Journal of Global Optimization **28:** 405–417, 2004.

A Linearization Method for Nondegenerate Variational Conditions

STEPHEN M. ROBINSON
Department of Industrial Engineering, University of Wisconsin–Madison, 1513 University Avenue, Madison, WI 53706-1572, USA (e-mail: smrobins@wisc.edu)

(Received and accepted 12 March 2003)

Abstract. We employ recent results about constraint nondegeneracy in variational conditions to design and justify a linearization algorithm for solving such problems. The algorithm solves a sequence of affine variational inequalities, but the variational condition itself need not be a variational inequality: that is, its underlying set need not be convex. However, that set must be given by systems of differentiable nonlinear equations with additional polyhedral constraints. We show that if the variational condition has a solution satisfying nondegeneracy and a standard regularity condition, and if the linearization algorithm is started sufficiently close to that solution, the algorithm will produce a well defined sequence that converges Q-superlinearly to the solution.

Key words: Linearization method, Nondegeneracy, Variational condition

1. Introduction

This paper develops a linearization method for solving certain variational conditions involving differentiable functions. Here we use the term *variational condition* in the sense of [10, Example 6.13], to mean an inclusion of the form

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

where S is a subset of $\mathbb{R}^n$ defined by

$$S = \{x \in P \cap X_0 \mid h(x) \in Q\}, \tag{1.2}$$

where P and Q are polyhedral convex sets in $\mathbb{R}^n$ and $\mathbb{R}^m$ respectively and f is a function from an open subset X_0 of $\mathbb{R}^n$ to $\mathbb{R}^n$. The set $N_S(x)$ is the *normal cone* of S at x; it is empty if $x \notin S$. We use the symbols $T_A(z)$ and $N_A(z)$ to denote tangent and normal cones to a set A at a point $z \in A$ in the sense of [10, Sections 6.A, 6.B]. If the sets involved are convex, these tangent and normal cones coincide with the ones used in convex analysis [9].

The key requirement we impose here is that (1.1) should satisfy the condition of *constraint nondegeneracy* studied in [8]. In that paper we identified a number of stability and sensitivity properties of variational conditions satisfying the constraint-nondegeneracy requirement. Here we illustrate an application of that general theory by showing how to use it to prove local Q-superlinear convergence of a linearization method for solving (1.1). We focus on the way in which application of the theory promotes conceptual simplicity of the arguments, rather than on obtaining the maximum possible generality.

The rest of this paper is organized in four sections. The next, Section 2, reviews enough of the material from [8] to provide the tools and notation that we will need. We then explain the linearization method in Section 3, and establish its convergence in Section 4. Finally, in Section 5 we illustrate the application of the method to a small example, and comment on issues that could benefit from additional research.

2. Review of nondegeneracy

This section reviews some of the material from [8] that we will need to use in the statement and analysis of the computational method. In defining and analyzing our method we will need to employ a variational inequality whose underlying set changes at each iteration, so we first examine some properties of the representation of such a set.

The framework of [8] included the introduction of perturbations into S to produce a set $S(u)$ depending on a parameter u in a Banach space U, with a point u_0 representing the unperturbed problem. The precise dependence of $S(u)$ on u was expressed by the definition

$$S(u)=\{x\in P\cap X_0 \mid h(x,u)\in Q\}=P\cap h(\cdot,u)^{-1}(Q), \tag{2.3}$$

where P and Q were as described in Section 1. The function h was assumed to be C^k $(k\geqslant 1)$ from a product of open neighborhoods X_0 and U_0 of points $x_0\in\mathbb{R}^n$ and $u_0\in U$ respectively, to $\mathbb{R}^m$. It is thus possible to write the set $S(u)$ as the solution set of a system of finitely many (generally nonlinear) inequalities or equations, though expressing it explicitly in such a form may require additional rewriting and may in some cases be cumbersome.

We then carried the perturbations into the variational condition by letting f be a function from $X_0\times U_0$ to $\mathbb{R}^n$ and considering the inclusion

$$0\in f(x,u)+N_{S(u)}(x). \tag{2.4}$$

Accordingly, in this setup both the function and the set appearing in the variational condition (2.4) depend on the parameter u.

