Discussiones Mathematicae
Differential Inclusions, Control and Optimization 20 (2000) 209–244

GENERALIZED NEWTON AND NCP- METHODS: CONVERGENCE, REGULARITY, ACTIONS

Bernd Kummer

Humboldt-Universität zu Berlin
Mathematisch-Naturwissenschaftliche Fakultät II
Institut für Mathematik

e-mail: kummer@mathematik.hu-berlin.de

Abstract

Solutions of several problems can be modelled as solutions of nonsmooth equations. Then, Newton-type methods for solving such equations induce particular iteration steps (actions) and regularity requirements in the original problems. We study these actions and requirements for nonlinear complementarity problems (NCP's) and Karush–Kuhn–Tucker systems (KKT) of optimization models. We demonstrate their dependence on the applied Newton techniques and the corresponding reformulations. In this way, connections to SQP-methods, to penalty-barrier methods and to general properties of so-called NCP-functions are shown. Moreover, direct comparisons of the hypotheses and actions in terms of the original problems become possible. Besides, we point out the possibilities and bounds of such methods in dependence of smoothness.

Keywords: nonsmooth functions, generalized Newton methods, critical points, complementarity, SQP methods, inverse mappings, regularity.

1991 Mathematics Subject Classification: 90C31, 90C48, 70H30.

1 Introduction

During the last fifteen years, several solution methods for nonsmooth equations have been studied and applied to variational inequalities, generalized equations, Karush–Kuhn–Tucket (KKT) systems or nonlinear complementarity problems (NCP's), c.f. [2, 5, 7, 8, 9, 12, 18, 19, 21, 27, 26, 29, 32, 34].

Accordingly, one finds various conditions for convergence of nonsmooth Newton methods (mainly written in terms of semismoothness) and may reformulate identical problems by means of different (nonsmooth) equations. Especially for complementarity problems, a big number of so-called NCP functions have been applied in order to obtain such a description as an equation.

In this paper, we want to help to elaborate those properties of approximations and NCP functions which are important for solving KKT systems or NCP's as nonsmooth equations. Thus we compare in detail the regularity assumptions and the content of a Newton step in terms of the original data for optimization problems in standard formulation. We show how the Newton steps are related to second order steps for penalty and barrier functions and how regularity requirements depend on smoothness of NCP-functions in related models.

The general idea can be simplified as follows:

For $f \in C^1(R^n, R^n)$, the injectivity of $Df(z^*)$ is crucial for both superlinear local convergence of Newton's method and regularity in the inverse-function sense. For f being only locally Lipschitz (or for multifunctions) such unifying condition does not exist. Injectivity conditions – based on different "reasonable" generalized derivatives (and only applicable if these derivatives may be determined!) – describe still different desirable properties of f^{-1} (called e.g. strong, metric, upper regularity, calmness...), but they may be completely useless for Newton's method (based on solving linear equations).

For the latter, one does not necessarily need any of the "well-established" derivatives, but

(i) a condition like *continuous differentiability* for the used "derivative" Rf, and

(ii) the regularity condition which requires that potential Newton-matrices $Rf(x)$ have uniformly bounded inverses (*Newton-regularity*).

Both conditions induce properties of f^{-1} depending essentially on Rf and on the type of the nonsmooth function f.

Therefore, we compare these properties with standard regularity notions and want to understand what a Newton step means in terms of the original (KKT or NCP) problem.

The common properties of all these methods become visible by considering a particular Lipschitzian perturbation of Kojima's system being assigned to KKT-points.

In §2, we present a brief summary of few (generalized) derivatives and their relations to regularity and Lipschitzian perturbations.

In §3 we discuss the classical (nonsmooth) Newton approach based on linear auxiliarly problems. We define *Newton maps* via a (multivalued) generalization of continuous differentiability, consider *pseudo-smooth* and *locally* PC^1*-functions*, and show how Newton's method and Newton maps are related with each other. Our definitions are mainly justified by Lemma 3.2 and the Theorems 3.3, 3.6, 4.3 and 4.4.

In §4, we consider NCP's of the form

$$u(x) \geq 0, v(x) \geq 0, \langle u(x), v(x) \rangle = 0 \tag{1.1}$$

and elaborate those properties of *NCP-functions* $g : R^2 \to R$ ($g = 0 \Leftrightarrow$ (s,t) ≥ 0 and $st = 0$) which are important for solving the equivalent equation

$$f_i(x) := g(u_i(x)v_i(x)) = 0. \tag{1.2}$$

by Newton's method. Particularly, we will obtain:

Whenever g belongs to class pNCP and g as well as u, v are *locally* PC^1, the Newton step at x finds a zero ξ (to put $x_{\text{new}} = x + \xi$) of some weighted combination of linearizations

$$L^g(\xi) := a_i^g(u_i(x) + Du_i(x)\xi) + b_i^g(v_i(x) + Dv_i(x)\xi) = 0. \tag{1.3}$$

Here either $c_i^g = (a_i^g, b_i^g)$ coincides with $Dg(\sigma_i)$, where $\sigma_i = (u_i(x), v_i(x))$ or, if g is not C^1 near σ_i, the vector c_i^g is a limit of gradients as (s,t) $\to \sigma_i$. Similarly, one may interpret Du and Dv at certain non-C^1-points x of the function $z = (u, v)$. The behavior of the coefficients as x tends to a solution x^* can be generally characterized. Theorem 4.3 clarifies the content of Newton-regularity in terms of smoothness of g. For several modified NCP-functions, c.f. [34], where $g = G+h$ with $G \in$ pNCP and h is "locally small", i.e. $|h_i(\sigma_i)| \leq o(x-x^*)$ and $\|c_i^h\| \leq O(x-x^*)$, the method can be seen as an approximation of the Newton-process by means of G (with the same local convergence behavior).

The applicability and the concrete actions of Newton steps for equations, assigned to KKT-points (via NCP-or Kojima-functions), are considered and compared in §5. There, the close connections between the methods mentioned, concrete problems of sequentially quadratic programming (SQP) and penalty-barrier methods become obvious.

Example. To illustrate forthcoming definitions, let us first mention a real Lipschitz function f, presented in [19]. It consists of a countable number of linear pieces and has the following properties:

(i) f and the inverse f^{-1} are real-valued, strongly increasing, directionally differentiable and globally Lipschitz.

(ii) f is not Fréchet-diffenrentiable on a countable set N_D with cluster point $0 \notin N_D$.

(iii) $f(0) = 0; Df(0) = 1, f$ is C^1 on the open and dense set $\Theta^1 = R \backslash (N_D \cup \{0\})$.
There are exactly 2 limits of derivatives $Df(x)$, as $x \to 0, x \in \Theta^1$, both different from $Df(0)$.

(iv) Newton's method with start at any $x^0 \in \Theta^1$, always generates an alternating sequence in Θ^1. Note that Θ^1 has *full* Lebesgue measure and f is strongly regular (cf. below).

To construct f, consider intervals $I(k) = [k^{-1}, (k-1)^{-1}] \subset R$ for integers $k \geq 2$, put

$$\begin{aligned} c(k) &= \tfrac{1}{2}[k^{-1} + (k-1)^{-1}] \quad \text{(the center of } I(k)), \\ c(2k) &= \tfrac{1}{2}[(2k)^{-1} + (2k-1)^{-1}] \quad \text{(the center of } I(2k)) \end{aligned}$$

and define

$$\begin{aligned} g_k(x) &= a_k(x + c(k)), \quad \text{where } a_k = (k-1)^{-1}/[(k-1)^{-1} + c(k)], \\ h_k(x) &= b_k(x - c(2k)), \quad \text{where } b_k = k^{-1}/[k^{-1} - c(2k)]. \end{aligned}$$

For $x > 0$ define f by $f(x) = \min\{g_k(x), h_k(x)\}$ if $x \in I(k)$ and $f(x) = g_2(x)$ if $x > 1$. Finally, put $f(0) = 0$ and $f(x) = -f(-x)$ for $x < 0$. The related properties can be elementary shown, we omit the details.

Notations. Every space X, considered here, is (at least) a real Banach space. For a subset A and C of X and $r \in R$, we denote by $A + rC$ the Minkowski sum $\{a + rc / a \in A, c \in C\}$ and identify singletons and points. The closed unit ball of X is denoted by B_x, so $x + rB_x$ is the closed ball around x of radius r. If the space is evident, we omit the subscript. Having a set M of linear operators, we put $Mu = \{Au / A \in M\}$.

Given a set-valued map $F : X \rightrightarrows Y$, i.e. $F(x) \subset Y$, the set $H(F, x, \Omega)$ is the (possibly empty) upper Hausdorff-limit of F at x with respect to $\Omega \subset X$: $H(F, x, \Omega) := \lim\ \sup_{\Omega \ni \xi \to x} F(\xi) := \{y / y = \lim \eta$ for certain $(\xi, \eta) \in$

$(\Omega, F(\xi))$, $\xi \to x\}$. We write $F \subset G$ if $F(x) \subset G(x)$ for all x. In particular, F may be a function on $\Omega \subset X$; then $\eta = F(\xi)$, and $F(\xi) = \emptyset$ for $\xi \in X \backslash \Omega$. By $C^{0.1}$ we denote the family of locally Lipschitz functions, while $f \in C^{1.1}$ says that the first (Fréchet) derivative belongs to $C^{0.1}$. Troughout, $o(\cdot)$ is a function with $o(0) = 0$ and $o(u) \parallel u \parallel^{-1} \to 0$ as $u \to 0$, while $O(\cdot)$ satisfies $\parallel O(u) \parallel \to 0$ as $u \to 0$. If $O(u)$ and $o(u) \in R^+$, we suppose, without loss of generality, that these functions are upper semicontinuous (u.s.c.). Otherwise we can take $o_{\sup}(u) = \limsup_{u' \to u} o(u')$. Finally, we say that any property holds *near* x if it holds for all x' in some neighborhood (nbhd) of x.

2 Transformations of nonsmooth equations

To show how Newton's method can be applied to KKT-points or NCP's under different approaches, we write the related conditions as an equation $F(z) = 0$. In §4, we will see (starting from the Newton-regularity condition (3.7)) that *strong* regularity of F plays a crucial role. Though there are various characterizations of this property in the literature, we need an analytical one, related to the derivatives in Newton's method. Moreover, to compare and to understand the content of the Newton steps in all approaches, we have to deal with and to interpret solutions of perturbed equations $F^t(z) = 0$ $(F^0 = F)$ where $F^t - F^0$ is a "small" Lipschitz function.

For these reasons, we present here the necessary analytical background as some kind of a crash course on analysis of sensitivity. A certain overview on conditions for strong regularity has been given in [15].

Some generalized derivatives and function classes

Let $f \in C^{0.1}(R^n, R^m)$. We consider the following generalized derivatives (at x in direction u). They are based on contingent derivatives [1], Thibault's limit sets [35] and Clarke's generalized Jacobians [3]:

$$
\begin{aligned}
Cf(x)(u) &= \{w / w = \lim t^{-1}[f(x+tu') - f(x)] \text{ for certain } t \downarrow 0 \& u' \to u\} \\
Tf(x)(u) &= \{w / w = \lim t^{-1}[f(x'+tu') - f(x')] \text{ for certain } t \downarrow 0 \& (x', u') \\
&\qquad \to (x, u)\} \\
\delta f(x)(u) &= \{w / w = Au, A \in \delta f(x)\}.
\end{aligned}
$$

Let $\Theta = \{x \in R^n / Df(x)$ exists as Fréchet derivative$\}$ and put, following Clarke, $\delta_0 f(x) = H(Df, x, \Theta)$. Then $\delta f(x) = \operatorname{conv} \delta_0 f(x)$. Often, $\delta_0 f(x)$

is called the B-subdifferential and denoted by δ_B. Notice that $Cf(x) \subset Tf(x) \subset \delta f(x)$, and the inclusions may be strict. For $Tf \neq \delta f$, see [20].
Next *we copy* Clarke's definition to define $D^0 f(x)$ (by considering C^1-points only) and add some elementary facts.

