On the Newton method for set-valued maps

S. Dias[a], G. Smirnov[b]

[a] *Department of Mathematics, School of Sciences, Oporto University, Rua do Campo Alegre, 687 4169-007 Porto, Portugal*
[b] *Centre of Physics, Department of Mathematics and Applications University of Minho, Campus de Gualtar, 4710-057 Braga, Portugal*

Abstract

The Newton method is one of the most powerful tools used to solve systems of nonlinear equations. Its set-valued generalization, considered in this work, allows one to solve also nonlinear equations with geometric constraints and systems of inequalities in a unified manner. The emphasis is given to systems of linear inequalities. The study of the well-posedness of the algorithm and of its convergence is fulfilled in the framework of modern variational analysis.

Key words: Newton's method, Set-valued maps, Variational analysis
2000 MSC: 49J53, 65K15, 90C30

1. Introduction

The Newton method is one of the most powerful tools used to solve systems of nonlinear equations

$$f(x) = 0, \tag{1}$$

where $f : R^n \to R^n$ is a continuously differentiable map [15]. The method generates the following sequence of points

$$x^{k+1} = x^k + \bar{x}^k, \quad k = 0, 1, \dots, \tag{2}$$

where $\bar{x}^k$ is a solution to the system of linear equations

$$\nabla f(x^k)\bar{x}^k = -f(x^k). \tag{3}$$

Email addresses: `valentina.susana@gmail.com` (S. Dias), `smirnov@math.uminho.pt` (G. Smirnov)

Preprint submitted to Nonlinear Analysis *March 23, 2011*

In this paper we study Newton-type methods suitable to solve inclusions

$$0 \in F(x), \tag{4}$$

where $F : R^n \rightrightarrows R^n$ is a set-valued map. The methods generate sequences of points (2), where $\bar{x}^k$ is a solution to an inclusion, a generalization of linear system (3). This generalization involves set-valued map derivatives [2]. Namely, instead of (3) we solve an inclusion

$$-v^k \in \Lambda(x^k, v^k)(\bar{x}^k), \tag{5}$$

where v^k is a nearest to zero point belonging to the set $F(x^k)$, and the graph of the set-valued map $\Lambda(x^k, v^k)(\cdot)$ is a cone tangent in some sense to the graph of F at the point (x^k, v^k). The condition of non-singularity of the matrix $\nabla f(x^k)$, essential to solve linear system (3) and to prove convergence theorems, is substituted by the condition of metric regularity [14, 18] (see also the survey [3]). This property in the case of linear operators goes back to the open mapping theorem and in the case of smooth maps to the Lyusternik theorem. It was successfully applied to justify the well-posedness and convergence of the Newton method for nonsmooth equations [11]. In this paper we show that the metric regularity of Lipschitzian set-valued maps is equivalent to the well-posedness of a Newton method for perturbed maps.

Similar issues are discussed in [6, 8, 9, 10, 12, 13] for so-called generalized equations

$$y \in f(x) + F(x). \tag{6}$$

For example, in [8] the equivalence of the Aubin continuity of the map $(f + F)^{-1}$ and the existence of Newton sequences defined by

$$y \in f(x^k) + \nabla f(x^k)(x^{k+1} - x^k) + F(x^{k+1})$$

and converging to a solution of (6), is established. Note that the extension of Newton's method to generalized equations operates with linearization of the smooth function f while leaving F untouched. Our approach involves "linearization" of F and is close to the one from [4], where the derivative of F is defined via the Clarke tangent cone [7] and the Newton method is applied to prove an open mapping theorem for set-valued maps. A continuous version of Newton's method involving set-valued map derivatives can be found in [19] (see also [1]).

The Newton-type method developed in this paper is a general tool suitable to solve in a unified manner systems of nonlinear equations with geometric constraints and systems of nonlinear inequalities (see Sec. 5). It also solves (exactly) a generic system of linear inequalities in a finite number of iterations (see Sec. 6).

The paper is organized in the following way. In the second section we briefly review some constructions and results from set-valued and variational analysis. In the third section, for Lipschitzian set-valued maps, we give a characterization of metric regularity in terms of the Newton method well-posedness. The rate of convergence of Newton's method for set-valued maps is studied in Sec. 4. In the fifth section we apply this method to a system of equations and inequalities. Finally, in the last section the Newton method is applied to a system of linear inequalities.

2. Set-valued derivatives and metric regularity

Throughout this paper we denote by R^n the real n-dimensional space and by $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$ the usual inner product and Euclidean norm, respectively. We use the notation $B_n = \{x \in \mathbb{R}^n \mid \|x\| \leq 1\}$ for the closed unit ball in R^n. The convex hull and the closure of a subset $C \subset R^n$ are denoted by $\text{co}C$ and $\text{cl}C$, respectively. The distance between a point x and the set C is denoted by $d(x, C) = \inf\{\|x - c\| \mid c \in C\}$. The projection of a vector x onto C is defined by $\pi(x, C) = \{c \in C \mid \|x - c\| = d(x, C)\}$. Let A be a matrix. Its transposed is denoted by A^T.

Recall some basic definitions from set-valued and variational analysis. Let $F : R^n \rightrightarrows R^m$ be a set-valued map. Its graph is denoted $\text{gph}F$ and is defined by

$$\text{gph}F = \{(x, v) \in R^n \times R^m \mid v \in F(x)\}.$$

The inverse map $F^{-1} : R^m \rightrightarrows R^n$ is defined by

$$F^{-1}(v) = \{x \in R^n \mid (x, v) \in \text{gph}F\}.$$

We say that F is Lipschitzian if there exists $L \geq 0$ such that

$$F(x_1) \subset F(x_2) + L\|x_1 - x_2\|B_m,$$

for all $x_1 \in R^n$ and $x_2 \in R^n$. A set-valued map F is said to be locally Lipschitzian if for any $x \in R^n$ there exist $\epsilon > 0$ and $L > 0$ such that

$$F(x_1) \subset F(x_2) + L\|x_1 - x_2\|B_m$$

for all $x_1, x_2 \in x + \epsilon B_n$.

