Before stating the main theorem, we shall try to motivate its hypotheses. Recall that in the classical inverse-function theorem the key assumption is that the linearization of the function being considered, about a point x_0 in the inverse image of 0, should be regular: specifically, that the inverse image of 0 under the linearized function should be a singleton (in fact, the point x_0 itself). In our situation, since we may be dealing with solution sets rather than points, we have to linearize about each point in a set. The first assumption in the theorem is that there is a nonempty bounded set X_0 (analogous to the point x_0 in the classical case) such that the inverse image of 0, under an appropriate kind of linearization performed at any point of X_0, is X_0 itself together with, possibly, points outside some neighborhood of X_0. There is also an assumption of uniform upper Lipschitz continuity, which is automatically true in the classical case. Finally, there is an assumption that the inverse image of any point near 0, under the linearization previously mentioned, has a convex component in the neighborhood of X_0 within which we are working. In the classical case this is equivalent to the first hypothesis, but not so here.

We shall show, below and in Part II, that many problems of practical interest satisfy these hypotheses. In particular we show in Proposition 1 that the third assumption can be replaced by an assumption of positive semidefiniteness which is often satisfied in applications.

In the following theorem, we use f_2 to denote the partial Fréchet derivative, with respect to the second argument, of a function $f(p, x)$ of two variables; B denotes the unit ball in $\mathbf{R}^n$ with respect to the Euclidean norm, which is used throughout the remainder of the paper.

Theorem 1. *Let P be a topological space, Ω an open set in $\mathbf{R}^n$ and T a closed multifunction from $\mathbf{R}^n$ into itself. Let f be a continuous function from $P \times \Omega$ into $\mathbf{R}^n$ such that f_2 is continuous on $P \times \Omega$. Let $p_0 \in P$; write $Lf_{x_0}(x)$ for $f(p_0, x_0) + f_2(p_0, x_0)(x - x_0)$. Suppose that there are a nonempty, bounded convex set X_0 and constants λ, $\gamma > 0$ and $\eta > 0$ with $X_\gamma := X_0 + \gamma B \subset \Omega$, such that for each $x_0 \in X_0$:*

(i) $X_\gamma \cap (Lf_{x_0} + T)^{-1}(0) = X_0$;
(ii) $X_\gamma \cap (Lf_{x_0} + T)^{-1}$ *is* U.L.(λ) *at* 0 *with respect to* ηB;
(iii) *for each* $y \in \eta B$, $X_\gamma \cap (Lf_{x_0} + T)^{-1}(y)$ *is convex and nonempty.*

Then there exist a number $\delta \in (0, \gamma]$ *and a neighborhood* $U(p_0)$ *such that with*

$$\Sigma(p) := \begin{cases} \{x \in X_0 + \delta B \mid 0 \in f(p, x) + T(x)\} & p \in U, \\ \emptyset & p \notin U, \end{cases}$$

one has:

(1) Σ *is upper semicontinuous from* U *to* $\mathbf{R}^n$;
(2) $\Sigma(p_0) = X_0$; *and*
(3) *For each* $\epsilon > 0$, *for some neighborhood* $U_\epsilon(p_0)$ *and for each* $p \in U_\epsilon$,

$$\emptyset \neq \Sigma(p) \subset \Sigma(p_0) + (\lambda + \epsilon)\alpha_0(p)B,$$

where

$$\alpha_0(p) := \max\{\|f(p, x) - f(p_0, x)\| \mid x \in X_0\}.$$

Note that if P is actually a normed linear space and if $f(p, x)$ is Lipschitzian in p uniformly over $x \in X_0$, then for some constant μ and each $p \in U_\epsilon$ we have

$$\Sigma(p) \subset \Sigma(p_0) + (\lambda + \epsilon)\mu\|p - p_0\|B,$$

so that Σ is locally U.L.$[(\lambda + \epsilon)\mu]$ at p_0.

Proof. Choose $x_0 \in X_0$; denote $Lf_{x_0} + T$ by $Q(x_0)$. Let $\theta \in (0, \eta]$ with $\lambda\theta \leq \gamma$ and let $y \in \theta B$; then $X_\gamma \cap Q(x_0)^{-1}(y) \subset X_0 + \lambda\|y\|B \subset X_\gamma$. Hypothesis (iii), together with closure of $Q(x_0)$, implies that for each $y \in \theta B$, $X_\gamma \cap Q(x_0)^{-1}(y)$ is non-empty, compact and convex. In particular, X_0 is a compact convex set.

