MATHEMATICS OF OPERATIONS RESEARCH
Vol. 5, No. 1, February 1980
Printed in U.S.A.

STRONGLY REGULAR GENERALIZED EQUATIONS*†

STEPHEN M. ROBINSON

University of Wisconsin-Madison

This paper considers generalized equations, which are convenient tools for formulating problems in complementarity and in mathematical programming, as well as variational inequalities. We introduce a regularity condition for such problems and, with its help, prove existence, uniqueness and Lipschitz continuity of solutions to generalized equations with parametric data. Applications to nonlinear programming and to other areas are discussed, and for important classes of such applications the regularity condition given here is shown to be in a certain sense the weakest possible condition under which the stated properties will hold.

1. Introduction. This paper deals with *generalized equations*; that is, inclusions of the form

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

where f is a function from a subset Ω of a normal linear space X to its (topological) dual X', C is a nonempty closed convex set in X, and the notation $\partial\psi_C$ is the normal cone operator:

$$\partial\psi_C(x) := \begin{cases} \{y \in X' \mid y(c - x) \leq 0 \quad \text{for all } c \in C\} & \text{if } x \in C, \\ \emptyset, & \text{if } x \notin C. \end{cases}$$

Such generalized equations conveniently formulate many problems of interest in applications, both in the infinite-dimensional and in the finite-dimensional cases. For example, when $X = \mathbf{R}^n$ and C is of the form $\mathbf{R}^k \times \mathbf{R}^l_+$, with $k + l = n$, then (1.1) can be used to formulate problems in complementarity and in nonlinear programming. Examples are given in [12] as well as in §4 of this paper.

In [12], we studied the behavior of solutions of (1.1) when the function f was subjected to small perturbations. The conditions imposed upon the generalized equation in [12] were such as to permit sets of solutions, rather than isolated or unique solutions. Of course, for purposes of analysis it is sometimes convenient to deal with solutions which are unique, or at least unique in some neighborhood. Therefore we investigate here the problem of finding appropriate conditions for solutions of (1.1) and problems "close" to it to have locally unique solutions, and for those solutions to have good continuity properties when regarded as functions of the perturbations introduced into (1.1). Assuming that f is Fréchet differentiable in some neighborhood of a solution x_0 of (1.1), we obtain a condition on the linearization of (1.1) about x_0, i.e., on the linear generalized equation

$$0 \in f(x_0) + f'(x_0)(x - x_0) + \partial\psi_C(x), \tag{1.2}$$

which ensures good behavior of the solutions of the nonlinear problem (1.1). This condition, which we call *strong regularity*, is analogous to the nonsingularity condition

* Received September 1, 1978.
AMS 1970 subject classification. Primary 49A50. Secondary 90C30.
IAOR 1973 subject classification. Main: Nonlinear programming.
Key words. Nonlinear programming, complementarity, variational inequalities.
† Sponsored by the U. S. Army under Contract No. DAAG29-75-C-0024 and by the National Science Foundation under Grant No. MCS74-20584 A02.

0364-765X/80/0501/0043$01.25

imposed in the usual implicit-function theorem for nonlinear equations, and indeed it reduces to that condition if C is the whole space (so that (1.1) reduces to $f(x) = 0$). We show also that, in a certain sense, these conditions are the best possible if only the information contained in (1.2) is used.

The reader will see in what follows that, throughout the paper, the linearization (1.2) plays a fundamental role, not only in the conditions just mentioned, but also in the sensitivity analysis of (1.1) with respect to some family of perturbations. For example, we show in Theorem 2.3 below that (1.2) contains the appropriate local information to carry out a sensitivity analysis of (1.1) for small perturbations, without solving the nonlinear problem (1.1) over again for each perturbation considered. Thus, (1.2) summarizes important first-order information about the behavior of (1.1) near x_0, in a form which is fairly convenient for analysis and computation.

The arrangement of topics in the rest of the paper is as follows: in §2 we define strong regularity and prove the main results about the nonlinear problem (1.1); in §3 we study the linearization (1.2) and exploit its structure in specific cases to develop criteria for strong regularity to hold. We also discuss briefly a reduction procedure which is very useful when C is polyhedral (as is the case in important areas of applications); the details of the reduction for general polyhedral C are deferred to the appendix. Finally, in §4 we apply the results previously developed to the important special cases of complementarity (briefly) and of nonlinear programming (in more detail). In particular, we show that satisfactory analysis of a nonlinear programming problem can be carried out without the usual assumption of strict complementary slackness.

We close this introduction by indicating briefly some similarities and differences between this paper and others which have been written previously. The standard implicit-function theorem was used in [3], [4], [11] to analyze perturbed nonlinear programming problems under stronger hypotheses than those we use here. Extensions of that approach to generalized equations will be reported in the forthcoming dissertation of A. Reinoza in the Computer Sciences Department of the University of Wisconsin-Madison. Levitin [7]–[9] has investigated stability aspects of infinite-dimensional optimization problems, and Mangasarian [10] has obtained uniqueness results in linear programming using hypotheses which seem to be closely related to those used here.

Finally, we mention two papers treating the stability of convex quadratic programming problems: Daniel [2] used assumptions similar to those made in [12], while Hager [5] imposed conditions similar to those used in this paper. However, Hager's method of analysis is quite different from ours: for example, he assumes that unique solutions will exist for all of the perturbed problems that he considers, and he then proceeds to investigate the continuity properties of these solutions. We prefer to identify conditions on the unperturbed problem which will permit us to prove that solutions of the perturbed problems will in fact exist, and which in addition enable us to analyze their continuity properties.

Many other papers have, of course, been written on stability questions related to those we investigate. The above papers represent only a sample of the literature; many additional references can be found in them.

2. Strong regularity and local solvability. In this section we define a condition, called *strong regularity*, which can be satisfied by a generalized equation at a solution point. We prove a basic solvability theorem which says, roughly speaking, that if a generalized equation is strongly regular at a solution point then it is invertible near that point and the inverse function is Lipschitzian; further, any generalized equation which is close, in a suitable sense, to the one with which we are working will share these desirable properties. We then show how this basic solvability theorem may be applied, first to the problem of parametric sensitivity and then to the extension of Banach's inversion lemma to linear generalized equations. Both of these applications depend upon strong regularity, which we now proceed to define; conditions to ensure that strong regularity holds at a point will be developed in §3. The definition and results given here are stated for spaces more general than $\mathbf{R}^n$, but the material in the remaining sections is developed for $\mathbf{R}^n$.

DEFINITION. Let X be a normed linear space, and let Ω be an open subset of X containing a point x_0. Let C be a closed convex set in X, and let $f : \Omega \to X'$ (the topological dual of X) be Fréchet differentiable at x_0. Suppose that the generalized equation

$$0 \in f(x) + \partial\psi_C(x) \tag{2.1}$$

has x_0 as a solution, and define, for $x \in X$,

$$Tx := f(x_0) + f'(x_0)(x - x_0) + \partial\psi_C(x).$$

We say that (2.1) is *strongly regular at x_0 with associated Lipschitz constant* λ if there exist neighborhoods U of the origin in X' and V of x_0 such that the restriction to U of $T^{-1} \cap V$ is a single-valued function from U to V which is Lipschitzian on U with modulus λ.

Note that if C is the whole space X (so that $\partial\psi_C(x) \equiv 0$) then (2.1) becomes $f(x) = 0$, and strong regularity then amounts to the assumption that $f'(x_0)^{-1}$ is a continuous linear operator, which is the regularity condition that one would normally impose in such a case.

We can now prove the main result, a type of implicit-function theorem for generalized equations satisfying the strong regularity condition.