The contingent cone to a set $C \subset R^n$ at a point $x \in C$ is defined by

$$T(x, C) = \{v \mid \liminf_{\lambda \downarrow 0} \lambda^{-1} d(x + \lambda v, C) = 0\}.$$

Denote by

$$\hat{N}(x, C) = \{v^* \in R^n \mid \langle v^*, v \rangle \leq 0, \ \ v \in T(x, C)\}$$

the polar cone to C at x. If $x \notin C$, we put $\hat{N}(x, C) = \emptyset$. The Mordukhovich normal cone to C at x [14, 18] is defined by

$$N(x, C) = \operatorname*{Limsup}_{x' \to x} \hat{N}(x', C),$$

where the upper limit of a set-valued map $F : R^n \rightrightarrows R^m$ is given by

$$\operatorname*{Limsup}_{x' \to x} F(x') = \{v = \lim_{n \to \infty} v_k \mid (x_k, v_k) \in \mathrm{gph}F, \ \ x_k \to x\}.$$

The set-valued map $DF(\hat{x}, \hat{v}) : X \rightrightarrows R^m$ defined by

$$\mathrm{gph}DF(\hat{x}, \hat{v}) = T((\hat{x}, \hat{v}), \mathrm{gph}F)$$

is called the contingent derivative of F at the point $(\hat{x}, \hat{v}) \in \mathrm{gph}F$ [2]. In other words $v \in DF(\hat{x}, \hat{v})(x)$ if and only if $(x, v) \in T((\hat{x}, \hat{v}), \mathrm{gph}F)$. The Mordukhovich coderivative of F at the point $(\hat{x}, \hat{v}) \in \mathrm{gph}F$ [14, 18] is defined by

$$D^*F(\hat{x}, \hat{v})(v^*) = \{x^* \in R^n \mid (x^*, -v^*) \in N((\hat{x}, \hat{v}), \mathrm{gph}F)\}.$$

Recall the notion of metric regularity and its coderivative characterization. A set-valued map $F : R^n \rightrightarrows R^m$ is metrically regular around $(\hat{x}, \hat{v}) \in \mathrm{gph}F$ if there exists $\epsilon > 0$ as well as a number $\mu > 0$ such that

$$d(x, F^{-1}(v)) \leq \mu d(v, F(x)), \ \ x \in \hat{x} + \epsilon B_n, \ \ v \in \hat{v} + \epsilon B_m.$$

Recall also the following coderivative characterization of the metric regularity property for set-valued maps [14, 18].

Theorem 2.1 (Mordukhovich criterion)**.** *A set-valued map* $F : R^n \rightrightarrows R^m$ *with closed graph is metrically regular around* $(\hat{x}, \hat{v}) \in \mathrm{gph}F$ *if and only if the inclusion* $0 \in D^*F(\hat{x}, \hat{v})(v^*)$ *implies that* $v^* = 0$.

3. Newton's method and metric regularity

In this section we show that for Lipschitzian set-valued map $F : R^n \rightrightarrows R^m$ the metric regularity can be characterized in terms of well-posedness of the Newton method for perturbed set valued maps $F(x) - \tilde{v}$, where the vectors $\tilde{v} \in R^m$ have a sufficiently small norm.

We need the following definition. We say that the Newton method for perturbed set-valued map F is well-posed around $(\hat{x}, \hat{v}) \in \mathrm{gph}F$ with modulus μ if there exists $\eta > 0$ such that for all $x \in \hat{x} + \eta B_n$ and $\tilde{v} \in \hat{v} + \eta B_m$ there is $v \in \pi(0, F(x) - \tilde{v})$ satisfying the condition

$$DF^{-1}(v + \tilde{v}, x)(-v) \cap \mu\|v\|B_n \neq \emptyset. \tag{7}$$

This condition implies that the largest possible Newton inclusion (5) for the perturbed map $x \to F(x) - \tilde{v}$ (the generalization of Newton's equation (3)) has at least one solution, $\bar{x}$, satisfying the boundedness condition $\bar{x} \in \mu B_n$.

Now we establish the principal result of this section.

Theorem 3.1. *Assume that the set-valued map $F : R^n \rightrightarrows R^m$ with closed values is Lipschitzian with the constant $L_F > 0$ in a neighbourhood of a point $\hat{x}$. Then the following conditions are equivalent:*

1. *The map F is metrically regular around $(\hat{x}, \hat{v}) \in \mathrm{gph}F$.*
2. *Newton's method for perturbed set-valued map F is well-posed around $(\hat{x}, \hat{v}) \in \mathrm{gph}F$.*

Proof. Let F be metrically regular around $(\hat{x}, \hat{v}) \in \mathrm{gph}F$ with modulus μ. Set $\eta = \epsilon/(4 + 2L_F)$, where ϵ is from the definition of metric regularity. Consider $x \in \hat{x} + \eta B_n$, $\tilde{v} \in \hat{v} + \eta B_m$, and $v \in \pi(0, F(x) - \tilde{v})$. Then we have

$$\|v + \tilde{v} - \hat{v}\| \leq \|\tilde{v} - \hat{v}\| + \|v\| = \|\tilde{v} - \hat{v}\| + d(\tilde{v}, F(x))$$
$$\leq 2\|\tilde{v} - \hat{v}\| + L_F\|x - \hat{x}\| \leq \frac{\epsilon}{2}.$$

Condition (7) is satisfied due to the following lemma that is a set-valued version of Proposition 3.2 from [11].

Lemma 3.2. *If F is metrically regular around $(\hat{x}, \hat{v}) \in \mathrm{gph}F$ with modulus μ, then*

$$DF^{-1}(w, x)(\bar{v}) \cap \mu\|\bar{v}\|B_n \neq \emptyset,$$

whenever $x \in \hat{x} + \frac{\epsilon}{2}B_n$, $w \in \left(\hat{v} + \frac{\epsilon}{2}B_m\right) \cap F(x)$, and $\bar{v} \in R^m$.