The basic idea of the proof is to approximate the inverse of the operator $f(p, x) + T(x)$ by the inverse of the operator

$$\begin{aligned} Q(\pi(x))(z) &:= Lf_{\pi(x)}(z) + T(z) \\ &= f(p_0, \pi(x)) + f_2(p_0, \pi(x))(z - \pi(x)) + T(z), \end{aligned}$$

where $\pi(x)$ is the closest point to x in X_0, just as one approximates the inverse of a function in the classical inverse-function theorem by the inverse of its linearization about some point. We then apply a fixed-point theorem; in proving the inverse-function theorem one usually uses the contraction principle, but here we have to use the Kakutani theorem. Observe that the "linearized" operator appearing here is of the type we discussed above in considering linear generalized equations; this illustrates our comment that these operators play a rôle in the analysis of generalized equations analogous to that of linear operators in classical analysis.

Of course, during this approximation it will be necessary to be careful that we work with the correct component of the inverse image (i.e., that lying in X_γ), and this adds a certain amount of complexity to the notation.

Define, for two subsets A and C of $\mathbf{R}^n$ and a point $x \in \mathbf{R}^n$, $d[x, C] := \inf(\|x - c\| \mid c \in C\}$ and $d[A, C] = \sup\{d[a, C] \mid a \in A\}$, where the supremum and infimum of $\emptyset$ are defined to be $-\infty$ and $+\infty$ respectively. Denote by π the projection from $\mathbf{R}^n$ onto X_0; π is well known to be nonexpansive, hence *a fortiori* continuous.

Using continuity and compactness, one can show that the function

$$\beta(\delta) := \max\{\|f_2(p_0, x) - f_2(p_0, \pi(x))\| \mid x \in X_0 + \delta B\}$$

is well defined for small δ, and is continuous at 0 with $\beta(0) = 0$. Thus, we can choose a $\delta \in (0, \gamma]$ such that $\lambda\beta(\delta) \leq \frac{1}{2}$ and $\delta\beta(\delta) \leq \frac{1}{2}\theta$. It is not difficult to show that for this fixed δ the function

$$\alpha_\delta(p) := \max\{\|f(p, x) - f(p_0, x)\| \mid x \in X_\delta\}$$

is well defined for all $p \in P$, and is continuous at p_0 with $\alpha_\delta(p_0) = 0$. Thus, we can choose a neighborhood $U(p_0)$ such that for each $p \in U$, $\alpha_\delta(p) < \frac{1}{2}\theta$ and $\lambda\alpha_\delta(p) \leq \frac{1}{2}\delta$. Now choose any $p \in U$, and define a multifunction F_p from X_δ into $\mathbf{R}^n$ by

$$F_p(x) := X_\gamma \cap Q(\pi(x))^{-1}[Lf_{\pi(x)}(x) - f(p, x)].$$

If x is any point of X_δ, we have

$$\|Lf_{\pi(x)}(x) - f(p, x)\| \leq \|f(p, x) - f(p_0, x)\| + \|f(p_0, x) - Lf_{\pi(x)}(x)\|. \tag{2.1}$$

Now define (for this fixed x) a function of one real variable τ by

$$g(\tau) := f(p_0, \tau x + (1 - \tau)\pi(x)) - Lf_{\pi(x)}(\tau x + (1 - \tau)\pi(x)).$$

We find that

$$\|f(p_0, x) - Lf_{\pi(x)}(x)\| = \|g(1) - g(0)\| \leq \sup(\|g'(\tau)\| \mid 0 < \tau < 1\}.$$