Let Θ^1 consist of all x such that f is C^1 near x (C^1-points) and let $D^0 f(x) = H(Df, x, \Theta^1)$. The pair $(D^0 f, \Theta^1)$ fulfils $D^0 f \equiv Df$ on Θ^1, it holds $D^0 f(x) \subset \delta_0 f(x) \subset Tf(x)$ and, by continuity arguments only, one sees that $D^0 f(x) = H(Df, x, \Omega)$ for each open and dense subset Ω of Θ^1. However, the open set Θ^1 and $D^0 f(x)$ may be empty for arbitrary $f \in C^{0.1}(R^n, R^m)$.

If Θ^1 is *dense* in R^n, we call f *pseudo-smooth*. In our example, f obeys this property, and $Df(0) = 1$, $D^0 f(0) = \{\frac{1}{2}, 2\}, \delta_0 f(0) = \{\frac{1}{2}, 1, 2\}$, and $\delta f(0) = [\frac{1}{2}, 2]$.

Further, we recall the class of piecewise C^1 functions: f belongs to PC^1 if there is a finite family of C^1-functions f^s such that the *sets of active indices* $I(x) := \{s / f(x) = f^s(x)\}$ are not empty for all $x \in R^n$. We also write $f = PC^1[f^1, ..., f^N]$. The max-norm of R^n belongs to PC^1 while the Euclidean norm does not.

Kojima's function and Karush–Kuhn–Tucker points and NCP's

Given an optimization problem,

$$(2.1) \qquad \min f(x) \text{ s.t. } g_i(x) \le 0 \; i = 1, ..., m; \quad f, g_i \in C^2(R^n, R),$$

the function $F : R^{n+m} \to R^{n+m}$, used and perhaps first introduced by Kojima [17], as

$$\begin{aligned} F_1(x,y) &= Df(x) + \Sigma y_i^+ Dg_i(x) \qquad & y_i^+ &= \max\{0, y_i\} \\ F_{2j}(x,y) &= g_j(x) - y_i^- \qquad & y_i^- &= \min\{0, y_i\} \end{aligned}$$

characterizes the Karush–Kuhn–Tucker points (KKT-points) (x, y) via

$$\begin{aligned} &(x,y) \text{ is a KKT-point } \Rightarrow (x, y + g(x)) \text{ is a zero (critical point) of } F \\ \text{and } \; &(x,y) \text{ is a zero of } F \quad \Rightarrow (x, y^+) \text{ is a KKT-point.} \end{aligned}$$

Defining the $(1+2m)$-vector $N(y) = (1, y^+, y^-)^T$, and the $(n+m, 1+2m)$-matrix $M(x)$ by

$$M(x) = \begin{bmatrix} Df(x) & Dg_1(x) & .... & Dg_m(x) & 0....0 \\ g(x) & 0 & .... & 0 & -E_m \end{bmatrix}$$

($E_m = (m,m)$-unit matrix) the PC^1-function F becomes

$$F(x,y) = M(x)N(y). \tag{2.2}$$

The same settings are possible for additional equality constraints, we omit them for the sake of brevity. Replacing Df and Dg_i by other functions Φ and Ψ_i of related dimension and smoothness, F has been called in [16] the *generalized Kojima function.* For details on such functions, applications and proofs of the following facts, we refer to [20, 15, 16]. For studying F in the framework of PC^1 equations, we refer to [30].

Given $u, v : R^n \to R^n$, the complementarity problem (1.1) claims to find x such that

$$u(x) \geq 0, v(x) \geq 0 \text{ and } \langle u(x), v(x)\rangle = 0. \tag{2.3}$$

With $y \in R^n$, this can be written as

$$F_1 := u(x) - y^+ = 0; \quad F_2 := -v(x) - y^- = 0. \tag{2.4}$$

Here F is a generalized Kojima function, the matrix M has the form

$$M = \begin{bmatrix} u & -E & 0 \\ -v & 0 & -E \end{bmatrix}$$

and $y^* = u(x^*) - v(x^*)$ holds at any solution x^*.

Derivatives of Kojima's function

The usual product rule of differential calculus is a key property of generalized Kojima functions. More precisely, if $M \in C^{0.1}$ then

$$TF(x,y)(u,v) = [TM(x)(u)]N(y) + M(x)[TN(y)(v)] \tag{2.5}$$

(for CF, replace T by C). Note that (2.5) is not true for products of arbitrary Lipschitz functions or multifunctions. Here, the equation holds because N is *simple* in the following sense:

Given $\mu \in TN(y)(v)$ and any sequence of $\lambda \downarrow 0$,
there are $y' \to y$ such that
$\mu = \lim t^{-1}(N(y' + \lambda v) - N(y'))$.

For details we refer to [20] and [16]. The *simple*-property is also fulfilled for our perturbed Kojima functions below. Replacing TN by CN and setting $y' = y$, then being simple just means directional differentiability.

To find TN or CN, one has only to deal with the functions $c_i(y_i) = (y_i^+, y_i^-)$ $= (y_i^+, y_i - y_i^+)$, where $y_i^+ = \frac{1}{2}(y_i + |y_i|)$ is as difficult as the absolute value function. So one easily sees that $TN = \delta N$ since $Tc_i = \delta c_i$.
The assumption $M \in C^{0.1}$ allows the study of problems (2.1) with $f, g_i \in C^{1.1}(R^n, R)$ which is a proper generalization since Hessians do not exist.

First, let $M \in C^1$. Now (2.5) yields $TF = \delta F$, and shows, after the related calculation, that $\delta F(x, y)$ consists of all matrixes $J(r)$ of the type

$$\text{(2.6)} \quad \begin{bmatrix} D_x F_1(x,y) & r_1 Dg_1(x) & \dots & r_i Dg_i(x) & \dots & r_m Dg_m(x) \\ Dg_i(x)^{ad} & 0 & \dots & -(1-r_i) & \dots & 0 \end{bmatrix}$$

where $r_i = 0$ if $y_i < 0$, $r_i = 1$ if $y_i > 0$ and $r_i \in [0,1]$ if $y_i = 0$; briefly $r \in R_T(y)$.

Note that the given r_i form just δy_i^+ at the current point y_i. The products $w = J(r)(u,v)^T, r \in R_T(y)$ form precisely the set $TF(x,y)(u,v)$. Concerning first investigations of δF we refer to [11]. For the NCP, these matrices $J(r)$ attain the same form (we write down the rows):

$$\text{(2.7)} \quad \begin{array}{l} Du_i(x) \quad 0 \dots - r_i \qquad\quad \dots 0 \text{ (row } i, \quad -r_i \text{ at column } n+i), \\ -Dv_i(x)\; 0 \dots - (1-r_i) \dots 0 \text{ (row } n+i, -(1-r_i) \text{ at column } n+i), \end{array}$$

again with $r \in R_T(y), y = u(x) - v(x)$. Setting $R_c(y,v) = \{r / r \in R_T(y)$ and $r_i = 1$ if $(y_i = 0$ and $v_i > 0)$, $r_i = 0$ if $(y_i = 0 \,\&\, v_i \le 0)\}$, the same products $w = J(r)(u,v)^T$, for $r \in R_c(y,v)$, form the set $CF(x,y)(u,v)$, which is a singleton (the usual directional derivative) since $M \in C^1$. Having $M \in C^{0.1}$, the elements w (for fixed r) become *sets* according to (2.5). The Hessian matrix $D_x F_1$ in (2.6) must be replaced by $T_x F_1$ (or $C_x F_1$), and as already mentioned, $TF \ne \delta F$ may happen.

Regularity conditions

Strong regularity of $h \in C(R^n, R^m)$ at $x \in R^n$ in Robinson's sense [31] (being regularity in [3]) requires that, for certain nbhds U and V of x and $h(x)$, respectively, the restricted inverse $h^{-1} : V \to U$ is well-defined and locally Lipschitz (this implies $m = n$).

If, less restrictive,

$$\operatorname{dist}(x', h^{-1}(y')) \le L \operatorname{dist}(y', h(x'))\ \forall\, x' \in U \text{ and } y' \in V$$

holds with some fixed L, then h is called *metrically regular* at x.

One says that h^{-1} is *locally upper Lipschitz* at x, if L, U and V exist in such a manner that

$$U \cap h^{-1}(y') \subset x + L\|y' - h(x)\|B \quad \forall y' \in V.$$

Strong regularity of F, assigned to (2.1) or (2.3), claims (locally) the existence, uniqueness and Lipschitz behavior of the primal-dual solutions $(x_{a,b}, y_{a,b})$ of

$$\min f(x) - \langle a, x\rangle \text{ s.t. } g(x) \le b$$

or of the solutions $x_{a,b}$ of

$$u(x) \ge a, v(x) \ge b \;\&\; \langle u(x) - a, v(x) - b\rangle = 0,$$

respectively. In this case, we also call the related problem strongly regular at the given point.

Theorem 21. *Let $h \in C^{0.1}(R^n, R^n)$.*

(i) *h is strongly regular at x if and only if $Th(x)$ is injective (i.e. $0 \notin Th(x)(R^n\backslash\{0\})$*

(ii) *h^{-1} is locally upper Lipschitz at x if and only if $Ch(x)$ is injective (in the same sense)*

Concerning statement (i), we refer to [20]; concerning (ii), we refer to [14] where also the multivalued case has been considered.

Upper Lipschitz criteria for maps h^{-1} which assign, to a parameter, the *stationary* points of a $C^{1.1}$ optimization problem (2.1), have been derived in [16]. Conditions for *metric regularity* (also called openness with linear rate [28]) can be found in [1, Chapter 7.5] in terms of Ch, in [25] in terms of co-derivatives and in [23] (where both derivatives have been used).

Let us return to $h = F$ now.

For $M \in C^1$, injectivity of Th means that all matrices $J(r), r \in R_T(y)$ in (2.6) and (2.7), respectively, are non-singular. This is the sufficient condition of Clarke's inverse function Theorem [3]. In complementarity theory, one usually works with smaller matrices $C(r)$, defined by combinations of Du_i and Dv_i. The bridge to these matrices establishes the following lemma.

Lemma 22.

(i) *For any $r \in R^n$, the matrix $J(r)$ in (2.7) is singular if and only if the matrix $C(r)$ with rows $C_i(r_i) = (1 - r_i)Du_i(x) + r_iDv_i(x)$ is singular.*

(ii) *The NCP is strongly regular at x^* if and only if the related matrices $C(r)$ are non-singular for all $r \in R_T(u(x^*) - v(x^*))$.*

The ***proof*** of (i) requires only to substitute nontrivial zeros, while (ii) follows via Theorem 2.1. ■

For $M \in C^{0.1}$, injectivity of TF is weaker than non-singularity of δF. In addition, *metric* and *strong regularity* of F coincide as long as $M \in C^1$, [6], but not for $M \in C^1$, [22].

Lipschitzian perturbations and penalty-barrier functions

Metric and strong regularity are persistent under small Lipschitzian perturbations of any continuous function h (even for quite general multifunctions). We consider here equations $h + g^1 = 0$, $h + g^2 = 0$, where $h \in C(R^n, R^m)$ and $g^1, g^2 \in C^{0.1}(R^n, R^m)$.

