Serdica

Mathematical Journal

Сердика

Математическо списание

Serdica Math. J. **22** (1996), 385-398

Serdica
Mathematical Journal

Institute of Mathematics
and Informatics, Bulgarian
Academy of Sciences

UNIFORM CONVERGENCE OF THE NEWTON METHOD FOR AUBIN CONTINUOUS MAPS*

Asen L. Dontchev

Communicated by S. L. Troyanski

Abstract. In this paper we prove that the Newton method applied to the generalized equation $y \in f(x) + F(x)$ with a C^1 function f and a set-valued map F acting in Banach spaces, is locally convergent uniformly in the parameter y if and only if the map $(f+F)^{-1}$ is Aubin continuous at the reference point. We also show that the Aubin continuity actually implies uniform Q-quadratic convergence provided that the derivative of f is Lipschitz continuous. As an application, we give a characterization of the uniform local Q-quadratic convergence of the sequential quadratic programming method applied to a perturbed nonlinear program.

This paper is about the Newton method for solving equations involving set-valued maps and parameters. Such "equations", commonly known as *generalized equations,* are of the form:

$$\text{(1)} \qquad \text{Find} \quad x \in X \quad \text{such that} \quad y \in f(x) + F(x),$$

where y is a parameter, f is a function and F is a map, possibly set-valued. Throughout X and Y are Banach spaces, $y \in Y$, $f : X \mapsto Y$ is C^1 on X and $F : X \mapsto 2^Y$ has closed graph. The generalized equation (1) is an abstract model for various problems

1991 *Mathematics Subject Classification:* 90C30, 47H04, 49M37, 65K10
Key words: generalized equation, Newton's method, sequential quadratic programming.
∗ This work was supported by National Science Foundation grant DMS 9404431.

including equations ($F = 0$), inequalities (F is the positive ortant in $\mathbb{R}^m$) and variational inequalities (F is the normal cone to a convex and closed set in X). In particular, (1) may represent first-order necessary conditions in optimization, e.g., the Karush-Kuhn-Tucker conditions in mathematical programming (see the discussion at the end of this paper) or the Pontryagin maximum principle in optimal control.

We study the following Newton-type method for solving (1): If x_k is the current iterate, the next iterate x_{k+1} is found from the relation

$$y \in f(x_k) + \nabla f(x_k)(x_{k+1} - x_k) + F(x_{k+1}), \qquad \text{for } k = 0, 1, \cdots, \tag{2}$$

where $\nabla f(x)$ denotes the Fréchet derivative of f at x and x_0 is a given starting point. For $F = \{0\}$ the procedure (2) is the classical Newton method for solving the equation $y = f(x)$. In the case when F is a (continuous) function, (2) is a Newton-type procedure using a partial linearization. If (1) represents the variational system associated with an optimization problem, then (2) is the corresponding version of the sequential quadratic programming method.

We use the following notation: All norms are denoted by $\| \cdot \|$. Defining the distance from a point $x \in X$ to a set $A \subset X$ as $\text{dist}(x, A) = \inf\{\|x - y\| : y \in A\}$, the excess e from the set A to the set C is given by $e(C, A) = \sup\{\text{dist}(x, A) : x \in C\}$. The inverse F^{-1} of a map F is defined as $F^{-1}(y) = \{x \in X : y \in F(x)\}$ and graph$F$ is the set $\{(x, y) \in X \times Y : y \in F(x)\}$. We denote by $B_a(x)$ the closed ball centered at x with radius a.

Recall that a set-valued map Γ from Y to the subsets of X is *Aubin continuous* at $(y_0, x_0) \in \text{graph}\Gamma$ with constants a, b and M if for every $y_1, y_2 \in B_b(y_0)$ and for every $x_1 \in \Gamma(y_1) \cap B_a(x_0)$ there exists an $x_2 \in \Gamma(y_2)$ with

$$\|x_1 - x_2\| \leq M \| y_1 - y_2 \| .$$

The constant M is called the modulus of Aubin continuity. The Aubin continuity of Γ is equivalent to the openness with linear rate of Γ^{-1} (the covering property) and to the metric regularity of Γ^{-1} (a basic well-posedness property in optimization). If f is a function which is strictly differentiable at some x_0, then the Aubin continuity of f^{-1} at $(f(x_0), x_0)$ is equivalent to the surjectivity of $\nabla f(x_0)$, by the Graves theorem [10] which is a stronger version of the Lyusternik theorem [13], for a discussion see [6]. Both Lyusternik and Graves theorems are consequences of the general observation that the Aubin property is "robust under (non)linearization". Namely, the following result was established in [7]: Let $f : X \to Y$ be a function which is strictly differentiable at x^*, let F be a set-valued map from X to the subsets of Y with closed graph, and let

$y^* \in f(x^*) + F(x^*)$. Then the Aubin continuity of the map $(f+F)^{-1}$ at (y^*, x^*) is equivalent to the Aubin continuity of $[f(x^*) + \nabla f(x^*)(\cdot - x^*) + F(\cdot)]^{-1}$ at (y^*, x^*).