THEOREM 2.1. *Let X, X', C, Ω and x_0 be as in the definition. Let P be a topological space, $p_0 \in P$, and let $f : P \times \Omega \to X'$. Suppose that the partial Fréchet derivative of f with respect to the second variable, written $f'(\cdot, \cdot)$, exists on $P \times \Omega$, that both $f(\cdot, \cdot)$ and $f'(\cdot, \cdot)$ are continuous at (p_0, x_0), and that x_0 solves*

$$0 \in f(p_0, x) + \partial\psi_C(x). \tag{2.2}$$

If (2.2) *is strongly regular at x_0, with associated Lipschitz constant λ, then for any $\epsilon > 0$ there exist neighborhoods N_ϵ of p_0 and W_ϵ of x_0, and a single-valued function $x : N_\epsilon \to W_\epsilon$, such that for any $p \in N_\epsilon$, $x(p)$ is the unique solution in W_ϵ of the inclusion*

$$0 \in f(p, x) + \partial\psi_C(x). \tag{2.3}$$

Further, for each p and q in N_ϵ one has

$$\|x(p) - x(q)\| \leqslant (\lambda + \epsilon)\|f(p, x(q)) - f(q, x(q))\|. \tag{2.4}$$

PROOF. Suppose that a positive ϵ has been prescribed; choose a positive number δ so small that $\lambda\delta < \epsilon/(\lambda + \epsilon)$. By strong regularity, there exist neighborhoods U of the origin and V of x_0 such that if we define, for $x \in X$,

$$L(x) := f(p_0, x_0) + f'(p_0, x_0)(x - x_0) + \partial\psi_C(x),$$

then the restriction to U of $L^{-1} \cap$ V is single-valued and Lipschitzian with modulus λ. Let

$$r(p, x) := f(p_0, x_0) + f'(p_0, x_0)(x - x_0) - f(p, x)$$

for $(p, x) \in P \times \Omega$, and choose a neighborhood W of p_0 and a closed ball V_ϵ of radius ρ about x_0 so that $V_\epsilon \subset V$ and for each $p \in W$ and $x \in V_\epsilon$ one has $r(p, x) \in U$ and

$\|f'(p,x)-f'(p_0,x_0)\| \leqslant \delta$. Now shrink W, if necessary, to obtain a neighborhood N_ϵ of p_0 with $N_\epsilon \subset W$ and having the property that for each $p \in N_\epsilon$,

$$\lambda\|f(p_0,x_0)-f(p,x_0)\| \leqslant (1-\lambda\delta)\rho.$$

Choose any $p \in N_\epsilon$, and define an operator Φ_p from V_ϵ to V by

$$\Phi_p(x) := V \cap L^{-1}[r(p,x)].$$

Note that $x \in V_\epsilon \cap \Phi_p(x)$ if and only if $x \in V_\epsilon$ and $0 \in f(p,x) + \partial\psi_C(x)$.

Now let x_1, x_2 be any two points of V_ϵ. We have, using the assumption of strong regularity,

$$\begin{aligned}\|\Phi_p(x_1)-\Phi_p(x_2)\| &\leqslant \lambda\|r(p,x_1)-r(p,x_2)\|\\ &\leqslant \lambda\|x_1-x_2\|\sup\{\|r'(p,(1-\mu)x_1+\mu x_2)\| \mid 0<\mu<1\}\\ &\leqslant \lambda\delta\|x_1-x_2\|,\end{aligned}$$

since $r'(p,x)=f'(p_0,x_0)-f'(p,x\ )$. Since $\lambda\delta<1$, Φ_p is strongly contractive on V_ϵ. Further, since $x_0 = V \cap L^{-1}(0)$ and $\Phi_p(x_0) = V \cap L^{-1}[r(p,x_0)]$, we have

$$\begin{aligned}\|\Phi_p(x_0)-x_0\| &\leqslant \lambda\|r(p,x_0)\|\\ &= \lambda\|f(p_0,x_0)-f(p,x_0)\| \leqslant (1-\lambda\delta)\rho,\end{aligned}$$

and therefore for any $x \in V_\epsilon$,

$$\begin{aligned}\|\Phi_p(x)-x_0\| &\leqslant \|\Phi_p(x)-\Phi_p(x_0)\| + \|\Phi_p(x_0)-x_0\|\\ &\leqslant \lambda\delta\|x-x_0\| + (1-\lambda\delta)\rho \leqslant \rho,\end{aligned}$$

so that Φ_p is a self-map of V_ϵ. By the contraction principle, Φ_p then has a unique fixed point $x(p) \in V_\epsilon$, and for each $x \in V_\epsilon$ one has the bound

$$\|x(p)-x\| \leqslant (1-\lambda\delta)^{-1}\|\Phi_p(x)-x\|. \tag{2.5}$$

It follows from our earlier observation that $x(p)$ is the unique solution of (2.3) in V_ϵ. Thus we have established the existence of the function $x: N_\epsilon \to V_\epsilon$; to obtain the bound (2.4) we take any p and q in N_ϵ and apply (2.5) with $x = x(q)$ to get

$$\|x(p)-x(q)\| \leqslant (1-\lambda\delta)^{-1}\|\Phi_p(x(q))-x(q)\|.$$

If we now recall that $x(q) = \Phi_q(x(q))$ and employ the bound

$$\begin{aligned}\|\Phi_p(x(q))-\Phi_q(x(q))\| &\leqslant \lambda\|r(p,x(q))-r(q,x(q))\|\\ &= \lambda\|f(p,x(q))-f(q,x(q))\|,\end{aligned}$$

we have

$$\|x(p)-x(q)\| \leqslant \lambda(1-\lambda\delta)^{-1}\|f(p,x(q))-f(q,x(q))\|.$$

Observing that $\lambda(1-\lambda\delta)^{-1} \leqslant \lambda+\epsilon$, we find that (2.4) holds, so that the proof is complete.

By imposing a more restrictive continuity condition we can show that $x(\cdot)$ is locally Lipschitzian at p_0.

COROLLARY 2.2. *Assume the notation and hypotheses of Theorem* 2.1. *Suppose further that P is a subset of a normed linear space and that for some constant ν and for* *each $p, q \in N_\epsilon$ and each $x \in V_\epsilon$, one has*

$$\|f(p, x) - f(q, x)\| \leqslant \nu\|p - q\|.$$

Then $x(\cdot)$ is Lipschitzian on N_ϵ with modulus $\nu(\lambda + \epsilon)$.

PROOF. Immediate from (2.4).

It may be worth pointing out here that the condition of strong regularity is the weakest possible condition which can be imposed on the value of a function f and its derivative at a point x_0, so that for each perturbation structure satisfying the hypotheses of Therorem 2.1 and Corollary 2.2, a function $x(\cdot)$ will exist having the properties stated in those results. To see this, one has only to consider a function $f : \Omega \to X'$ which is Fréchet differentiable at x_0 and which satisfies $0 \in f(x_0) + \partial\psi_C(x_0)$. Let P be a neighborhood of the origin in X', and let

$$f(p, x) := f(x_0) + f'(x_0)(x - x_0) - p,$$

with $p_0 = 0$. Fix some positive ϵ. If neighborhoods N_ϵ and V_ϵ, and a function $x(\cdot)$, exist having the properties asserted in Theorem 2.1 and Corollary 2.2, then with

$$Tx := f(x_0) + f'(x_0)(x - x_0) + \partial\psi_C(x),$$

we see that the restriction to N_ϵ of $T^{-1} \cap V_\epsilon$ is a single-valued, Lipschitzian function: that is, the generalized equation $0 \in f(x) + \psi_C(x)$ is strongly regular at x_0.

One of the by-products of Theorem 2.1 is a useful formula for parametric sensitivity analysis, which we give in the next theorem.

THEOREM 2.3. *Assume the notation and hypotheses of Theorem* 2.1 *and Corollary* 2.2. *Then for each $\epsilon > 0$ there exists a function $\alpha_\epsilon : N_\epsilon \to \mathbf{R}$, with $\lim_{p \to p_0} \alpha_\epsilon(p) = 0$, such that for any $p \in N_\epsilon$ one has*

$$\|x(p) - \Phi_p(x_0)\| \leqslant \alpha_\epsilon(p)\|p - p_0\|.$$

Before proving this result it may be helpful to interpret it. To compute $\Phi_p(x_0)$ is to find the unique solution in V_ϵ of the *linear* generalized equation

$$0 \in f(p, x_0) + f'(p_0, x_0)(x - x_0) + \partial\psi_C(x), \tag{2.6}$$

and the theorem says that this solution, for values of p near p_0, will be very close to the solution $x(p)$ of the *nonlinear* generalized equation

$$0 \in f(p, x) + \partial\psi_C(x). \tag{2.7}$$

It is easy to verify that in the case $C = X$ (i.e., when one is solving $f(p, x) = 0$) this corresponds to the result that if $f(\cdot, x_0)$ is Fréchet differentiable at p_0, then so is $x(\cdot)$, with

$$x'(p_0) = -f'(p_0, x_0)^{-1}\partial f(p_0, x_0)/\partial p.$$

In many applications, one might find (2.6) significantly easier to solve than (2.7). For example, in finite-dimensional applications involving nonlinear problems in complementarity or in mathematical programming, (2.6) is, respectively, a linear complementarity problem or a quadratic programming problem. Particularly if one is interested in sensitivity analysis of a computed solution to a nonlinear problem, one may already have at hand much of the information needed to solve (2.6) quickly for different values of p (e.g., using parametric linear complementarity or parametric quadratic programming techniques). Thus, in such cases Theorem 2.3 could provide a relatively cheap way to find a good approximation to $x(p)$ for p near p_0.