Second, we required the representation (2.3) to satisfy a certain property called *(constraint) nondegeneracy* that led to very favorable behavior under perturbations. If we represent partial derivatives in x and in u on $X_0\times U_0$ by the commonly used symbols $h_x(x,u)$ and $h_u(x,u)$, then we say that the representation (2.3) is *nondegenerate* at x_0 for $u=u_0$ if

$$h_x(x_0,u_0)[\operatorname{lin} T_P(x_0)]-\operatorname{lin} T_Q(h(x_0,u_0))=\mathbb{R}^m, \tag{2.5}$$

and *degenerate* there otherwise. Here the symbol $\operatorname{lin} C$ denotes the *lineality space* of a convex set C, defined to be the set of points v such that $C+v=C$.

This nondegeneracy condition uses only information about the derivative of h at the point (x_0,u_0), so that (except for the underlying differentiability requirement) we do not need any information about perturbed versions $h(\cdot,u)$ of h. In fact, given the underlying sets P and Q, to determine whether nondegeneracy holds we need only know the point x_0, the function value $y_0=h(x_0,u_0)$, and the partial derivative $h_x(x_0,u_0)$. Thus, nondegeneracy is a property defined by the linearized constraint system. The original introduction of nondegeneracy (without the requirement of polyhedrality) was in [7]. More recently, Bonnans and Shapiro have considered another approach in [1, Section 4.6.1]. They noted in [1, Remark 4.72] that for the present case the two definitions both reduce to the requirement above.

Nondegeneracy is sometimes used to describe a different property of variational inequalities, which has nothing to do with the way we use the term here. Specifically, some authors describe a solution c_0 of a variational inequality specified by a function f and a polyhedral convex set C as nondegenerate if it lies in the relative interior of the critical face of C (that is, the face $\{c\in C \mid \langle f(c_0), c-c_0\rangle = 0\}$): see, e.g., [2, Definition 3.4.1]. That definition involves the position of $f(c_0)$ with respect to C, whereas the concept with which we deal here involves only the representation of the underlying set.

In [8] we showed how to combine nondegeneracy with a known regularity property that has been shown to imply good behavior for variational inequalities. In the next section we will need this combination, so we give the relevant theorem here. It is a combination of Theorems 3.1 and 5.3 of [8].

THEOREM 2.1. *Let P and Q be polyhedral convex subsets of $\mathbb{R}^n$ and $\mathbb{R}^m$ respectively. Let U be a Banach space, u_0 a point of U, and U_0 an open subset of U containing u_0. Let X_0 be an open subset of $\mathbb{R}^n$ and x_0 a point in $P\cap X_0$, and let h be a C^2 function from $X_0\times U_0$ to $\mathbb{R}^m$. Let $y_0=h(x_0,u_0)$ and assume that $y_0\in Q$.*

For $u\in U_0$ define $S(u)$ by (2.3), and assume that the representation (2.3) is nondegenerate at (x_0,y_0) for $u=u_0$. Let f be a C^1 function from $X_0\times U_0$ to $\mathbb{R}^n$, and suppose that (x_0,u_0) solves (2.4). Let Z be the kernel in $\mathbb{R}^{m+n}$ of the linear transformation $G=[h_x(x_0,u_0)\ \ -I]$, let M be an m-dimensional subspace of $\operatorname{lin} T_P(x_0)\times \operatorname{lin} T_Q(y_0)$ whose image under G is $\mathbb{R}^m$, and let G^- be the unique linear generalized inverse of G with respect to the direct-sum decomposition $\mathbb{R}^{n+m}=Z\oplus M$. Let N be the linear operator from $\mathbb{R}^{n+m}$ to $\mathbb{R}^n$ defined by $N(x,y)=x$.

Define a linear transformation $L:\mathbb{R}^n\to\mathbb{R}^n$ by

$$L(z)=d_x f(x_0,u_0)(z)-f(x_0,u_0)NG^- h_{xx}(x_0,u_0)(z), \tag{2.6}$$

and polyhedral convex cones T and K in $\mathbb{R}^n$ by

$$T=\{z\in T_P(x_0)\mid h_x(x_0,u_0)z\in T_Q(y_0)\},$$

$$K=\{z\in T\mid \langle f(x_0,u_0),z\rangle=0\}.$$

If the normal map L_K is a homeomorphism, then there is a positive λ_0 such that for each $\lambda > \lambda_0$ there are neighborhoods X'' of x_0 and U'' of u_0, and a function $x: U'' \to \mathbb{R}^n$ that is Lipschitzian with modulus λ, such that $x(u_0) = x_0$ and such that for each $u \in U''$ $x(u)$ is the unique point in X'' satisfying (2.4)*. Further, the function x is B-differentiable at u_0 with*

$$dx(u_0)(w_u) = \Pi_K \circ (L_K)^{-1} \circ [-d_u p(u_0)(w_u)] - NG^- h_u(x_0, u_0)(w_u), \tag{2.7}$$

