Set-Valued and Variational Analysis
https://doi.org/10.1007/s11228-022-00651-2

On the Application of the SCD Semismooth* Newton Method to Variational Inequalities of the Second Kind

Helmut Gfrerer[1] · **Jiří V. Outrata**[2,3] · **Jan Valdman**[2,4]

Received: 15 December 2021 / Accepted: 3 November 2022

Abstract
The paper starts with a description of SCD (subspace containing derivative) mappings and the SCD semismooth* Newton method for the solution of general inclusions. This method is then applied to a class of variational inequalities of the second kind. As a result, one obtains an implementable algorithm which exhibits locally superlinear convergence. Thereafter we suggest several globally convergent hybrid algorithms in which one combines the SCD semismooth* Newton method with selected splitting algorithms for the solution of monotone variational inequalities. Finally, we demonstrate the efficiency of one of these methods via a Cournot-Nash equilibrium, modeled as a variational inequality of the second kind, where one admits really large numbers of players (firms) and produced commodities.

Keywords Newton method · Semismoothness* · Superlinear convergence · Global convergence · Generalized equation · Coderivatives

Mathematics Subject Classification (2010) 65K10 · 65K15 · 90C33

✉ Helmut Gfrerer
helmut.gfrerer@jku.at

Jiří V. Outrata
outrata@utia.cas.cz

Jan Valdman
jan.valdman@utia.cas.cz

[1] Institute of Computational Mathematics, Johannes Kepler University Linz, A-4040 Linz, Austria

[2] Institute of Information Theory and Automation, Czech Academy of Sciences, 18208 Prague, Czech Republic

[3] Centre for Informatics and Applied Optimization, Federation University of Australia, POB 663, Ballarat, Vic 3350, Australia

[4] Department of Applied Mathematics, Faculty of Information Technology, Czech Technical University in Prague, Thákurova 9, 16000 Prague 6, Czech Republic

Published online: 28 November 2022

1 Introduction

In [5] the authors proposed the so-called semismooth* Newton method for the numerical solution of a general inclusion

$$0 \in F(x),$$

where $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is a closed-graph multifunction. This method has been further developed in [6], where it has coined the name SCD (subspace containing derivative) semismooth* Newton method. Compared to the original method of [5], the new variant requires a slightly stronger approximation of the limiting coderivative of H, but exhibits locally superlinear convergence under substantially less restrictive assumptions. The aim of this paper is to work out this Newton-type method for the numerical solution of the *generalized equation* (GE)

$$0 \in H(x) := f(x) + \partial q(x), \tag{1.1}$$

where $f : \mathbb{R}^n \to \mathbb{R}^n$ is continuously differentiable, $q : \mathbb{R}^n \to \overline{\mathbb{R}}$ is proper convex and lower-semicontinuous (lsc) and ∂ stands for the classical Moreau-Rockafellar subdifferential. It is easy to see that GE (1.1) is equivalent to the variational inequality (VI):

Find $\bar{x} \in \mathbb{R}^n$ such that

$$\langle f(\bar{x}), x - \bar{x}\rangle + q(x) - q(\bar{x}) \geq 0 \ \text{ for all } \ x \in \mathbb{R}^n. \tag{1.2}$$

Model (1.2) has been introduced in [8] and one speaks about the *variational inequality (VI) of the second kind*. It is widely used in the literature dealing with equilibrium models in continuum mechanics cf., e.g., [10] and the references therein. For the numerical solution of GE (1.1), several methods can be used ranging from nonsmooth optimization methods (applicable when ∇f is symmetric) over various types of the Josephy-Newton method [12, 13] to a broad family of splitting methods (usable when F is monotone), cf. [3, Chapter 12]. If GE (1.1) amounts to stationarity condition for a Nash game, then also a simple coordinate-wise optimization technique can be used, cf. [15] and [22]. Concerning the Newton-type methods, let us mention also the possibility to write GE (1.1) as an equation on a monotone graph, which enables us to apply the Newton procedure from [25]. However, note that the subproblems to be solved in the Josephy-Newton method or in this approach are typically rather difficult. In other papers the authors reformulate the problem as a (standard) nonsmooth equation which is then solved by the classical semismooth Newton method, see, e.g., [11, 29].

As mentioned above, in this paper, we will investigate the numerical solution of GE (1.1) using the SCD semismooth* Newton method developed in [6]. In contrast to the Newton methods of Josephy, in this method (as well as in its original variant from [5]) the multi-valued part of (1.1) is also approximated and, differently from some other Newton-type methods, this approximation is provided by means of a linear subspace belonging to the graph of the limiting coderivative of ∂q. In this way, the computation of the Newton direction reduces to the solution of a linear system of equations. To ensure locally superlinear convergence, two properties must be fulfilled. The first one is a weakening of the semismooth* property from [5] and pertains to the subdifferential mapping ∂q. The second one, called SCD regularity, concerns the mapping H and amounts, roughly speaking, to the strong metric subregularity of the considered GE *around* the solution.