PROOF OF THEOREM 2.3. We know that $x(p) = \Phi_p(x(p))$; thus by strong regularity we have

$$\begin{aligned}\|x(p) - \Phi_p(x_0)\| &\leqslant \lambda\|r(p, x(p)) - r(p, x_0)\| \\ &\leqslant \lambda\|x(p) - x_0\|\sup\{\|r'(p, (1-\mu)x(p) + \mu x_0)\| \mid 0 < \mu < 1\} \\ &\leqslant \lambda\nu(\lambda + \epsilon)\|p - p_0\|\sup\{\|f'(p_0, x_0) - f'(p, (1-\mu)x(p) \\ &\qquad + \mu x_0)\| \mid 0 < \mu < 1\}. \qquad (2.8)\end{aligned}$$

The quantity in brackets approaches zero as p approaches p_0, by the continuity of $x(\cdot)$ and of $f'(\cdot, \cdot)$, and this completes the proof.

Note that if we knew that for some constant β and all $p \in P$, $x \in \Omega$, we had

$$\|f'(p, x) - f'(p_0, x_0)\| \leqslant \beta(\|p - p_0\| + \|x - x_0\|),$$

then we could obtain from (2.8) a bound of the form

$$\|x(p) - \Phi_p(x_0)\| \leqslant \gamma\|p - p_0\|^2,$$

for some constant γ.

We illustrate next another application of Theorem 2.1, this time to the establishment of an analogue of the Banach perturbation lemma for linear operators.

THEOREM 2.4. *Let X be a Banach space, let a_0 be a point of the dual space X', let C be a closed convex set in X and let A_0 belong to $L(X, X')$ (the space of bounded linear operators from X to X'). Suppose that x_0 is a point of X which satisfies the generalized equation*

$$0 \in A_0x + a_0 + \partial\psi_C(x). \qquad (2.9)$$

If (2.9) is strongly regular at x_0 with associated Lipschitz modulus λ, then there exist neighborhoods M of A_0 in $L(X, X')$, N of a_0 and W of the origin in X', and V of x_0, such that if for $A \in M$, $a \in N$ and $x \in V$ one defines

$$T(A, a, x) := Ax + a + \partial\psi_C(x),$$

then $T(A, a, \cdot)^{-1} \cap V$ is a single-valued function on W and is Lipschitzian there with modulus $\lambda(1 - \lambda\|A - A_0\|)^{-1}$.

PROOF. Apply Theorem 2.1 with $P := L(X, X') \times X'$, $p_0 := (A_0, a_0)$, and $f(p, x) := Ax + a$ (with any positive ϵ) to produce neighborhoods N_1 of (A_0, a_0) and V of x_0, with a single-valued function $x : N_1 \to V$ having the property that for each $(A, a) \in N_1$, $x(A, a)$ is the unique solution in V of $0 \in Ax + a + \partial\psi_C(x)$. Note that (2.4) implies $\|x(A, a) - x_0\| \leqslant (\lambda + \epsilon)\|(A - A_0)x_0 + (a - a_0)\|$ for such (A, a). Choose neighborhoods M of A_0, N of a_0, and W of the origin in X', such that

(1) for each $A \in M$, $\lambda\|A - A_0\| < 1$,

(2) $M \times (N - W) \subset N_1$,

and

(3) for each $A \in M$, $a \in N$ and $y \in W$, the point $y + (A_0 - A)x(A, a) + (a_0 - a)$ lies in the neighborhood U of the origin given in the definition of strong regularity for (2.9). Evidently for any $A \in M$, $a \in N$ and $y \in W$, the generalized equation $y \in Ax + a + \partial\psi_C(x)$ is uniquely solvable in V (by $x(A, a - y)$). Let y_1 and y_2 belong to W, and let x_1 and x_2 be the solutions associated with y_1 and y_2.

Then for $i = 1, 2$ we have

$$y_i + (A_0 - A)x_i + (a_0 - a) \in A_0x_i + a_0 + \partial\psi_C(x_i).$$

Thus, by strong regularity,

$$\|x_1 - x_2\| \leqslant \lambda\|[y_1 + (A_0 - A)x_1 + (a_0 - a)] - [y_2 + (A_0 - A)x_2 + (a_0 - a)]\|$$
$$\leqslant \lambda\|y_1 - y_2\| + \lambda\|A_0 - A\|\,\|x_1 - x_2\|.$$

But as $\lambda\|A_0 - A\| < 1$, we finally obtain

$$\|x_1 - x_2\| \leqslant \lambda(1 - \lambda\|A_0 - A\|)^{-1}\|y_1 - y_2\|,$$

which completes the proof.

It should be noted that this result says, among other things, that if a generalized equation is strongly regular at a solution x_0, then any sufficiently "close" generalized equation will be strongly regular at its solution near x_0 (which must exist and be unique by Theorem 2.1); further, the neighborhoods involved in the definition of strong regularity can be taken to be the same for all nearby generalized equations. This is a property which is not available under the weaker hypotheses used in [12].

3. Conditions for strong regularity. We have seen that nonlinear generalized equations can be expected to behave in desirable ways if their linearizations are strongly regular at the points in question. In this section we develop a general condition which is sufficient for strong regularity, as well as a sharper condition, designed for the case most frequently seen in applications, which we show to be both necessary and sufficient.

To begin with, we suppose that we are considering a generalized equation of the form $0 \in Ax + a + \partial\psi_C(x)$, where A is $n \times n$, $a \in \mathbf{R}^n$, and C is a nonempty polyhedral convex set in $\mathbf{R}^n$. Let x_0 be a solution of this equation, and consider the inclusion

$$y \in Ax + a + \partial\psi_C(x), \tag{3.1}$$

for y near 0 and x near x_0. We claim that this can actually be reduced to the consideration of

$$z \in Bw + \partial\psi_{\mathbf{R}^r \times K}(w), \tag{3.2}$$

for some nonnegative integers r, s with $r + s \leqslant n$, a square matrix B of dimension $(r + s)$, a pointed polyhedral convex cone K in $\mathbf{R}^s$ of dimension s, and points z, w near the origin in $\mathbf{R}^{r+s}$. This reduction is carried out in detail in the appendix to this paper, but here we indicate how it may be done in the case found most frequently in applications: namely, that in which $C = \mathbf{R}^l \times \mathbf{R}^m_+$. This case includes the standard linear complementarity problem ($l = 0$, $m = n$) and the problem of quadratic programming without implicit constraints, such as nonnegativity (with l the number of variables and equality constraints, and m the number of inequality constraints). If $C = \mathbf{R}^l \times \mathbf{R}^m_+$, we know that for $i = l + 1, \ldots, l + m = n$ we must have $(Ax_0 + a)_i \geqslant 0$, with equality if $(x_0)_i > 0$. Assume that the variables and the elements of $Ax_0 + a$ have been reordered, if necessary, so that

for $i = l + 1, \ldots, l + j$, $(Ax_0 + a)_i = 0$ and $(x_0)_i > 0$;
for $i = l + j + 1, \ldots, l + j + s$, $(Ax_0 + a)_i = 0$ and $(x_0)_i = 0$;
for $i = l + j + s + 1, \ldots, n$, $(Ax_0 + a)_i > 0$ and $(x_0)_i = 0$.

It is quite clear that if x and y satisfy (3.1), with x near x_0 and y near zero, then for $i = l + j + s + 1, \ldots, n$ we shall have $(Ax + a)_i > 0$ and thus $x_i = 0$; similarly, for $i = l + 1, \ldots, l + j$, $x_i > 0$ and so $(Ax + a)_i = 0$. Thus if we let $r := l + j$, partition elements $v \in \mathbf{R}^n$ as (v^1, v^2, v^3) with $v^1 \in \mathbf{R}^r$, $v^2 \in \mathbf{R}^s$, $v^3 \in \mathbf{R}^{n-r-s}$, and partition A

conformably as

$$A = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{bmatrix},$$

we see that we really only have to consider solutions of

$$\begin{aligned} &y^1 = A_{11}x^1 + A_{12}x^2 + a^1, \\ &y^2 \leqslant A_{21}x^1 + A_{22}x^2 + a^2, \\ &x^2 \geqslant 0, \qquad \langle x^2, A_{21}x^1 + A_{22}x^2 + a^2 - y^2 \rangle = 0; \end{aligned}$$

that is, of

$$\begin{bmatrix} y^1 \\ y^2 \end{bmatrix} \in \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} x^1 \\ x^2 \end{bmatrix} + \begin{bmatrix} a^1 \\ a^2 \end{bmatrix} + \partial\psi_{\mathbf{R}^r \times \mathbf{R}^t_+} \begin{bmatrix} x^1 \\ x^2 \end{bmatrix}. \tag{3.3}$$