where $d_u p(u_0)(w_u)$ is given by

$$\begin{aligned}\langle d_u p(u_0)(w_u), z\rangle &= \langle f_x(x_0, u_0)[-NG^- h_u(x_0, u_0)(w_u)] + f_u(x_0, u_0)(w_u), z\rangle \\ &+ \langle f(x_0, u_0), -NG^- h_{xx}(x_0, u_0)[-NG^- h_u(x_0, u_0)(w_u)][z] \\ &- NG^- h_{ux}(x_0, u_0)[w_u][z]\rangle. \end{aligned} \tag{2.8}$$

This section has presented some foundational material, taken from [8], about the analysis of variational conditions satisfying certain conditions for good behavior. In the next section we build on that foundation to develop and justify a linearization method for solving (1.1).

3. The linearization method

This section suggests a way of constructing a linearization method for numerical solution of (1.1), and comments on some issues that arise in the construction. The next section establishes conditions under which that method will converge at least Q-superlinearly in a neighborhood of a solution x_* of (1.1) at which certain conditions hold, including the constraint nondegeneracy condition described in Section 2.

To motivate the method, we observe first that the normal-map analysis in Theorem 2.1 is directed at a certain affine variational inequality, namely

$$d \in L(z) + N_K(z),$$

whose solutions give substantial amounts of information about the nonlinear problem. Such affine approximations have been studied before, e.g., by Bonnans and Shapiro [1, Eq. (5.29)] and [11, Eqs. (4.6), (4.7)]. In these cases the affine problem appears with multipliers as explicit variables. Facchinei and Pang [2, Chapter 5] also employ a similar idea.

However, if we do not yet know the solution of the variational condition we are analyzing, then we will not be able to identify the critical cone K. Therefore we might think of taking the kind of approach employed in nonlinear programming by [12, 5], and in a series of works by Pshenichnyj and co-authors, summarized in [4]. The basic idea in all of these was to linearize the set on which one works and then to compensate for the linearization by introducing into the objective function a term (usually quadratic) incorporating information about the curvature of the boundary of that set near the point of interest. One then solves the subproblem thus generated, and uses the information produced by that solution to generate the next subproblem.

To implement this idea in our situation we can look to the linear operator L given in Theorem 2.1 for guidance in choosing the quadratic compensation term. Let us start by defining a generic variable $z=(x,q)\in\mathbb{R}^n\times\mathbb{R}^m$. We will use various additional symbols to distinguish specific points, for example $z'=(x',q')$. Now assume for the moment that x_* is a solution of (1.1) that satisfies the conditions of Theorem 2.1, and that f and h have enough differentiability near x_*; we shall be precise in the next section. Let $q_* := -f(x_*)NG^-(x_*)$, where G^- and N are as defined in Theorem 2.1, and write $z_*=(x_*,q_*)$.

Given a point $z'=(x',q')$, we define a variational inequality VI(z') by

$$0\in F(x,z')+N_{S(z')}(x), \tag{3.9}$$

where

$$F(x,z')=f(x')+f_x(x')(x-x')+q'h_{xx}(x')(x-x'), \tag{3.10}$$

and

$$S(z')=\{x\in P \mid H(x,z')\in Q\}, \tag{3.11}$$

with $H(x,z')=h(x')+h_x(x')(x-x')$. Note that VI($z'$) is an affine variational inequality defined over a polyhedral convex set. By solving (3.9) (if that can be done) we obtain a new point x and, via the expression of $-F(x,z')$ as an element of the normal cone

$$N_{S(z')}(x)=\{qh_x(x')+p \mid p\in N_P(x),\ q\in N_Q[h(x')+h_x(x')(x-x')]\}, \tag{3.12}$$

we obtain a new set of multipliers q, which we shall show to be unique. Now we have a new point $z=(x,q)$ and we repeat the process using z in place of z'.

Since this process uses the multipliers q, one might ask two questions at this point. First, as we shall see below, there is an explicit formula for the multipliers in terms of the associated point x forming the rest of the pair $z=(x,q)$. Thus we may ask why q has to appear at all: perhaps we could write the procedure in terms of x alone. The answer is that the explicit formula uses the subspace M appearing in Theorem 2.1, and this subspace is unknown until we determine the solution x_*. Therefore during the computation we actually have to obtain q from normal-cone information generated at each step.