Theorem 23. *Let h be metrically regular at a zero x^0 and let $g^k (k = 1, 2)$ have on some nbhd U of x^*, (smallest) Lipschitz rank $L(g^k)$ and sup-norm $S(g^k)$. Then, provided that the local $C^{0.1}$-norms $|g^k|_U = \max\{L(g^k), S(g^k)\}$ are small enough, there is a second nbhd Ω of x^* and a constant K such that, to each zero x^1 of $h + g^1$ in Ω, there is a zero x^2 of $h + g^2$ satisfying $\|x^1 - x^2\| \le K \|g^1(x^1) - g^2(x^1)\|$.*

For proofs and estimates of $K\&\Omega$, cf. [23] and (a bit less general) [4] and [5]. If h is even strongly regular, then x^1 and x^2 are unique whenever $|g^k|_U$ are small enough. Thus, the solutions $x = x(g)$ of $h + g = 0$ are locally Lipschitz, measured by the sup-norm $S(g^2 - g^1)$ on U. This follows also (by the proofs) from [31].

Perturbations of Kojima's function may be induced by parametrizations of problems (2.1). Then, only $M(\cdot)$ will vary. In the following we change N,

$$N_t(y) = (1,\ y_1^+, \dots, y_m^+,\ y_1^- + t_1\, y_1^+, \dots, y_m^- + t_m\, y_m^+)^{ad}.$$

This leads us, for (2.1), to a parametric Kojima function F^t and system

$$\text{(2.8)} \qquad \begin{aligned} F_1 &= Df(x) + \sum y_i^+ Dg_i(x) = 0, \\ F_{2i}^t &= g_i(x) - y_i^- - t_i y_i^+ = 0. \end{aligned}$$

For applying Theorem 2.3 to the current perturbations, it suffices to suppose $f, g_i \in C^1$. For computing, with fixed t, the derivatives of F^t by the rule

(2.5), one needs $f, g_i \in C^{1.1}$ to ensure that $M \in C^{0.1}$. Compared with $J(r)$ in (2.6), now the terms $-(1-r_i)$ in the lower right diagonal must be replaced by $-(1-r_i+t_ir_i)$, only. This will be used in §5.

Quadratic penalties: Suppose $t_i > 0$ for all i.

Let (x, y) solve (2.8).

$$\begin{array}{llcl} \text{If } y_i \le 0, \text{ then it follows} & y_i^+ = 0 & \text{and} & g_i(x)^+ = 0. \\ \text{If } y_i > 0, \text{ then it follows} & g_i(x) = t_i y_i^+ & \text{and} & y_i^+ = t_i^{-1} g_i(x)^+. \end{array}$$

Hence, we obtain in both cases $0 = F_1 = Df(x) + \sum t_i^{-1} g_i(x)^+ Dg_i(x)$, i.e. x is a stationary point of the penalty function $P_t(x) = f(x) + \frac{1}{2}\sum t_i^{-1}[g_i(x)^+]^2$. Conversely, if x is stationary for $P_t(x)$, then (x, y) with

$$y_i = t_i^{-1} g_i(x) \text{ for } g_i(x) > 0 \text{ and } y_i = g_i(x) \text{ for } g_i(x) \le 0$$

solves (2.8).

Logarithmic barriers: Let $t_i < 0$ for all i.

Now, the second equation of (2.8), $g_i(x) = y_i^- + t_i\, y_i^+ (\le 0)$, implies feasibility of x in (2.1). Let (x, y) solve (2.8).

$$\begin{array}{lrclcl} \text{If } y_i \le 0, \text{ then} & g_i(x) & = & y_i^- & \text{and} & y_i^+ = 0. \\ \text{If } y_i \le 0, \text{ then} & g_i(x) & = & t_i\, y_i^+ & \text{and} & y_i^+ = t_i^{-1} g_i(x)^-. \end{array}$$

Setting $J = \{i / y_i > 0\}$ we thus observe

$$0 = F_1 = Df(x) + \sum_{i \in J} t_i^{-1} g_i(x)^- Dg_i(x).$$

Hence, the point x is feasible for (2.1), fulfils $g_i(x) < 0 \ \forall i \in J$, and is stationary (not necessarily minimal !) for the function

$$Q_i(x) = f(x) + \frac{1}{2} \sum_{i \in J} t_i^{-1}\, [g_i\,(x)^-]^2.$$

Conversely, having the latter properties, the point (x, y) with

$$y_i = t_i^{-1} g_i(x)^- (i \in J) \text{ and } y_i = g_i(x)\ (i \notin J)$$

solves (2.8). The following transformation, due to A. Ponomarenko, establishes the bridge to usual logarithmic barrier function:

For $i \in J$, the terms $g_i(x)^- Dg_i(x)$ coincide with $g_i(x)^2 D(\text{In } (-g_i(x))$. So we see that

$$t_i^{-1} g_i(x)^- Dg_i(x) = t_i^{-1} g_i(x)^2 D(\text{In } (-g_i(x)) = t_i y_i^2 D(\text{In } (-g_i(x)).$$

Accordingly, the actual x is also stationary for the function

$$B_t(x) = f(x) - \sum_{i \in J} |t_i|\, y_i^2 \text{ In } (-g_i(x)).$$

In this manner, zeros of the perturbed Kojima quation (2.8) and critical points of well-known auxiliarly functions find a natural interpretation.

Under strong regularity of (2.1) at a critical point (x^*, y^*), we can say something more:

(i) The solutions (x_t, y_t) of (2.8) are, for small $\|t\|$, locally unique and Lipschitz since the maps $y_i \mapsto t_i y_i^+$ are small Lipschitz functions in the sense of Theorem 2.3. So, it holds

$$\|(x_{s,}, y_s - (x_t, y_t)\| \le L\|s - t\| \text{ for all } s, t \text{ near the origin.}$$

This inequality now compares solutions of *different methods* in a Lipschitzian manner.

(ii) Further, one may mix the signs of the t -components and obtains similarly stationary points for auxiliary functions containing both penalty and barrier terms. For example, given x, y, it is quite natural to put $t_i < 0$ if $g_i(x) < 0$ and $t_i > 0$ if $g_i(x) > 0$ with absolute values depending on $\|F(x, y)\|$.

Moreover, similar arguments lead us to estimates of not unique critical points (x_t, y_t) under metric regularity of F at (x^*, y^*) or to estimates of $(x_t, y_t) - (x^*, y^*)$ under the upper Lipschitz property of F^{-1} at this point.

3 Continuous differentiability, Newton's method and semismoothness

Newton maps

If f is *continuously differentiable* near x^*, the two approximations

$$f(x) - f(x^*) - Df(x^*)(x - x^*) = o_1(x - x^*)$$

and

$$f(x) - f(x^*) - Df(x)(x - x^*) = o_2(x - x^*)$$

may be replaced by each other, because both, o_1 and o_2 satisfy $o_k(u)/\|u\| \to 0$. For $f(x) = x^2 \sin x^{-1}$ ($f(0) = 0), o_1$ exists, not so o_2. For $f(x) = |\, x\,|$, the reverse situation occurs. When applying solution methods, we need (or have) Df at points x *near* a solution x^*. So the other approximation becomes important and, if $f \notin C^1$, the condition must be adapted.

Let X, Y be Banach spaces, $f : X \to Y$ be any function and $Rf : X \to \mathrm{Lin}(X, Y)$ be locally bounded. We say that Rf is a *Newton function of f at x^* if*

$$f(x^* + u) - f(x^*) - Rf(x^* + u)u \in o(u)B. \tag{3.1}$$

The notation will be motivated by Lemma 3.2. At this moment, we regard the actual property as a version of *continuous differentiability for nonsmooth functions.*

Notice that the function Rf may be arbitrary at the point x^* and is not uniquely defined at $x \neq x^*$, too.

If Rf satisfies (3.1), then it is a Newton function for all g at x^*, whenever $g(x) = f(x) + o(x - x^*)$. Here, $o = g - f$ is not necessarily small in the $C^{0.1}$-norm used in Theorem 2.3.

Newton functions at x^* are selections of locally bounded maps $M : X \rightrightarrows \mathrm{Lin}\ (X, Y)$ such that

$$\begin{aligned} \emptyset \neq M(x^* + u)u &:= \{Au/A \in M(x^* + u)\} \\ &\subset f(x^* + u) - f(x^*) + o(u)B. \end{aligned} \tag{3.2}$$

Accordingly, we call M a *Newton map.* This property is invariant if one forms the union or the convex hull of two Newton maps.

Examples. If $f \in C^1(R^n, R^m)$ and B^{nm} denotes the unit ball of (n, m)-matrices, then

$$M(x) = \{Df(x)\} \text{ and } M(x) = Df(x) + \|f(x)\|B^{nm}$$

are Newton maps at x^*. For $f = PC^1[f^1, \ldots, f^N]$ and $f(x^*) = 0$, one may put

$$M(x) = \{Df^i(x)/i \in J(x)\},$$

where $J(x) = \{i / \|f^i(x) - f(x)\| \leq \|f(x)\|^2\}$. Indeed, for $\|u\|$ sufficiently small, the index sets fulfil $J(x^* + u) \subset J(x^*)$. Thus,

$$\begin{aligned}
&f(x^*+u) - f(x^*) - Df^i(x^*+u)u \\
&\in f^i(x^*+u) - f^i(x^*) - Df^i(x^*+u)u + \|f(x^*+u)\|^2 B \\
&\subset o_i(u)B + L^2\|u\|^2 B.
\end{aligned}$$

So $o(u) = L^2\|u\|^2 + \max_i o_i(u)$ satisfies (3.2).

Particular statements are valid for $f \in C^{0.1}(R^n, R^m)$:

(i) To define a Newton map M_0, it suffices to know a locally bounded map $M : X \rightrightarrows \text{Lin}\,(X,Y)$ satisfying (3.2) for all u in a dense subset $\Omega \subset R^n$, because $M_0(x) := H(x, M, \Omega)$ satisfies (3.2) for all u by continuity arguments (with $o = o_{\sup}$) after applying (3.2) to u' with $x^* + u' \in \Omega$.

(ii) Moreover, due to $f(x^*+u) - f(x^*) \subset Cf(x^*)(u) + o(u)B$ (this can be easily shown by using finite dimension) and by the relations between C, T and δ, one sees that (3.2) implies, with possibly new o-type function,

$$\begin{aligned}
&M(x^*+u)u \subset Cf(x^*)(u) + o(u)B \\
&\subset Tf(x^*)(u) + o(u)B \subset \delta f(x^*)(u) + o(u)B.
\end{aligned} \tag{3.3}$$

However, f is not necessarily directionally differentiable (see Lemma 3.1), and M has not to be a so-called approximate Jacobian [10]. Condition (3.1) is a weak one, and Newton functions satisfy a common chain rule.

Lemma 31. (*existence and chain rule for Newton functions*)

(i) *Every $C^{0.1}$-function $f : X \to Y$ (Banach spaces) possesses, at each x^*, a Newton function Rf being (locally) bounded by a local Lipschitz constant L for f near x^*.*

(ii) *Let $h : X \to Y$ and $g : Y \to Z$ be $C^{0.1}$ with Newton functions Rh at x^* and Rg at $h(x^*)$. Then, $Rf(x) = Rg(h(x))Rh(x)$ is a Newton function of $f(\cdot) = g(h(\cdot))$ at x^*.*

Proof.