Proof. Consider a sequence $t_j \downarrow 0$. From the metric regularity condition we have

$$d(x, F^{-1}(w + t_j \bar{v})) \leq \mu d(w + t_j \bar{v}, F(x)) \leq \mu t_j \|\bar{v}\|.$$

Therefore there exists a sequence

$$\bar{x}_j \in t_j^{-1}(F^{-1}(w + t_j \bar{v}) - x),$$

satisfying the inequality

$$\|\bar{x}_j\| \leq \mu \|\bar{v}\|.$$

Without loss of generality $\bar{x}_j$ converges to a vector

$$\bar{x} \in DF^{-1}(w, x)(\bar{v}) \cap \mu \|\bar{v}\| B_n.$$

This ends the proof. □

To prove that the well-posedness of Newton's method for perturbed set-valued maps implies the metric regularity, fix $\tilde{v} \in R^m$ and consider the function

$$\rho(x) = d(0, F(x) - \tilde{v}).$$

It suffices to show that if $\rho(\tilde{x}) > 0$, $\tilde{x} \in \hat{x} + \frac{\eta}{4(1+\mu+\mu L)} B_n$, and $\tilde{v} \in \hat{v} + \frac{\eta}{4(1+\mu+\mu L)} B_m$, then there exists $x' \in F^{-1}(\tilde{v})$ such that $\|\tilde{x} - x'\| \leq 2\mu\rho(\tilde{x})$. (Here $\eta > 0$ is from the definition of the well-posedness of the Newton method for perturbed set-valued map $F(x) - \tilde{v}$.) Indeed, if such x' exists, then the metric regularity of F follows from the inequality

$$d(\tilde{x}, F^{-1})(\tilde{v}) \leq \|\tilde{x} - x'\| \leq 2\mu\rho(\tilde{x}) = 2\mu d(\tilde{v}, F(\tilde{x})).$$

We need some auxiliary results. The following lemma contains an estimate for the lower Dini derivative of $\rho(\cdot)$ along the Newton's direction and, to certain extent, can be considered as a set-valued version of Proposition 3 from [5].

Lemma 3.3. *Let $v \in \pi(0, F(x) - \tilde{v})$ and $\bar{x} \in DF^{-1}(v + \tilde{v}, x)(-v)$. Then the inequality*

$$\liminf_{t \downarrow 0} t^{-1}(\rho(x + t\bar{x}) - \rho(x)) \leq -\rho(x) \tag{8}$$

holds.

Proof. Since $(\bar{x}, -v) \in T((x, v+\tilde{v}), \mathrm{gph}F)$ and F is Lipschitzian, we see that there exists

$$p(t) \in F(x + t\bar{x}) - (v + \tilde{v} - tv)$$

satisfying

$$\liminf_{t \downarrow 0} t^{-1}\|p(t)\| = 0.$$

Observe that

$$\rho(x + t\bar{x}) = d(0, F(x + t\bar{x}) - \tilde{v})$$
$$\leq \|v - tv + p(t)\| \leq (1 - t)\rho(x) + \|p(t)\|.$$

Therefore we have

$$t^{-1}(\rho(x + t\bar{x}) - \rho(x)) \leq -\rho(x) + t^{-1}\|p(t)\|.$$

Since the function $\rho(\cdot)$ is Lipschitzian, we obtain (8). □

Set $x^0 = \tilde{x}$.

Lemma 3.4. *Assume that* $\rho(x^0) > 0$, $x^0 \in \hat{x} + \frac{\eta}{4(1+\mu+\mu L)} B_n$, *and* $\tilde{v} \in \hat{v} + \frac{\eta}{4(1+\mu+\mu L)} B_m$. *Then there exists a sequence generated by the Newton method*

$$x^{k+1} = x^k + t^k \bar{x}^k, \;\; v^k \in \pi(0, F(x^k) - \tilde{v}), \;\; k = 0, 1, \ldots \tag{9}$$

where

$$\bar{x}^k \in DF^{-1}(v^k + \tilde{v}, x^k)(-v^k) \cap \mu\|v^k\|B_n, \;\; t^k > 0, \tag{10}$$

such that $\|v^k\| > \|v^{k+1}\|$, $k = 0, 1, \ldots$, *and there exists the limit* $\lim_{k\to\infty} x^k = x^\omega$ *satisfying the inequalities*

$$\|x^0 - x^\omega\| \leq 2\mu(\rho(x^0) - \rho(x^\omega)) \tag{11}$$

and

$$\|\hat{x} - x^\omega\| \leq (1 + 2\mu L_F)\|x^0 - \hat{x}\| + 2\mu\|\tilde{v} - \hat{v}\| - 2\mu\rho(x^\omega) < \eta. \tag{12}$$

Proof. By Lemma 3.3 there exists $t^k > 0$ such that

$$\|v^{k+1}\| = \rho(x^{k+1}) \leq \left(1 - \frac{t^k}{2}\right)\rho(x^k) = \left(1 - \frac{t^k}{2}\right)\|v^k\|.$$

Therefore we have $\|v^k\| > \|v^{k+1}\|$, and $t^k\|v^k\| \le 2(\|v^k\| - \|v^{k+1}\|)$, $k = 0, 1, \ldots$. Set $\nu = \lim_{k\to\infty} \|v^k\|$. Observe that

$$\|x^{k+1} - x^k\| = t^k\|\bar{x}^k\| \le t^k\mu\|v^k\| \le 2\mu(\|v^k\| - \|v^{k+1}\|).$$

From this we obtain

$$\|x^{k+p} - x^k\| \le \sum_{j=k}^{k+p-1} \|x^{j+1} - x^j\| \le 2\mu \sum_{j=k}^{k+p-1} (\|v^j\| - \|v^{j+1}\|)$$

$$= 2\mu(\|v^k\| - \|v^{k+p}\|) \le 2\mu(\|v^k\| - \nu).$$

Therefore there exists the limit $\lim_{k\to\infty} x^k = x^\omega$. Putting $k = 0$ and passing to the limit as p goes to infinity, we get (11). From (11) and the inequality

$$\rho(x^0) = d(\tilde{v}, F(x^0)) \le \|\tilde{v} - \hat{v}\| + L_F\|\hat{x} - x^0\|$$

we obtain

$$\begin{aligned} \|\hat{x} - x^\omega\| &\le \|\hat{x} - x^0\| + \|x^0 - x^\omega\| \\ &\le \|\hat{x} - x^0\| + 2\mu(\rho(x^0) - \rho(x^\omega)) \\ &\le (1 + 2\mu L_F)\|\hat{x} - x^0\| + 2\mu\|\tilde{v} - \hat{v}\| - 2\mu\rho(x^\omega) < \eta. \end{aligned} \tag{13}$$