However, if we recall that

$$Ax_0 + a = \begin{bmatrix} A_{11} \\ A_{21} \\ A_{31} \end{bmatrix} x_0^1 + \begin{bmatrix} a^1 \\ a^2 \\ a^3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ A_{31}x_0^1 + a^3 \end{bmatrix},$$

and that $x_0^2 = 0$, then by writing

$$w := \begin{bmatrix} x^1 - x_0^1 \\ x^2 - x_0^2 \end{bmatrix}$$

we can write (3.3) as

$$\begin{bmatrix} y^1 \\ y^2 \end{bmatrix} \in \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} w + \partial\psi_{\mathbf{R}^r \times \mathbf{R}^t_+}(w), \tag{3.4}$$

and with

$$z := \begin{bmatrix} y^1 \\ y^2 \end{bmatrix}$$

this is in the form (3.2). This reduction, though quite simple, is very useful in identifying that portion of the problem (*viz*., the matrix $[\begin{smallmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{smallmatrix}]$) to which we have to attach conditions in order for the original generalized equation to be strongly regular at x_0. In quadratic programming problems with linear constraints but without implicit constraints on the variables, this reduction procedure amounts to (1) eliminating the constraints which are inactive at the point in question, and (2) regarding as equations those (active) inequality constraints whose associated multipliers are positive.

Before proceeding to establish conditions ensuring that (3.4) will behave well, we comment on the question of continuity for such problems. If K is any polyhedral convex set in $\mathbf{R}^s$, the operator

$$Tw := \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} w + \partial\psi_{\mathbf{R}^r \times K}(w)$$

is an example of a polyhedral multifunction [12], and T^{-1} is also polyhedral. We have from [12, Proposition 2] that for some fixed λ, T^{-1} is then locally upper Lipschitzian with modulus λ at each point $y_0 \in \mathbf{R}^{r+s}$: that is, for each such y_0 there exists a neighborhood V of y_0 such that for each $y \in V$,

$$T^{-1}(y) \subset T^{-1}(y_0) + \lambda\|y - y_0\| B,$$

where B is the unit ball. It is not difficult to show that if T^{-1} is also single-valued on a convex set $D \cap \mathbf{R}^{r+s}$, then T^{-1} will necessarily be Lipschitzian there with modulus λ. Thus, if we can show that such an operator is single valued on some convex set, we can conclude immediately that it is also Lipschitzian there.

In the following theorem, we develop conditions for single-valuedness of such an operator. We use the idea of a Schur complement: if a matrix A is partitioned as

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

with A_{11} and A_{22} square and A_{11} nonsingular, then the *Schur complement of* A_{11} *in* A, written (A/A_{11}), is defined to be $A_{22} - A_{21}A_{11}^{-1}A_{12}$. An interesting treatment of this idea may be found in [1].

THEOREM 3.1. *Let r and s be positive integers, and let $K \subset \mathbf{R}^s$ be a nonempty closed convex set. Let A be an $(r+s) \times (r+s)$ matrix*:

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix},$$

where A_{11} is $r \times r$. For $w \in \mathbf{R}^{r+s}$ define

$$Tw := Aw + \partial\psi_{\mathbf{R}^r \times K}(w).$$

For T^{-1} to be a Lipschitzian function defined on all of $\mathbf{R}^{r+s}$, it suffices that:

(1) *A_{11} be nonsingular*

and

(2) *(A/A_{11}) be positive definite.*

In the special case in which $K = \mathbf{R}^S_+$, (2) may be replaced by

(2′) *(A/A_{11}) have positive principal minors,*

and conditions (1) *and* (2′) *are then both necessary and sufficient.*

PROOF. For the first assertion, suppose that (1) and (2) hold. Let

$$y^i := \begin{pmatrix} y_1^i \\ y_2^i \end{pmatrix}, \qquad y_1^i \in \mathbf{R}^r, \quad y_2^i \in \mathbf{R}^s, \quad i = 1, 2,$$

and consider the system

$$y^i \in Aw^i + \partial\psi_{\mathbf{R}^r \times K}(w^i). \tag{3.5}$$

This is equivalent to

$$y_1^i = A_{11}w_1^i + A_{12}w_2^i,$$
$$y_2^i \in A_{21}w_1^i + A_{22}w_2^i + \partial\psi_K(w_2^i).$$

As A_{11} is nonsingular, we can see that

$$w_1^i = A_{11}^{-1}y_1^i - A_{11}^{-1}A_{12}w_2^i, \tag{3.6}$$

so that (3.5) holds if and only if (3.6) holds and

$$z^i := y_2^i - A_{21}A_{11}^{-1}y_1^i \in (A/A_{11})w_2^i + \partial\psi_K(w_2^i). \tag{3.7}$$

The operator defined on $\mathbf{R}^s$ by

$$S(w) := (A/A_{11})w + \partial\psi_K(w)$$

is maximal monotone; as (A/A_{11}) is positive definite, $S(w)$ is also strongly monotone, and thus its inverse is Lipschitzian on all of $\mathbf{R}^s$. Thus, for some fixed L,

$$\|w_2^1 - w_2^2\| \leqslant L\|z^1 - z^2\| \leqslant L\|[-A_{21}A_{11}^{-1}\ I]\|\ \|y^1 - y^2\|.$$

This inequality, together with (3.6), implies that for a constant M independent of y^1 and y^2,

$$\|w^1 - w^2\| \leqslant M\|y^1 - y^2\|,$$

which proves the sufficiency of (1) and (2).

Now assume that $K = \mathbf{R}_+^s$ and replace (2) by (2′). For sufficiency, note that (3.7) is the linear complementarity problem

$$\begin{aligned}(A/A_{11})w_2^i - z^i &\geqslant 0,\\ w_2^i &\geqslant 0,\\ \langle w_2^i, (A/A_{11})w_2^i - z^i\rangle &= 0,\end{aligned} \tag{3.8}$$

where $\langle\cdot,\cdot\rangle$ denotes the standard inner product. It is known [6] that (3.8) has a unique solution for each $z^i \in \mathbf{R}^s$ if and only if (A/A_{11}) has positive principal minors. In view of (3.6) and our earlier comments about polyhedral multifunctions, this is enough to prove sufficiency. For necessity, suppose that T^{-1} is Lipschitzian on all of $\mathbf{R}^{r+s}$. If A_{11} is singular, let u be a point of $\mathbf{R}^r$ not in the range of A_{11}. Let v be a point of $\mathbf{R}^s$ with all of its components strictly negative, and for any nonnegative λ define $y_\lambda := \binom{\lambda u}{v}$. We know that $y_0 \in T(0)$, and by the assumed properties of T^{-1} we know that for all small positive λ the point

$$T^{-1}(y_\lambda) =: \begin{pmatrix} w_1^\lambda \\ w_2^\lambda \end{pmatrix}$$

will be near the origin. However, we have

$$v - A_{21}w_1^\lambda - A_{22}w_2^\lambda \in \partial\psi_{\mathbf{R}_+^s}(w_2^\lambda),$$

and for small λ the left-hand side will be strictly negative in all components, implying that $w_2^\lambda = 0$. But then

$$\lambda u = A_{11}w_1^\lambda + A_{12}w_2^\lambda = A_{11}w_1^\lambda,$$

contradicting our choice of u. Thus A_{11} must be nonsingular, so (A/A_{11}) is well defined. Now by our previous analysis $T^{-1}(y)$ will be a singleton for each $y \in \mathbf{R}^s$ if and only if the complementarity problem (3.8) is uniquely solvable for each $z^i \in \mathbf{R}^s$. We have already remarked that this is equivalent to (2′), and this completes the proof.

We observe that if K is a cone in Theorem 3.1, then the operator T is positively homogeneous (i.e., for $\lambda > 0$, $T(\lambda x) = \lambda T(x)$). In that case, for T^{-1} to have a property such as unique solvability on all of $\mathbf{R}^n$ it is necessary and sufficient that that property hold on any neighborhood of the origin. Also, the conclusions of Theorem 3.1 are not invalidated if either of the spaces $\mathbf{R}^r$ and $\mathbf{R}^s$ does not appear (i.e., if the set $\mathbf{R}^r \times K$ is $\mathbf{R}^r$ itself, or is K). In fact, the arguments are simplified in that case; of course, the obvious changes in the conditions (1), (2) and (2′) must be made.

The following corollary relates the conditions just developed to the property of strong regularity which we used in §2.