(i) Given $u \in X\backslash\{0\}$ there is a linear operator $\Phi_u : X \to Y$ with $\Phi_u(u) = f(x^*+u) - f(x^*)$. By Hahn-Banach arguments (extension of Φ_u, from the line $ru, r \in R$ onto the whole space), Φ_u exists with bounded norm $\|\Phi_u\| \leq \|f(x^*+u) - f(x^*)\| / \|u\|$. Hence $\|\Phi_u\| \leq L$ for small $\|u\|$, for other u define $\Phi_u = 0$. So it suffices to put $Rf(x^*+u) = \Phi_u$ and $o(u) = 0$.

(ii) By taking the "derivatives" at x instead of x^*, the straightforward proof is the same as for Fréchet derivatives. So we omit the details. ∎

The function Rf, defined in this proof does not use any local behavior of f near x, and Rf depends on x^* which is often an unknown solution. So one cannot directly apply statement (i) of Lemma 3.1 for solution methods. One has to find Rf satisfying (3.1) without using x^*. Nevertheless, *having* Rf, it can be applied like Df for Newton's method.

Newton's method based on linear auxiliary problems

For computing a zero x^* of h, Newton's method is determined by the iterations

$$x^{k+1} = x^k - A_k^{-1} h(x^k),$$

where $A_k = Dh(x^k)$ is supposed to be invertible. The locally superlinear convergence means that, for $\|x^0 - x^*\|$ small enough, we have

$$x^{k+1} - x^* = o(x^k - x^*), \tag{3.4}$$

which is, after substituting x^{k+1} and multiplying with A_k,

$$A_k(x^k - x^*) - A_k o(x^k - x^*) = h(x^k) - h(x^*). \tag{3.5}$$

The equivalence between (3.4) and (3.5) is still true if one defines,

$$x^{k+1} = x^k - A^{-1} h(x^k), \quad A \in M(x^k), \tag{3.6}$$

where $M(x^k) \neq \emptyset$ is any given set of invertible linear maps. Then, x^{k+1} depends on A. So we should state more precisely that (3.4) should hold *independently* on the choice of $A \in M(x^k)$. Having *uniformly bounded* $\|A\| \leq K^+$ and writing $x = x^k$, (3.5) implies that h satisfies a *pointwise Lipschitz condition* at x^*:

$$\|h(x) - h(x^*)\| \leq (1 + K^+)\|x - x^*\| \quad \text{ for } x \text{ near } x^*.$$

Having *uniformly bounded* $\|A^{-1}\| \leq K^-$, now (3.5) implies

$$\|h(x) - h(x^*)\| \geq (1 + K^-)^{-1}\|x - x^*\| \quad \text{ for } x \text{ near } x^*.$$

This restricts h in a canonical manner and tells us that h^{-1} is locally upper Lipschitz at $(0, x^*)$. In what follows we suppose that constants K^+ and K^- exist such that

$$\|A\| \leq K^+ \text{ and } \|A^{-1}\| \leq K^- \text{ for all } A \in M(x^* + u) \text{ and small } \|u\|. \tag{3.7}$$

Then, interpreting $o(\cdot)$ as a real-valued, non-negative function and setting $u = x^k - x^*$, condition (3.5) takes the equivalent form

$$Au \in h(x^* + u) - h(x^*) + o(u)B \text{ for all } A \in M(x^* + u) \tag{3.8}$$

and describes – again equivalently – the local convergence of method (3.6) with order

$$\|x^{k+1} - x^*\| \leq K^- o(x^k - x^*) \text{ for all initial points } x^0 \text{ sufficiently close to } x^*. \tag{3.9}$$

But (3.8) is condition (3.2): M has to be a *Newton map of* h at x^*.

Lemma 32. (*convergence*) *Supposing* (3.7) *and* $M(\cdot) \neq \emptyset$, *the method* (3.6) *fulfils condition* (3.9) *if and only if* M *satisfies* (3.8) (*with the same* o). *The latter means that* M *is a Newton map of* h *at* x^*.

Proof. Note that the norms of $A_k o(x^k - x^*)$ in (3.5) are just bounded by $o(u)$ in (3.8). ■

To investigate convergence of Newton's method for $h \in C^{0.1}(X, Y)$, maps M satisfying (3.2) and Lemma 3.2 have been used in [19]. There, and in [32, 21], neither relations between M and δh nor the existence of $h'(x^*; \cdot)$ or finite dimension were needed for the interplay of the conditions (3.7), (3.8), (3.9) in accordance with Lemma 3.2.

Semismoothness

This notion, based on Mifflin [24], has been introduced for $h \in C^{0.1}(R^n, R^m)$: h is *semismooth* at x^* if $M = \delta h$ is a Newton map at x^*, c.f. [26] and [29] and many subsequent papers.

Often, directional derivatives $h'(x^*; u)$ (provided they exist) replace $h(x^* + u) - h(x^*)$ in (3.2) which yields equivalenty (e.g. in [7]) the condition $\delta h(x^* + u)u \subset h'(x^*; u) + o(u)B$. In other papers, M is a map that approximates δh and h satisfying the related condition (3.2) is called *weakly* semismooth. By the Lemma, we have to determine those functions which allow us to find a computable Newton map M, in particular the semismooth ones. The related concrete function classes, studied in the recent literature, are not very big: PC^1-functions and NCP-functions (mainly composed by norms and PC^1-functions). Before showing how Newton's method can be applied to the class locPC^1 defined below, we recall conditions for semismoothness given in [24, Proposition 3, Theorem 2].

Theorem 33. *Convex functions $f : R^n \to R$ and maximum functions $f(x) = \max_{y\in Y} g(x,y)$ of C^1-functions g over compact Y are semismooth.*

As a consequence, each DC-functional f (difference of convex functions) is semismooth. The same is valid (cf. Lemma 3.1 (ii)) whenever $f : R^n \to R^m$ has DC components since

$$\emptyset \neq \delta f(x) \subset (\delta f_i(x), \dots, \delta f_m(x))$$

However, the example in the introduction demonstrates that being pseudo-smooth is not enough for semismoothness.

Dense subsets and approximations

If M satisfies (3.7) and (3.8) for all u in a *dense subset* U of R^n, then $M_0(x) = H(M, x, U)$ is a Newton map which also fulfils (3.7). Again, evidently, if some map M satisfies (3.7) and (3.8) then (3.7) holds for each M' with $\emptyset \neq M' \subset M$, and (3.8) holds for each M' with $\emptyset \neq M' \subset \operatorname{conv} M$.

Further, one may replace M satisfying (3.7) and (3.8) by any map N as far as

$$\emptyset \neq N(x) \subset M(x) + O(x - x^*)B_{L(X,X)},$$

where $B_{L(X,Y)}$ denotes the unit ball in $\operatorname{Lin}(X,Y)$. In particular, let us consider

$$N(x) = M(x) + \|h(x)\| B_{L(X,Y)}, \tag{3.10}$$

which permits us to approximate elements of $M(x)$ with accuracy $\|h(x)\|$. Let L be a Lipschitz rank of h near x^*.

Remark. Using N, condition (3.7) is still satisfied with each $K_N^- > K^-$. The function $o(\cdot)$ in (3.8) changes only by $L\|\cdot\|^2$. Thus, the replacement (3.10) will not disturb locally quadratic (or worse) convergence of method (3.6).

Indeed, both calculations are elementary:

Let $A \in N(x)$ and let x be close to x^* and such that $\|h(x)\| < 1/K^-$. Then $v = Au$ yields, by writing $A = A_M + A_h$ with $A_M \in M(x)$ and $\|A_h\| \leq \|h(x)\|$:

$$\|v\| \geq ((1/K^-) - \|h(x)\|)\, \|u\|,$$

hence

$$\|A^{-1}\| \leq ((1/K^-) - \|h(x)\|)^{-1} = K^-(1 - K^-\|h(x)\|)^{-1}.$$

The latter is smaller than K_N^- for x near x^*. Further, (3.8) applied to M ensures, for every $A \in M(x^* + u)$ and $C \in B_{L(X,Y)}$:

$$\begin{aligned} Au + \|h(x)\|Cu &\in h(x^* + u) - h(x^*) + (o(u) + \|h(x)\| \, \|u\|)B \\ &\subset h(x^* + u) - h(x^*) + (o(u) + L\|u\|^2)B. \end{aligned}$$

We are now going to describe further functions having applicable Newton maps.

Pseudo-smoothness and $D^0 f$

Let $f \in C^{0.1}(R^n, R^m)$ be pseudo-smooth and Θ^1 be its C^1-set. Then, selections $Rf \in D^0 f$ are natural candidates for being Newton functions, and $D^0 f = Df$ on Θ^1.

Lemma 34. (*selections of* $D^0 f$) *If f is pseudo-smooth and some selection Rf of $D^0 f$ is a Newton function at x^*, then $D^0 f$ is a Newton map at x^* and*

$$Cf(x^*)(u) \subset D^0 f(x^*)u. \tag{3.11}$$

Proof. The first statement holds again by continuity arguments (using $o = o_{sup}$). We prove (3.11). Let $a \in Cf(x^*)(u)$, i.e. $a = \lim a(t)$ where $a(t) = t^{-1}[f(x^* + tu) - f(x^*)]$ for certain $t \downarrow 0$. The point $a(t)$ can be approximated by $b(t) := t^{-1}[f(x^* + tu(t)) - f(x^*)]$ such that

$$\|u(t) - u\| < t, \; x^* + tu(t) \in \Theta^1 \text{ and } \|b(t) - a(t)\| < t.$$

Because of (3.1), it holds $b(t) \in Df(x^* + tu(t))u(t) + t^{-1}o(tu(t))B$, which yields the assertion since $a = \lim a(t) = \lim b(t) \in D^0 f(x^*)u$ as $t \downarrow 0$. ■

Our example presents a pseudo-smooth, directionally differentiable real function such that $D^0 f(x^*) \neq \delta_0 f(x^*)$, (3.11) fails to hold though $Df(x^*)$ exists, and neither $D^0 f$ nor $\delta_0 f$ does contain a Newton function at $x^* = 0$. By $f(x) = |x|$ one sees that (3.11) does not hold as equation.

Locally PC^1 functions

Let f be pseudo-smooth. We call f *locally* PC^1 (*and write* $f \in \text{loc } PC^1$) if there is an open and dense subset $\Omega \subset R^n$ such that f is C^1 on Ω and the following holds: There exists a finite collection of open sets $U^s \subset R^n$ and of continuous functions $f^s : R^n \to R^m$ satisfying

(i) f^s is C^1 on U^s, and $Df^s(\cdot)$ is uniformly continuous on $U^s \cap K$ for bounded sets K, and

(ii) for each $x \in R^n$ there exists $r > 0$ such that, given $y \in \Omega_r := \Omega \cap (x + rB)$, one finds some s with rel int conv $\{x, y\} \subset U^s, f^s(x) = f(x), f^s(y) = f(y)$ and $Df^s(y) = Df(y)$.

In comparison with (proper) PC^1 functions, we do not claim that f^s is C^1 on the whole space.

Lemma 35. *The Euclidean norm of a linear function $f(y) = \|Ay\|$ and all functions $f \in PC^1$ are locally PC^1.*
A pseudo-smooth function f is locally PC^1 if there is a covering $\{P^s/s = 1, \dots, N\}$ of R^n by convex polyhedrons P^s such that f is C^1 and Df is uniformly continuous on int$P^s \cap K$ *for all bounded sets $K \subset R^n$.*
In addition, if g and h are locally PC^1 and $\Phi \in C^1$, then $f(x) = \Phi(g(x), h(x))$ is again locally PC^1 (provided that g, h, Φ are of appropriate dimension).