There are several characterizations of the Aubin continuity in the literature. Rockafellar [18] showed that a closed-valued map Γ is Aubin continuous at a point $(x_0, y_0) \in \text{graph}\Gamma$ if and only if the distance function $\text{dist}(y, \Gamma(x))$ is Lipschitz continuous in both x and y around (x_0, y_0). (This result is proved in [18] in a finite-dimensional setting, but the proof can be easily extended to very general spaces). In finite dimensions, Mordukhovich [14] gave a characterization of the Aubin property in terms of a "coderivative". The Aubin continuity of maps with convex and closed graphs is characterized by the Robinson-Urcescu theorem. In particular, the Aubin continuity of the solution set of a system of equalities and inequalities in finite-dimensions is equivalent to the Mangasarian-Fromovitz condition. In a recent paper [9] we showed that the solution set of a nonlinear variational inequality over a convex polyhedral set in $\mathbb{R}^n$ is Aubin continuous if and only if it is locally single-valued and Lipschitz continuous (strongly regular in the sense of Robinson). This basically means that for such maps the (extended) Graves theorem is equivalent to the (extended) Robinson implicit function theorem (for a discussion of these extensions see [7]). We also gave in [9] a characterization of the Aubin continuity (or, equivalently, strong regularity) of the Karush-Kuhn-Tucker map in nonlinear programming.

In the last two decades, a number of papers have appeared dealing with Newton-type methods for nonsmooth equations and variational inequalities; for motivations and recent state-of-the-art works see e.g. Bonnans [3], Kummer [12], Pang [15], Qi [16], Robinson [17] and Xiao and Harker [19]. In particular, Kummer [11] gave a necessary and sufficient condition for superlinear convergence of the Newton method, which is mainly designed for derivative-type approximations of a nonsmooth function around an *isolated* zero. We also mention a recent paper by Azé and Chou [1] who considered a general inclusion problem in infinite dimensions, using a condition for a kind of derivative of a set-valued map which implies both the Aubin property and convergence of a Newton-type iterative procedure.

Our approach here is somewhat different; it rests on the idea that the Newton method can be used to prove open mapping and inverse function type theorems which is present already in the proofs of Lyusternik and Graves. In the main step of his original proof in [10], Graves used a perturbed version of the (modified) Newton method showing that the surjectivity of the derivative implies convergence of the Newton iterates which is *uniform* in a neighborhood of the reference point. Essentially the same procedure was also used by Lyusternik in [13] and a similar fact is contained in his proof. Here we develop further this idea identifying the Aubin property as a necessary and sufficient condition for a kind of convergence of the Newton method which is "stable" under perturbations (and thus more meaningful for practical computations). We also retain the original form of the problem as a generalized equation and in this way our results address directly the motivating applications, e.g. variational inequalities, without employing derivative-type approximations.

In a previous paper [4] we considered the generalized equation (1) without a parameter (e.g., $y = 0$) and proved that if x^* is a solution, then the Aubin continuity of the map $(f+F)^{-1}$ at $(0, x^*)$ is a sufficient condition for local Q-quadratic convergence of the Newton method (2). The paper [5] is a continuation of [4], where we report on results of similar type for approximate and nonsmooth versions of the Newton method as well as a Kantorovich-type theorem. In the present paper we give, generally speaking, an answer to the question what kind of "well-posedness" of the Newton method would correspond to the well-posedness represented by the Aubin property. First, we consider the perturbed generalized equation (1) and prove (Theorem 1) that the Aubin property of $(f+F)^{-1}$ at (y^*, x^*) is equivalent to a kind of convergence of the Newton method which is uniform in the sense that the attraction region does not depend on small variations of the value of the parameter y near y^* and for such values of y the method finds a solution x which is at distance from x^* proportional to the variation of y. Then we show in Theorem 2 that the Aubin property actually implies uniform Q-quadratic convergence, provided that the derivative of f is Lipschitz continuous. This is a generalization of the result of [4]. As an illustration of the the results obtained we consider a parametric nonlinear program, obtaining a characterization of the uniform local convergence of the sequential quadratic programming method.

The main result of the paper follows.