The second question naturally follows from this; it is why, if we really need multipliers, we do not just write out (1.1) in extended form, with variables and multipliers, as a variational inequality over a polyhedral convex set, and apply known methods to it. The answer to that question is that the extended form requires the identity $N_Q = N_{Q^\circ}^{-1}$, which in turn requires Q to be a cone (whose polar is Q°). In our formulation we have not made this restriction, so the extended form is not available to us.

Of course, in order to have any confidence in the solution procedure we have outlined, one would have to know that the subproblems it generates had solutions, preferably at least locally unique, and that the sequence of points $\{z_k\}$, obtained from repetitive applications of this idea starting with some given z_0, converged to (x_*, q_*). For efficient numerical solution one would also want to know that the rate of convergence was at least Q-superlinear; see [3].

We show in the next section that under the conditions of Theorem 2.1 these properties will hold, and moreover that their establishment is conceptually fairly direct and simple once we apply the analytical tools given by that theorem to the auxiliary variational inequality (3.9) that we constructed above.

4. Convergence analysis

This section establishes Q-superlinear convergence of the repeated linearization algorithm described in Section 3. The proof takes three steps. First we show that under suitable regularity conditions, for z' close enough to $z_* = (x_*, q_*)$ the subproblem (3.9) has a locally unique solution $x(z')$ that is Lipschitzian and that, moreover, $x(z')$ has a B-derivative at $z' = z_*$ that is zero (hence actually a Fréchet derivative). Then we show that if we define the associated multiplier $q(z')$ by the formula established in [8], the function $q(z')$ also has a zero F-derivative at $z' = z_*$. Combining these two results we conclude that the function $z(z') = (x(z'), q(z'))$ has a zero F-derivative at $z' = z_*$, and we use that fact to prove the convergence theorem.

Here is the first of the three steps.

THEOREM 4.1. *Assume that* (1.1) *satisfies the conditions of Theorem* 2.1 *at a solution x_*, and that the functions f and h are respectively C^2 and C^3 near x_*. Let $q_* = q(x_*)$ and $z_* = (x_*, q_*)$.*

Then there is a positive λ_0 such that for each $\lambda > \lambda_0$ there are neighborhoods X_* of x_* and Z_* of z_*, and a function $x: Z_* \to \mathbb{R}^n$ that is Lipschitzian with modulus λ, such that $x(z_*) = x_*$ and such that for each $z' \in Z_*$, $x(z')$ is the unique point in X_* satisfying the variational inequality $VI(z')$ given by (3.9)*. Moreover, the function x is F-differentiable at z_* with $dx(z_*) = 0$.*

Proof. We first show that x_* solves VI(z_*). The fact that x_* satisfies (1.1) implies that $x_* \in S$, so $x_* \in P$ and $h(x_*) \in Q$. It follows from the definition of $S(z)$ that then $x_* \in S(z_*)$. We have $F(x_*, z_*) = f(x_*)$, and by using 3.12 we obtain

$$N_{S(z_*)}(x_*) = \{q^* h_x(x_*) + p^* \mid p^* \in N_P(x_*),\ q^* \in N_Q[h(x_*)]\} = N_S(x_*),$$

so that

$$0 \in f(x_*) + N_S(x_*) = F(x_*, z_*) + N_{S(z_*)}(x_*).$$

Therefore x_* solves VI(z_*).

For local uniqueness, we apply [8, Theorem 5.3] to the perturbed variational inequality (3.9), taking the base value of the perturbation z' to be z_*. The nondegeneracy condition on the representation (3.11) of $S(z)$ at the point (x_*, z_*) is that

$$h_x(x_*)[\operatorname{lin} T_P(x_*)] - \operatorname{lin} T_Q[h(x_*)] = \mathbb{R}^m,$$

which is the same as the nondegeneracy condition on S at x_* that is part of the hypothesis. We also have to verify a regularity condition on (3.9) at z_*. To do so, define a linear transformation $L(w)$ by

$$\begin{aligned} L(w) &= F_x(x_*, z_*)(w) - F(x_*, z_*)NG^- H_{xx}(x_*, z_*)(w) \\ &= f_x(x_*)(w) - f(x_*)NG^- h_{xx}(x_*)(w). \end{aligned} \tag{4.13}$$