The plan of the paper is as follows. After the preliminary Section 2, where we provide the background needed from modern variational analysis, Section 3 is devoted to the broad class of SCD mappings, which is the basic framework for the application of the used

method. In particular, the subdifferential of a proper convex lsc function is an SCD mapping. In Section 4 the SCD semismooth* Newton method is described and its convergence is analyzed. Thereafter, in Section 5 we develop an implementable version of the method for the solution of GE (1.1) and show its locally superlinear convergence under mild assumptions. Section 6 deals with the issue of global convergence. First, we suggest a heuristic modification of the method from the preceding section which exhibits very good convergence properties in the numerical experiments. Thereafter, we show global convergence for a family of hybrid algorithms, where one combines the semismooth* Newton method with various frequently used splitting methods. Finally, in Section 7 we demonstrate the efficiency of the developed methods using a modification of a Cournot-Nash equilibrium problem from [21] which can be modeled in the form of GE (1.1). In contrast to the numerical approach in [22], we may work here with "arbitrarily" large numbers of players (firms) and commodities,

The following notation is employed. Given a matrix A, rge A and ker A denote the range space and the kernel of A, respectively, and $\|A\|$ represents its spectral norm. For a set Ω, $\operatorname{dist}(x, \Omega) := \inf_{a\in\Omega} \|x - a\|$ signifies the distance from x to Ω and $\operatorname{ri}\Omega$ is the relative interior of Ω. Furthermore, $L^{\perp}$ denotes the orthogonal complement of a linear subspace L and $\operatorname{diag}(A, B)$ means a block diagonal matrix with matrices A, B as diagonal blocks. $\mathscr{B}_\delta(x)$ denotes the closed ball around x with radius δ.

2 Preliminaries

Throughout the paper, we will frequently use the following basic notions of modern variational analysis.

Definition 2.1 Let A be a set in $\mathbb{R}^s$, $\bar{x} \in A$ and A be locally closed around $\bar{x}$. Then

(i) The *tangent (contingent, Bouligand) cone* to A at $\bar{x}$ is given by

$$T_A(\bar{x}) := \operatorname*{Lim\,sup}_{t\downarrow 0} \frac{A - \bar{x}}{t}.$$

(ii) The set

$$\widehat{N}_A(\bar{x}) := \big(T_A(\bar{x})\big)^{\circ}$$

is the *regular (Fréchet) normal cone* to A at $\bar{x}$, and

$$N_A(\bar{x}) := \operatorname*{Lim\,sup}_{x \xrightarrow{A} \bar{x}} \widehat{N}_A(x)$$

is the *limiting (Mordukhovich) normal cone* to A at $\bar{x}$.

In this definition "Lim sup" stands for the Painlevé-Kuratowski *outer (upper) set limit*, see, e.g., [27]. The above listed cones enable us to describe the local behavior of set-valued maps via various generalized derivatives. Let $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ be a (set-valued) mapping with the domain and the graph

$$\operatorname{dom} F := \{x \in \mathbb{R}^n \mid F(x) \neq \emptyset\}, \quad \operatorname{gph} F := \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^m \mid y \in F(x)\}.$$

Definition 2.2 Consider a (set-valued) mapping $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ and let $\operatorname{gph} F$ be locally closed around some $(\bar{x}, \bar{y}) \in \operatorname{gph} F$.

(i) The multifunction $DF(\bar{x}, \bar{y}) : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$, given by $\operatorname{gph} DF(\bar{x}, \bar{y}) = T_{\operatorname{gph} F}(\bar{x}, \bar{y})$, is called the *graphical derivative* of F at $(\bar{x}, \bar{y})$.

(ii) The multifunction $D^*F(\bar{x}, \bar{y}) : \mathbb{R}^m \rightrightarrows \mathbb{R}^n$, defined by

$$\operatorname{gph} D^*F(\bar{x}, \bar{y}) = \{(y^*, x^*) \mid (x^*, -y^*) \in N_{\operatorname{gph} F}(\bar{x}, \bar{y})\},$$

is called the *limiting (Mordukhovich) coderivative* of F at $(\bar{x}, \bar{y})$.

Let us now recall the following regularity notions.

Definition 2.3 Let $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ be a a (set-valued) mapping and let $(\bar{x}, \bar{y}) \in \operatorname{gph} F$.

1. F is said to be *metrically subregular at* $(\bar{x}, \bar{y})$ if there exists $\kappa \geq 0$ along with some neighborhood X of $\bar{x}$ such that

$$\operatorname{dist}(x, F^{-1}(\bar{y})) \leq \kappa \operatorname{dist}(\bar{y}, F(x)) \; \forall x \in X.$$

2. F is said to be *strongly metrically subregular at* $(\bar{x}, \bar{y})$ if it is metrically subregular at $(\bar{x}, \bar{y})$ and there exists a neighborhood X' of $\bar{x}$ such that $F^{-1}(\bar{y}) \cap X' = \{\bar{x}\}$.
3. F is said to be *metrically regular around* $(\bar{x}, \bar{y})$ if there is $\kappa \geq 0$ together with neighborhoods X of $\bar{x}$ and Y of $\bar{y}$ such that

$$\operatorname{dist}(x, F^{-1}(y)) \leq \kappa \operatorname{dist}(y, F(x)) \; \forall (x, y) \in X \times Y.$$

4. F is said to be *strongly metrically regular around* $(\bar{x}, \bar{y})$ if it is metrically regular around $(\bar{x}, \bar{y})$ and F^{-1} has a single-valued localization around $(\bar{y}, \bar{y})$, i.e., there are open neighborhoods Y' of $\bar{y}$, X' of $\bar{x}$ and a mapping $h : Y' \to \mathbb{R}^n$ with $h(\bar{y}) = \bar{x}$ such that $\operatorname{gph} F \cap (X' \times Y') = \{(h(y), y) \mid y \in Y'\}$.