Theorem 1. *Let x^* be a solution of* (1) *for $y = 0$, let f be a function which is Fréchet differentiable in an open neighborhood $\mathcal{O}$ of x^*, and let its derivative ∇f be continuous in $\mathcal{O}$. Let F have closed graph. Then the following are equivalent:*

(i) *The map $(f+F)^{-1}$ is Aubin continuous at $(0, x^*)$;*

(ii) *There exist positive constants σ, b and c such that for every $y \in B_b(0)$ and for every $x_0 \in B_\sigma(x^*)$ there exists a Newton sequence x_k starting from x_0 which is convergent to a solution x of (1) for y, moreover, if x_0 is a solution of (1) for y_0, then the limit x satisfies $\|x - x_0\| \le c\|y - y_0\|$.*

The implication (ii) $\Rightarrow$ (i) follows directly from the definition of Aubin continuity. In the proof of (i) $\Rightarrow$ (ii) we use the following lemma:

Lemma 1. *Let $(x^*, y^*) \in graph(f+F)$, let f be a function which is Fréchet differentiable in an open neighborhood of x^*, let its derivative ∇f be continuous at x^*,* *and let F have closed graph. Suppose that the map $(f+F)^{-1}$ is Aubin continuous at (y^*,x^*). Then there exist positive constants α,β and M such that for every $x\in B_\alpha(x^*)$, if $P_x=[f(x)+\nabla f(x)(\cdot-x)+F(\cdot)]^{-1}$, then*

$$e(P_x(y')\cap B_\alpha(x^*),P_x(y''))\le M\|y'-y''\|$$

for every $y',y''\in B_\beta(y^*)$.

Proof. From a result in [7], the map $Q=[f(x^*)+\nabla f(x^*)(\cdot-x^*)+F(\cdot)]^{-1}$ is Aubin continuous at (y^*,x^*); let a,b and M' be the associated constants. Choose $\varepsilon>0$ such that $M'\varepsilon<1$ and $\alpha>0$ such that $\|\nabla f(x)-\nabla f(x^*)\|\le\varepsilon$ for every $x\in B_\alpha(x^*)$. Take $\alpha>0$ smaller if necessary so that $2\alpha\le a$ and $4\varepsilon\alpha<b$. Further, choose $\beta>0$ such that

$$\beta+4\varepsilon\alpha\le b\quad\text{and}\quad\frac{2M'\beta}{1-M'\varepsilon}\le\alpha. \tag{3}$$

Let $x\in B_\alpha(x^*)$, let $y',y''\in B_\beta(y^*)$ and let

$$x'\in P_x(y')\cap B_\alpha(x^*)=[f(x)+\nabla f(x)(\cdot-x)+F(\cdot)]^{-1}(y')\cap B_\alpha(x^*).$$

Denote $x_1=x'$. Then

$$x_1\in Q(y'-f(x)-\nabla f(x)(x_1-x)+f(x^*)+\nabla f(x^*)(x_1-x^*))\cap B_a(x^*)$$

and

$$\|x-x_1\|\le\|x-x^*\|+\|x^*-x_1\|\le 2\alpha.$$

Using (3) we obtain

$$\begin{aligned}&\|y'-f(x)-\nabla f(x)(x_1-x)+f(x^*)+\nabla f(x^*)(x_1-x^*)-y^*\|\\&\le\|y'-y^*\|+\|f(x)-f(x^*)-\nabla f(x^*)(x-x^*)\|\\&\quad+\|(\nabla f(x^*)-\nabla f(x))(x-x_1)\|\\&\le\beta+\varepsilon(\|x-x^*\|+\|x-x_1\|)\le\beta+3\varepsilon\alpha\le b.\end{aligned}$$

This same inequality holds for y''. From these estimates and from the Aubin continuity of Q we obtain that there exists an

$$x_2\in Q(y''-f(x)-\nabla f(x)(x_1-x)+f(x^*)+\nabla f(x^*)(x_1-x^*));$$

that is,

$$y''\in f(x)+\nabla f(x)(x_1-x)+\nabla f(x^*)(x_2-x_1)+F(x_2),$$

and such that

$$\|x_2-x_1\|\le M'\|y'-y''\|.$$

Proceeding by induction, suppose that there exist an integer $n > 2$ and points $x_2, x_3, \cdots, x_n$ with

$$y'' \in f(x) + \nabla f(x)(x_{i-1} - x) + \nabla f(x^*)(x_i - x_{i-1}) + F(x_i),$$

and

$$\|x_i - x_{i-1}\| \le M'\|y' - y''\|(M'\varepsilon)^{i-2}, \quad i = 3, 4, \cdots, n.$$