Noting that $q_* = -f(x_*)NG^-$ by [8, Theorem 5.6], we see that

$$L(w) = f_x(x_*)(w) + q_* h_{xx}(x_*)(w).$$

The cones T and K involved in the statement of [8, Theorem 5.3] for (3.9) are

$$T = \{w \in T_P(x_*) \mid h_x(x_*)w \in T_Q[h(x_*)]\}$$

and its critical cone at $(0, -f(x_*))$, namely

$$K = \{w \in T \mid \langle -f(x_*), w\rangle = 0\}.$$

But these quantities L, T, and K are the same as the corresponding quantities for the regularity condition applicable to (1.1) at x_*, and we have assumed the homeomorphism property for them. Therefore [8, Theorem 5.3] tells us that there is a positive λ_0 such that for each $\lambda > \lambda_0$ there are neighborhoods X_* of x_* and Z_* of z_* and a function $x: Z_* \to \mathbb{R}^n$ that is Lipschitzian with modulus λ, such that $x(z_*) = x_*$ and such that for each $z' \in Z_*$, $x(z')$ is the unique point in X_* satisfying (3.9). Moreover, the function x is B-differentiable at z_* with

$$dx(z_*)(w_z) = \Pi_K \circ (L_K)^{-1} \circ [-p(x_*, z_*)(w_z)] - NG^- H_z(x_*, z_*)(w_z), \tag{4.14}$$

where $p(x_*, z_*)(w_z)$ is defined by

$$\begin{aligned} \langle p(x_*, z_*)(w_z), v\rangle &= \langle F_x(x_*, z_*)[-NG^- H_z(x_*, z_*)(w_z)] + F_z(x_*, z_*)(w_z), v\rangle \\ &+ \langle F(x_*, z_*), -NG^- H_{xx}(x_*, z_*)[-NG^- H_z(x_*, z_*)(w_z)][v] \\ &- NG^- H_{zx}(x_*, z_*)[w_z][v]\rangle. \end{aligned} \tag{4.15}$$

We will show that this B-derivative is zero (hence is an F-derivative), and to do so it will be convenient to begin by considering the quantity $H_z(x_*, z_*)$. As $H(x,z') = h(x') + h_x(x')(x-x')$, where as before $z' = (x', q')$, we find that (using w_z as a placeholder for the argument on which the derivative operates)

$$\begin{aligned} H_z(x,z')(w_z) &= h_x(x')(w_{x'}) + h_{xx}(x')(w_{x'})(x-x') - h_x(x')(w_{x'}) \\ &= h_{xx}(x')(w_{x'})(x-x'), \end{aligned} \tag{4.16}$$

as there is no component in q. Accordingly $H_z(x_*, z_*) = 0$, so the last term in (4.14) will vanish. However, this quantity also appears in two places in (4.15), so we can simplify what is given there to the form

$$\begin{aligned} \langle p(x_*, z_*)(w_z), v\rangle = {} & \langle F_z(x_*, z_*)(w_z), v\rangle \\ & + \langle F(x_*, z_*), -NG^- H_{zx}(x_*, z_*)[w_z][v]\rangle. \end{aligned} \tag{4.17}$$

From (3.10) we have

$$F(x,z') = f(x') + f_x(x')(x-x') + q' h_{xx}(x')(x-x'),$$

so that

$$\begin{aligned} F_z(x,z')(w_z) = {} & f_{xx}(x')(w_{x'})(x-x') + q' h_{xxx}(x')(w_{x'})(x-x') \\ & - q' h_{xx}(x')(w_{x'}) + (w_{q'}) h_{xx}(x')(x-x'), \end{aligned}$$

and therefore

$$\langle F_z(x_*, z_*)(w_z), v\rangle = -q_* h_{xx}(x_*)(w_{x'})[v]. \tag{4.18}$$

Returning to (4.17), we take the other remaining term and evaluate

$$\langle F(x_*, z_*), -NG^- H_{zx}(x_*, z_*)[w_z][v]\rangle.$$

We already found in (4.16) that $H_z(x,z')(w_z) = h_{xx}(x')(w_{x'})(x-x')$, so we have $H_{zx}(x_*, z_*)[w_z] = h_{xx}(x_*)(w_{x'})$. Then

$$\begin{aligned} \langle F(x_*, z_*), -NG^- H_{zx}(x_*, z_*)[w_z][v]\rangle &= -F(x_*, z_*) NG^- h_{xx}(x_*)(w_{x'})[v] \\ &= q_* h_{xx}(x_*)(w_{x'})[v]. \end{aligned} \tag{4.19}$$

Comparing (4.18) and (4.19), we find that the two quantities differ only in sign. Accordingly, when added together in (4.17) they cancel, and therefore we have shown that $dx(z_*) = 0$. □

Theorem 4.1 carries out the first step of the convergence analysis by showing that, as a function of the variable z, the solution x of (3.9) is F-differentiable at z_* with F-derivative zero. But x is only part of the problem; at each step we will allow the problem to determine a new value of the multiplier q, and then use the resulting pair $z = (x, q)$ to set up the next problem. The next theorem provides the second step of the analysis by showing that q is F-differentiable as a function of z, with $\mathrm{d}q(z_*)=0$.