It is easy to see that the strong metric regularity around $(\bar{x}, \bar{y})$ implies the strong metric subregularity at $(\bar{x}, \bar{y})$ and the metric regularity around $(\bar{x}, \bar{y})$ implies the metric subregularity at $(\bar{x}, \bar{y})$. To check the metric regularity one often employs the so-called Mordukhovich criterion, according to which this property around $(\bar{x}, \bar{y})$ is equivalent with the condition

$$0 \in D^*F(\bar{x}, \bar{y})(y^*) \;\Rightarrow\; y^* = 0. \tag{2.3}$$

For pointwise characterizations of the other stability properties from Definition 2.3 the reader is referred to [6, Theorem 2.7].

We end up in this preparatory section with a definition of the semismooth* property which paved the way both to semismooth* Newton method in [5] as well as to the SCD semismooth* Newton method in [6].

Definition 2.4 We say that $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is *semismooth** at $(\bar{x}, \bar{y}) \in \operatorname{gph} F$ if for every $\varepsilon > 0$ there is some $\delta > 0$ such that the inequality

$$|\langle x^*, x - \bar{x}\rangle - \langle y^*, y - \bar{y}\rangle| \leq \varepsilon \|(x, y) - (\bar{x}, \bar{y})\| \|(x^*, y^*)\|$$

holds for all $(x, y) \in \operatorname{gph} F \cap \mathscr{B}_\delta(\bar{x}, \bar{y})$ and all (y^*, x^*) belonging to $\operatorname{gph} D^*F(x, y)$.

3 On SCD Mappings

3.1 Basic Properties

In this section we want to recall the basic definitions and features of the SCD property introduced in the recent paper [6].

In what follows, we denote by $\mathscr{Z}_n$ the metric space of all n-dimensional subspaces of $\mathbb{R}^{2n}$ equipped with the metric

$$d_{\mathscr{Z}}(L_1, L_2) := \|P_{L_1} - P_{L_2}\|,$$

where P_{L_i} is the symmetric $2n \times 2n$ matrix representing the orthogonal projection on L_i, $i = 1, 2$.

Sometimes we will also work with bases for the subspaces $L \in \mathscr{Z}_n$. Let $\mathscr{M}_n$ denote the collection of all $2n \times n$ matrices with full column rank n and for $L \in \mathscr{Z}_n$ we define

$$\mathscr{M}(L) := \{Z \in \mathscr{M}_n \mid \operatorname{rge} Z = L\},$$

i.e., the columns of $Z \in \mathscr{M}(L)$ are a basis for L.

We treat every element of $\mathbb{R}^{2n}$ as a column vector. In order to keep the notation simple, we write (u, v) instead of $\begin{pmatrix} u \\ v \end{pmatrix} \in \mathbb{R}^{2n}$ when this does not lead to confusion. To refer to the components of the vector $z = \begin{pmatrix} u \\ v \end{pmatrix}$ we set $\pi_1(z) := u,\ \pi_2(z) = v$.

Let $L \in \mathscr{Z}_n$ and consider $Z \in \mathscr{M}(L)$. Then we can partition Z into two $n \times n$ matrices A and B and we will write $Z = (A, B)$ instead of $Z = \begin{pmatrix} A \\ B \end{pmatrix}$. It follows that

$$\operatorname{rge}(A, B) := \{(Au, Bu) \mid u \in \mathbb{R}^n\} \doteq \left\{ \begin{pmatrix} Au \\ Bu \end{pmatrix} \mid u \in \mathbb{R}^n \right\} = L.$$

Similarly as before, we will also use $\pi_1(Z) := A$, $\pi_2(Z) := B$ to refer to the two $n \times n$ parts of Z.

Furthermore, for every $L \in \mathscr{Z}_n$ we can define the *adjoint* space

$$L^* := \{(-v^*, u^*) \mid (u^*, v^*) \in L^\perp\},$$

where $L^\perp$ denotes as usual the orthogonal complement of L. Then it can be shown that $(L^*)^* = L$ and $d_{\mathscr{Z}}(L_1, L_2) = d_{\mathscr{Z}}(L_1^*, L_2^*)$. Thus, the mapping $L \to L^*$ defines an isometry on $\mathscr{Z}_n$.

We denote by S_n the $2n \times 2n$ orthogonal matrix

$$S_n := \begin{pmatrix} 0 & -I \\ I & 0 \end{pmatrix}$$

so that $L^* = S_n L^\perp$.

Definition 3.1 Consider the mapping $F : R^n \rightrightarrows \mathbb{R}^n$.