Then

$$\begin{aligned} \|x_n - x^*\| &\le \sum_{j=2}^{n} \|x_j - x_{j-1}\| + \|x_1 - x^*\| \\ &\le 2M'\beta \sum_{j=2}^{n} (M'\varepsilon)^j + \alpha \\ &\le \frac{2M'\beta}{1 - M'\varepsilon} + \alpha \le 2\alpha \le a \end{aligned}$$

and

$$x_n \in Q(y'' - f(x) - \nabla f(x)(x_{n-1} - x) + f(x^*) + \nabla f(x^*)(x_{n-1} - x^*)) \cap B_a(x^*).$$

Taking into account that

$$\begin{aligned} \|x_n - x\| &\le \sum_{j=2}^{n} \|x_j - x_{j-1}\| + \|x_1 - x\| \\ &\le \frac{2M'\beta}{1 - M'\varepsilon} + 2\alpha \le 3\alpha, \end{aligned}$$

we obtain for both $y = y'$ and $y = y''$,

$$\|y - f(x) - \nabla f(x)(x_n - x) + f(x^*) + \nabla f(x^*)(x_n - x^*) - y^*\| \le \beta + 4\varepsilon\alpha \le b.$$

Then there exists an

$$x_{n+1} \in Q(y'' - f(x) - \nabla f(x)(x_n - x) + f(x^*) + \nabla f(x^*)(x_n - x^*));$$

that is,

$$y'' \in f(x) + \nabla f(x)(x_{n+1} - x) + \nabla f(x^*)(x_{n+1} - x_n) + F(x_{n+1}), \tag{4}$$

and such that

$$\begin{aligned} \|x_{n+1} - x_n\| &\le M'\|(\nabla f(x^*) - \nabla f(x))(x_n - x_{n-1})\| \\ &\le M'\varepsilon\|x_n - x_{n-1}\| \le M'\|y' - y''\|(M'\varepsilon)^n. \end{aligned}$$

The induction step is complete. Thus x_n is a Cauchy sequence, hence there exists x'' such that $x_n \to x''$ as $n \to \infty$. Moreover, passing to the limit in (4), we get

$$x'' \in P_x(y'') = [f(x) + \nabla f(x)(\cdot - x) + F(\cdot)]^{-1}(y'')$$

and

$$\begin{aligned} \|x' - x''\| &\leq \limsup_{n\to\infty} \sum_{i=2}^{n} \|x_i - x_{i-1}\| \\ &\leq \limsup_{n\to\infty} \sum_{i=2}^{n} (M'\varepsilon)^{i-2} M' \|y' - y''\| \\ &\leq \frac{M'}{1 - M'\varepsilon} \|y' - y''\|. \end{aligned}$$

Hence, the lemma holds with $M = M'/(1 - M'\varepsilon)$. □

Proof of Theorem 1. Let α, β and M be the constants in Lemma 1 and let $P_x = [f(x) + \nabla f(x)(\cdot - x) + F(\cdot)]^{-1}$ for $x \in B_\alpha(x^*)$. Let $\varepsilon > 0$ satisfy $M\varepsilon < 1$ and choose $a > 0$ such that $B_a(x^*) \subset \mathcal{O}$ and $\|\nabla f(x') - \nabla f(x'')\| \leq \varepsilon$ whenever $x', x'' \in B_a(x^*)$. Choose $\sigma > 0$ such that

$$\sigma \leq \alpha, \ \frac{2\sigma}{1 - M\varepsilon} < a \ \text{ and } \ 2\varepsilon\sigma < \beta,$$

and let $b > 0$ satisfy

$$b(1 + M\varepsilon) + 2\varepsilon\sigma \leq \beta \ \text{ and } \ \frac{Mb + 2\sigma}{1 - M\varepsilon} \leq a.$$

Let $x_0 \in B_\sigma(x^*)$. Then

$$x^* \in P_{x_0}(-f(x^*) + f(x_0) + \nabla f(x_0)(x^* - x_0)) \cap B_\alpha(x_0).$$

Further,

$$\begin{aligned} &\|f(x^*) - f(x_0) - \nabla f(x_0)(x^* - x_0)\| \\ &\quad \leq \|\int_0^1 (\nabla f(x^* + t(x_0 - x^*)) - \nabla f(x_0))(x_0 - x^*)dt\| \\ &\quad \leq \varepsilon\|x_0 - x^*\| \leq \varepsilon\sigma \leq \beta. \end{aligned}$$

Let $y \in B_b(0)$. From Lemma 1 it follows that there exists $x_1 \in P_{x_0}(y)$, i.e.,

$$y \in f(x_0) + \nabla f(x_0)(x_1 - x_0) + F(x_1),$$

such that

$$\begin{aligned}\|x_1 - x^*\| &\le M\|y + f(x^*) - f(x_0) - \nabla f(x_0)(x^* - x_0)\| \\ &\le M(\|y\| + \varepsilon\|x_0 - x^*\|) \le Mb + M\varepsilon\sigma \le Mb + \sigma \le a.\end{aligned}$$