Proof. Euclidean norm: If $A \neq 0$ put $\Omega = R^n \setminus \ker A, U^1 = \Omega, f^1 = f, r = 1$ if $x \in \ker A$ and $r = \frac{1}{2}$ dist $(x, \ker A)$ otherwise.

PC^1 : Let $f = PC^1[f^1, \dots, f^N]$ and $I(y) = \{s/f^s(y) = f(y)\}$. It suffices to put $\Omega = \cup_s$ int $I^{-1}(s)$ and $U^s = R^n$. The density of Ω can be shown by contradiction since $R^n = \cup_s I^{-1}(s)$.

Covering: Define $f^s = f, U^s =$ int $P^s, \Omega = \cup U^s$ and take r small enough such that, for $0 < \varepsilon < r$, the set $S(\varepsilon) := \{s/(x + \varepsilon B) \cap U^s \neq \emptyset\}$ is constant. The existence of r is ensured since all P^s are polyhedrons.

Φ : With the related sets and radii assigned to g and h, one may put $\Omega = \Omega(g) \cap \Omega(h)$, $U^{s\sigma} = U^s(g) \cap U^\sigma(h)$, $f^{s\sigma} = \Phi(g^s, h^\sigma)$ and $r = \min\{r(g), r(h)\}$. ■

The main motivation of the above definitions presents

Theorem 36. (*Newton maps of locally PC^1 functions*)
Let f be a locally PC^1 function and $x^* \in R^n$. Then

(i) *$M = D^0 f$ is a Newton map of f at x^*.*

(ii) *The function $o(\cdot)$ in (3.2) can be taken as $o(u) = \|u\| O(\|u\|)$ provided that both $O(\|u\|)$ is a modulus of uniform continuity for all functions $Df^s(\cdot)$ on U^s near x^* and $O(\cdot)$ is continuous.*

(iii) *For the composition $f = g(h(x))$ of locally PC^1 functions g and h, $M(x) = D^0 g(h(x)) D^0 h(x)$ is a Newton map of f at x^*.*

Remark. *Modulus of uniform continuity* means $\|Df^s(x') - Df^s(x'')\| \leq O(\|x' - x''\|) \forall x', x'' \in U^s$ near x^*. In particular, if all Df^s are globally Lipschitz on U^s, then $o(u) \leq K\|u\|^2$ holds for small $\|u\|$.

***Proof of Theorem* 3.6.** (i) and (ii): Given x^* let r define the ball x^*+rB in the definition of locPC^1 and let $y = x^* + u \in \Omega_r$. Using s according to the definition, we can integrate and estimate

$$\begin{aligned} f(y) - f(x^*) = f^s(y) - f^s(x^*) &= \int_0^1 Df^s(x^* + tu)\, u\, dt \\ \in \int_0^1 Df^s(y)u dt + \|u\| \sup_{0<t<1} &\|Df^s(x^* + tu) - Df^s(y)\| B. \end{aligned}$$

The supremum is bounded by $O(\|u\|)$. Since $Df(y) = Df^s(y)$, this guarantees

$$f(x^* + u) - f(x^*) - Df(x^* + u)u \in \|u\| O(\|u\|) B \ \ \forall x^* + u \in \Omega_r. \tag{3.12}$$

So (3.1) holds true, as far as $x^* + u$ belongs to a dense subset of $x^* + rB$. By density of Ω in Θ^1, (3.12) also holds for $D^0 f(x^* + u')$ at all $x^* + u' \in x^* + rB$, i.e.

$$f(x^* + u) - f(x^*) - D^0 f(x^* + u')u' \subset \|u'\| O(\|u'\|) B,$$

which verifies (i) and (ii). Finally, knowing (i), statement (iii) follows from Lemma 3.1. ■

Generalized and usual Newton method for PC^1 functions

Condition (3.8) also holds for all PC^1-functions h, if we put

$$M(x) = \{Dh^s(x) / s \in I(x)\}; \quad I(x) = \{s / h^s(x) = h(x)\}.$$

Condition (3.7) now means regularity of all matrices $Dh^s(x^*), s \in I(x^*)$. In that case, x^* is obviously an isolated zero of each C^1-function $h^s, s \in I(x^*)$. So, one may apply the *usual* Newton method to *any* fixed generating function $g = h^s, s \in I(x^0)$, provided that $\|x^0 - x^*\|$ is small enough such that $I(x^0) \subset I(x^*)$. This simplification is possible, if all generating functions h^s are *explicitly* known, e.g. for all NCP's with $(u, v) \in C^1$.

4 Some properties of NCP-functions

Preliminaries

NCP-functions are functions $g : R^2 \to R$ with $g^{-1}(0) = \{(s,t) \geq 0 / st = 0\}$.

They are used in order to formulate the NCP (2.3) as an equation

$$h(x) = 0 \quad ; \quad h_i(x) := g(z_i(x)) \tag{4.1}$$

where $z = (u,v)$ describes an NCP. The NCP is said to be (*strongly*) *monotone* if

$$\langle u(y) - u(x), v(y) - v(x) \rangle \geq \lambda \|y - x\|^2 \quad \forall x, y \in R^n,$$

where $\lambda \geq 0$ $(\lambda > 0)$ is a fixed constant. *A standard NCP* is defined by $v(x) = x$. Throughout, we suppose (at least)

$$\begin{aligned} &g \in \mathrm{loc}PC^1 \text{ with } C^1\textit{-set } \Theta^1(g) \\ &\text{and } z \in \mathrm{loc}PC^1 \text{ with } C^1\textit{-set } \Theta^1(z). \end{aligned} \tag{4.2}$$

By g_s, g_t we denote the partial derivatives of g on $\Theta^1(g)$. If $x \in \Theta^1(z)$, monotonicity yields (via $y = x + w$ and first-order approximation):

$$\lambda \|w\|^2 \leq \sum\nolimits_i (Dv_i(x)w)(Du_i(x)w).$$

The same remains true (consider limits for $x' \to x, x' \in \Theta^1(z)$) if the pairs $(Du_i(x), Dv_i(x)) = (R_i u(x), R_i v(x))$ are components of $Rz(x) \in D^0 z(x)$, i.e.

$$\lambda \|w\|^2 \leq \sum\nolimits_i (R_i u(x), w)(R_i v(x), w). \tag{4.3}$$

There are two principal possibilities of solving (4.1).

(i) minimize a so-called merit function, e.g.

$$q(x) = \frac{1}{2} \sum\nolimits_i h_i(x)^2 \tag{4.4}$$

by a descent method or

(ii) solve (4.1) directly by a Newton method.

Though also combinations of both ideas are possible, we regard these cases separately because they require different properties of g.

Case (i). Having $z \in C^1$, the function g should ensure that $q \in C^1$. This is true if g satisfies

$$\Theta^1(g) \cup g^{-1}(0) = R^2. \tag{4.5}$$

As a second requirement, $Dq(x) = 0$ should imply $q(x) = 0$. The latter cannot be ensured for all problems, but at least for *monotone standard NCP's.* Clearly, then g has to be monotone in a certain sense, too.

We call g *strongly monotone*, if $a\,b > 0\,\forall\,(a,b) \in D^0 g\,(s,t)$ and $(s,t) \in R^2\backslash g^{-1}(0)$.

Lemma 41. *Let g fulfil* (4.5) *and be strongly monotone. Further, let the NCP be monotone, $z \in C^1$ and $Dv(x)$ be regular. Then $Dq(x) = 0$ implies $q(x) = 0$.*

Proof. Given $\sigma = z(x)$, define w by

$$Dv_i(x)w \;=\; h_i(x)g_s(\sigma_i) \qquad\qquad (\text{if } h_i = 0,\ \text{put } h_i g_s = 0).$$

Then,

$$\begin{aligned} Dq(x)w &= \textstyle\sum h_i(x)g_s(\sigma_i)Du_i(x)w \;+\; \sum h_i(x)g_t(\sigma_i)Dv_i(x)w \\ &= \textstyle\sum (Dv_i(x)w)(Du_i(x)w) \;+\; \sum h_i(x)^2 g_t(\sigma_i)g_s(\sigma_i). \end{aligned} \tag{4.6}$$

The first sum is non-negative by (4.3), the second one is positive if and only if $q(x) > 0$. ■

Remarks.

(i) For *strongly monotone NCP's*, the same is true if g is *monotone* in the weaker sense:

$$ab \ge 0 \text{ and } a \ne 0 \quad \forall (a,b) \in D^0 g(s,t) \text{ and } (s,t) \in R^2 \backslash g^{-1}(0), \tag{4.7}$$

because now (4.6) and $h(x) \ne 0$ ensure $w \ne 0$ and $0 < \lambda\|w\|^2 \le \sum(Dv_i(x)w)(Du_i(x)w)$.

(ii) For $z \in \mathrm{loc}PC^1$, one may replace Du and Dv by a Newton function as in (4.3) and may define $Rq(x) := \sum h_i(x)[g_s(\sigma_i)R_i u(x) + g_t(\sigma_i)R_i v(x)]$. Then $Rq(x) = 0$ implies $q(x) = 0$ by the same arguments.

(iii) Without supposing the smoothness (4.5) one can replace $(g_s(\sigma_i), g_t(\sigma_i))$ by pairs $(a_i, b_i) \in D^0 g(\sigma_i)$ and comes to the same conclusion.

Knowing that $q = 0$ if $Dq = 0$, all first order methods for minimizing a C^1- or a $C^{1.1}$-function may be applied to q. NCP-functions g satisfying the assumptions of the Lemma can be chosen arbitrarily smooth. One may also apply methods of nonsmooth convex optimization for minimizing $Q(x) = \sum_i |h_i(x)|$ as long as $G = |g|$ is sublinear and the NCP is monotone. Then we have at C^1 points:

$$\begin{aligned} Q &= \sum_i \langle DG(z_i(x)), z_i(x)\rangle, \\ DQ(x) &= \sum_i DG(z_i(x))Dz_i(x) \end{aligned}$$

and $Q+DQ(x)w = \sum_i \langle DG(z_i(x)), z_i(x)+Dz_i(x)w\rangle$. Directions w satisfying $\langle DG(z_i(x)), z_i(x) + Dz_i(x)w\rangle = 0\,\forall\, i$ will just appear as Newton directions in the next case, c.f. formula (4.17).

Case (ii). Now we require that the NCP function g satisfies

$$g \in \ \text{loc}PC^1 \text{ and } 0 \notin \Theta^1 \tag{4.8}$$

$$g \quad \text{is positively homogenous} \tag{4.9}$$

$$g^{-1}(0)\backslash\{0\} \subset \Theta^1 \tag{4.10}$$

$$Dg(\sigma) \geq 0 \quad \text{and} \quad Dg(\sigma) \neq 0 \qquad \forall \sigma \in \Theta^1, \tag{4.11}$$

where $\Theta^1 = \Theta^1(g)$.

If $0 \in \Theta^1$, then $Dg(0) = 0$, hence $Dh_i(x) = 0$ if $z_i(x) = 0$ and $z \in C^1$. So system (4.1) degenerates if strict complementarity ($z_i(x^*) \neq 0\ \forall i$) does not hold. By (4.9), g belongs to the simplest functions satisfying $0 \notin \Theta^1(g)$. Condition (4.10) guarantees that h is C^1 at strictly complementary solutions x^*. Condition (4.11), consistent with the assumption of Lemma 4.1, avoids singular derivatives of h for strictly monotone NCP's, c.f. Theorem 4.4.