The lemma is proved. □

The point x^ω (here ω stands for the least infinite ordinal) constructed in the proof of Lemma 3.4 may be not a solution to the inclusion $0 \in F(x) - \tilde{v}$. In this case we apply the same procedure with $x^0 = x^\omega$ and construct $x^{2\omega}$, and so on. If $\rho(x^{n\omega}) > 0$, then from (11) we have

$$\|x^{(n+p)\omega} - x^{n\omega}\| \le 2\mu(\rho(x^{n\omega}) - \rho(x^{(n+p)\omega})).$$

Therefore the sequence $x^{n\omega}$, $n = 0, 1, \ldots$, converges to a point x^{ω^2}. If we have already constructed a point x^α, where α is a countable ordinal, then using the above procedure we can construct $x^{\alpha+\omega}$ and $x^{\omega\alpha}$, etc. More generally, let A be a set of ordinals. If $x^{\alpha_1} \in \mathrm{cl}\{x^\alpha \mid \alpha \in A\}$, $\rho(x^{\alpha_1}) = \inf\{\rho(x^\alpha) \mid \alpha \in A\} > 0$, and $x^{\alpha_1} \in \hat{x} + \eta B_n$, then we can construct an element x^{α_1+1} defined by

$$x^{\alpha_1+1} = x^{\alpha_1} + t^{\alpha_1}\bar{x}^{\alpha_1}, \quad v^{\alpha_1} \in \pi(0, F(x^{\alpha_1}) - \tilde{v}), \tag{14}$$

$$\bar{x}^{\alpha_1} \in DF^{-1}(v^{\alpha_1} + \tilde{v}, x^{\alpha_1})(-v^{\alpha_1}) \cap \mu\|v^{\alpha_1}\|B_n, \quad t^{\alpha_1} > 0. \tag{15}$$

Thus we obtain a net (Moore-Smith sequence) $\{x^\alpha \mid \alpha < \omega_1\}$ generated by the Newton method. Here ω_1 is the first uncountable ordinal. If $\rho(x^\alpha) = 0$ the process stops. To show that this net is correctly defined we have to prove the inclusion $x^\alpha \in \hat{x} + \eta B_n$, whenever $\tilde{x} \in \hat{x} + \frac{\eta}{4(1+\mu+\mu L)} B_n$, and $\tilde{v} \in \hat{v} + \frac{\eta}{4(1+\mu+\mu L)} B_m$. This can be done using transfinite induction. Let $\{x^\alpha\}$ be a net generated by Newton's method. Assume that

$$\|\tilde{x} - x^\alpha\| \leq 2\mu(\rho(\tilde{x}) - \rho(x^\alpha)), \tag{16}$$

whenever $\alpha < \alpha'$. If α' is a successor, i.e. there exists α_1 such that $\alpha' = \alpha_1 + 1$, then by the induction hypothesis we have

$$\|\tilde{x} - x^{\alpha_1}\| \leq 2\mu(\rho(\tilde{x}) - \rho(x^{\alpha_1})),$$

and, as in the proof of Lemma 3.4, we obtain

$$\|x^{\alpha_1} - x^{\alpha'}\| \leq 2\mu(\rho(x^{\alpha_1}) - \rho(x^{\alpha'})),$$

where x^{α_1+1} is defined by (14) and (15). Adding the last two inequalities we get

$$\|\tilde{x} - x^{\alpha'}\| \leq 2\mu(\rho(\tilde{x}) - \rho(x^{\alpha'})). \tag{17}$$

Therefore (see (13)) we have

$$\|\hat{x} - x^{\alpha'}\| \leq \|\hat{x} - x^0\| + 2\mu(\rho(x^0) - \rho(x^{\alpha'})) < \eta. \tag{18}$$

If α' is a limit ordinal, then $x^{\alpha'} = \lim_{j\to\infty} x^{\alpha_j}$, $\alpha_j < \alpha'$. Passing to the limit in the inequality

$$\|\tilde{x} - x^{\alpha_j}\| \leq 2\mu(\rho(\tilde{x}) - \rho(x^{\alpha_j})),$$

we obtain (17) and, as a consequence, (18). Thus, the net generated by Newton's method is well-defined. Define a partial order in the set of all points generated by Newton's method from the initial point $x^0 = \tilde{x}$. We say that x^{α_1} is less than x^{α_2} if x^{α_1} and x^{α_2} belong to the same net generated by Newton's method and $\alpha_1 < \alpha_2$. Obviously a net is a totally ordered subset. By the Hausdorff maximal principle in any partially ordered set, every totally ordered subset is contained in a maximal totally ordered subset. Let $\{x^\alpha\}_{\alpha\in A}$ be a maximal net generated by Newton's method from the initial point $\tilde{x}$. There exists a point $x' = x^{\alpha_1} \in \mathrm{cl}\{x^\alpha\}$ satisfying $\rho(x^{\alpha_1}) = \inf\{\rho(x^\alpha) \mid \alpha \in A\}$. If $\rho(x^{\alpha_1}) > 0$, then since $\mathrm{cl}\{x^\alpha\}_{\alpha\in A} \subset \hat{x} + \eta B_n$, the point x^{α_1+1} defined by (14) and (15) is greater than $\{x^\alpha\}_{\alpha\in A}$. This contradicts the maximality of the net $\{x^\alpha\}_{\alpha\in A}$. Thus $\rho(x') = 0$. This ends the proof. □

4. Newton's method for set-valued maps: convergence analysis

Let $F : R^n \rightrightarrows R^m$ be a set-valued map with closed values. We shall study the convergence of the following set-valued version of Newton's method. Given a point $(x^k, v^k) \in \text{gph}F$, we define the next iterate as

$$x^{k+1} = x^k + t^k \bar{x}^k, \quad v^{k+1} \in \pi(0, F(x^{k+1})), \tag{19}$$

where $\bar{x}^k$ is a solution to the inclusion

$$-v^k \in \Lambda(x^k, v^k)(\bar{x}^k), \tag{20}$$

and $\Lambda(x, v) : R^n \rightrightarrows R^m$ is a positively homogeneous set-valued map satisfying the inclusion $\Lambda(x, v)(\bar{x}) \subset DF(x, v)(\bar{x})$ for all $(x, v) \in \text{gph}F$ and $\bar{x}$. The vector $\bar{x}^k$ is chosen from the condition

$$\bar{x}^k \in \pi(0, \Lambda(x^k, v^k))^{-1}(-v^k)). \tag{21}$$

If $\Lambda(x^k, v^k))^{-1}(-v^k)$ is convex, (21) uniquely defines $\bar{x}^k$.