1. We call F *graphically smooth of dimension* n at $(x, y) \in \operatorname{gph} F$, if $T_{\operatorname{gph} F}(x, y) = \operatorname{gph} DF(x, y) \in \mathscr{Z}_n$. Further, we denote by $\mathscr{O}_F$ the set of all points where F is graphically smooth of dimension n.

2. We associate with F the four mappings $\widehat{\mathscr{S}}F : \operatorname{gph} F \rightrightarrows \mathscr{Z}_n$, $\widehat{\mathscr{S}^*}F : \operatorname{gph} F \rightrightarrows \mathscr{Z}_n$, $\mathscr{S}F : \operatorname{gph} F \rightrightarrows \mathscr{Z}_n$, $\mathscr{S}^*F : \operatorname{gph} F \rightrightarrows \mathscr{Z}_n$, given by

$$\widehat{\mathscr{S}}F(x,y) := \begin{cases} \{\operatorname{gph} DF(x,y)\} & \text{if } (x,y) \in \mathscr{O}_F, \\ \emptyset & \text{else,} \end{cases}$$

$$\widehat{\mathscr{S}^*}F(x,y) := \begin{cases} \{\operatorname{gph} DF(x,y)^*\} & \text{if } (x,y) \in \mathscr{O}_F, \\ \emptyset & \text{else,} \end{cases}$$

$$\begin{aligned} \mathscr{S}F(x,y) &:= \limsup_{(u,v) \xrightarrow{\operatorname{gph} F} (x,y)} \widehat{\mathscr{S}}F(u,v) \\ &= \left\{ L \in \mathscr{Z}_n \mid \exists (x_k,y_k) \xrightarrow{\mathscr{O}_F} (x,y) \colon \lim_{k\to\infty} d_{\mathscr{Z}}\big(L, \operatorname{gph} DF(x_k,y_k)\big) = 0 \right\}, \\ \mathscr{S}^*F(x,y) &:= \limsup_{(u,v) \xrightarrow{\operatorname{gph} F} (x,y)} \widehat{\mathscr{S}^*}F(u,v) \\ &= \left\{ L \in \mathscr{Z}_n \mid \exists (x_k,y_k) \xrightarrow{\mathscr{O}_F} (x,y) \colon \lim_{k\to\infty} d_{\mathscr{Z}}\big(L, \operatorname{gph} DF(x_k,y_k)^*\big) = 0 \right\}. \end{aligned}$$

3. We say that F has the *SCD (subspace containing derivative) property at* $(x,y) \in \operatorname{gph} F$, if $\mathscr{S}^*F(x,y) \neq \emptyset$. We say that F has the SCD property *around* $(x,y) \in \operatorname{gph} F$, if there is a neighborhood W of (x,y) such that F has the SCD property at every $(x',y') \in \operatorname{gph} F \cap W$. Finally, we call F an *SCD mapping* if F has the SCD property at every point of its graph.

Since $L \to L^*$ is an isometry on $\mathscr{Z}_n$ and $(L^*)^* = L$, the mappings $\mathscr{S}^*F$ and $\mathscr{S}F$ are related via

$$\mathscr{S}^*F(x,y) = \{L^* \mid L \in \mathscr{S}F(x,y)\}, \ \mathscr{S}F(x,y) = \{L^* \mid L \in \mathscr{S}^*F(x,y)\}.$$

The name SCD property is motivated by the following statement.

Lemma 3.2 (cf.[6, Lemma 3.7]) *Let* $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ *and let* $(x,y) \in \operatorname{gph} F$. *Then* $L \subseteq \operatorname{gph} D^*F(x,y)\ \forall L \in \mathscr{S}^*F(x,y)$.

The mappings $\mathscr{S}F(x,y)$ and $\mathscr{S}^*F(x,y)$ can be considered as a generalization of the B-Jacobian for single-valued mappings to multifunctions. In fact, if the single-valued mapping $F : \mathbb{R}^n \to \mathbb{R}^n$ is Lipschitz continuous around x, then the subspaces $L \in \mathscr{S}F(x,F(x))$ ($L \in \mathscr{S}^*F(x,F(x))$) are exactly the graphs of the linear mappings (adjoint mappings) induced by the matrices from the B-Jacobian of F at x, cf. [6, Lemma 3.11].

Next, we turn to the notion of SCD regularity.

Definition 3.3 1. We denote by $\mathscr{Z}_n^{\mathrm{reg}}$ the collection of all subspaces $L \in \mathscr{Z}_n$ such that

$$(y^*,0) \in L \ \Rightarrow \ y^* = 0.$$

2. A mapping $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is called *SCD regular around* $(x,y) \in \operatorname{gph} F$, if F has the SCD property around (x,y) and

$$(y^*,0) \in L \Rightarrow \ y^* = 0\ \forall L \in \mathscr{S}^*F(x,y), \tag{3.4}$$

i.e., $L \in \mathscr{Z}_n^{\mathrm{reg}}$ for all $L \in \mathscr{S}^*F(x,y)$. Further, we will denote by

$$\operatorname{scd\,reg} F(x,y) := \sup\{\|y^*\| \mid (y^*,x^*) \in L, L \in \mathscr{S}^*F(x,y), \|x^*\| \le 1\}$$

the *modulus of SCD regularity* of F around (x, y).