Then

$$\|x_1 - x_0\| \le \|x_1 - x^*\| + \|x^* - x_0\| \le Mb + M\varepsilon\sigma + \sigma. \tag{5}$$

Note that

$$x_1 \in P_{x_1}(y + f(x_1) - f(x_0) - \nabla f(x_0)(x_1 - x_0)) \cap B_\alpha(x_1)$$

and

$$\begin{aligned}&\|y + f(x_1) - f(x_0) - \nabla f(x_0)(x_1 - x_0)\| \\ &\le b + \varepsilon\|x_1 - x_0\| \le b + \varepsilon(Mb + M\varepsilon\sigma + \sigma) \le \beta.\end{aligned}$$

Then from Lemma 1 there exists an

$$x_2 \in P_{x_1}(y) = [f(x_1) + \nabla f(x_1)(\cdot - x_1) + F(\cdot)]^{-1}(y)$$

such that

$$\|x_2 - x_1\| \le M\|f(x_1) - f(x_0) - \nabla f(x_0)(x_1 - x_0)\| \le M\varepsilon\|x_1 - x_0\|.$$

Further,

$$\begin{aligned}\|x_2 - x^*\| &\le \|x_2 - x_1\| + \|x_1 - x_0\| + \|x_0 - x^*\| \\ &\le (1 + M\varepsilon)\|x_1 - x_0\| + \|x_0 - x^*\| \\ &\le \frac{Mb + M\varepsilon\sigma + \sigma}{1 - M\varepsilon} + \sigma \le \frac{Mb + 2\sigma}{1 - M\varepsilon} \le a.\end{aligned} \tag{6}$$

Suppose that for some integer $n > 2$ the points $x_2, x_3, \cdots, x_n$ are obtained by the Newton method (2), that is, $x_i \in P_{x_{i-1}}(y)$, and

$$\|x_i - x_{i-1}\| \le (M\varepsilon)^{i-1}\|x_1 - x_0\| \quad i = 3, 4, \cdots, n.$$

Then, by repeating the argument in (6) we obtain that $x_i \in B_a(x^*)$. Further, we have

$$\begin{aligned}&\|y + f(x_n) - f(x_{n-1}) - \nabla f(x_{n-1})(x_n - x_{n-1})\| \\ &\le b + \varepsilon\|x_n - x_{n-1}\| \le b + \varepsilon\|x_1 - x_0\| \le \beta.\end{aligned}$$

Then from

$$x_n \in P_{x_n}(y + f(x_n) - f(x_{n-1}) - \nabla f(x_{n-1})(x_n - x_{n-1})) \cap B_\alpha(x_n)$$

and from Lemma 1 we conclude that there exists a Newton iterate

$$x_{n+1} \in P_{x_n}(y) = [f(x_n) + \nabla f(x_n)(\cdot - x_n) + F(\cdot)]^{-1}(y) \tag{7}$$

satisfying

$$\|x_{n+1} - x_n\| \leq M\varepsilon\|x_n - x_{n-1}\| \leq (M\varepsilon)^n\|x_1 - x_0\|. \tag{8}$$

This completes the induction step. Hence, there exists a Newton sequence x_n which is a Cauchy sequence, and, passing to the limit in (7), we obtain that x_n is geometrically convergent to a solution $x \in (f+F)^{-1}(y)$.

Let $y_0 \in B_b(0)$ and $x_0 \in (f+F)^{-1}(y_0) \cap B_\sigma(x^*)$. Then $x_0 \in P_{x_0}(y_0) \cap B_\alpha(x_0)$. From Lemma 1 we obtain that there exists $x_1 \in P_{x_0}(y)$ such that $\|x_1 - x_0\| \leq M\|y - y_0\|$. By repeating the argument between (5) and (8) we obtain a Newton sequence x_n satisfying (7) and (8) and convergent to a solution $x \in (f+F)^{-1}(y)$. Moreover,

$$\begin{aligned}\|x_n - x_0\| &\leq \sum_{i=1}^{n}\|x_i - x_{i-1}\| \leq \sum_{i=1}^{n}(M\varepsilon)^i M\|y - y_0\| \\ &\leq \frac{M}{1 - M\varepsilon}\|y - y_0\|.\end{aligned}$$

Passing to the limit with n and taking $c = M/(1 - M\varepsilon)$ we complete the proof. □

In the following theorem we show that if the derivative of f is Lipschitz continuous around x^*, then the Aubin continuity implies the existence of a Q-quadratically convergent Newton sequence.