In the sequel this method is called Newton method, if $t^k = 1$. If $t^k \in]0, 1]$, the method is called damped Newton method. For example, the step-length t^k can be chosen from the condition

$$d(0, F(x^k + t^k \bar{x}^k)) = \min_{t \in [0,1]} d(0, F(x^k + t\bar{x}^k)). \tag{22}$$

Let $x \notin F^{-1}(0)$, $v \in \pi(0, F(x))$, and $\bar{x} \in (\Lambda(x, v))^{-1}(\bar{v})$. Then there exists $p(t) \in R^n$ such that

$$v + t\bar{v} + p(t) \in F(x + t\bar{x})$$

and

$$\|p(t)\| = t\rho(x, v, \bar{x}, \bar{v}, t),$$

where

$$\rho(x, v, \bar{x}, \bar{v}, t) = t^{-1} d(v + t\bar{v}, F(x + t\bar{x})). \tag{23}$$

Note that if F is Lipschitzian, then

$$\liminf_{t \downarrow 0} \rho(x, v, \bar{x}, \bar{v}, t) = 0.$$

If $\bar{v} = -v$, then we have

$$d(0, F(x + t\bar{x})) \leq \|(1 - t)v + p(t)\| \leq (1 - t)\|v\| + t\rho(x, v, \bar{x}, -v, t). \tag{24}$$

The rate of convergence of Newton's method depends on the properties of function ρ.

Theorem 4.1. *Assume that the following conditions are satisfied:*

1. *There exists $\mu > 0$ such that for all $x \in R^n$, $v \in \pi(0, F(x))$ and $\bar{v} \in R^m$ the set $(\Lambda(x, v))^{-1}(\bar{v}) \cap \mu\|\bar{v}\|B_n$ is nonempty.*
2. *There exists a monotone increasing function $\omega : [0, +\infty[\to [0, +\infty[$ such that $\lim_{\alpha\downarrow 0} \omega(\alpha) = 0$ and $\rho(x, v, \bar{x}, -v, t) \leq \|v\|\omega(t\|v\|)$ for all $x \in R^n \setminus F^{-1}(0)$, $v \in \pi(0, F(x))$, and $\bar{x} \in \pi(0, (\Lambda(x, v))^{-1}(-v)$.*

Then for any initial point x^0 there exists a monotone non-decreasing sequence $t^k > 0$, such that $t^k = 1$ for large k, and the corresponding damped Newton method/Newton method, starting at x^0, converges to a point $x^0 + 2\mu d(0, F(x^0))B_n$. If the gph$F$ *is closed, then $x^\infty \in F^{-1}(0)$ and the convergence is R-superlinear. If $\omega(\alpha) = O(\alpha)$, $\alpha \downarrow 0$, then the convergence is R-quadratic. If, in addition, F is locally Lipschitzian, then the convergence is Q-quadratic.*

Proof. From the first condition of the theorem we see that the Newton method is well-defined. Suppose that the points $(x^j, v^j, \bar{x}^j)$, $j = \overline{0, k}$, are already generated by the Newton method. To construct the point x^{k+1}, put

$$\rho^k(t) = \rho(x^k, v^k, \bar{x}^k, -v^k, t).$$

Since $\rho^k(t) \leq \|v^k\|\omega(t\|v^k\|)$, we have $\lim_{t\downarrow 0} \rho^k(t) = 0$. There exists a vector $p^k(t)$ satisfying

$$(1 - t)v^k + p^k(t) \in F(x^k + t\bar{x}^k), \quad \|p^k(t)\| \leq t\rho^k(t),$$

(see (23) and (23)). Set $t^k = \min\{1, \omega^{-1}(1/2)/\|v^k\|\}$ and $x^{k+1} = x^k + t^k\bar{x}^k$. Observe that

$$d(0, F(x^{k+1})) = \|v^{k+1}\| \leq \|(1 - t^k)v^k + p^k(t^k)\|$$

$$\leq (1 - t^k)\|v^k\| + \rho^k(t^k) \leq (1 - t^k)\|v^k\| + t^k\|v^k\|\omega(t^k\|v^k\|)$$
$$= (1 - t^k + t^k\omega(t^k\|v^k\|))\|v^k\|.$$

If $t^k < 1$, then we obtain

$$\|v^{k+1}\| \leq (1 - t^k/2)\|v^k\|. \tag{25}$$

If $t^k = 1$, then the inequality

$$\|v^{k+1}\| \leq \|v^k\| \omega(\|v^k\|) < \|v^k\|/2 \tag{26}$$

holds. From (25) we see that the sequence $\|v^k\|$ is monotone decreasing and the sequence t^k is monotone increasing, whenever $t^k < 1$. Thus $t^k = 1$ for large k.

Since $\|\bar{x}^k\| \leq \mu \|v^k\|$, we have

$$\|x^{k+1} - x^k\| = t^k \|\bar{x}^k\| \leq \mu t^k \|v^k\| \leq 2\mu(\|v^k\| - \|v^{k+1}\|).$$

By induction we obtain

$$\|x^{k+p} - x^k\| \leq 2\mu(\|v^k\| - \|v^{k+p}\|). \tag{27}$$

This implies that there exists the limit $x^\infty = \lim_{k\to\infty} x^k$. If the graph of F is closed, then the function $x \to d(0, F(x))$ is lower semi-continuous and we have

$$0 = \lim_{k\to\infty} \|v^k\| = \lim_{k\to\infty} d(0, F(x^k)) \geq \liminf_{x' \to x^\infty} d(0, F(x')) \geq d(0, F(x^\infty)).$$