COROLLARY 3.2. *Let C be a polyhedral convex set in $\mathbf{R}^n$, let A be $n \times n$, $a \in \mathbf{R}^n$, and let x_0 solve*

$$0 \in Ax + a + \partial\psi_C(x). \tag{3.9}$$

Let the reduced form of (3.9) *at x_0, if not vacuous, be*

$$0 \in Bw + \partial\psi_{\mathbf{R}^r \times K}(w), \qquad B = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}, \tag{3.10}$$

where K is a polyhedral convex cone in $\mathbf{R}^s$. For (3.9) *to be strongly regular at x_0 it suffices that* (i) *the reduced form be vacuous or that* (ii) *B_{11} be nonsingular and (B/B_{11}) be positive definite. If $K = \mathbf{R}^s_+$ then for* (3.8) *to be strongly regular at x_0 it is necessary and sufficient that* (i) *the reduced form be vacuous, or* (ii) *B_{11} be nonsingular and (B/B_{11}) have positive principal minors.*

PROOF. Immediate from Theorem 3.1 together with Theorem A.4 in the appendix.

4. Applications and examples. Having developed the theoretical aspects of strong regularity in §§2 and 3, we apply these developments here to problems in complementarity and in nonlinear programming. We first give an example from complementarity: consider the linear complementarity problem

$$\begin{aligned} Ax + a &\geqslant 0, \\ x &\geqslant 0, \\ \langle x, Ax + a \rangle &= 0, \end{aligned}$$

in which

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 4 & 5 & 1 \\ 7 & 1 & 1 \end{bmatrix}, \qquad a = \begin{bmatrix} -1 \\ -4 \\ 0 \end{bmatrix}.$$

This is equivalent to the generalized equation $0 \in Ax + a + \partial\psi_{\mathbf{R}^3_+}(x)$. The matrix A is neither positive definite nor a P-matrix (i.e., a matrix with positive principal minors), but the problem has a solution with $x_0^T = (1, 0, 0)$. Noting that the third component of $Ax_0 + a$ is positive, as is the first component of x_0, we can apply the reduction procedure to obtain the problem in reduced form as

$$0 \in \begin{bmatrix} 1 & 1 \\ 4 & 5 \end{bmatrix}\begin{bmatrix} z_1 \\ x_2 \end{bmatrix} + \partial\psi_{\mathbf{R} \times \mathbf{R}_+}\begin{bmatrix} z_1 \\ x_2 \end{bmatrix},$$

where $z_1 := x_1 - 1$. The conditions for strong regularity thus reduce to the requirements that the submatrix [1] be nonsingular and that its Schur complement be a P-matrix. As the Schur complement is $[5] - [4][1]^{-1}[1] = [1]$, we see that this problem is strongly regular at x_0. The reader may wish to check that if the entry of 5 appearing in the matrix is changed to 4, the problem is no longer strongly regular at x_0, although x_0 remains a solution. In fact, with that change one finds that if the second component of a is perturbed to $-4 - \epsilon$ ($\epsilon > 0$), the resulting problem cannot have any solution near x_0, although it has the solution $x_\epsilon^T = (0, 1 + \epsilon/4, 0)$. Even with $\epsilon = 0$ the solution set consists of the line segment $\{(1 - \lambda)(1, 0, 0)^T + \lambda(0, 1, 0)^T \mid 0 \leqslant \lambda \leqslant 1\}$, so that x_0 is not an isolated solution.

We next examine the standard nonlinear programming problem

$$\begin{aligned} &\text{minimize} && \theta(x) \\ &\text{subject to} && g(x) \leqslant 0, \\ & && h(x) = 0, \end{aligned} \tag{4.1}$$

where θ, g and h are Fréchet differentiable functions from some open set $\Omega \subseteq \mathbf{R}^n$ into $\mathbf{R}$, $\mathbf{R}^p$ and $\mathbf{R}^q$, respectively. The optimality conditions for (4.1) are

$$\begin{aligned} \mathcal{L}'(x, u, v) &= 0, \\ g(x) &\leqslant 0, \\ h(x) &= 0, \\ u &\geqslant 0, \\ \langle u, g(x)\rangle &= 0, \end{aligned} \tag{4.2}$$

where u and v are points in $\mathbf{R}^p$ and $\mathbf{R}^q$, respectively, and where $\mathcal{L}(x, u, v) := \theta(x) + \langle u, g(x)\rangle + \langle v, h(x)\rangle$ and the prime indicates differentiation with respect to the first variable (x). The conditions (4.2) can be written more conveniently as the generalized equation

$$0 \in \begin{bmatrix} \mathcal{L}'(x, u, v) \\ -g(x) \\ -h(x) \end{bmatrix} + \partial\psi_{\mathbf{R}^n \times \mathbf{R}^p_+ \times \mathbf{R}^q} \begin{bmatrix} x \\ u \\ v \end{bmatrix}, \tag{4.3}$$

and we shall consider the question of proving strong regularity at a solution of (4.3). If the components of such a solution are denoted by x_0, u_0 and v_0, we can partition the vector $g(x_0)$ into smaller vectors $g^+(x_0)$, $g^0(x_0)$ and $g^-(x_0)$, of dimensions r, s and t, respectively, and partition u_0 conformably into u_0^+, u_0^0, and u_0^- so that

$$\begin{aligned} g^+(x_0) &= 0, & u_0^+ &> 0, \\ g^0(x_0) &= 0, & u_0^0 &= 0, \\ g^-(x_0) &< 0, & u_0^- &= 0, \end{aligned}$$

where the ordering is componentwise. The linearization of (4.3) about the solution we are examining can, after suitable rearrangement, be written as

$$\begin{aligned} 0 \in &\begin{bmatrix} \mathcal{L}'' & H^T & G^{+T} & G^{0T} & G^{-T} \\ -H & 0 & 0 & 0 & 0 \\ -G^+ & 0 & 0 & 0 & 0 \\ -G^0 & 0 & 0 & 0 & 0 \\ -G^- & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x - x_0 \\ v - v_0 \\ u^+ - u_0^+ \\ u^0 - u_0^0 \\ u^- - u_0^- \end{bmatrix} \\ &+ \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ -g^-(x_0) \end{bmatrix} + \partial\psi_{\mathbf{R}^n \times \mathbf{R}^q \times \mathbf{R}^r_+ \times \mathbf{R}^s_+ \times \mathbf{R}^t_+} \begin{bmatrix} x \\ v \\ u^+ \\ u^0 \\ u^- \end{bmatrix}, \end{aligned} \tag{4.4}$$

where $\mathcal{L}''$ denotes $\mathcal{L}''(x_0, u_0, v_0)$, H denotes $h'(x_0)$, G^+ denotes $g^{+\prime}(x_0)$, etc. One can check that the reduced form of (4.4) is

$$0 \in \left[\begin{array}{ccc|c} \mathcal{L}'' & H^T & G^{+T} & G^{0T} \\ -H & 0 & 0 & 0 \\ -G^+ & 0 & 0 & 0 \\ \hline -G^0 & 0 & 0 & 0 \end{array}\right] \begin{bmatrix} y \\ v \\ w^+ \\ u^0 \end{bmatrix} + \partial\psi_{\mathbf{R}^{n+q+r} \times \mathbf{R}^s_+} \begin{bmatrix} y \\ v \\ w^+ \\ u^0 \end{bmatrix};$$

where $w^+ := u^+ - u_0^+$; the dotted lines in the matrix indicate the appropriate partitioning for the analysis of §3. Using the results of that section, we see that necessary and sufficient conditions for strong regularity of (4.3) at the solution in question are that the matrix

$$\begin{bmatrix} \mathcal{L}'' & H^T & G^{+T} \\ -H & 0 & 0 \\ -G^+ & 0 & 0 \end{bmatrix} \tag{4.5}$$

be nonsingular, and that its Schur complement

$$\begin{bmatrix} G^0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \mathcal{L}'' & H^T & G^{+T} \\ -H & 0 & 0 \\ -G^+ & 0 & 0 \end{bmatrix}^{-1} \begin{bmatrix} G^{0T} \\ 0 \\ 0 \end{bmatrix} \tag{4.6}$$

be a P-matrix. Of course, if $g^0(x_0)$ is vacuous then the matrix in (4.6) does not appear, and in that case one may apply the standard implicit-function theorem as was done in, e.g., [3], [4], and [11]. If $g^0(x_0)$ is not vacuous, then the results given in this paper permit one to carry out a similar analysis even though the classical implicit-function theorem does not apply.