Let $pNCP$ be the cone of NCP-functions g satisfying (4.8) – (4.11).

Properties and construction of $g \in pNCP$

Due to (4.9), we have

$$Dg(\sigma) = Dg(\lambda\sigma) \qquad \forall \lambda > 0 \qquad \forall \sigma \in \Theta^1. \tag{4.12}$$

Hence, one easily derives that

$$\begin{gathered} D^0g(0) = clDg(\Theta^1), g(\sigma) = Dg(\sigma)\sigma\forall\sigma \in \Theta^1 \\ \text{and } \{g(\sigma)\} = D^0g(\sigma)\sigma. \end{gathered} \tag{4.13}$$

As a consequence, there is a positive lower bound for all gradient norms:

(4.14) $\exists p > 0$ such that $\| Dg(\sigma)\| \geq p \,\forall\, \sigma \in \Theta^1$ and $\inf \|D^0 g(0)\| \geq p$.

Moreover, $Dg(e^1) = \lambda e^2$ and $Dg(e^2) = \mu e^1$ hold with certain $\lambda, \mu > 0$.

Examples. Put $g = g_{\min}(s,t) := \min\{s,t\}$, an often used concave standard function, or $g = g_{dist}(s,t) := \operatorname{dist}((s,t), M)$ and, to satisfy (4.11), change the sign of g on $R^2 \backslash R^{2+}$.

One can define g via any norm of R^2, such that its unit sphere bd$B$ is piecewise smooth, has no kinks at the positive axes and fulfils $e^1 + e^2 \notin B$, $B \subset e^1 + e^2 - R^{2+}$ and $\{e^1, e^2\} \subset \operatorname{bd}B$. Setting $\Psi(p) = e^1 + e^2 - p$ for $p \in \operatorname{bd}B$ and $g(\lambda p) = \lambda \langle \Psi(p), p\rangle$ for $\lambda \geq 0$, one easily infers that g belongs to $pNPC$. With the Euclidean ball, one obtains the strongly monotone concave function $g_2(s,t) := s + t - \|(s,t)\|$, used e.g. in [13] (for penalization), [7] and [12]. In addition, g can be defined (and each $g \in pNCP$ can be written) by means of a real 2π-periodic locPC^1 function ϕ with zeros at 0 and $\pi/2$: $g(s,t) = r\phi(\omega)$, where (r,ω) are the polar coordinates of (s,t). Then

$$Dg(s,t) = r^{-1}(s\phi(\omega) - tD\phi(\omega), t\phi(\omega) + sD\phi(\omega))$$
$$\text{for radius } r > 0 \text{ at } (s,t) \in \Theta^1.$$

In particular, the natural setting $\phi(\omega) = \sin(2\omega)$ for $0 \leq \omega \leq \pi/2$ with the symmetric extension $\phi(\omega) = -3\phi((2\pi - \omega/3)$ for $\pi/2 \leq \omega \leq 2\pi$ defines a function g_Φ which satisfies, like g_2, all the already mentioned conditions.

Lemma 42. *For $g \in pNCP$, it holds $\lim g_s(\sigma)/g(\sigma) = 0$ as $\sigma \to e^1$ in Θ^1 and $\lim g_t(\sigma)/g(\sigma) = 0$ as $\sigma \to e^2$ in Θ^1.*

Proof. We apply the polar representation of Dg, put $\sigma = (s,t) = r(\cos\omega, \sin\omega)$ and study the first limit; $\omega \to 0, t \to 0$. Due to (4.10), ϕ is C^1 near 0, so one may write

$$\phi(\omega) = D\phi(0)\omega + o(\omega) \text{ and } D\phi(\omega) = D\phi(0) + O(\omega),$$

where $D\phi(0) \neq 0$ by (4.11). Hence

$$\begin{aligned} & g_s(\sigma)/g(\sigma) \\ & = r^{-2}(s\phi(\omega) - tD\phi(\omega))/\phi(\omega) \\ & = r^{-2}s - r^{-2}t(D\phi(0) + O(\omega))/(D\phi(0)\omega + o(\omega)) \\ & = r^{-2}s - r^{-1}\omega^{-1}\sin\omega(D\phi(0) + O(\omega)/(D\phi(0) + o(\omega)/\omega). \end{aligned}$$

Since $s \to 1, r \to 1$ and $\omega^{-1} \sin \omega \to 1$, we obtain the first assertion, the other one is left to the reader. ∎

While $g = g_{\min}$ does not fulfil the requirements of Lemma 4.1, it belongs (as we will see) to the best NCP-functions concerning the regularity hypothesis (3.7) for Newton's method.

Newton's method applied to complementarity problems

Let us now apply Newton's method to (4.1), $h_i := g(z_i(\cdot)) = 0$, under assumption (4.2). The maps $x \mapsto D^0 z(x)$ and $\sigma \mapsto D^0 g(\sigma)$ are Newton maps. Given some element $Rz(x) \in D^0 z(x)$ let $R_i z(x) = (R_i u(x), R_i v(x))$ denote its i-th component. Further, let $\Phi(x)$ consist of all matrices A having rows A_i of the form

$$A_i = G^i R_i z(x) \quad \text{where } G^i \in D^0 g(\sigma_i) \text{ and } \sigma_i = z_i(x). \tag{4.15}$$

The map Φ, contained in the product of $D^0 g(z_i(\cdot)) D^0 z_i(\cdot)$, is a Newton map for h, c.f. Theorem 3.6. Hence only condition (3.7), namely the existence of K^-, remains the problem for solving (4.1) with Newton steps

$$g(\sigma_i) + A_i w = 0 \quad \text{and} \quad x_{\text{new}} := x + w. \tag{4.16}$$

The equation means equivalently

$$\begin{array}{llll} & g(\sigma_i) + a_i R_i u(x) w + b_i R_i v(x) w & = 0, & (a_i, b_i) \in D^0 g(\sigma_i) \\ \text{(or)} & ([a_i/g(\sigma_i)] R_i u(x) + [b_i/g(\sigma_i)] R_i v(x)) w & = -1 & \text{if } g(\sigma_i) \neq 0 \\ \text{and} & (a_i R_i u(x) + b_i R_i v(x)) w & = 0 & \text{if } g(\sigma_i) = 0. \end{array}$$

By Lemma 4.2, we know that

$$\begin{array}{l} a_i/g(\sigma_i) \to 0 \text{ and } \liminf b_i > 0 \text{ if } x \to x^* \text{ with } u_i(x^*) > 0, \text{ as well as} \\ b_i/g(\sigma_i) \to 0 \text{ and } \liminf a_i > 0 \text{ if } x \to x^* \text{ with } v_i(x^*) > 0. \end{array}$$

Due to (4.13), we may write (4.16) as the "weighted equation" (see also (1.3)):

$$a_i(u_i(x) + R_i u(x) w) + b_i(v_i(x) + R_i v(x) w) = 0, (a_i, b_i) \in D^0 g(\sigma_i). \tag{4.17}$$

Theorem 43. (*regularity condition* (3.7) *for NCP*)
Let $g \in pNCP, z = (u, v) \in C^1$ and x^* be a solution of the NCP.

(i) *If condition* (3.7) *is fulfilled for the settings of method* (4.16)*, then* (3.7) *is also satisfied for the particular NCP function* $g_{\min}\{s,t\} = \min\{s,t\}$.

(ii) *Condition* (3.7) *is fulfilled if the NCP is strongly regular at* x^*.

(iii) *Condition* (3.7) *is equivalent with strong regularity of the NCP at* x^* *if the unit vectors of* R^2 *can be connected by an arc in* $\Theta^1(g)$.

Proof. Recalling (4.14), it holds $\max\{a_i, b_i\} \geq p$ for some $p > 0$. So, we see that the matrices A in (4.15) are regular if and only if the same holds for all matrices $C(r,x)$ with rows

$$C_i(r,x) := (1-r_i)R_i u(x) + r_i R_i v(x),$$

where $r_i = b_i[a_i+b_i]^{-1}, 1-r_i = a_i[a_i+b_i]^{-1}$ and $(a_i,b_i) \in D^0 g(u_i(x), v_i(x))$. For $z \in C^1$, these rows have the form

$$C_i(r,x) = (1-r_i)Du_i(x) + r_i Dv_i(x),$$

and the coefficients r_i form a subset $S_i(x) \subset [0,1]$. By continuity arguments, for showing (3.7), it suffices to consider $x = x^*$ only. So, (3.7) holds true if and only if all matrices $C(r,x^*)$ (which form a compact set) are invertible. This condition is as weaker as smaller the sets $S_i(x^*)$ are. To study $S_i(x^*)$, let $y^* = u(x^*) - v(x^*)$. If $y_i^* > 0$ then g is C^1 near $(u_i(x^*), v_i(x^*))$ and $a_i = 0, b_i > 0$. Therefore, $r_i = 1$. Similarly, $y_i^* < 0$ yields $r_i = 0$. Let $y_i^* = 0$. Now the pairs (a_i,b_i) vary in the whole set $\mathrm{cl}Dg(\Theta^1(g))$, and $S_i(x^*) = \{b_i[a_i+b_i]^{-1} / (a_i,b_i) \in \mathrm{cl}Dg(\Theta^1(g))\}$.

In the "smallest case", $S_i(x^*)$ contains 0 and 1 only. This is just the situation for $g = g_{\min}$. In the "largest case", the whole interval [0,1] belongs to $S_i(x^*)$ whenever $y_i^* = 0$. Then, nonsingularity of all $C(r,x^*)$ coincides, by Lemma 2.2, just with *strong regularity* of the NCP at x^*. So (i) and (ii) are true. Having an (continuous) arc in $\Theta^1(g)$ which connects e^1 and e^2, the set $S_i(x^*)$ is connected and contains 0 and 1 (for this conclusion, one needs $z \in C^1$ and the definition of $D^0 g$ via C^1-points of g). So, $S_i(x^*) = [0,1]$ for $y_i^* = 0$ is in fact true. ∎

Theorem 4.3 tells us that condition (3.7), with $M = D^0 g$, is strong if the function g is "smooth". In particular, the examples g_2 and g_ϕ need the strongest regularity condition to ensure local convergence of Newton's method.