Hence $0 \in F(x^\infty)$. Passing to the limit, as p goes to infinity, in (27), we get

$$\|x^k - x^\infty\| \leq 2\mu \|v^k\|$$

and

$$\|x^0 - x^\infty\| \leq 2\mu d(0, F(x^0)).$$

From (26) we obtain

$$\|x^{k+1} - x^\infty\| \leq 2\mu \|v^{k+1}\| \leq 2\mu \|v^k\| \omega(\|v^k\|),$$

i.e. the convergence is R-superlinear. If $\omega(\alpha) = O(\alpha)$, $\alpha \downarrow 0$, then there exists a constant $M > 0$ such that

$$\|x^{k+1} - x^\infty\| \leq 2\mu \|v^{k+1}\| \leq 2\mu M \|v^k\|^2,$$

and the convergence is R-quadratic. Now assume that F is Lipschitzian with the constant L_F, then we have

$$\|x^{k+1} - x^\infty\| \leq 2\mu M \|v^k\|^2 = 2\mu M d^2(0, F(x^k)) \leq 2\mu M L_F^2 \|x^k - x^\infty\|^2.$$

Thus the convergence is Q-quadratic. $\square$

Let $f : R^n \to R^n$ be a continuous function. Define the map $F(x) = \{f(x)\}$. In this case the second condition of the theorem implies

$$\|f(x + t\bar{x}) - f(x) + tf(x)\| = o(t), \quad t \downarrow 0,$$

for all $\bar{x} \in \pi(0, (\Lambda(x, f(x)))^{-1}(-f(x))$, i.e. the function f is directionally differentiable along Newton's directions. This assumption is quite natural (see Proposition 2 from [5]). For non-Lipschitzian functions, like $f(x) = x^{1/3}$ at $x = 0$, this condition generally does not hold. Note that in the single-valued case Theorem 4.1 is not contained in Theorems 3.3 and 3.4 from [11] and does not generalize them.

5. Systems of nonlinear equations with geometric constraints

Let $f : R^n \times R^l \to R^m$ be a continuously differentiable function. Its derivative is supposed to be Lipschitzian with a constant $L_{\nabla f}$. Consider the nonlinear equation

$$f(x, u) = 0. \tag{28}$$

The problem is to find a solution (x, u) such that the variable u satisfies the geometric constraints

$$u \in U, \tag{29}$$

where $U \subset R^l$ is a closed convex set. Let K be the recessive cone of U, i.e. $K = \{u \in R^l \mid u + U \subset U\}$ (see [17]).

Figure 1: Trajectories of the Newton method (Example 1).

Consider, for example, functions $g : R^n \to R^{m-l}$ and $h : R^n \to R^l$ and the system of equations and inequalities (cf. [16])

$$g(x) = 0, \quad h(x) \leq 0.$$

It can be rewritten as (28) and (29) in the following way:

$$g(x) = 0, \quad h(x) + u = 0, \quad u \geq 0.$$

In this case $U = K = \{u \in R^l \mid u \geq 0\}$.

System (28) and (29) is equivalent to the inclusion $0 \in F(x) = f(x, U)$. Let $v = f(x, u) \in \pi(0, F(x))$. To apply the Newton method to this inclusion, put

$$\Lambda(x, u)(\bar{v}) = \nabla_x f(x, u)\bar{x} + \nabla_u f(x, u)K.$$

Let $(\hat{x}, \hat{u}) \in R^n \times U$ be such that $f(\hat{x}, \hat{u}) = 0$. Assume that

$$((\nabla_u f(\hat{x}, \hat{u}))^T)^{-1} K^* \cap \ker \nabla_x f(\hat{x}, \hat{u}) = \{0\}.$$

Then from the Mordukhovich criterion (Theorem 2.1) we see that the set-valued map F is metrically regular around $(\hat{x}, 0)$. By Theorem 3.1 the Newton method is well defined around $(\hat{x}, 0)$ and

$$d(0, (\Lambda(x,u))^{-1}(\bar{v})) \leq L_\Lambda \|\bar{v}\|, \tag{30}$$

whenever $(x, u) \in (\hat{x}, \hat{u}) + \eta B_{n+l}$ and $\eta > 0$ is sufficiently small. Let $v \in \pi(0, F(x))$. There exist $\bar{x} \in R^n$ and $\bar{w} \in K$ such that

$$-v = \Lambda(x,u)(\bar{x}) = \nabla_x f(x,u)\bar{x} + \nabla_u f(x,u)\bar{w}$$

and

$$(\|\bar{x}\|^2 + \|\bar{w}\|^2)^{\frac{1}{2}} \leq L_\Lambda \|v\|$$

Since $u + t\bar{w} \in U$, $t > 0$, we have

$$\rho(x, v, \bar{x}, -v, t) = t^{-1} d(v - tv, F(x + t\bar{x})) \leq t^{-1} \|v - tv - f(x + t\bar{x}, u + t\bar{w})\|$$

$$\leq t L_{\nabla f}(\|\bar{x}\|^2 + \|\bar{w}\|^2) \leq t L_{\nabla f}(L_\Lambda \|v\|)^2.$$

Following the proof of Theorem 4.1 we see that the Newton method, starting in a sufficiently small neighbourhood of $\hat{x}$, converges Q-quadratically to the set $F^{-1}(0)$. As we can see from the following example, the neighbourhood, where the method converges quadratically, is really small even in very simple situations.

Example 1

Apply the Newton method to the system

$$x_1^2 + x_2^2 = 1 \quad x_1^2 - x_2 \leq 0.$$

Typical trajectories are shown in Fig. 1. At the initial stage the trajectories zigzag. Such behaviour corresponds to linear convergence. Only near the set $F^{-1}(0)$ we observe quadratic convergence. Damped Newton methods exhibit fast convergence in a wider area. We illustrate this in the next section where a system of linear inequalities is considered and this phenomenon is especially evident.

6. Systems of linear inequalities

Consider a system of linear inequalities

$$\langle a_i, x\rangle - b_i \le 0, \quad i = \overline{1,m}, \tag{31}$$

where $a_i \in R^n$ and $b_i \in R$, $i = \overline{1,m}$. It is assumed that $\|a_i\| = 1$, $i = \overline{1,m}$. Define the set-valued map

$$F(x) = \{v \in R \mid \langle a_i, x\rangle - b_i \le v, \quad i = \overline{1,m}\}.$$

System (31) and the inclusion

$$0 \in F(x) \tag{32}$$

are equivalent. Assume that the following hypotheses are satisfied:

(H1) For any subset of indices $1 \le i_1 \le \ldots \le i_n \le m$ the vectors $a_{i_j} \in R^n$, $j = \overline{1,n}$, are linearly independent.