To set the stage for the theorem, we recall that [8, Theorem 5.6] showed that under the nondegeneracy condition, (2.4) holds at a point (x,u) near (x_*,u_*) in the equivalent extended (multiplier) form

$$0=f(x,u)+qh_x(x,u)+p, \qquad p\in N_P(x), \qquad q\in N_Q[h(x,u)], \tag{4.20}$$

if and only if

$$p=-f(x,u)[I-NG(x,u)^- h_x(x,u)], \qquad q=-f(x,u)NG(x,u)^-, \tag{4.21}$$

and

$$p\in N_P(x), \qquad q\in N_Q[h(x,u)].$$

Accordingly, we have an explicit formula for the (necessarily unique) multiplier q, and we will exploit this fact in our analysis.

There is a somewhat subtle point here: as we pointed out above, the formula for the multiplier depends on the subspace M, which is unknown until we know x_*. Thus, we cannot use this formula directly in computation. This is why we will find successive estimates of the multiplier from the solutions to the intermediate subproblems in our algorithm. However, we are assuming that a nondegenerate solution *exists* (even though we do not yet know it), so it is perfectly acceptable to use the form of the multiplier in analytical work.

THEOREM 4.2. *Assume the hypotheses of Theorem* 4.1. *For $z'=(x',q')$ near z_* let $x(z')$ be a solution of VI(z') as given in* (3.9) *and let $q(z')$ be the associated multiplier. Then the function $q(z')$ is F-differentiable at z_* with $dq(z_*)=0$.*

Proof. As we already know that the point x comprising the other component of the solution is a function $x(z')$, we see from the comments preceding this theorem that the multiplier has the form

$$q(z')=-F(x(z'),z')NG(x(z'),z')^-. \tag{4.22}$$

The matrix $G(x(z'),z')^-$ is given by the formula

$$G(x,z')^-=B[G(x,z')B]^{-1}, \tag{4.23}$$

where B is an $(n+m)\times m$ matrix whose columns are a basis for the subspace M and where

$$G(x,z')=\begin{bmatrix}H_x(x,z') & -I\end{bmatrix}.$$

As x appears as an intermediate variable in (4.22), we can use the chain rule to express the F-derivative of q at z_* as

$$\mathrm{d}q(z_*) = q_x(x(z_*), z_*)dx(z_*) + q_z(x(z_*), z_*).$$

Under our hypotheses we have already shown that $dx(z_*) = 0$, so the first term will be zero at $z = z_*$. We also know that $x(z_*) = x_*$. Thus we need only find the partial derivative of q with respect to z, and then evaluate it at the pair (x_*, z_*).

Using (4.22) and (4.23), we find that

$$\begin{aligned} q_z(x, z')(w_z) &= -F_z(x, z')(w_z)NB[G(x, z')B]^{-1} \\ &\qquad - F(x, z')NB[G(x, z')B]^{-1}[-G_z(x, z')(w_z)B][G(x, z')B]^{-1}. \end{aligned} \tag{4.24}$$

We showed in (4.18) above that $F_z(x_*, z_*)(w_z) = -q_* h_{xx}(x_*)(w_x)$, so the value of the first term in (4.24) at $x = x_*$ and $z = z_*$ will be

$$q_* h_{xx}(x_*)(w_x)NG(x_*, z_*)^-. \tag{4.25}$$

Now use the formulas in (4.22) and (4.23) to rewrite the second term as

$$q(z')[-G_z(x, z')(w_z)B][G(x, z')B]^{-1}.$$

We have for any z' the formula

$$G(x, z') = \left[H_x(x, z') \ -I\right] = \left[h_x(x') \ -I\right],$$

so that

$$G_z(x, z')(w_z) = \left[h_{xx}(x')(w_x) \ 0\right] = h_{xx}(x')(w_x)N.$$

Using this in the expression for the second term we obtain

$$q(z')[-h_{xx}(x')(w_x)NB][G(x, z')B]^{-1} = -q(z')h_{xx}(x')(w_x)NG(x, z')^-,$$

so that at $x = x_*$ and $z = z_*$ the value of the second term will be

$$-q_* h_{xx}(x_*)(w_x)NG(x_*, z_*)^-. \tag{4.26}$$

Putting (4.25) and (4.26) into (4.24), we obtain $q_z(x_*, q_*) = 0$, and in light of our earlier comment this also shows that $dq(z_*) = 0$. □

The analysis of our linearization algorithm is now fairly simple, provided that we adopt the viewpoint that we proposed in [6]. That viewpoint consists in regarding the execution of the linearization algorithm as a simple iteration on the implicit function $x(z)$ identified in Theorem 4.1. The properties of that function, especially the zero B-derivatives at the point z_*, then yield the required properties for the sequence x_k of approximate solutions.