In the case in which $g^0(x_0)$ is vacuous, it is well known that certain standard assumptions on the problem suffice to guarantee nonsingularity of the matrix in (4.5) above. In terms of our notation, these are:

(a) The *second-order sufficient condition* [3, chapter 2]: For each nonzero y such that

$$G^+y = 0,$$

$$G^0y \leqslant 0,$$

$$Hy = 0$$

one has $\langle y, \mathcal{L}''y \rangle > 0$.

Note that G^0 appears in the general form of these conditions; if $g^0(x_0)$ were vacuous then of course G^0 would not appear.

(b) *Linear independence of gradients of the binding constraints*: The matrix

$$\begin{bmatrix} G^+ \\ G^0 \\ H \end{bmatrix}$$

has full row rank.

(c) *Strict complementary slackness*: $g^0(x_0)$ is vacuous.

Actually, only (a) and (b) are required for nonsingularity, but (c) is needed to make sure that (in our terminology) the reduced form of the problem contains only equations.

We shall show now that condition (b), together with a slightly strengthened form of (a), will suffice to guarantee that conditions (1) and (2) above are met, and thus that (4.3) is strongly regular at the solution we are considering. The strengthened form of (a) is:

(a′) The *strong second-order sufficient condition*: For each nonzero y with

$$G^+y = 0,$$

$$Hy = 0,$$

one has $\langle y, \mathcal{L}''y \rangle > 0$.

Of course, if $g^0(x_0)$ is vacuous (i.e., if (c) holds) then (a) and (a′) are the same, but in general (a′) is a stronger requirement than is (a). With this change, we can prove that even without strict complementary slackness the problem (4.3) will be strongly regular.

THEOREM 4.1. *Let θ, g and h be functions from an open set $\Omega \subset \mathbf{R}^n$ to $\mathbf{R}$, $\mathbf{R}^p$ and $\mathbf{R}^q$, respectively, which are twice differentiable at a point $x_0 \in \Omega$. Suppose that x_0, together with points $u_0 \in \mathbf{R}^p$ and $v_0 \in \mathbf{R}^q$, solves* (4.3). *If the strong second-order sufficient condition holds at (x_0, u_0, v_0) together with linear independence of the gradients of the binding constraints, then* (4.3) *is strongly regular there.*

PROOF. To prove that the matrix in (4.5) is nonsingular, suppose that a, b and c are such that

$$\begin{aligned} \mathfrak{L}''a + H^Tb + G^{+T}c &= 0, \\ -Ha &= 0, \\ -G^{+}a &= 0. \end{aligned} \tag{4.7}$$

Premultiplying the equations in (4.7) by a^T, b^T and c^T, respectively, and adding the results, we find that $\langle a, \mathfrak{L}''a\rangle = 0$. This, together with the second and third equations of (4.7) and the strong second-order sufficient condition, implies that $a = 0$; the first equation of (4.7) and the linear independence assumption now imply that b and c are also zero. Thus the matrix is nonsingular, so that the Schur complement shown in (4.6), which we shall denote by S, exists. To gain some additional information about its structure, we note that the equations

$$\begin{aligned} \mathfrak{L}''V + H^TA + G^{+T}B &= G^{0T}, \\ -HV &= 0, \\ -G^{+}V &= 0 \end{aligned} \tag{4.8}$$

uniquely define matrices $V(n \times s)$, $A(q \times s)$ and $B(r \times s)$. We then have

$$S = \begin{bmatrix} G^0 & 0 & 0 \end{bmatrix} \begin{bmatrix} V \\ A \\ B \end{bmatrix} = G^0V,$$

but upon premultiplying the first equation of (4.8) by V^T we find (since $\mathfrak{L}''$ is symmetric) that $S = V^T\mathfrak{L}''V$. Thus S is symmetric, so it will be a P-matrix if and only if it is positive definite. Suppose that $z \in \mathbf{R}^s$ with $\langle z, Sz\rangle \leqslant 0$. Then with $y := Vz$, we have $\langle y, \mathfrak{L}''y\rangle \leqslant 0$ and (from (4.8)) $Hy = 0$ and $G^{+}y = 0$. By the strong second-order sufficient condition we must now have $y = 0$. Postmultiplying the first equation of (4.8) by z and using $Vz = 0$ we have $H^TAz + G^{+T}Bz = G^0z$, which implies $z = 0$ by the linear independence assumption. This completes the proof.

It is clear that the conditions of Theorem 4.1, although sufficient, are not in general necessary for strong regularity (consider the problem of minimizing the scalar function $-\xi^2$ with no constraints and let $\xi_0 = 0$!). It may not be so clear whether we could have used the standard second-order sufficient condition (as given in (a) above) in the hypothesis of Theorem 4.1 instead of the somewhat stronger condition (a′). To see that this cannot be done, consider the following example:

$$\begin{aligned} &\text{minimize } \tfrac{1}{2}\left(x_1^2 - x_2^2\right) - px_1 \\ &\text{subject to } -x_1 + 2x_2 \leqslant 0, \\ &\qquad\qquad\quad -x_1 - 2x_2 \leqslant 0, \end{aligned} \tag{4.9}$$

where p represents a perturbation parameter. We can write the necessary optimality conditions for (4.9) as the linear generalized equation

$$0 \in \begin{bmatrix} 1 & 0 & -1 & -1 \\ 0 & -1 & 2 & -2 \\ 1 & -2 & 0 & 0 \\ 1 & 2 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ u_1 \\ u_2 \end{bmatrix} + \begin{bmatrix} -p \\ 0 \\ 0 \\ 0 \end{bmatrix} + \partial\psi_{\mathbf{R}^2 \times \mathbf{R}^2_+} \begin{bmatrix} x_1 \\ x_2 \\ u_1 \\ u_2 \end{bmatrix}, \tag{4.10}$$

and we see that for $p = 0$, (4.10) is already in reduced form with its unique solution at the origin in $\mathbf{R}^4$. However, for each positive p there are *three* solutions, as follows:

$$x = \begin{bmatrix} p \\ 0 \end{bmatrix}, \qquad u = 0 \quad \text{(saddle point);} \tag{i}$$

$$x = \tfrac{2}{3} p \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \qquad u = \tfrac{1}{3} p \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \text{(local minimum);} \tag{ii}$$

$$x = \tfrac{2}{3} p \begin{bmatrix} 2 \\ -1 \end{bmatrix}, \qquad u = \tfrac{1}{3} p \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad \text{(local minimum).} \tag{iii}$$

By making p sufficiently small, all of these solutions can be brought within any preassigned neighborhood of the origin. It follows that (4.10) is not strongly regular there; however, one can verify easily that the solution of (4.9) for $p = 0$ satisfies both the standard second-order sufficient condition and the linear independence condition. Those conditions would therefore not suffice to establish Theorem 4.1.

We have dealt with the nonlinear programming problem in the form (4.1) because that form (with no implicit constraints on x) is frequently seen in the literature. However, it is perhaps worth pointing out here that for some problems one may do much better by formulating them with implicitly-constrained variables. For example, consider the problem of minimizing the expression $\langle c, x\rangle + \alpha\|x\|_\infty$ for $x \in \mathbf{R}^2$, where $\|x\|_\infty$ is the maximum absolute value of the components of x. If, for example, we take $c = [\frac{1}{2}, 0]$ and $\alpha = 1$, one way to formulate this problem using explicit linear constraints is:

$$\begin{aligned} &\underset{x_1, x_2, \eta}{\text{minimize}} && \eta + \frac{1}{2} x_1 \\ &\text{subject to} && \begin{bmatrix} I \\ -I \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - e\eta \leqslant 0, \end{aligned} \tag{4.11}$$

where I denotes the identity matrix and e is a vector of ones. The solution is evidently $\eta = 0$, $x = 0$, and all four constraints are binding; thus one cannot use Theorem 4.1 to show that the problem is strongly regular. In fact, this problem is *not* strongly regular: if the vector 0 is perturbed to $(-\epsilon, 0, -\epsilon, 0)^T$ for $\epsilon > 0$, then the optimal solution set becomes $\{(0, \lambda, \epsilon) \mid \lambda \in [-\epsilon, \epsilon]\}$. However, if one formulates the problem as

$$\begin{aligned} &\underset{x_1, x_2, \eta}{\text{minimize}} && \eta + \frac{1}{2} x_1 \\ &\text{subject to} && \begin{pmatrix} x_1 \\ x_2 \\ \eta \end{pmatrix} \in C := \left\{ \begin{pmatrix} x_1 \\ x_2 \\ \eta \end{pmatrix} \middle| \, \eta \geqslant \|x\|_\infty \right\}, \end{aligned} \tag{4.12}$$

then it can be expressed by the generalized equation

$$0 \in \begin{bmatrix} \frac{1}{2} \\ 0 \\ 1 \end{bmatrix} + \partial\psi_C \begin{bmatrix} x_1 \\ x_2 \\ \eta \end{bmatrix}, \tag{4.13}$$

and the reduction procedure alone suffices to show that (4.13) is strongly regular (see the Appendix; here the reduced form is vacuous, since in this case the face F is just the origin). Of course, this problem can be solved by inspection, but it conveniently illustrates the point that formulation can make a critical difference in the stability properties of a problem. Here the formulation in (4.11) makes even the structure of the function $\|\cdot\|_\infty$ subject to perturbation, whereas in the original problem the data subject to perturbation might reasonably have been only α and the components of c. This feature of the problem is appropriately reflected in the formulation given in (4.12), and this illustrates the fact that the standard form (4.1) may sometimes be seriously inadequate for proper representation of a problem.