Theorem 2. *Let x^* be a solution of* (1) *for $y = 0$, let f be a function which is Fréchet differentiable in an open neighborhood $\mathcal{O}$ of x^*, and let its derivative ∇f be Lipschitz continuous in $\mathcal{O}$. Let F have closed graph and let the map $(f+F)^{-1}$ be Aubin continuous at $(0, x^*)$. Then there exist positive constants σ, b and γ such that for every $y \in B_b(0)$ and for every $x_0 \in B_\sigma(x^*)$ there exists a Newton sequence x_k starting from x_0 which is Q-quadratically convergent with a constant γ to a solution x of* (1) *for y, that is,*

$$\|x_{k+1} - x\| \leq \gamma\|x_k - x\|^2, \quad k = 0, 1, 2, \cdots.$$

Proof. Let $(f+F)^{-1}$ be Aubin continuous at $(0, x^*)$ with modulus c. Then, from the very definition, there exists $\delta > 0$ such that for every $y \in B_\delta(0)$ there exists $x \in (f+F)^{-1}(y) \cap B_{c\|y\|}(x^*)$. To prove the existence of a Newton sequence which convergence quadratically to x uniformly in y we use induction as in the proof of Theorem 1, but the idea is different. Let α, β and M be the constant in the statement of Lemma 1 and let L be the Lipschitz constant of ∇f in $\mathcal{O}$. Choose positive σ and b such that $B_{cb}(x^*) \subset \mathcal{O}$,

$$\sigma \le \alpha/2, \quad b \le \min\{\beta/2, \delta\}, \quad cb \le \alpha/2,$$

$$L(cb+\sigma)^2 \le \beta, \quad ML(cb+\sigma)^2 \le \alpha, \quad \frac{1}{2}ML(cb+\sigma) \le 1.$$

Let $x_0 \in B_\sigma(x^*), y \in B_b(0)$, and let $x \in (f+F)^{-1}(y) \cap B_{c\|y\|}(x^*)$. Then $\|x - x^*\| \le cb \le \alpha$. Note that

$$x \in P_{x_0}(y - f(x) + f(x_0) + \nabla f(x_0)(x - x_0)) \cap B_\alpha(x^*)$$

and

$$\begin{aligned}&\|y - f(x) + f(x_0) + \nabla f(x_0)(x - x_0)\| \\ &\le b + \frac{1}{2}L\|x - x_0\|^2 \le b + \frac{1}{2}L(cb+\sigma)^2 \le \beta.\end{aligned}$$

Hence, from Lemma 1 there exists an $x_1 \in P_{x_0}(y)$, that is,

$$y \in f(x_1) + \nabla f(x_0)(x_1 - x_0) + F(x_1),$$

such that

$$\|x - x_1\| \le M\|f(x) - f(x_0) - \nabla f(x_0)(x - x_0)\| \le \frac{1}{2}ML\|x - x_0\|^2.$$

Then

$$\begin{aligned}\|x_1 - x^*\| &\le \|x_1 - x\| + \|x - x^*\| \\ &\le \frac{1}{2}ML\|x - x_0\|^2 + cb \\ &\le \frac{1}{2}ML(cb+\sigma)^2 + cb \le \alpha.\end{aligned}$$

Further,

$$x \in P_{x_1}(y - f(x) + f(x_1) + \nabla f(x_1)(x - x_1)) \cap B_\alpha(x^*),$$

and from

$$\|x - x_1\| \le \frac{1}{2}ML(cb+\sigma)^2 \le \alpha$$

we have

$$\begin{aligned}&\|y - f(x) + f(x_1) + \nabla f(x_1)(x - x_1)\| \\ &\le b + \frac{1}{2}L\|x - x_1\|^2 \le b + \frac{1}{2}L(\frac{1}{2}ML(cb+\sigma)^2)^2 \le \beta.\end{aligned}$$

Then there exists a Newton iterate $x_2 \in P_{x_1}(y)$ with

$$\|x_2 - x\| \leq M\|f(x) - f(x_1) - \nabla f(x_1)(x - x_1)\| \leq \frac{1}{2}ML\|x_1 - x\|^2.$$

The induction step is fairly obvious. Taking $\gamma \geq ML/2$ we are done. □

By assuming that the function f is merely continuously differentiable near x^*, one can obtain an analogous result where the uniform Q-quadratic convergence is replaced by the weaker uniform superlinear convergence. Note that the argument we use is purely metric and goes through for more general spaces, e.g. X complete metric and Y Fréchet. Also, note that the Newton sequence in Theorem 1 may be different from the one obtained in Theorem 2. By replacing the Aubin continuity with the stronger requirement the map $(f + F)^{-1}$ be locally single-valued and Lipschitz, one can obtain analogous results where for every starting point there exists a unique Newton sequence.