Since the elements of $\mathscr{S}^*F(x, y)$ are contained in $\operatorname{gph} D^*F(x, y)$, it follows from the Mordukhovich criterion (2.3) that SCD regularity is weaker than the metric regularity.

In the following proposition we state some basic properties of subspaces $L \in \mathscr{Z}_n^{\text{reg}}$.

Proposition 3.4 (cf. [6, Proposition 4.2]) *Given a $2n \times n$ matrix Z, there holds $\operatorname{rge} Z \in \mathscr{Z}_n^{\text{rge}}$ if and only if the $n \times n$ matrix $\pi_2(Z)$ is nonsingular. Thus, for every $L \in \mathscr{Z}_n^{\text{rge}}$ there is a unique $n \times n$ matrix C_L such that $L = \operatorname{rge}(C_L, I)$. Furthermore, $L^* = \operatorname{rge}(C_L^T, I) \in \mathscr{Z}_n^{reg}$,*

$$\langle x^*, C_L^T v\rangle = \langle y^*, v\rangle \ \forall (y^*, x^*) \in L \forall v \in \mathbb{R}^n.$$

and

$$\|y^*\| \leq \|C_L\| \|x^*\| \ \forall (y^*, x^*) \in L.$$

Note that for every $L \in \mathscr{Z}_n^{\text{reg}}$ and every $(A, B) \in \mathscr{M}(L)$ the matrix B is nonsingular and $C_L = AB^{-1}$.

Combining [6, Equation (34), Lemma 4.7 and Proposition 4.8] we obtain the following lemma.

Lemma 3.5 *Assume that $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is SCD regular around $(\bar{x}, \bar{y}) \in \operatorname{gph} F$. Then*

$$\operatorname{scd\,reg}\ F(\bar{x}, \bar{y}) = \sup\{\|C_L\| \ \mid \ L \in \mathscr{S}^*F(\bar{x}, \bar{y})\} < \infty.$$

Moreover, F is SCD regular around every $(x, y) \in \operatorname{gph} F$ sufficiently close to $(\bar{x}, \bar{y})$ and

$$\limsup_{(x,y) \xrightarrow{\operatorname{gph} F} (\bar{x}, \bar{y})} \operatorname{scd\,reg}\ F(x, y) \leq \operatorname{scd\,reg}\ F(\bar{x}, \bar{y}).$$

3.2 On the SCD Property of the Subdifferential of Convex Functions

Theorem 3.6 (cf. [6, Corollary 3.28]) *For every proper convex lsc function $q : \mathbb{R}^n \to \overline{\mathbb{R}}$ the subdifferential mapping ∂q is an SCD mapping and for every $(x, x^*) \in \operatorname{gph} \partial q$ and for every $L \in \mathscr{S}^*\partial q(x, x^*) = \mathscr{S}\partial q(x, x^*)$ there is a symmetric positive semidefinite $n \times n$ matrix B with $\|B\| \leq 1$ such that $L = \operatorname{rge}(B, I - B) = L^*$.*

The representation of L via the matrix B is only one possibility. E.g., if q is twice continuously differentiable then $\operatorname{rge}\big(I, \nabla^2 q(x)\big) = \operatorname{gph} D^*\partial q\big(x, \nabla q(x)\big)$ and the relation between B and $\nabla^2 q(x)$ is given by $B = \big(I + \nabla^2 q(x)\big)^{-1}$, $I - B = \big(I + \nabla^2 q(x)\big)^{-1}\nabla^2 q(x)$ and $\nabla^2 q(x) = B^{-1}(I - B)$.

Example 3.7 Assume that $q(x) = \|x\|$ so that

$$\partial q(x) = \begin{cases} \mathscr{B} & \text{for } x = 0 \\ \frac{x}{\|x\|} & \text{otherwise.} \end{cases}$$

By virtue of Theorem 3.6, ∂q is an SCD mapping. When considering a pair $(\bar{x}, \bar{x}^*) \in \operatorname{gph} \partial q$ with $\bar{x} = 0$ and $\|\bar{x}^*\| < 1$, then it is easy to see that ∂q is graphically smooth of dimension n at $(\bar{x}, \bar{x}^*)$ and, by Definition 3.1,

$$\mathscr{S}\partial q(\bar{x}, \bar{x}^*) = \mathscr{S}^*\partial q(\bar{x}, \bar{x}^*) = \big\{\{0\} \times \mathbb{R}^n\big\}.$$

In this case we have representation $\{0\} \times \mathbb{R}^n = \operatorname{rge}(B, I - B)$ with $B = 0$. If $x \neq 0$ then q is even twice continuously differentiable near x and, as pointed out below Theorem 3.6, with $\bar{x}^* = \frac{x}{\|x\|}$ one has

$$\mathscr{S}\partial q(\bar{x}, \bar{x}^*) = \mathscr{S}^*\partial q(\bar{x}, \bar{x}^*) = \operatorname{rge}(B_x, I - B_x) = \operatorname{rge}\big(I, \nabla^2 q(x)\big)$$

with

$$B_x = \big(I + \nabla^2 q(x)\big)^{-1} = \left(I + \frac{1}{\|x\|}\Big(I - \frac{xx^T}{\|x\|^2}\Big)\right)^{-1} = \left(\frac{\|x\|+1}{\|x\|}\Big(I - \frac{xx^T}{\|x\|^2(1+\|x\|)}\Big)\right)^{-1}.$$