(H2) For any subset of indices $1 \le i_1 \le \ldots \le i_{n+1} \le m$ the vectors $(a_{i_j}, -1) \in R^{n+1}$, $j = \overline{1,n+1}$, are linearly independent.

(H3) For any subset of indices $1 \le i_1 \le \ldots \le i_{n+1} \le m$ the set $\{x \in R^n \mid \langle a_{i_j}, x\rangle = b_{i_j}, j = \overline{1,n+1}\}$ is empty.

(H4) For any subset of indices $1 \le i_1 \le \ldots \le i_{n+2} \le m$ the set $\{(x,v) \in R^{n+1} \mid \langle a_{i_j}, x\rangle - v = b_{i_j}, j = \overline{1,n+2}\}$ is empty.

Figure 2: Newton's method for a system of linear inequalities (Example 2).

Figure 3: Damped Newton's method for a system of linear inequalities, I.

Figure 4: Damped Newton's method for a system of linear inequalities, II.

Figure 5: Damped Newton's method for a system of linear inequalities, III.

An arbitrary system of inequalities can be transformed to a system satisfying (H1) - (H4) by a small perturbation of the data. Let $x \in R^n$. Put

$$v(x) = \max_{i=\overline{1,m}} (\langle a_i, x\rangle - b_i)$$

and

$$I(x) = \{i \mid v(x) = \langle a_i, x\rangle - b_i\}.$$

By $|I(x)|$ we denote the cardinality of the set $I(x)$. Note that the hypotheses (H2) and (H4) imply that $|I(x)| \leq n+1$ and that the number of points x satisfying $|I(x)| = n+1$ is finite.

Lemma 6.1. *Let $x \in R^n$ be such that $v(x) > 0$ and let $I(x) = \{i_1, \dots, i_{n+1}\}$. Then either there exists $l = \overline{1, n+1}$ such that the solution $\bar{x}$ of the system $\langle a_{i_j}, \bar{x}\rangle = 1$, $j \neq l$, satisfies the inequality $\langle a_{i_l}, \bar{x}\rangle > 0$, or system (31) has no solution.*

Proof. Assume that for any $l = \overline{1, n+1}$ the solution $\bar{x}^{(l)}$ to the system $\langle a_{i_j}, \bar{x}\rangle = 1$, $j \neq l$ (by hypothesis (H1) $\bar{x}^{(l)}$ exists and is unique) satisfies the inequality

$$\langle a_{i_l}, \bar{x}^{(l)}\rangle \leq 0. \tag{33}$$

Show that system (31) has no solution. Set $y^{(l)} = x - v(x)\bar{x}^{(l)}$. From (33) we have $\langle a_{i_l}, x - \bar{y}^{(l)}\rangle \leq 0$. From this we obtain

$$0 < v(x) = \langle a_{i_l}, x\rangle - b_{i_l} \leq \langle a_{i_l}, y^{(l)}\rangle - b_{i_l}, \tag{34}$$

for any $l = \overline{1, n+1}$. Show that if the set $Y = \{y \mid \langle a_{i_j}, y\rangle \leq b_{i_j},\ j = \overline{1, n+1}\}$ is non-empty, then it contains at least one point $y^{(l)}$. Let $y \in Y$. Since $x \notin Y$, we see that there exists a non-empty set $I_0 \subset I(x)$ and a point $y_0 \in Y$ belonging to the segment connecting the points y and x, such that $\langle a_{i_j}, y_0\rangle = b_{i_j}$, $i_j \in I_0$. If $|I_0| = n$, then y_0 coincides with one of the points $y^{(l)}$, $l = \overline{1, n+1}$. Otherwise consider a non-zero vector $z \in \{z \mid \langle a_{i_j}, z\rangle = 0,\ i_j \in I_0\}$. If for all $t \in R$ the inequalities $\langle a_{i_j}, y_0 + tz\rangle < 0$, $i_j \in I(x) \setminus I_0$ hold, then $\langle a_{i_j}, z\rangle = 0$, $j = \overline{1, n+1}$. By hypothesis (H1) we have $z = 0$, a contradiction. Therefore there exists a set $I_1 \supset I_0$, $|I_1| > |I_0|$, and a number t_0 such that the point $y_1 = y_0 + t_0 z \in Y$ satisfies the equalities $\langle a_{i_j}, y_1\rangle = b_{i_j}$, $i_j \in I_1$.

By induction we construct a set of indices I_s, $|I_s| \geq n$ and a point $y_s \in Y$ satisfying the equalities $\langle a_{i_j}, y_s \rangle = b_{i_j}$, $i_j \in I_s$. By hypothesis (H3) $|I_s| = n$. Thus y_s coincides with one of the vectors $y^{(l)}$, $l = \overline{1, n+1}$, and $y^{(l)} \in Y$. On the other hand, from (34) we have $\langle a_{i_l}, y^{(l)} \rangle > b_{i_l}$, a contradiction. The lemma is proved. □

Let $x \in R^n$ be such that $v(x) > 0$. If $|I(x)| \leq n$, then we set $J(x) = I(x)$. If $|I(x)| = n+1$, then by Lemma 6.1 either there exists l such that $\langle a_{i_l}, \bar{x} \rangle > 0$, where $\bar{x}$ is a unique solution of the system $\langle a_{i_j}, \bar{x} \rangle = 1$, $j \neq l$, or the set of solutions of system (31) is empty. Put $J(x) = I(x) \setminus \{i_l\}$. We set

$$\Lambda(x, v(x))(\bar{x}) = \{\bar{v} \mid \langle a_i, \bar{x} \rangle = \bar{v}, \;\; i \in J(x)\}.$$

It is easy to see that the set $\Lambda(x, v)(\bar{x})$ is contained in $DF(x, v)(\bar{x})$.

Example 2

The Newton method (19) and (21) applied to inclusion (32) usually generates an iterative process with a linear rate of convergence. Consider the system of two inequalities $x_2 \geq 3x_1$ and $x_2 \geq -3x_1$. The trajectory starting at $(-1, -10)$ is shown in Fig. 2.