More formally, let a point $z_0 = (x_0, q_0) \in \mathbb{R}^{n+m}$ be given. For $n \geqslant 0$ let x_{n+1} solve the affine variational inequality

$$0 \in F(y, z_n) + N_{S(z_n)}(y), \tag{4.27}$$

where $F(y, z_n)$ and $S(z_n)$ are defined by (3.10) and (3.11). Let q_{n+1} be the multipliers q obtained from the expression

$$\begin{aligned} N_{S(z_n)}(x_{n+1}) = \{ & q h_x(x_n) + p \mid p \in N_P(x_{n+1}), \\ & q \in N_Q[h(x_n) + h_x(x_n)(x_{n+1} - x_n)]\}, \end{aligned} \tag{4.28}$$

and define z_{n+1} to be (x_{n+1}, q_{n+1}). The following theorem shows that if the problem has the regularity properties that we have discussed above, and if we choose z_0 close enough to z_*, then this procedure is well defined and in fact yields a sequence $\{z_n\}$ that converges Q-superlinearly to z_*.

THEOREM 4.3. *Assume the notation and hypotheses of Theorem* 4.1. *There is then a neighborhood Z' of z_* such that for any choice of z_0 in Z' the sequence $\{z_n\}$ is uniquely defined, remains in Z', and converges Q-superlinearly to z_*.*

Proof. We start with the neighborhoods Z_* and X_* produced by Theorem 4.1, and recall that if $z' \in Z_*$ then $x(z')$ is the unique solution in X_* of (3.9) for this choice of z'. Moreover, we find from [8, Theorem 5.6]smr:cnv that for some neighborhoods X_0 of x_* and Z_0 of Z_*, if $x \in X_0$ and $z' \in Z_0$ with x a solution of (3.9) for the parameter z', then the multipliers q in (4.28) are not only unique, but are given by the differentiable functions $q(z')$ in (4.22). Finally, as we have shown the B-derivatives $dx(z')$ and $dq(z')$ at $z' = z_*$ to be zero, for any positive ϵ we can find a neighborhood Z_ϵ of z_* such that if $z \in Z_\epsilon$ then

$$\|x(z') - x_*\| \leqslant \epsilon \|z' - z_*\|, \qquad \|q(z') - q_*\| \leqslant \epsilon \|z' - z_*\|, \tag{4.29}$$

where we use the Euclidean norm.

Now choose a positive ρ such that the closed ball $B(z_*, \rho)$ about z_* satisfies

$$B(z_*, \rho) \subset Z_* \cap Z_0 \cap x^{-1}(X_0) \cap Z_{\epsilon_0},$$

where $\epsilon_0 = 2^{-3/2}$. Consider any n for which $z_n \in B(z_*, \rho)$. Then the point x_{n+1} is well defined since $z_n \in Z_*$, x_{n+1} belongs to X_0 because $z_n \in x^{-1}(X_0)$, and q_{n+1} is well defined because $z_n \in Z_0$ and $x_{n+1} \in X_0$. Therefore the point $z_{n+1} = (x_{n+1}, q_{n+1})$ is well defined. Moreover, as $z_n \in Z_{\epsilon_0}$ we have

$$\begin{aligned} \|z_{n+1} - z_*\|^2 &= \|x_{n+1} - x_*\|^2 + \|q_{n+1} - q_*\|^2 \\ &\leqslant (2^{-3/2}\|z_n - z_*\|)^2 + (2^{-3/2}\|z_n - z_*\|)^2 \\ &= 2^{-2}\|z_n - z_*\|^2, \end{aligned} \tag{4.30}$$