Appendix. *Conversion of a linear generalized equation to reduced form.* We shall deal here with the linear generalized equation

$$0 \in Ax + a + \partial\psi_C(x), \tag{A.1}$$

where C is a nonempty polyhedral convex set in $\mathbf{R}^n$, A is $n \times n$ and $a \in \mathbf{R}^n$. If x_0 is a solution of (A.1), we shall show that in the general case one can find integers r and s, with $r + s \leqslant n$, a pointed polyhedral cone $K \subset \mathbf{R}^s$ of dimension s, and an $(r + s) \times (r + s)$ matrix B, such that there is a natural correspondence between solutions of

$$y \in Ax + a + \partial\psi_C(x) \tag{A.2}$$

for small y and for x near x_0, and solutions of the reduced form

$$z \in Bw + \partial\psi_{\mathbf{R}^r \times K}(w) \tag{A.3}$$

for z and w near zero. In the case $C = \mathbf{R}^k \times \mathbf{R}^l_+$ this is clear, as has already been remarked in the main part of the paper: it corresponds to removal of the "inactive constraints." In other special cases, either K or $\mathbf{R}^r$ (or both) might be vacuous, so that the problem would become even simpler. For the general case, we need some properties of polyhedral convex sets, which we shall list here without proof. We use the notation $T_C(x_0)$ for the tangent cone to C at x_0 (i.e., for $\partial\psi_C(x_0)^\circ$); for general information about tangent cones, polyhedrality, etc., see [13]. In all of the following statements, C is a nonempty polyhedral convex set in $\mathbf{R}^n$.

PROPOSITION A.1. *Let $x_0 \in C$. Then there exists a neighborhood U of the origin such that $(C - x_0) \cap U = T_C(x_0) \cap U$.*

PROPOSITION A.2. *Let $y_0 \in \mathbf{R}^n$, and let $F := \partial\psi_C^*(-y_0)$. Then for each $x \in F$, $\partial\psi_F(x) = \partial\psi_C(x) + y_0\mathbf{R}_+ := \{y + \alpha y_0 \mid y \in \partial\psi_C(x), \alpha \geqslant 0\}$. Further, for each y near y_0, $\partial\psi_C^*(-y) \subset F$.*

The next proposition provides the key to establishing a correspondence between solutions of (A.2) and those of a problem equivalent to the reduced form (A.3).

PROPOSITION A.3. *Let y_0 and F be as in Proposition A.2. Let $x_0 \in F$, and write $T := T_F(x_0)$. Let L be the subspace parallel to F. Then there exist neighborhoods U and V of the origin, such that for each $h \in U$ and each $k \in V$,*

$$0 \in (y_0 + k) + \partial\psi_C(x_0 + h)$$

if and only if

$$0 \in P_L k + \partial\psi_T(h),$$

where P_L denotes the orthogonal projector on L.

PROOF. Applying Proposition A.1 to F, we can find an open neighborhood U of the origin such that $(F - x_0) \cap U = T \cap U$. By Proposition A.2, we can find a neighborhood V_0 of the origin such that for each $k \in V_0$, if $0 \in (y_0 + k) + \partial\psi_C(x)$ then $x \in F$. Also, there is a neighborhood V_1 of zero such that for any $x \in F$,

$$(\partial\psi_C(x) + y_0\mathbf{R}_+) \cap V_1 \subset \partial\psi_C(x) + y_0[0, 1] := \{y + \alpha y_0 \mid y \in \partial\psi_C(x),\quad 0 \leqslant \alpha \leqslant 1\}.$$

To see this, note that $\partial\psi_C(\cdot)$ has only finitely many values on F; for each of these values, say a polyhedral convex cone P, let σ be a simplex containing the origin in its interior. The set $(P + y_0\mathbf{R}_+) \cap \sigma$ is a bounded polyhedral convex set, so it can be written as the convex hull of points $q_1, \ldots, q_m$ in $\mathbf{R}^n$. For each i, $q_i = p_i + \alpha_i y_0$, with $p_i \in P$ and $\alpha_i \geqslant 0$. With $\sigma_0 := \sigma/\max\{1, \alpha_1, \ldots, \alpha_m\}$ it follows that $(P + y_0\mathbf{R}_+) \cap \sigma_0 \subset P + y_0[0, 1]$. Repeating this procedure for each value of $\partial\psi_C(\cdot)$ on F and intersecting the resulting simplices, we obtain the required V_1. Finally, we let $V := V_0 \cap (-V_1)$. Now choose any $h \in U$ and $k \in V$.

(only if): Suppose $0 \in (y_0 + k) + \partial\psi_C(x_0 + h)$. As $k \in V_0$, we have $x_0 + h \in F$, and thus $h \in (F - x_0) \cap U$, so actually $h \in T$. Let $t \in T$; we shall prove that $\langle P_L k, t - h\rangle \geqslant 0$. To begin with, note that we may replace $t - h$ by $\lambda(t - h) = [(1 - \lambda)h + \lambda t] - h$ for any small positive λ; thus lose no generality by assuming that $t \in T \cap U$ (recall U was open, so $h \in \operatorname{int} U$). Then $x_0 + t \in F$, and of course $x_0 + h \in F$, so

$$t - h = (x_0 + t) - (x_0 + h) \in L.$$

As y_0 is orthogonal to L by construction, we have $\langle y_0 + P_{L^\perp} k, t - h\rangle = 0$. Therefore, since $k = P_L k + P_{L^\perp} k$, we have

$$\langle P_L k, t - h\rangle = \langle y_0 + k, (x_0 + t) - (x_0 + h)\rangle \geqslant 0,$$

since we assumed that $0 \in (y_0 + k) + \partial\psi_C(x_0 + h)$. It follows that $0 \in P_L k + \partial\psi_T(h)$.

(if): Suppose that $0 \in P_L k + \partial\psi_T(h)$. As $h \in T \cap U$, we have $x_0 + h \in F$; as $h \in \operatorname{int} U$ we have also that $\partial\psi_T(h) = \partial\psi_F(x_0 + h) = \partial\psi_C(x_0 + h) + y_0\mathbf{R}_+$, where we have used Propositions A.1 and A.2. Noting that $L^\perp \subset \partial\psi_F(x_0 + h)$, we see that since $k = P_L k + P_{L^\perp} k$,

$$0 \in P_L k + \partial\psi_F(x_0 + h) = k + \partial\psi_F(x_0 + h) = k + y_0\mathbf{R}_+ + \partial\psi_C(x_0 + h).$$

But then $-k \in (\partial\psi_C(x_0 + h) + y_0\mathbf{R}_+) \cap V_1$, so for some $\alpha \in [0, 1]$,

$$-k - \alpha y_0 \in \partial\psi_C(x_0 + h),$$

and since $x_0 + h \in F$, we have

$$-y_0 \in \partial\psi_C(x_0 + h).$$

Multiplying the second inclusion by $(1 - \alpha)$ and adding it to the first we obtain

$$0 \in (y_0 + k) + \partial\psi_C(x_0 + h),$$

which completes the proof.

Using Proposition A.3, we can now construct the reduced form of (A.1) for general polyhedral C. To do so, suppose that x_0 solves (A.1). Define $y_0 := Ax_0 + a$, and let F, T and L be defined as above. Let M be the lineality space of T and choose orthonormal bases $b_1, \ldots, b_r$ for M, $b_{r+1}, \ldots, b_{r+s}$ for $L \cap M^\perp$, and $b_{r+s+1}, \ldots, b_n$ for $L^\perp$ (note: we assume here for generality that none of these spaces is of zero dimension; if one or more are zero-dimensional, the analysis is only simplified). Assume that A and all vectors in $\mathbf{R}^n$ are written in terms of the basis $b_1, \ldots, b_n$; we shall write, for example, $x = (x^1, x^2, x^3)$ with $x^1 \in M$, $x^2 \in L \cap M^\perp$, and $x^3 \in L^\perp$.