We claim that

$$B_x = \frac{\|x\|}{\|x\|+1}\Big(I + \frac{xx^T}{\|x\|^3}\Big). \tag{3.5}$$

Indeed,

$$\begin{aligned}\Big(I - \frac{xx^T}{\|x\|^2(1+\|x\|)}\Big)\Big(I + \frac{xx^T}{\|x\|^3}\Big) &= I + xx^T\Big(\frac{1}{\|x\|^3} - \frac{1}{\|x\|^2(1+\|x\|)} - \frac{\|x\|^2}{\|x^5\|(1+\|x\|)}\Big)\\ &= I + xx^T\Big(\frac{1+\|x\|-\|x\|-1}{\|x\|^3(1+\|x\|)}\Big) = I,\end{aligned}$$

and so formula (3.5) holds true.

Finally, consider the point $(\bar{x}, \bar{x}^*)$ with $\bar{x} = 0$ and $\|\bar{x}^*\| = 1$. By Definition 3.1 and Theorem 3.6 one has that

$$\mathscr{S}\partial q(\bar{x}, \bar{x}^*) = \mathscr{S}^*\partial q(\bar{x}, \bar{x}^*) = \big\{\{0\} \times \mathbb{R}^n\big\} \cup \limsup_{\substack{x \to 0, x \neq 0\\ \frac{x}{\|x\|} \to \bar{x}^*}} \operatorname{rge}(B_x, I - B_x).$$

Since the matrices B_x are bounded, the above Lim sup amounts to $\operatorname{rge}(B, I - B)$ where, taking into account (3.5),

$$B = \lim_{\substack{x \to 0, x \neq 0\\ \frac{x}{\|x\|} \to \bar{x}^*}} B_x = \bar{x}^*\bar{x}^{*T}.$$

However, note that

$$\lim_{\substack{x \to 0, x \neq 0\\ \frac{x}{\|x\|} \to \bar{x}^*}} \nabla^2 q(x)$$

does not exist. Finally note that, at points $(\bar{x}, \bar{x}^*)$ with $\bar{x} = 0$ and $\|\bar{x}^*\| = 1$, one has

$$\operatorname{gph} D^*\partial q(\bar{x}, \bar{x}^*) = \mathscr{S}^*\partial q(\bar{x}, \bar{x}^*) \cup \{(s, s^*) \mid s \in \mathbb{R}_-\{\bar{x}^*\}, \langle s^*, \bar{x}^*\rangle \leq 0\},$$

where the last term is generated by sequences $(0, x^*) \to (0, \bar{x}^*)$ with $\|x^*\| = 1$. Therefore, in this situation, the mapping $\mathscr{S}^*\partial q(\bar{x}, \bar{x}^*)$ has a simpler structure than the limiting coderivative $D^*\partial q(\bar{x}, \bar{x}^*)$ (similarly as in [6, Example 3.29]).

In our numerical experiments we will use convex functions with some separable structure, which carries over to $\mathscr{S}^*\partial q$.

Lemma 3.8 *If $q(x_1, x_2) = q(x_1) + q_2(x_2)$ for lsc convex functions $q_i : \mathbb{R}^{n_i} \to \overline{\mathbb{R}}$, $i = 1, 2$, then for every $\big((\bar{x}_1, \bar{x}_2), (\bar{x}_1^*, \bar{x}_2^*)\big) \in \operatorname{gph} \partial q$ there holds*

$$\begin{aligned}\mathscr{S}\partial q\big((\bar{x}_1, \bar{x}_2), (\bar{x}_1^*, \bar{x}_2^*)\big) = \Big\{\big\{\big((u_1, u_2), (u_1^*, u_2^*)\big) \,\big|\, (u_i, u_i^*) \in L_i,\\ i = 1, 2\big\} \,\Big|\, L_i \in \mathscr{S}\partial q_i(\bar{x}_i, \bar{x}_i^*),\ i = 1, 2\Big\}.\end{aligned}$$

Proof We claim that $\mathscr{O}_{\partial q} = \big\{\big((x_1, x_2), (x_1^*, x_2^*)\big) \,\big|\, (x_i, x_i^*) \in \mathscr{O}_{\partial q_i},\ i = 1, 2\big\}$ and that

$$T_{\operatorname{gph}\partial q}\big((x_1, x_2), (x_1^*, x_2^*)\big) = \big\{\big((u_1, u_2), (u_1^*, u_2^*)\big) \,\big|\, (u_i, u_i^*) \in T_{\operatorname{gph}\partial q_i}(x_i, x_i^*),\ i = 1, 2\big\} \tag{3.6}$$