Theorem 44. (*uniform regularity and strict monotonicity*)
Let the NCP be strictly monotone. Then, for $g \in pNCP$ and $z \in \mathrm{loc}PC^1$, every matrix A in $\Phi(x)$ (4.15) *is regular. Moreover, for each bounded set $X \subset R^n$, it holds*

$$\|A^{-1}\|_\infty \leq nc(\lambda p)^{-1} \qquad \forall x \in X$$

where $c := \max_i \sup\{\|Dz_i(x)\|_\infty / x \in X \cap \Theta^1(z)\}$, $\lambda = \lambda(u, v)$ is the strict monotonicity constant of NCP and $p = p(g)$ is the constant from (4.14) *taken with the max-norm.*

Proof. Suppose, one finds $\varepsilon > 0, x \in X, A \in \Phi(x)$ and some $w \in \mathrm{bd}B$ (Euclidean sphere) such that

$$\varepsilon \geq |A_i\, w| \text{ for all rows } A_i \text{ of } A. \tag{4.18}$$

This corresponds to the fact that A is singular or $\|A^{-1}\|_\infty \geq 1/\varepsilon$ in terms of the maximum-norm in the image space. By definition of Φ, it holds for certain $(a_i, b_i) \in D^0 g(z_i(x))$

$$A_i w = a_i R_i u(x) w + b_i R_i v(x) w.$$

Due to (4.2), we know that $\frac{1}{2}\lambda \leq \sum_i (R_i u(x), w)(R_i v(x), w)$. Let P_i denote these products and let $P_k = \max_i P_i$. Since

$$P_k \geq \lambda / n, \tag{4.19}$$

The factors $\varphi_u = R_k u(x) w, \varphi_v = R_k v(x) w$ have the same (non-zero) sign. Further, $\max\{|\varphi_u|, |\varphi_v|\}$ is bounded by c. So, $\varphi_u \varphi_v = P_k \geq \lambda/n$ ensures $\min\{|\varphi_u|, |\varphi_v|\} \geq \lambda c^{-1}/n$. Returning to (4.18), and recalling that $a_k \geq 0$ and $b_k \geq 0$, the latter yields by (4.14),

$$\varepsilon \geq |\, a_k\, \varphi_u + b_k\, \varphi_v \,| \geq [\lambda\, c^{-1}/n] \max\{a_k, b_k\} \geq [\lambda c^{-1}/n] p.$$

Therefore $\|A^{-1}\|_\infty \geq 1/\varepsilon$ implies $1/\varepsilon \leq ([\lambda c^{-1}/n]p)^{-1}$ as asserted. ∎

With $u \equiv 0$ and $v_i(x) = x_i > 0$, Theorem 4.4 fails to hold for a monotone standard NCP: We obtain $A_i = 0$ because $Dg(z_i(x)) = (D_s g(0, x_i), 0)$ and $Dz_i(x) = (0, e^i)$. On the other hand, the Theorem holds without strong monotonicity of NCP whenever (4.19) remains true for some $\lambda = \lambda(z) > 0$ and some $k = k(x, w)$. Moreover, if g has *locally Lipschitz* derivatives on $\Theta^1(g)$ and if $z \in C^{1.1}$, then one obtains quadratic convergence because $o(\cdot)$ in (3.8) now fulfils $o(\cdot) \leq L\, \|\cdot\|^2$.

5 Particular realizations and assigned SQP methods

Let us assume that h in §3 coincides with Kojima's function $F(x,y)$, assigned to our standard C^2 optimization problem (2.1),

$$F_1(x,y) = Df(x) + \sum y_i^+ Dg_i(x), \qquad F_{2j}(x,y) = g_j(x) - y_j^- .$$

Then F is PC^1, and all the mentioned derivatives will satisfy (3.2). Depending on the choice of M (as a Newton map), we discuss condition (3.7), imposed for points $z = (x,y)$ near a zero z^*, and the kind of the related problems (3.6). In all cases, we assume that $z = (x,y)$ is the current iteration point and investigate the meaning of (u,v) defined by the Newton step

$$(x,y)_{new} = (x,y) + (u,v).$$

Case 1. Apply the *usual* Newton method to *any* fixed generating function F^S of F being active at the initial point (x^0,y^0). The functions F^S are defined by an index set $S \subset \{1,...,m\}$ as

$$\begin{array}{rcll} F_1^S(x,y) &=& Df(x) + \sum_{i\in S} y_i Dg_i(x) & \\ F_{2i}^S(x,y) &=& g_i(x) & \text{if } i \in S \\ F_{2j}^S(x,y) &=& g_j(x) - y_j & \text{if } j \in \{1,...,m\}\backslash S. \end{array}$$

Here, we assigned, to (y_i^+, y_i^-), the function $y_i \to (y_i, 0)$ if $i \in S$ and $y_i \to (0, y_i)$ otherwise. The initial set S^0 has to be active at (x^0,y^0), i.e. $i \in S^0$ if $y_i^0 > 0$ and $j \notin S^0$ if $y_j^0 < 0$. Because $S = S^0$ is fixed during all steps, the iterations require

$$F^S(x,y) + DF^S(x,y)(u,v)^T = 0.$$

The equations releated to F_{2j}^S for $j \notin S$ have the form $g_j(x) + Dg_j(x)u = y_j + v_j = y_j^{new}$, so v_j, which does not appear in other equations, may be deleted. Thus, we solve the problem

$$P(S^0) \quad \min f(x) \text{ s.t. } g_i(x) = 0 \text{ for } i \in S^0$$

by linearization of the related C^1-Karush–Kuhn–Tucker system. Condition (3.7) requires regularity of the Jacobians $DF^S(z^*)$ for all S, active at z^*. This is strong regularity of all the related problems $P(S)$ at the assigned point (x^*, y_s^*). So condition (3.7) is weaker than strong regularity of the original problem at the solution.

Case 2. With the Kojima–Shindo approach, one selects some set S being active at (x, y), and makes next a Newton-step based on (changing) S as above. The condition (3.7) is the former one.

Case 3. Applying the generalized Jacobian $M = \delta F(= TF$ since $f, g \in C^2)$, one may take any matrix $J(r) \in \delta F(z)$, c.f. (2.6), for the Newton step $F(z) + J(r)(u, v)^T = 0$.

Condition (3.7) requires just strong regularity of (2.1) at (x^*, y^*).
We study the Newton steps for the original Kojima system and the perturbed equation (2.8) at once by considering any $t_i \in R$ in (2.8) and dealing with the Newton equation

$$(5.1) \qquad F^t(z) + J(r,t)(u,v)^T = 0, \qquad J(r,t) \in \delta F^t(z).$$

Recall that this setting represents a mixed penalty-barrier approach (§ 2) for solving (2.1). Let $z = (x, y)$ and t be fixed. Practically, t may depend on z (in each step). Then, to obtain locally superlinear convergence, it suffices to ensure that

$$\|t\| = o(F(z)),$$

c.f. (3.10) and take into account that also F (the original function) has been changed. We abbreviate $Df = Df(x), Dg_i = Dg_j(x)$ and $F = F(x, y)$. Given $r \in R_T(y)$ (§2) we put

$$b_i = 1 - r_i + t_i r_i \qquad \forall i$$

in accordance with the "derivative" of $y_i^- + t_i y_i^+$. We define index sets I^+, I^- and I^0 depending on the signs of y_i. Below, $D_x^2 L(z)$ will stand for $D_x F_1(z)$, so $L = f + \langle y^+, g\rangle$ does not depend on $y_i \leq 0$ (in contrast to the next Case 4). Finally, put

$$J = \{i / b_i \neq 0\}, \quad K = \{k / b_k = 0\}.$$

If $y_i < 0$ then we have $r_i = 0$; hence $b_i = 1, i \in J$, and our weights w_i below are zero. We show:

A Newton step (5.1) means to find a KKT-point (u, μ) of the problem

$$(5.2) \qquad \min_u (Df + \sum_{k \in K} y_k Dg_k)u + \frac{1}{2} u^T D_x^2 L u + \frac{1}{2} \sum_{i \in J} w_i (g_i + Dg_i u)^2$$

s.t. $g_k + Dg_k u = 0 \quad \forall k \in K$, where $w_i = r_i b_i^{-1}$ and $L = f + \langle y^+, g\rangle$. The vector v in (5.1) is then given by $v_k = \mu_k(1-t_k)(k \in K)$ and $v_i = b_i^{-1}(g_i + Dg_i u - (y_i^- + t_i y_i^+))(i \in J)$.

Proof. The linearized equations $F_{2i}^t = 0$ require (equivalently), by the product rule (2.5),

$$g_i + Dg_i u - b_i v_i = y_i^- + t_i y_i^+,$$

$$\text{i.e.}\quad v_i = b_i^{-1}[g_i + Dg_i u - (y_i^- + t_i y_i^+)](i \in J) \text{ and } g_k + Dg_k u = 0 (k \in K).$$

Substituting v_J in the linearized equation $F_1 = 0$, i.e. in

$$F_1 + D_x^2 Lu + \sum_{k\in K} r_k v_k Dg_k + \sum_{i \in J} r_i v_i Dg_i = 0$$

and setting $\mu_k = r_k v_k = (1-t_k)^{-1} v_k$ yields

$$\begin{aligned} 0 &= F_1 + D_x^2 Lu &&+ \sum_{k\in K} r_k v_k Dg_k + \sum_{i\in J} w_i [g_i + Dg_i u - [y_i^- + t_i y_i^+)] Dg_i \\ &= D_x^2 Lu + Df &&+ \sum_{k\in K} (y_k^+ + \mu_k) Dg_k \\ & &&+ \sum_{i\in J} (y_i^+ + w_i[g_i + Dg_i u - (y_i^- + t_i y_i^+)]) Dg_i \\ &= D_x^2 Lu + Df &&+ \sum_{k\in K} (y_k^+ + \mu_k) Dg_k \\ & &&+ \sum_{i\in J} ((1 - w_i t_i) y_i^+ - w_i y_i^- + w_i[g_i + Dg_i u]) Dg_i. \end{aligned}$$

For $i \in J$, we have

$$(1 - w_i t_i) y_i^+ - w_i y_i^- = 0.$$

Indeed, if $y_i < 0$, we know that $y_i^+ = 0$ and $w_i = 0$; if $y_i > 0$, we have $r_i = 1, w_i = b_i^{-1}$ and $b_i = t_i$.

So the F_1-Newton equation becomes

$$0 = D_x^2 Lu + Df + \sum_{k\in K} y_k Dg_k + \sum_{i\in J} w_i (g_i + Dg_i u) Dg_i + \sum_{k \in K} \mu_k Dg_k$$

and has the form

$$0 = D_u Q(u, r) + \sum_{k\in K} \mu_k Dg_k,$$

where $Q = (Df + \sum_{k\in K} y_k Dg_k)u + \frac{1}{2} u^T D_x^2 Lu + \frac{1}{2}\sum_{i\in J} w_i (g_i + Dg_i u)^2$. This proves the assertion ■

Having $t_i \neq 0\ \forall i$, the case of $K = \emptyset$ can be forced by setting $r_i = 1$ whenever $y_i \geq 0$. Now, the equality constraints disappear.

Let $t_i > 0$. Then, if $y_i > 0$, the weights $w_i = t_i^{-1}$ are just the penalty factors. For $y_i = 0$ and $t_i = 0$, all $r_i \in [0{,}1]$ are allowed. So w_i may attain all non-negative values.

Let $t_i < 0$. If $y_i > 0$, now $w_i = t_i^{-1}$ is negative, and stationary u are not necessarily minimizer of problem (5.2). If $y_i = 0$, it holds $0 \geq w_i \geq t_i^{-1}$.

Case 4. Application of NCP functions. To solve the KKT-system of the C^2-problem (2.1) via $G \in pNCP$, require the usual Lagrange condition (with $L = f + \langle \cdot\,, g\rangle$)

$$\Phi_1(x,y) := D_x L(x,y) := Df(x) + \sum_i y_i Dg_i(x) = 0$$

and write the remaining conditions as

$$\Phi_{2i}(x,y) := G(-g_i(x), y_i) = 0.$$

Using D^0G and (4.13) we have to solve

$$D_x L(x,y) + D_x^2 L(x,y)u + \sum_i v_i Dg_i(x) = 0, \tag{5.3}$$

$$-a_i(g_i(x) + Dg_i(x)u) + b_i(y_i + v_i) = 0, \tag{5.4}$$

with

$$(a_i, b_i) \in D^0G(-g_i(x), y_i).$$

Let

$$J = \{i / b_i \neq 0\}, K = \{k / b_k = 0\}.$$

Now, the Newton equation means again (5.2), only a_i and r_i for $i \in J$ must be identified.