As an illustration of our general results we consider the following nonlinear programming problem with canonical perturbations:

$$\text{minimize } g_0(w, x) + \langle v, x \rangle \text{ in } x \text{ subject to} \tag{9}$$

$$g_i(w, x) - u_i \begin{cases} = 0 & \text{for } i \in [1, r], \\ \leq 0 & \text{for } i \in [r + 1, m], \end{cases}$$

where the functions $g_i : \mathbb{R}^d \times \mathbb{R}^n \to \mathbb{R}$, $i = 0, 1, \ldots, m$ are twice continuously differentiable and the vectors $w \in \mathbb{R}^d$, $v \in \mathbb{R}^n$ and $u = (u_1, \ldots, u_m) \in \mathbb{R}^m$ are parameters. Denote $p = (v, u, w)$. Introducing the Lagrangian

$$L(w, x, \lambda) = g_0(w, x) + \lambda_1 g_1(w, x) + \cdots + \lambda_m g_m(w, x),$$

the Karush-Kuhn-Tucker (KKT) conditions for (9) are of the form

$$\begin{cases} \quad v + \nabla_x L(w, x, \lambda) = 0, \\ -u + \nabla_\lambda L(w, x, \lambda) \in N_D(\lambda) \quad \text{for } \; D = \mathbb{R}^r \times \mathbb{R}^{m-r}_+, \end{cases} \tag{10}$$

where $N_D(\lambda)$ is the normal cone to the set D at the point λ. These can be written as the variational inequality

$$0 \in (v, u) + f(w, x, \lambda) + N_C(x, \lambda), \tag{11}$$

where

$$f(w, x, \lambda) = (\nabla_x L(w, x, \lambda), -\nabla_\lambda L(w, x, \lambda)), \qquad C = \mathbb{R}^n \times D.$$

The sequential quadratic programming (SQP) algorithm for solving (9), for a recent survey see [2], is a direct application of the Newton method to the variational inequality

(11). Since the linearization of (11) is a KKT system of a quadratic programming problem, the Newton iteration can be obtained in the following way: If (x_k, λ_k) denotes the current iterate of the SQP algorithm, then the new iterate (x_{k+1}, λ_{k+1}) is obtained by computing x_{k+1} as a local minimizer of the following quadratic program:

$$\text{(12)} \qquad \text{minimize } \frac{1}{2}\langle B_k(x - x_k), x - x_k\rangle + \langle v + g_0(x_k, w), x - x_k\rangle$$

subject to

$$-u_i + g_i(w, x_k) + \nabla_x g_i(w, x_k)(x - x_k) \begin{cases} = 0 & \text{for } i \in [1, r], \\ \leq 0 & \text{for } i \in [r+1, m], \end{cases}$$

where B_k is the Hessian of the Lagrangian evaluated at $(x_k, \lambda_k, v, u, w)$. The new λ_{k+1} is the multiplier associated with x_{k+1} for (12).

Let (x^*, λ^*) satisfy the KKT conditions (10) or equivalently the variational inequality (11) for given u^*, v^*, and w^*, and let the index sets I_1, I_2 in $\{1, 2, \ldots, m\}$ be defined as:

$$\begin{aligned} I_1 &= \{i \in [r+1, m] \,|\, g_i(w^*, x^*) - u_i^* = 0,\ \lambda_i^* > 0\} \cup \{1, \ldots, r\}, \\ I_2 &= \{i \in [r+1, m] \,|\, g_i(w^*, x^*) - u_i^* = 0,\ \lambda_i^* = 0\}. \end{aligned}$$

By combining the results obtained in the present paper with Theorems 3 and 6 in [9] we obtain the following corollary:

Corollary 1. *Let x^* be a solution of (9) for $p^* = (v^*, u^*, w^*)$ with an associate multiplier λ^*. Then the following are equivalent:*

(i) *There exist constants σ, β, γ and c such that for every $p = (v, u, w) \in B_\beta(p^*)$ and for every initial point $(x_0, \lambda_0) \in B_\sigma(x^*, \lambda^*)$ there exists a unique sequence of iterates (x_k, λ_k) of the SQP algorithm which is Q-quadratically convergent with the constant γ to a point (x, λ) where x is a solution of* (9) *for p and λ is an associated Lagrange multiplier, moreover, with the property that if $z_0 = (x_0, \lambda_0)$ is a primal-dual pair for* (9) *for some p_0, then the limit $z = (x, \lambda)$ satisfies $\|z - z_0\| \leq c\|p - p_0\|$;*