holds for all $\big((x_1, x_2), (x_1^*, x_2^*)\big) \in \mathscr{O}_{\partial q}$. In fact, if $\big((x_1, x_2), (x_1^*, x_2^*)\big) \in \mathscr{O}_{\partial q}$ then $(x_1, x_1^*) \in \mathscr{O}_{\partial q_1}$ due to $\big\{\big((u_1, 0), (u_1^*, 0)\big) \,\big|\, (u_1, u_1^*) \in \partial q_1(x_1, x_1^*)\big\} \subseteq T_{\operatorname{gph}\partial q}\big((x_1, x_2), (x_1^*, x_2^*)\big)$ and, analogously, $(x_2, x_2^*) \in \mathscr{O}_{\partial q_2}$. This proves $\mathscr{O}_{\partial q} \subseteq \big\{\big((x_1, x_2), (x_1^*, x_2^*)\big) \,\big|\, (x_i, x_i^*) \in \mathscr{O}_{\partial q_i},\ i = 1, 2\big\}$. To show the reverse inclusion, consider $(x_i, x_i^*) \in \mathscr{O}_{q_i}$, $i = 1, 2$. Taking into account [6, Corollary 3.28, Remark 3.18], the sets $\operatorname{gph}\partial q_i$ are geometrically derivable at points $(x_i, x_i^*) \in \mathscr{O}_{\partial q_i}$, $i = 1, 2$, and therefore

$$T_{\operatorname{gph}\partial q_1 \times \operatorname{gph}\partial q_2}\big((x_1, x_1^*), (x_2, x_2^*)\big) = T_{\operatorname{gph}\partial q_1}(x_1, x_1^*) \times T_{\operatorname{gph}\partial q_2}(x_2, x_2^*) \tag{3.7}$$

by [7, Proposition 1]. Thus, $T_{\operatorname{gph}\partial q_1 \times \operatorname{gph}\partial q_2}\big((x_1, x_1^*), (x_2, x_2^*)\big)$ is an $n_1 + n_2$ dimensional subspace and, since the tangent cones in (3.6) and (3.7) coincide up to a reordering of the elements, $\big((x_1, x_2), (x_1^*, x_2^*)\big) \in \mathscr{O}_{\partial q}$ and (3.6) hold true. Hence, our claim is true, and the assertion of the lemma follows from the definition. □

Clearly, the assertion of Lemma 2.10 can be extended to the general case where the sum defining q has an arbitrary finite number of terms.

4 On semismooth* Newton Methods for SCD Mappings

In this section we recall the general framework for the semismooth* Newton method introduced in [5] and adapted to the SCD mappings in [6]. Consider the inclusion

$$0 \in F(x), \tag{4.8}$$

where $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is a mapping having the SCD property around some point $(\bar{x}, 0) \in \operatorname{gph} F$.

Definition 4.1 We say that $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is *SCD semismooth** at $(\bar{x}, \bar{y}) \in \operatorname{gph} F$ if F has the SCD property around $(\bar{x}, \bar{y})$ and for every $\varepsilon > 0$ there is some $\delta > 0$ such that the inequality

$$|\langle x^*, x - \bar{x}\rangle - \langle y^*, y - \bar{y}\rangle| \le \varepsilon \|(x, y) - (\bar{x}, \bar{y})\| \|(x^*, y^*)\|$$

holds for all $(x, y) \in \operatorname{gph} F \cap \mathscr{B}_\delta(\bar{x}, \bar{y})$ and all (y^*, x^*) belonging to any $L \in \mathscr{S}^* F(x, y)$.

Clearly, every mapping with the SCD property around $(\bar{x}, \bar{y}) \in \operatorname{gph} F$ which is semismooth* at $(\bar{x}, \bar{y})$ is automatically SCD semismooth* at $(\bar{x}, \bar{y})$. Therefore, the class of SCD semismooth* mappings is even richer than the class of semismooth* maps. In particular, it follows from [14, Theorem 2] that every mapping whose graph is a closed subanalytic set is SCD semismooth* , cf. [6].

The following proposition provides the key estimate for the semismooth* Newton method for SCD mappings.

Proposition 4.2 (cf. [6, Proposition 5.3]) *Assume that $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is SCD semismooth* at $(\bar{x}, \bar{y}) \in \operatorname{gph} F$. Then for every $\varepsilon > 0$ there is some $\delta > 0$ such that the estimate*

$$\|x - C_L^T(y - \bar{y}) - \bar{x}\| \le \varepsilon \sqrt{n(1 + \|C_L\|^2)} \|(x, y) - (\bar{x}, \bar{y})\|$$

holds for every $(x, y) \in \operatorname{gph} F \cap \mathscr{B}_\delta(\bar{x}, \bar{y})$ and every $L \in \mathscr{S}^* F(x, y) \cap \mathscr{Z}_n^{reg}$.