We can write $T = M + (T \cap M^{\perp})$, and the cone $K := T \cap M^{\perp}$ is pointed and contained in $L \cap M^{\perp}$; in fact, since $L = \operatorname{aff} T$, K has the dimension of $L \cap M^{\perp}$. Thus, if $h \in T$ we have

$$h = (h^1, h^2, 0), \quad h^1 \in \mathbf{R}^r, \quad h^2 \in K,$$

$$\partial\psi_T(h) = \{0\} \times \partial\psi_K(h^2) \times \mathbf{R}^{n-r-s}.$$

Now apply Proposition A.3, with $y_0 := Ax_0 + a$, to obtain neighborhoods U and V of the origin; construct additional neighborhoods N of x_0 and W of 0 such that if $x \in N$ and $y \in W$ then

$$h := x - x_0 \in U,$$

$$k := (-y + Ax + a) - (Ax_0 + a) = Ah - y \in V.$$

Choose any $x \in N$ and any $y \in W$. Now Proposition A.3 tells us that (A.2) holds for x and y if and only if

$$0 \in P_L(Ah - y) + \partial\psi_T(h). \tag{A.4}$$

Writing

$$y = \begin{bmatrix} y^1 \\ y^2 \\ y^3 \end{bmatrix}, \quad h = \begin{bmatrix} h^1 \\ h^2 \\ h^3 \end{bmatrix}, \quad A = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{bmatrix},$$

we see that (A.4) is equivalent to

$$h^3 = 0, \quad \begin{bmatrix} y^1 \\ y^2 \end{bmatrix} \in \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} h^1 \\ h^2 \end{bmatrix} + \begin{bmatrix} \{0\} \\ \partial\psi_K(h^2) \end{bmatrix}, \tag{A.5}$$

and with

$$z := \begin{bmatrix} y^1 \\ y^2 \end{bmatrix}, \quad w := \begin{bmatrix} h^1 \\ h^2 \end{bmatrix}, \quad B := \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix},$$

we see that the second relation in (A.5) is of the form shown in (A.3). We have thus established the required correspondence.

This construction, together with Proposition A.3, now permits us to state a general criterion for strong regularity of the generalized equation (A.1).

THEOREM A.4. *Let C be a polyhedral convex set in $\mathbf{R}^n$; let A be $n \times n$ and $a \in \mathbf{R}^n$. Suppose that $x_0 \in \mathbf{R}^n$ solves* (A.1). *Then* (A.1) *is strongly regular at x_0 if and only if its reduced form* (i) *is vacuous, or* (ii) *has an inverse which is single-valued on all of L.*

PROOF. We assume the notation used in the earlier part of the appendix.

(if): First, suppose the reduced form of (A.1) is vacuous. This means that $L = \{0\}$, so that F is the singleton $\{x_0\}$ and $T = \{0\}$. Proposition A.3 now implies that for any y near zero, x_0 itself is the unique solution of (A.2) near x_0. Thus (A.1) is strongly regular. On the other hand, suppose that the inverse of the reduced form is single-valued on all of $\mathbf{R}^n$ (and therefore Lipschitzian there, as we have previously pointed out). Consider the neighborhoods W and N constructed above, and find a neighborhood W_0 of the origin with $W_0 \subset W$ and having the property that if $y \in W_0$ and if

$[{y^1 \atop y^2}]$ and some $[{h^1 \atop h^2}]$ satisfy the second relation in (A.5), then

$$x_0 + \begin{bmatrix} h^1 \\ h^2 \\ 0 \end{bmatrix} \in N$$

(here we have used the continuity of the reduced form's inverse). Choose any $y \in W_0$; by hypothesis the second relation in (A.5) is solvable by a unique $[{h_1 \atop h_2}]$. With

$$x := x_0 + \begin{bmatrix} h_1 \\ h_2 \\ 0 \end{bmatrix},$$

we have $x \in N$, and therefore x and y solve (A.2). Further, the solution x is unique in N because the solution $[{h_1 \atop h_2}]$ of (A.5) is unique. Thus the restriction to W_0 of $N \cap [A(\cdot) + a + \partial\psi_C(\cdot)]^{-1}$ is a single-valued function; the Lipschitzian property now follows from [12, Proposition 2] and our earlier remarks. Therefore (A.1) is strongly regular at x_0.

(only if): Suppose that (A.1) is strongly regular at x_0, and that its reduced form is nonvacuous; choose neighborhoods S of 0 and Q of x_0 so that for each $y \in S$, (A.2) has a solution $x \in Q$ which is unique in Q (here we are again using continuity). Let $S_0 := S \cap W$ and $Q_0 := Q \cap N$, and choose any $[{y_1 \atop y_2}]$ in the projection of S_0 on L (a neighborhood of zero in L). Let $y \in S_0$ with $P_L y = [{y_1 \atop y_2}]$. By construction, a solution $x \in Q$ exists for (A.2); define $h := x - x_0$. By the discussion preceding this theorem, we see that $h_3 = 0$ and that $[{y_1 \atop y_2}]$ and $[{h_1 \atop h_2}]$ satisfy the inclusion in (A.5). Let

$$G := \left\{ \begin{bmatrix} g_1 \\ g_2 \end{bmatrix} \middle| \begin{bmatrix} g_1 \\ g_2 \\ 0 \end{bmatrix} \in Q - x_0 \right\};$$

this is a neighborhood of the origin in L, and it contains $[{h_1 \atop h_2}]$. If there were another solution $[{h_1' \atop h_2'}]$ of (A.5) in G, then

$$x_0 + \begin{bmatrix} h_1' \\ h_2' \\ 0 \end{bmatrix}$$

would belong to Q and would solve (A.2), contradicting strong regularity. Therefore the reduced form of (A.1) is uniquely solvable for each $[{y_1 \atop y_2}]$ in the neighborhood $P_L S_0$; but the reduced form is positively homogeneous so it uniquely solvable for each $[{y_1 \atop y_2}] \in L$. This completes the proof.

Note added in proof. For algorithmic applications of the results given here, see Norman H. Josephy, *Newton's Method for Generalized Equations and the PIES Energy Model*, Ph.D. Dissertation, Department of Industrial Engineering, University of Wisconsin–Madison, May 1979.

In addition, existence and continuity results for the special case of nonlinear programming, under the assumptions of Theorem 4.1, have been obtained by K. Jittorntrum (*Sequential Algorithms in Nonlinear Programming*, Dissertation, The Australian National University, 1978).

References

[1] Cottle, R. W. (1974). Manifestations of the Schur Complement. *Linear Algebra and Appl.* **8** 189–211.

[2] Daniel, J. W. (1973). Stability of the Solution of Definite Quadratic Programs. *Math. Programming* **5** 41–53.

[3] Fiacco, A. V. and McCormick, G. P. (1968). *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*. Wiley, New York.

[4] Fiacco, A. V. (1976). Sensitivity Analysis for Nonlinear Programming Using Penalty Methods. *Math. Programming* **10** 287–311.

[5] Hager, W. W. (1979). Lipschitz Continuity for Constrained Processes. *SIAM J. Control Optimization*. **17** 321–338.

[6] Lemke, C. E. (1970). Recent Results on Complementarity Problems. In: *Nonlinear Programming*, J. B. Rosen, O. L. Mangasarian and K. Ritter, eds. pp. 349–384. Academic Press, New York.

[7] Levitin, E. S. (1974). On Differential Properties of the Optimum Value of Parametric Problems of Mathematical Programming. *Soviet Math. Dokl.* **15** 603–608.

[8] ———. (1975). On the Local Perturbation Theory of a Problem of Mathematical Programming in a Banach Space. *Soviet Math. Dokl.* **16** 1354–1358.

[9] ———. (1976). Differentiability with Respect to a Parameter of the Optimal Value in Parametric Problems of Mathematical Programming. *Cybernetics* **12** 46–64.

[10] Mangasarian, O. L. (1979). Uniqueness of Solution in Linear Programming. *Linear Algebra and Appl.* **25** 151–162.

[11] Robinson, S. M. (1974). Perturbed Kuhn-Tucker Points and Rates of Convergence for a Class of Nonlinear-Programming Algorithms. *Math. Programming* **7** 1–16.

[12] ———. (1979). Generalized Equations and Their Solutions, Part I: Basic Theory. *Math. Programming Stud.* **10** 128–141.

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

MATHEMATICS RESEARCH CENTER, UNIVERSITY OF WISCONSIN–MADISON, MADISON, WISCONSIN 53706.