Now we describe a damped Newton method (19) and (21) solving (32) in a finite number of iterations. Let $v(x^k) > 0$. Put

$$t^k = \min\{1, \min\{t_i(x^k, \bar{x}^k) \mid i \notin I(x^k), \; v(x^k) + \langle a_i, \bar{x}^k \rangle > 0\}\}, \qquad (35)$$

where

$$t_i(x^k, \bar{x}^k) = \frac{v(x^k) - \langle a_i, x^k \rangle + b_i}{v(x^k) + \langle a_i, \bar{x}^k \rangle}.$$

It is easy to see that $t^k > 0$.

Theorem 6.2. *Under hypothesis (H1) - (H4) damped Newton method (19), (21), and (35) solves system (31) in a finite number of iterations.*

Proof. If the number of elements in the set $I(x^k)$ does not exceed n, then from (21) we obtain

$$\bar{x}^k = -A_k^T (A_k A_k^T)^{-1} v(x^k) e_{I(x^k)},$$

where $e_{I(x^k)} = (1, \ldots, 1) \in R^{I(x^k)}$ and the $I(x^k) \times n$-matrix A_k has the rows a_i, $i \in I(x^k)$. If $t^k = 1$, then we have

$$\langle a_j, x^k + t\bar{x}^k \rangle - b_j < \langle a_i, x^k + t\bar{x}^k \rangle - b_i$$

$$= v(x^k)(1 - t), \quad i \in I(x^k), \quad j \notin I(x^k), \quad t \in [0, 1[.$$

Therefore $\langle a_i, x^k + \bar{x}^k \rangle - b_i \leq 0$, $i = \overline{1, m}$. If $t^k < 1$, then the number of elements in the set $I(x^{k+1})$ is greater than in the set $I(x^k)$. Thus, either after at most n iterations the method solves the system of inequalities, or it arrives at a point x^k such that the set $I(x^k)$ contains $n+1$ elements (see Figs. 3 and 4). If the set $J(x^k)$ from the definition of Λ does not exist, then the set of solutions of system (31) is empty (see Fig.5). Otherwise we can construct the next iterate x^{k+1} (see Fig. 4). Show that $v(x^{k+1}) < v(x^k)$. Indeed, if $I(x^k) = \{i_1, \ldots, i_{n+1}\}$ and $i_l \notin J(x^k)$, then we have $\langle a_{i_j}, \bar{x}^k \rangle = -v(x^k)$, $j \neq l$ and $\langle a_{i_l}, \bar{x}^k \rangle < 0$. Therefore we get

$$v(x^{k+1}) = \max\{(1 - t^k)v(x^k),\ v(x^k) + t^k \langle a_{i_l}, \bar{x}^k \rangle\}.$$

Thus $v(x^{k+1}) < v(x^k)$. Set $\hat{v} = \min\{v(x) \mid |I(x)| = n + 1\}$. After a finite number of iterations the algorithm arrives at a point x^k satisfying $v(x^k) < \hat{v}$. After that it needs at most n iterations to solve system (31). □

Acknowledgements

The authors are grateful to Oleg Burdakov, Boris Mordukhovich, and Vera Roshchina for their valuable comments. This research was supported by the Portuguese Foundation for Science and Technologies (FCT), the Portuguese Operational Programme for Competitiveness Factors (COMPETE), the Portuguese National Strategic Reference Framework (QREN), and the European Regional Development Fund (FEDER).

References

[1] H. Attouch, B. F. Svaiter, *A continuous dynamical Newton-like approach to solving monotone inclusions*, SIAM J. Control and Optimization (to appear).

[2] J.-P. Aubin, I. Ekeland, *Applied nonlinear analysis*, Wiley, New York, 1984.

[3] D. Azé, *Unified theory for metric regularity of multifunctions*, J. Convex Anal., Vol. 13, (2006), pp 225252.

[4] D. Azé, C. C. Chou, *On a Newton type iterative method for solving inclusions*, Math. Oper. Res., Vol. 20, (1995), pp 790800.

[5] O. Burdakov, *On properties of Newton's method for smooth and nonsmooth equations*, In: R.P. Agarwal (Ed.), Recent Trends in Optimization Theory and Applications (1995), World Scientific, pp. 17-24.

[6] C. Cabuzel, A. Pietrus, *Local convergence of Newton's method for subanalytic variational inclusions*, Positivity, Vol. 12, (2008), pp 525-533.

[7] F. H. Clarke, *Optimization and nonsmooth analysis*, Wiley, New York, 1983.

[8] A. L. Dontchev, *Uniform convergence of the Newton method for Aubin continuous maps*, Serdica Math. J. Vol. 22, (1996), pp 385-398.

[9] A. L. Dontchev, R. T. Rockafellar, *Newton's method for generalized equations: a sequential implicit function theorem*. Math. Program., Vol. 123(1), (2010), pp 139-159.

[10] A. L. Dontchev, R. T. Rockafellar, *Implicit functions and solution mappings*. Springer, Dordrecht, 2010.

[11] T. Hoheisel, C. Kanzow, B. S. Mordukhovich, H. Phan, *Generalized Newton's method for nonsmooth equations based on graphical derivatives*, SIAM J. Optim. (to appear).

[12] C. Jean-Alexis, A. Pietrus, *On the convergence of some methods for variational inclusions*, Rev. R. Acad. Cien. Serie A. Mat., Vol. 102(2), (2008), pp 355-361.

[13] D. Klatte, B. Kummer, *Stability of inclusions: characterizations via suitable Lipschitz functions and algorithms*, Optimization, Vol. 55 (2006), pp 627-660.

[14] B. S. Mordukhovich, *Variational analysis and generalized differentiation, I. Basic theory*, Springer, Berlin, 2006.

[15] J. M. Ortega, W. C. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Pess, New York, 1970.

[16] B. N. Pshenichnyi, *Newton's method for solution of systems of equalities and inequalities*, Math. Notes, Vol. 8, (1970), pp 827-830.

[17] R. T. Rockafellar, *Convex analysis*, Princeton Univ. Press, Princeton, 1970.

[18] R. T. Rockafellar, R. J.-B. Wets, *Variational analysis*, Springer, Berlin, 2004.

[19] P. Saint-Pierre, *Newton and other continuation methods for multivalued inclusions*, Set-Valued Analysis, Vol. 3, (1995), pp 143-156.