However, for $\tau \in [0, 1]$,

$$g'(\tau) = [f_2(p_0, x_\tau) - f_2(p_0, \pi(x))][x - \pi(x)],$$

where $x_\tau := \tau x + (1 - \tau)\pi(x)$. We have by properties of the projection that $\pi(x_\tau) = \pi(x)$, so

$$\|f_2(p_0, x_\tau) - f_2(p_0, \pi(x))\| = \|f_2(p_0, x_\tau) - f_2(p_0, \pi(x_\tau))\| \leq \beta(\delta),$$

since $x_\tau \in X_\delta$. Hence

$$\|f(p_0, x) - Lf_{\pi(x)}(x)\| \leq \beta(\delta)\|x - \pi(x)\|. \tag{2.2}$$

As $\|f(p, x) - f(p_0, x)\| \leq \alpha_\delta(p)$, we have from (2.1) and (2.2)

$$\|Lf_{\pi(x)}(x) - f(p, x)\| \leq \alpha_\delta(p) + \beta(\delta)\|x - \pi(x)\| < \tfrac{1}{2}\theta + \tfrac{1}{2}\theta = \theta. \tag{2.3}$$

Hence, by our previous remarks $F_p(x)$ is a nonempty compact convex set for each $x \in X_\delta$. Also, using (i), (ii) and (2.3) we have for $x \in X_\delta$,

$$\begin{aligned} d[F_p(x), X_0] &= d[X_\gamma \cap Q(\pi(x))^{-1}[Lf_{\pi(x)}(x) - f(p,x)], X_\gamma \cap Q(\pi(x))^{-1}(0)] \\ &\le \lambda \|Lf_{\pi(x)}(x) - f(p,x)\| \qquad (2.4) \\ &\le \lambda\alpha_\delta(p) + \lambda\beta(\delta)\|x - \pi(x)\| \le \tfrac{1}{2}\delta + \tfrac{1}{2}\delta = \delta, \end{aligned}$$

so F_p carries X_δ into itself. We have

$$\begin{aligned} \text{graph } F_p &= \{(x,y) \mid x \in X_\delta,\ y \in X_\delta,\ Lf_{\pi(x)}(x) - f(p,x) \in Lf_{\pi(x)}(y) + T(y)\} \\ &= \{(x,y) \mid 0 \in f(p,x) + f_2(p_0, \pi(x))(y-x) + T(y)\} \cap (X_\delta \times X_\delta). \end{aligned}$$

Using the continuity of f, f_2, and π, together with the fact that T is closed, one can show without difficulty that graph F_p is closed in $X_\delta \times X_\delta$. We can thus apply the Kakutani fixed-point theorem [5, 9] to conclude that there is some $x_p \in X_\delta$ with $x_p \in F_p(x_p)$; that is,

$$Lf_{\pi(x_p)}(x_p) - f(p, x_p) \in Lf_{\pi(x_p)}(x_p) + T(x_p),$$

so

$$0 \in f(p, x_p) + T(x_p)$$

and thus $x_p \in \Sigma(p)$, which is therefore nonempty. We have

$$\text{graph } \Sigma = \{(p,x) \in U \times X_\delta \mid 0 \in f(p,x) + T(x)\};$$

this is closed in $U \times X_\delta$ by joint continuity of f and closure of T. However, the range of Σ is contained in the compact set X_δ; thus by [9, Lemma 4.4] Σ is actually upper semicontinuous from U to X_δ. If $x_0 \in X_0$ then by (i) one has $0 \in Lf_{x_0}(x_0) + T(x_0) = f(p_0, x_0) + T(x_0)$, so $x_0 \in \Sigma(p_0)$ and thus $\Sigma(p_0) \supset X_0$. On the other hand, if $x \in \Sigma(p_0)$ then $x \in X_\delta$ and $0 \in f(p_0, x) + T(x)$; therefore

$$Lf_{\pi(x)}(x) - f(p_0, x) \in Lf_{\pi(x)}(x) + T(x),$$

so that $x \in F_{p_0}(x)$. As $x \in X_\delta$, we have from (2.4) with $p = p_0$ that

$$d[x, X_0] \le d[F_{p_0}(x), X_0] \le \lambda \|Lf_{\pi(x)}(x) - f(p_0, x)\|.$$

But from (2.3) with $p = p_0$, we find that

$$\|Lf_{\pi(x)}(x) - f(p_0, x)\| \le \beta(\delta)\|x - \pi(x)\| = \beta(\delta) d[x, X_0].$$

Thus

$$d[x, X_0] \le \lambda\beta(\delta) d[x, X_0] \le \tfrac{1}{2} d[x, X_0],$$

implying that $x \in X_0$ since X_0 is closed. Thus we actually have $\Sigma(p_0) = X_0$.