(ii) *The constraint gradients $\nabla_x g_i(w^*, x^*)$ for $i \in I_1 \cup I_2$ are linearly independent and the strong second-order sufficient condition for local optimality holds for $(u^*, v^*, w^*, x^*, \lambda^*)$; that is,*

$$\langle x', \nabla^2_{xx} L(w^*, x^*, \lambda^*) x'\rangle > 0 \text{ for all } x' \neq 0 \quad \text{in the subspace}$$

$$M = \{x' \,|\, x' \perp \nabla_x g_i(w^*, x^*) \text{ for all } i \in I_1\}.$$

The general theory presented here can be also utilized for convergence analysis of Newton-type methods applied to discrete approximations in optimal control. Without going into this further, we only mention that if the SQP algorithm is applied to a problem obtained by a discretization of a nonlinear optimal control problem, then, under standard constraint qualification and second-order sufficient conditions, e.g., as in [8], the algorithm is locally superlinearly/quadratically convergent uniformly in the choice of the starting point and the size of the discretization grid.

REFERENCES

[1] D. Azé, C. C. Chou. On a Newton type iterative method for solving inclusions. *Math. Oper. Res.* **20** (1995), 790-800.

[2] P. T. Boggs, J. W. Tolle. Sequential quadratic programming. *Acta Numerica*, Cambridge Univ. Press, 1995, pp. 1-51.

[3] J. F. Bonnans. Local analysis of Newton-type methods for variational inequalities and nonlinear programming. *Appl. Math. Optim.*, **19** (1994), 161-186.

[4] A. L. Dontchev. Local convergence of the Newton method for generalized equations. *C. R. Acad. Sc. Paris Sér I Math.* **322** (1996), 327-331.

[5] A. L. Dontchev. Local analysis of a Newton-type method based on partial linearization. In Proc. of the Summer Seminar "Mathematics of Numerical Analysis: Real Number Algorithms", Park City, UT July 17-August 11, 1995, J. Renegar, M. Shub and S. Smale (eds.), to appear.

[6] A. L. Dontchev. The Graves theorem revisited. *J. Convex Analysis* **3** (1996), 1-9, to appear.

[7] A. L. Dontchev, W. W. Hager. An inverse function theorem for set-valued maps. *Proc. Amer. Math. Soc.* **121** (1994), 481-489.

[8] A. L. Dontchev, W. W. Hager, A. B. Poore, B. Yang. Optimality, stability and convergence in nonlinear control. *Appl. Math. Optim.* **31** (1995), 297-326.

[9] A. L. Dontchev, R. T. Rockafellar. Characterizations of strong regularity for variational inequalities over polyhedral convex sets. *SIAM J. Optim.*, to appear.

[10] L. M. Graves. Some mapping theorems. *Duke Math. J.* **17** (1950), 111-114.

[11] B. Kummer. Newton's method based on generalized derivatives for nonsmooth functions: convergence analysis. In Systems Modelling and Optimization, P. Kall (ed.), Lecture Notes in Control and Inform. Sci. vol. **180**, Springer 1992, pp. 3-16.

[12] B. Kummer. Approximations of multifunctions and superlinear convergence. In Recent Developments in Optimization, R. Durier and C. Michelot (eds.), Lecture Notes in Econom. and Math. Systems vol. **429**, Springer 1995, pp. 243-251.

[13] L. A. Lyusternik. On the conditional extrema of functionals. *Math. Sbornik* **41** (1934), 390-401.

[14] B. Mordukhovich. Stability theory for parametric generalized equations and variational inequalities via nonsmooth analysis. *Trans. Amer. Math. Soc.* **343** (1994), 609-657.

[15] J.-S. Pang. Convergence of splitting and Newton methods for complementarity problems: an application of some sensitivity results. *Math. Programming* **58** (1993), 149-160.

[16] L. Qi. Convergence analysis of some algorithms for solving non-smooth equations. *Math. Oper. Res.* **18** (1993), 227-244.

[17] S. M. Robinson. Newton's method for a class of nonsmooth functions. *Set-Valued Analysis* **2** (1994), 291-305.

[18] R. T. Rockafellar. Lipschitz properties of multifunctions. *Nonlinear Anal. TMA* **9** (1985), 867-885.

[19] B. Xiao, P. T. Harker. A nonsmooth Newton method for variational inequalities, I: Theory. *Math. Programming* **65** (1994), 151-194.

Mathematical Reviews
416 Fourth Street
Ann Arbor, MI 48107
USA
e-mail: `ald@math.ams.org`

Received March 31, 1996
Revised July 12, 1996