so that

$$\|z_{n+1}-z_*\| \leqslant (1/2)\|z_n-z_*\|. \tag{4.31}$$

If we take z_0 to be in $B(z_*,\rho)$, then by the above argument z_1 is well defined with $\|z_1-z_*\| \leqslant (1/2)\|z_0-z_*\|$, so that $z_1 \in B(z_*,2^{-1}\rho)$. Continuing this argument, we see that the entire sequence $\{z_n\}$ is well defined with $z_n \in B(z_*,2^{-n}\rho)$, so that the z_n must converge to z_*. However, for any positive ϵ there is an N_ϵ such that for $n \geqslant N_\epsilon$ the points z_n all belong to Z_ϵ. Accordingly, for $n \geqslant N_\epsilon$ we have

$$\|z_{n+1}-z_*\| \leqslant \epsilon\|z_n-z_*\|,$$

which implies Q-superlinear convergence. □

We have therefore established that under the appropriate regularity assumptions this linearization method is locally Q-superlinearly convergent. In the final section we give a very small example to illustrate the application of this method, and we make some comments about possible extensions of the results given here.

5. Example

This section gives a very small and simple example to illustrate application of the method, and it adds some comments about possibly sharper or more complete results.

For the trivial example in $\mathbb{R}^2$ given by

$$0 \in x+a+N_S(x),$$

with $a=(-5,0)$ and

$$S=\{(x_1,x_2) \mid x_1-0.5x_2^2 \leqslant 0\},$$

the algorithm gave the sequence of points shown in Table I; the initial estimates are shown with $n=0$. Computations were done to single precision in a spreadsheet. The table clearly shows the superlinear convergence to the solution $x_*=(0,0)$, $q_*=5$.

As observed in Section 1, the focus in this paper has been on the way in which we can use the general analytical tools of [8] to gain conceptual simplicity in analyzing a solution procedure. Several aspects of this procedure remain to be investigated; these include its numerical behavior for problems of realistic size, the choice of numerical methods for solution of the subproblems, and the use of higher-order expansions to sharpen the convergence analysis, as well as how one might adapt it to situations in which constraint nondegeneracy does not hold. We hope to report on some of these areas in the future.

Table I. Solution of problem in $\mathbb{R}^2$

n	$(x_n)_1$	$(x_n)_2$	q_n
0	1.00E-1	1.00E-1	1.00E+1
1	5.56E-4	5.56E-2	5.00E+0
2	-1.55E-3	-2.94E-5	5.00E+0
3	-4.33E-10	-1.14E-8	5.00E+0
4	-4.33E-10	-1.14E-8	5.00E+0

Acknowledgements

The research reported here was sponsored by the U. S. Army Research Office under Grants DAAG55-97-1-0324 and DAAD19-01-1-0502. The U. S. Government has certain rights in this material, and is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon. The views and conclusions contained herein are those of the author and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the sponsoring agency or the U. S. Government.

References

1. Bonnans, J. F. and Shapiro, A. (2000), *Perturbation Analysis of Optimization Problems*, Springer Series in Operations Research, Springer, New York.
2. Facchinei, F. and Pang, J.-S. (2003), *Finite-Dimensional Variational Inequalities and Complementarity Problems*, Springer Series in Operations Research, Springer, New York, Published in two volumes, paginated continuously.
3. Ortega, J. M. and Rheinboldt, W. C. (1970), *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York.
4. Pshenichnyj, B. N. (1994), *The Linearization Method for Constrained Optimization*, Springer Series in Computational Mathematics, Springer, Berlin. Translated by Stephen S. Wilson from the Russian original, *Metod linearizatsii,* Nauka, Moscow, 1983.
5. Robinson, S. M. (1972), A quadratically-convergent algorithm for general nonlinear programming problems, *Mathematical Programming* 3, 145–156.
6. Robinson, S. M. (1974), Perturbed Kuhn–Tucker points and rates of convergence for a class of nonlinear-programming algorithms, *Mathematical Programming*, 7, 1–16.
7. Robinson, S. M. (1984), Local structure of feasible sets in nonlinear programming, Part II: Nondegeneracy, *Mathematical Programming Studies* 22, 217–230.
8. Robinson, S. M. (2003), Constraint nondegeneracy in variational analysis, *Mathematics of Operations Research* 28, 201–232.
9. Rockafellar, R. T. (1970), *Convex Analysis*, Princeton University Press, Princeton, NJ.
10. Rockafellar, R. T. and Wets, R. J. (1998), *Variational Analysis*, Number 317 in Grundlehren der mathematischen Wissenschaften. Springer, Berlin.
11. Shapiro, A., Sensitivity analysis of generalized equations, *Journal of Mathematical Sciences*, Forthcoming.
12. Wilson, R. B. (1963), *A Simplicial Method for Concave Programming*, Doctoral dissertation, Graduate School of Business Administration, Harvard University, Boston, MA.