Now take any $\epsilon > 0$; find $\delta_\epsilon \in (0, \delta]$ such that for $\sigma \in [0, \delta_\epsilon]$ one has $\lambda\beta(\sigma) \le \frac{1}{2}\epsilon/(\lambda + \epsilon)$. One can show that the function

$$\gamma(p) := \max \|f_2(p,x) - f_2(p_0, x)\| \mid x \in X_0 + \delta_\epsilon B\}$$

is well defined on P and is continuous at p_0; choose a neighborhood $U_\epsilon(p_0) \subset U$ so that if $p \in U_\epsilon$ we have $\Sigma(p) \subset \Sigma(p_0) + \delta_\epsilon B$ and $\lambda\gamma(p) \le \frac{1}{2}\epsilon/(\lambda + \epsilon)$. Now choose any $p \in U_\epsilon$ and any $x \in \Sigma(p)$. Using (2.4) and the fact that $x \in F_p(x)$, we have

$$\begin{aligned} d[x, \Sigma(p_0)] &\le d[F_p(x), X_0] \\ &\le \lambda \|Lf_{\pi(x)}(x) - f(p, x)\| \\ &\le \lambda \|h(x) - h(\pi(x))\| + \lambda \|h(\pi(x))\| + \lambda \|f(p_0, x) - Lf_{\pi(x)}(x)\|, \end{aligned} \tag{2.5}$$

where $h(x) := f(p, x) - f(p_0, x)$. If we define, as before, $x_\tau := \tau x + (1 - \tau)\pi(x)$, we have

$$\|h(x) - h(\pi(x))\| \le \|x - \pi(x)\| \sup\{\|h'(x_\tau)\| \mid 0 < \tau < 1\}.$$

But $h'(x_\tau) = f_2(p, x_\tau) - f_2(p_0, x_\tau)$, so

$$\|h(x) - h(\pi(x))\| \le \gamma(p)\|x - \pi(x)\|.$$

Thus, using (2.2), (2.5) and the fact that $\|h(\pi(x))\| \le \alpha_0(p)$, we have

$$\begin{aligned} d[x, \Sigma(p_0)] &\le \lambda\gamma(p)\|x - \pi(x)\| + \lambda\alpha_0(p) + \lambda\beta(\delta_\epsilon)\|x - \pi(x)\| \\ &\le [\epsilon/(\lambda + \epsilon)]\|x - \pi(x)\| + \lambda\alpha_0(p). \end{aligned}$$

But $\|x - \pi(x)\| = d[x, \Sigma(p_0)]$, so if $\lambda > 0$ we obtain

$$[\lambda/(\lambda + \epsilon)]d[x, \Sigma(p_0)] \le \lambda\alpha_0(p)$$

and thus

$$d[x, \Sigma(p_0)] \le (\lambda + \epsilon)\alpha_0(p). \tag{2.6}$$

On the other hand, if $\lambda = 0$, then (2.5) implies that $d[x, \Sigma(p_0)] = 0$, in which case (2.6) holds trivially. In either case, therefore,

$$\Sigma(p) \subset \Sigma(p_0) + (\lambda + \epsilon)\alpha_0(p)B,$$

which completes the proof.

Verification of the hypotheses of this theorem in a particular case may be difficult; this is particularly true of (ii) and (iii). It is therefore desirable to look for classes of problems for which this verification may be easier. In the next section we exhibit such a class for hypothesis (ii); we do so for (iii) in the following proposition.

Proposition 1. *In Theorem* 1, *the hypothesis* (iii) *may be replaced by*
(iii)′ $f_2(p_0, x_0)$ *is positive semidefinite and T is maximal monotone.*