A Newton step means to find a KKT-Point (u, μ) of problem (5.2)

$$\min_u (Df + \sum_{k\in K} y_k Dg_k)u + \frac{1}{2}u^T D_x^2 L u + \frac{1}{2}\sum_{i\in J} w_i (g_i + Dg_i u)^2$$

$$\text{s.t. } g_k + Dg_k u = 0 \quad \forall k \in K,$$

where $w_i = a_i b_i^{-1} \geq 0$ and $L = f + \langle y, g\rangle$. The vector v in (5.3), (5.4) is then given by

$$v_k = \mu_k \ (k \in K) \quad \text{and} \quad v_i + w_i(g_i + Dg_i u) \ \ (i \in J)$$

Remark. In the current case, we have $y^* \geq 0$, and non-zero coefficients $w_i = a_i b_i^{-1}$ $(i \in J, y_i \geq 0)$ coincide with $w_i = r_i(1 - r_i + t_i r_i)^{-1}$ of Case 3 after setting $t_i = b_i a_i^{-1}$ and $r_i = 1$.

Proof. Since $(a_i, b_i) \neq 0$, (5.4) yields

$$\begin{aligned} g_k + Dg_k u &= 0 & (k \in K) \\ v_i &= -y_i + w_i(g_i + Dg_i u) & (i \in J). \end{aligned}$$

Raplacing v_J in (5.3) we obtain

$$\begin{aligned} 0 &= D_x^2 Lu + D_x L + \sum_{k \in K} v_k Dg_k + \sum_{i \in J} [-y_i + w_i(g_i + Dg_i u)] Dg_i \\ &= D_x^2 Lu + Df(x) + \sum_{k \in K} (y_k + v_k) Dg_k + \sum_{i \in J} w_i(g_i + Dg_i u) Dg_i. \end{aligned}$$

So the equivalence follows by the same arguments as in Case 3. ■

For $k \in K$, now $y_k < 0$ is possible. Further, $z \to z^*$ yields $w_i \to \infty$ if $y_i^* > 0$, and $w_i \downarrow 0$ if $g_i(x^*) < 0$. So, the method realizes basically a penalty approach.

Case 5. Perturbed generalized Jacobians. Let the Newton step be given by

(5.5) $$F(z) + J(r,t)(u,v)^T = 0, \qquad J(r,t) \in \delta F^t(z).$$

where F^t belongs again to the perturbed equation (2.8), $t_i \in R$. We are using an approximation (3.10) of the Newton map $M = \delta F(z)$ which is justified as long as

$$\|t\| \leq \|F(z)\|.$$

Compared with case 3, now the terms $t_i y_i^+$ do not appear, and the above proof leads us via

$$(1 - w_i t_i) y_i^+ - w_i y_i^- = y_i^+ - w_i y_i^- = y_i^+$$

directly to the modified objective

$$(Df + \sum_{i \in K \cup J} y_i^+ Dg_i)u + \frac{1}{2} u^T D_x^2 Lu + \frac{1}{2} \sum_{i \in J} w_i (g_i + Dg_i u)^2.$$

All the other conclusions of case 3 remain true after setting $t_i y_i^+ = 0$, i.e.

$$r \in R_T(y), \quad b_i = 1 - r_i + t_i r_i \quad \forall i \quad J = \{i / b_i \neq 0\}, \quad K = \{k / b_k = 0\}.$$

A Newton step (5.5) means to find a KKT-point (u, μ) of the problem

$$\begin{array}{llll} & \min_u & D_x Lu + \frac{1}{2} u^T D_x^2 Lu + \frac{1}{2} \sum_{i \in J} w_i (g_i + Dg_i u)^2 & \\ (5.6) & & & \\ & \text{s.t.} & g_k + Dg_k u = 0 \qquad \forall k \in K, & \end{array}$$

where $w_i = r_i b_i^{-1}$ and $L = f + \langle y^+, g \rangle$. The vector v in (5.5) is then given by

$$v_k = \mu(1 - t_k) \ (k \in K) \quad \text{and} \quad v_i = b_i^{-1}(g_i + Dg_i u - y_i^-) \ (i \in J)$$

In comparison with (5.2) now the derivative of the *full* Lagrangean appears in the objective. Setting particularly $t_i \neq 0 \, \forall i$ and selecting $r \in R_T(y)$ with $r_i = 1$ if $y_i \geq 0$, we obtain:

A Newton step (5.5) means to find a stationary point u of

$$D_x Lu + \frac{1}{2} u^T D_x^2 Lu + \frac{1}{2} \sum_{i \in J} t_i^{-1} (g_i + Dg_i u)^2 \tag{5.7}$$

and to put $v_i = t_i^{-1}(g_i + Dg_i u - y_i^-) \quad (i \in J)$.

Case 6. For completeness, we mention a case outside the scope of this paper. To solve auxiliary problems of Wilson-type

$$\min D_x Lu + \frac{1}{2} u^T D_x^2 Lu \quad \text{s.t.} \quad g(x) + Dg(x)u \leq 0$$

one has to apply Newton's method by using directional (or contingent-) derivatives of F:

$$F(x, y) + CF(x, y)(u, v) = 0.$$

The solutions (u, v) fulfil the same condition as in case 3 since $CF \subset TF$. The structure of R_c (§2) implies additionally that $r_i \in \{0, 1\}$ and $\mu_k = r_k v_k \geq 0$. So the constraints may be written as inequalities. The existence of a solution (u, v) is again ensured under strong regularity, because $CF(x, y)$ turns out to be surjective. Concerning this fact as well as convergence and regularity conditions in detail we refer to [21].

Acknowledgement

I would like to thank the referee for various helpful and constructive remarks and suggestions.

References

[1] J.P. Aubin and I. Ekeland, Applied Nonlinear Analysis, Wiley, New York 1984.

[2] S.C. Billups and M.C. Ferris, *QPCOMP: A quadratic programming based solver for mixed complementarity problems*, Math. Progr. B **76** (3) (1997), 533–562.

[3] F.H. Clarke, *On the inverse function theorem*, Pacific Journ. Math. **64** (1) (1976), 97–102.

[4] R. Cominetti, *Metric Regularity, Tangent Sets, and Second-Order Optimality Conditions*, Appl. Math. Optim. **21** (1990), 265–287.

[5] A.L. Dontchev, *Local convergence of the Newton method for generalized equations*, C.R. Acad. Sc. Paris 332 Ser. I (1996), 327–331.

[6] A.L. Dontchev and R.T. Rockafellar, *Characterizations of Strong Regularity for Variational Inequalities over Polyhedral Convex Sets*, SIAM J. Optimization **6** (1996), 1087–1105.

[7] A. Fischer, *Solutions of monotone complementarity problems with locally Lipschitzian functions*, Math. Progr. B **76** (3) (1997), 513–532.

[8] P.T. Harker and J.-S. Pang, *Finite-dimensional variational inequality and nonlinear complementarity problems: A survey of theory, algorithms and applications*, Mathematical Programming **48** (1990), 161–220.

[9] C.M. Ip and J. Kyparisis, *Local convergence of quasi-Newton methods for B-diffenrentialble equations*, MP **56** (1992), 71–89.

[10] V. Jeyakumar, D.T. Luc and S. Schaible, *Characterization of generalized monotone nonsmooth continuous map using approximate Jacobians*, J. Convex Analysis **5** (1) (1998), 119–132.

[11] H.Th. Jongen, D. Klatte and K. Tammer, *Implicit functions and sensitivity of stationary points*, Math. Programming **49** (1990), 123–138.

[12] C. Kanzow, N. Yamashita and M. Fukushima, *New NCP-function and their properties*, JOTA **94** (1997), 115–135.

[13] A. Kaplan, *On the convergence of the penalty function method*, Soviet Math. Dokl. **17** (4) (1976), 1008–1012.

[14] A. King and R.T. Rockafellar, *Sensitivity analysis for nonsmooth generalized equations*, MP **55** (1992), 341–364.

[15] D. Klatte and B. Kummer, *Strong stability in nonlinear programming revisited*, J. Australian Mathem. Soc. Ser. B **40** (1999), 336–352.

[16] D. Klatte and B. Kummer, *Generalized Kojima functions and Lipschitz stability of critical points*, Computational Otimization and Appl. **13** (1999), 61–85.

[17] M. Kojima, *Strongly stable stationary solutions in nonlinear programs*, in: Analysis and Computation of Fixed Points, S.M. Robinson ed., Academic Press, New York (1980), 93–138.

[18] M. Kojima and S. Shindo, *Extensions of Newton and quasi- Newton methods to systems of PC^1 equations*, Journ. of Operations Research Soc. of Japan **29** (1987), 352–374.

[19] B. Kummer, *Newton's method for non-differentiable functions*, in: Advances in Math. Optimization, J. Guddat et al. ed., Akademie Verlag Berlin, Ser. Mathem. Res. **45** (1988), 114–125.

[20] B. Kummer, *Lipschitzian Inverse Functions, Directional Derivatives and Application in $C^{1.1}$-Optimization*, Journal of Optimization Theory and Appl. **70** (3) (1991), 559–580.

[21] B. Kummer, *Newton's method based on generalized derivatives for nonsmooth functions: Convergence Analysis*, in: Lecture Notes in Economics and Mathematical Systems 382; Advances in Optimization; W. Oettli, D. Pallaschke (eds.), Springer, Berlin (1992), 171–194.

[22] B. Kummer, *Lipschitzian and Pseudo-Lipschitzian Inverse Functions and Applications to Nonlinear Optimization*, Lecture Notes in Pure and Applied Mathematics **195** (1997) (Math. Programming with Data Perturbations, ed. A.V. Fiacco), 201–222.

[23] B. Kummer, *Metric Regularity: Characterizations, Nonsmooth Variations and Successive Approximation*, Optimization **46** (1999), 247–281.

[24] R. Mifflin, *Semismooth and semiconvex functions in constrained optimization*, SIAM J. Control and Optim. **15** (1977), 957–972.

[25] B.S. Mordukhovich, *Complete characterization of opennes, metric regularity and Lipschitzian properties of maps*, Trans. Amer. Math. Soc. **340** (1993), 1–35.

[26] J.-S. Pang and Liqun Qi, *Nonsmooth equations: motivation and algorithms*, SIAM J. Optimization **3** (1993), 443–465.

[27] J.-S. Pang, *Newton's method for B-differentiable equations*, Mathematics of OR **15** (1990), 311–341.

[28] J.-P. Penot, *Metric Regularity, openness and Lipschitz behavior of multifunctions*, Nonlin. Analysis **13** (1989), 629–643.

[29] L. Qi and J. Sun, *A nonsmooth version of Newton's method*, Math. Programming **58** (1993), 353–367.

[30] D. Ralph and S. Scholtes, *Sensitivity analysis of composite piecewise smooth equations*, Math. Progr. B **76** (3) (1997), 593–612.

[31] S.M. Robinson, *Strongly regular generalized equations*, Math. Oper. Res. **5** (1980), 43–62.

[32] S.M. Robinson, *Newton's method for a class of nonsmooth functions*, Working Paper, (Aug. 1988) Univ. of Wisconsin-Madison, Department of Industrial Engineering, Madison, WI 53706; in Set-Valued Analysis **2** (1994), 291–305.

[33] R.T. Rockafellar, Convex Analysis, Princeton University Press, Princeton 1970.

[34] D. Sun and L. Qi, *On NCP functions*, Computational Optimization and Appl. **13** (1999), 201–220.

[35] L. Thibault, *Subdifferentials of compactly Lipschitzian vector-valued functions*, Ann. Mat. Pura Appl. (4) **125** (1980), 157–192.

Received 20 December 1999
Revised 19 May 2000