Proof. We shall show that (iii)′, together with the other hypotheses of Theorem 1, implies (iii). Choose any $x_0 \in X_0$; under (iii)′ the function Lf_{x_0} will be a maximal monotone operator. As T is also maximal monotone and as dom Lf_{x_0} (the effective domain of Lf_{x_0}) is all of $\mathbf{R}^n$, we have from [1, Corollary 2.7] that $Q(x_0)$ is maximal monotone; hence so is $Q(x_0)^{-1}$. The set $Q(x_0)^{-1}(0)$ is then convex, so that (i) implies that $Q(x_0)^{-1}(0) = x_0$. It follows that for $y \in \eta B$, $X_\gamma \cap Q(x_0)^{-1}(y) \subset X_0 + \lambda\|y\|B$ (by (ii)). Now let $\alpha \in (0, \eta]$ with $\lambda\alpha < \gamma$. If $y \in \alpha B$, the convexity of $Q(x_0)^{-1}(y)$ implies that $X_\gamma \cap Q(x_0)^{-1}(y) = Q(x_0)^{-1}(y)$, so $Q(x_0)^{-1}$ is locally U.L.(λ) at 0. But this, together with the boundedness of $Q(x_0)^{-1}(0)$, shows that $Q(x_0)^{-1}$ is locally bounded at 0; in fact, it must be locally bounded at every point of $\operatorname{int} \alpha B$, since the image of some ball around such a point will be contained in the image of αB, which in turn is contained in the bounded set $X_{\lambda\alpha} = X_0 + \lambda\alpha B$. But then from [12, Theorem 1] we have that $\operatorname{int} \alpha B$ cannot contain any boundary point of $\operatorname{dom} Q(x_0)^{-1}$; however, as $\operatorname{int} \alpha B$ meets $\operatorname{dom} Q(x_0)^{-1}$ (at 0) and is connected we finally conclude that $\operatorname{int} \alpha B \subset \operatorname{int} \operatorname{dom} Q(x_0)^{-1}$. Thus, for each y with $\|y\| < \alpha$ the set $Q(x_0)^{-1}(y)$ is nonempty, convex and contained in $X_{\lambda\alpha} \subset X_\gamma$. Now let η_0 be any positive number smaller than α. As hypothesis (ii) of Theorem 1 was true for η, and as $\alpha \le \eta$, that hypothesis will be satisfied also for η_0; as we have just seen, hypothesis (iii) also holds for η_0, and this proves Proposition 1.

The hypothesis (iii)′ is certainly simpler than is (iii); however, (iii) covers a more general class of problems. For example, consider the linear generalized equation

$$0 \in -\alpha x + \beta + \partial\psi_{[-1,1]}(x),$$

where $\alpha > 0$. This does not satisfy (iii)′; however, if $|\beta| \neq \alpha$ then each of its solutions (one if $|\beta| > \alpha$, three if $|\beta| < \alpha$) can be analyzed under (iii). If $|\beta| = \alpha$ then the solution at $-\operatorname{sgn} \beta$ can be so analyzed, but the solution at $\operatorname{sgn} \beta$ cannot (indeed, the conclusions of Theorem 1 fail for that solution).

3. Polyhedral multifunctions

In the last section, we exhibited a class of problems for which hypothesis (iii) of Theorem 1 always held. Here we do somewhat the same thing for hypothesis (ii): we show that for a class of multifunctions important in applications to optimization and equilibrium problems, local upper Lipschitz continuity holds at each point of the range space. The problem of verifying hypothesis (ii), in the case of such functions, then reduces to that of showing that the Lipschitz constants are uniformly bounded and that the continuity holds on a fixed neighborhood for each function in the family considered. For the application given in Section 4 this is trivial; some cases in which it is non-trivial are treated in Part II.

Definition 2. A multifunction $Q : \mathbf{R}^n \to \mathbf{R}^m$ is *polyhedral* if its graph is the union of a finite (possibly empty) collection of polyhedral convex sets (called *components*).

Here we use "polyhedral convex set" as in [13, Section 19].

It is clear that a polyhedral multifunction is always closed, and that its inverse is likewise polyhedral. Further, one can show without difficulty that the class of polyhedral multifunctions is closed under scalar multiplication, (finite) addition, and (finite) composition. The following proposition shows that they have good properties also with respect to upper Lipschitz continuity. For brevity, we omit the proofs of this proposition and the next; they may be found in [10].

Proposition 2. *Let F be a polyhedral multifunction from $\mathbf{R}^n$ into $\mathbf{R}^m$. Then there exists a constant λ such that F is locally* U.L.(λ) *at each $x_0 \in \mathbf{R}^n$.*

It is worth pointing out that λ depends only on F and not on x_0, although of course the size of the neighborhood of x_0 within which the continuity holds will in general depend on x_0.

The importance of polyhedral multifunctions for applications is illustrated by the following fact, in the statement of which we use the concepts of subdifferential and of a polyhedral convex function (one whose epigraph is a polyhedral convex set), which are discussed further in [13].

Proposition 3. *Let f be a polyhedral convex function from $\mathbf{R}^n$ into $(-\infty, +\infty]$. Then the subdifferential ∂f is a polyhedral multifunction.*

It follows from this proposition that subdifferentials of polyhedral convex functions display the upper Lipschitz continuity required in Theorem 1. In view of our earlier remarks about polyhedral multifunctions, this behavior is not lost if we combine these subdifferentials in various ways with other polyhedral multifunctions. For example, let C be a nonempty polyhedral convex set in $\mathbf{R}^n$ and let $\psi_C : \mathbf{R}^n \to (-\infty, +\infty]$ be its indicator function, defined by

$$\psi_C(x) := \begin{cases} 0, & x \in C, \\ +\infty, & x \notin C. \end{cases}$$

It is readily verified that ψ_C is a polyhedral convex function. Now, if A is a linear transformation from $\mathbf{R}^n$ into itself and $a \in \mathbf{R}^n$, then the operator $Ax + a + \partial\psi_C(x)$ and its inverse are, by Propositions 2 and 3, everywhere locally upper Lipschitzian. Hence, generalized linear equations have good continuity properties with respect to perturbations of the right-hand side; we shall exploit this fact in the next section.

This discussion also shows that, if the operator T in Theorem 1 is polyhedral, then the linearized operators $Lf_{x_0} + T$ have at least some of the continuity properties required in hypothesis (ii) of that theorem; it is still necessary to prove uniformity, but this is trivial if X_0 is a singleton, while in general it can often be done by using the structure of the problem (e.g., in nonlinear programming: see Part II of this paper).

4. An application: stability of a linear generalized equation

To illustrate an application of Theorem 1, we specialize it to analyze the behavior of the solution set of the linear generalized equation

$$0 \in Ax + a + \partial\psi_C(x), \tag{4.1}$$

where A is an $n \times n$ matrix, $a \in \mathbf{R}^n$, and C is a nonempty polyhedral convex set in $\mathbf{R}^n$. Such problems include, as special cases, the problems of linear and quadratic programming and the linear complementarity problem. We shall characterize stability of the solution set of (4.1) when the matrix A is positive semidefinite (but not necessarily symmetric); a more general (but more complicated) result could be obtained by dropping the assumption of positive semidefiniteness but assuming hypothesis (iii) of Theorem 1.

Theorem 2. *Let A be a positive semidefinite $n \times n$ matrix, C be a nonempty polyhedral convex set in $\mathbf{R}^n$ and $a \in \mathbf{R}^n$. Then the following are equivalent:*

(a) *The solution set of (4.1) is nonempty and bounded.*

(b) *There exists $\epsilon_0 > 0$ such that for each $n \times n$ matrix A' and each $a' \in \mathbf{R}^n$ with*

$$\epsilon' := \max\{\|A' - A\|, \|a' - a\|\} < \epsilon_0, \tag{4.2}$$

the set

$$S(A', a') := \{x \mid 0 \in A'x + a' + \partial\psi_C(x)\}$$

is nonempty.

Further, suppose these conditions hold; let μ be a bound on $S(A, a)$, and λ be a local upper Lipschitz constant for $[A(\cdot) + a + \partial\psi_C(\cdot)]^{-1}$ at 0 (which exists by the results of Section 3). Then for any open bounded set Ψ containing $S(A, a)$ there is some $\epsilon_1 > 0$ such that for each A', a' with $\max\{\|A' - A\|, \|a' - a\|\} < \epsilon_1$ we have

$$\emptyset \neq S(A', a') \cap \Psi \subset S(A, a) + \lambda\epsilon'(1 - \lambda\epsilon')^{-1}(1 + \mu)B. \tag{4.3}$$

Finally, if (A', a') are restricted to values for which $S(A', a')$ is known to be connected (in particular, if A' is restricted to be positive semidefinite), then Ψ can be replaced by $\mathbf{R}^n$.

Proof. (b$\Rightarrow$a) If (b) holds then in particular $S(A, a')$ is nonempty for all a' in some ball about a. This means that 0 belongs to the interior of the range of the operator $A(\cdot) + a + \partial\psi_C(\cdot)$, which is maximal monotone by [1, Corollary 2.7]. Accordingly, the inverse of this operator is locally bounded at 0 [1, Proposition 2.9] and so in particular $S(A, a)$ is bounded.

(a$\Rightarrow$b) We apply Theorem 1, taking P to be the normed linear space of pairs (A', a') of $n \times n$ matrices and points of $\mathbf{R}^n$, with the distance from (A', a') to (A'', a'') given by $\max\{\|A' - A''\|, \|a' - a''\|\}$; we take $p_0 := (A, a)$, $T := \partial\psi_C$, and $f[(A', a'), x] := A'x + a'$. The set X_0 is then $S(A, a)$; we let Ω be any open bounded set containing X_0, and since $Lf_{x_0}(x) = Ax + a$ for any x_0, it is clear that the hypotheses are satisfied (note that Proposition 1 implies that (iii) holds). We then find that for some $\delta > 0$, $\epsilon_0 > 0$ and all (A', a') with $\epsilon' < \epsilon_0$, we have $S(A', a') \cap [S(A, a) + \delta B]$ nonempty, which proves (b).

Now choose Ψ; without loss of generality we can suppose that Ω was taken to be this Ψ. As Ψ is bounded, we can find $\epsilon_1 \in (0, \epsilon_0]$ with $\lambda\epsilon_1 < 1$ and such that for each $x \in \Psi$, $\epsilon_1(1 + \|x\|) \leq \eta$, where η is the parameter appearing in Theorem 1. Now pick any (A', a') with $\epsilon' < \epsilon_1$; by the above discussion $S(A', a') \cap \Psi$ is nonempty, and we take x' to be any point of that intersection. We know that

$$0 \in A'x' + a' + \partial\psi_C(x'),$$

which is equivalent to

$$x' \in [A(\cdot) + a + \partial\psi_C(\cdot)]^{-1}[(A - A')x' + (a - a')].$$

But since $x' \in \Psi$,

$$\begin{aligned}\|(A - A')x' + (a - a')\| &\leq \max\{\|A - A'\|, \|a - a'\|\}(1 + \|x'\|)\\ &\leq \epsilon_1(1 + \|x'\|) \leq \eta,\end{aligned}$$

and so by upper Lipschitz continuity,

$$d[x', S(A, a)] \leq \lambda\|(A - A')x' + (a - a')\|.$$

Now let x_0 be the closest point to x' in $S(A; a)$; then

$$\begin{aligned}\|(A - A')x' + (a - a')\| &\leq \|(A - A')x_0 + (a - a')\| + \|(A - A')(x' - x_0)\|\\ &\leq \epsilon'(1 + \mu) + \epsilon'\|x' - x_0\|.\end{aligned}$$

Accordingly, as $\|x' - x_0\| = d[x', S(A, a)]$ we have

$$d[x', S(A, a)] \leq \lambda\epsilon'(1 + \mu) + \lambda\epsilon' d[x', S(A, a)],$$

yielding

$$d[x', S(A, a)] \leq \lambda\epsilon'(1 - \lambda\epsilon')^{-1}(1 + \mu).$$

Since x' was arbitrary in $S(A', a') \cap \Psi$, we have (4.3).

Finally, we observe that for all small ϵ', $S(A', a') \cap \Psi$ is contained in $S(A, a) + \delta B$ which is contained in Ψ. If $S(A', a')$ also met the complement of Ψ then it would be disconnected; thus if $S(A', a')$ is connected it must lie entirely in Ψ, so that we may replace Ψ by $\mathbf{R}^n$ in (4.3). In particular, if A' is positive semidefinite then $A'(\cdot) + a' + \partial\psi_C(\cdot)$ is maximal monotone, so that $S(A', a')$ is convex as the inverse image of 0 under this operator. This completes the proof.

One might wonder, since the boundedness of Ψ is used at only one place in the proof, whether a refinement of the technique would permit replacement of Ψ by $\mathbf{R}^n$ in all cases. The following example shows that this cannot be done even for $n = 1$: take $C = \mathbf{R}_+$, $A = [0]$ and $a = [1]$, so that the problem is

$$0 \in [0]x + [1] + \partial\psi_{\mathbf{R}_+}(x),$$

whose solution set is $S([0],[1])=\{0\}$. However, it is readily checked that for any $\epsilon>0$, $S([-\epsilon],[1])=\{0,\epsilon^{-1}\}$; thus we cannot take $\Psi=\mathbf{R}$ in this case.

Theorem 2 provides, in particular, a complete stability theory for convex quadratic programming (including linear programming) and for linear complementarity problems with positive semidefinite matrices; this extends earlier work of Daniel [3] on strictly convex quadratic programming, and of the author [11] on linear programming. Stability results for more general nonlinear programming problems are developed in Part II of this paper.

It might be worth pointing out that the strong form of Theorem 2 (i.e., with A' restricted to be positive semidefinite) can sometimes be shown to hold because of the form of the problem. For example, consider the quadratic programming problem

$$\begin{aligned} &\text{minimize} \quad \tfrac{1}{2}\langle x, Qx\rangle + \langle q, x\rangle + \langle p, y\rangle, \\ &\text{subject to} \quad Bx + Dy \le d \end{aligned} \tag{4.4}$$

(we could also have added equality constraints, constrained variables, etc. but have omitted these for simplicity). Here Q is $m\times m$, B is $r\times m$ and D is $r\times s$. The formulation of (4.4) as a generalized equation is (taking Q to be symmetric)

$$0\in\begin{bmatrix} Q & 0 & B^{\mathrm{T}} \\ 0 & 0 & D^{\mathrm{T}} \\ -B & -D & 0\end{bmatrix}\begin{bmatrix} x\\ y\\ u\end{bmatrix}+\begin{bmatrix} q\\ p\\ d\end{bmatrix}+\partial\psi_C(x,y,u), \tag{4.5}$$

where $C=\mathbf{R}^m\times\mathbf{R}^s\times\mathbf{R}^r_+$. The matrix shown in (4.5) is then the matrix A of Theorem 2; it is positive semidefinite if and only if Q is positive semidefinite (i.e., if and only if the problem (4.4) is convex). Now, if Q is actually positive definite, then for all small perturbations of the data of (4.4) (i.e., of Q, q, p, B, D, and d) the matrix in (4.5) will remain positive semidefinite and the strong form of Theorem 2 will hold. The point here is that the structure of the problem prevents the type of perturbation which could destroy the positive semidefiniteness of A. This comment, of course, applies in particular to all linear programming problems [11].

References

[1] H. Brézis, *Opérateurs maximaux monotones* (North-Holland, Amsterdam, 1973).

[2] R.W. Cottle, "Nonlinear programs with positively bounded Jacobians", *SIAM Journal on Applied Mathematics* 14 (1966) 147–158.

[3] J.W. Daniel, "Stability of the solution of definite quadratic programs", *Mathematical Programming* 5 (1973) 41–53.

[4] G.B. Dantzig and R.W. Cottle, "Positive (semi-) definite programming", in: J. Abadie, ed., *Nonlinear programming* (North-Holland, Amsterdam, 1968) 55–73.

[5] S. Kakutani, "A generalization of Brouwer's fixed point theorem", *Duke Mathematical Journal* 8 (1941) 457–459.

[6] O.L. Mangasarian, *Nonlinear programming* (McGraw-Hill, New York, 1969).
[7] J.J. Moré, "Coercivity conditions in nonlinear complementarity problems", *SIAM Review* 16 (1974) 1–16.
[8] J.J. Moré, "Classes of functions and feasibility conditions in nonlinear complementarity problems", *Mathematical Programming* 6 (1974) 327–338.
[9] H. Nikaido, *Convex structures and economic theory* (Academic Press, New York and London, 1968).
[10] S.M. Robinson, "An implicit-function theorem for generalized variational inequalities", Technical Summary Report No. 1672, Mathematics Research Center, University of Wisconsin-Madison, 1976; available from National Technical Information Service under Accession No. AD A031952.
[11] S.M. Robinson, "A characterization of stability in linear programming", *Operations Research* 25 (1977) 435–447.
[12] R.T. Rockafellar, "Local boundedness of nonlinear, monotone operators", *Michigan Mathematical Journal* 16 (1969) 397–407.
[13] R.T. Rockafellar, *Convex analysis* (Princeton University Press, Princeton, NJ, 1970).
[14] R.T. Rockafellar, "Monotone operators and the proximal point algorithm", *SIAM Journal on Control and Optimization* 14 (1976) 877–898.
[15] H. Scarf, *The computation of economic equilibria* (Yale University Press, New Haven and London, 1973).