Now we describe the SCD variant of the semismooth* Newton method. Given a solution $\bar{x} \in F^{-1}(0)$ of (4.8) and some positive scalar, we define the mappings $\mathscr{A}_{\eta,\bar{x}} : \mathbb{R}^n \rightrightarrows \mathbb{R}^n \times \mathbb{R}^n$ and $\mathscr{N}_{\eta,\bar{x}} : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ by

$$\mathscr{A}_{\eta,\bar{x}}(x) := \{(\hat{x}, \hat{y}) \in \operatorname{gph} F \mid \|(\hat{x}, \hat{y}) - (\bar{x}, 0)\| \le \eta \|x - \bar{x}\|\},$$
$$\mathscr{N}_{\eta,\bar{x}}(x) := \{\hat{x} - C_L^T \hat{y} \mid (\hat{x}, \hat{y}) \in \mathscr{A}_{\eta,\bar{x}}(x), L \in \mathscr{S}^* F(\hat{x}, \hat{y}) \cap \mathscr{Z}_n^{\mathrm{reg}}\}.$$

Proposition 4.3 *Assume that F is SCD semismooth* at $(\bar{x}, 0) \in \operatorname{gph} F$ and SCD regular around $(\bar{x}, 0)$ and let $\eta > 0$. Then there is some $\bar{\delta} > 0$ such that for every $x \in \mathscr{B}_{\bar{\delta}}(\bar{x})$ the mapping F is SCD regular around every point $(\hat{x}, \hat{y}) \in \mathscr{A}_{\eta,\bar{x}}(x)$. Furthermore, for every $\varepsilon > 0$ there is some $\delta \in (0, \bar{\delta}]$ such that*

$$\|z - \bar{x}\| \le \varepsilon \|x - \bar{x}\| \ \forall x \in \mathscr{B}_\delta(\bar{x}), \forall z \in \mathscr{N}_{\eta,\bar{x}}(x).$$

Proof Let $\kappa := \operatorname{scd\,reg} F(\bar{x}, 0)$. Then, by Lemma 3.5 there is some $\delta' > 0$ such that F is SCD regular with $\operatorname{scd\,reg} F(x, y) \le \kappa + 1$ around any $(\hat{x}, \hat{y}) \in \operatorname{gph} F \cap \mathscr{B}_{\delta'}(\bar{x}, 0)$ and the first assertion follows with $\bar{\delta} := \delta'/\eta$. Now consider $\varepsilon > 0$ and set $\tilde{\varepsilon} := \varepsilon/(\eta\sqrt{n(1 + (1 + \kappa)^2)})$. By Proposition 4.2 there is some $\tilde{\delta} \in (0, \delta']$ such that the inequality

$$\|\hat{x} - C_L^T \hat{y}\| \le \tilde{\varepsilon}\sqrt{n(1 + \|C_L\|^2)}\|(\hat{x}, \hat{y}) - (\bar{x}, 0)\|$$

holds for every $(\hat{x}, \hat{y}) \in \operatorname{gph} F \cap \mathscr{B}_{\tilde{\delta}}$ and every $L \in \mathscr{S}^* F(\hat{x}, \hat{y}) \cap \mathscr{Z}_n^{\mathrm{reg}}$. Set $\delta := \tilde{\delta}/\eta$ and consider $x \in \mathscr{B}_\delta(\bar{x})$. For every $(\hat{x}, \hat{y}) \in \mathscr{A}_{\eta,\bar{x}}(x)$ we have $\|(\hat{x}, \hat{y}) - (\bar{x}, 0)\| \le \eta\|x - \bar{x}\| \le \tilde{\delta} \le \delta'$ and consequently

$$\|C_L\| \le \operatorname{scd\,reg} F(\hat{x}, \hat{y}) \le \kappa + 1 \ \forall L \in \mathscr{S}^* F(\hat{x}, \hat{y}).$$

Thus

$$\|\hat{x} - C_L^T \hat{y}\| \le \tilde{\varepsilon}\sqrt{n(1 + (1 + \kappa)^2)}\|(\hat{x}, \hat{y}) - (\bar{x}, 0)\| \le \varepsilon\|x - \bar{x}\|$$

and the second assertion follows. □

Assuming that we are given some iterate $x^{(k)}$, the next iterate is formally given by $x^{(k+1)} \in \mathscr{N}_{\eta,\bar{x}}(x^{(k)})$. Let us take a closer look at this rule. Since we cannot expect in general that $F(x^{(k)}) \neq \emptyset$ or that 0 is close to $F(x^{(k)})$, even if $x^{(k)}$ is close to a solution $\bar{x}$, we first perform some step which yields $(\hat{x}^{(k)}, \hat{y}^{(k)}) \in \operatorname{gph} F$ as an approximate projection of $(x^{(k)}, 0)$ onto $\operatorname{gph} F$. We require that

$$\|(\hat{x}^{(k)}, \hat{y}^{(k)}) - (\bar{x}, 0)\| \le \eta\|x^{(k)} - \bar{x}\| \tag{4.9}$$

for some constant $\eta > 0$, i.e., $(\hat{x}^{(k)}, \hat{y}^{(k)}) \in \mathscr{A}_{\eta,\bar{x}}(x^{(k)})$. For instance, if

$$\|(\hat{x}^{(k)}, \hat{y}^{(k)}) - (x^{(k)}, 0)\| \le \beta \operatorname{dist}((x^{(k)}, 0), \operatorname{gph} F)$$

holds with some $\beta \ge 1$, then

$$\begin{aligned}\|(\hat{x}^{(k)}, \hat{y}^{(k)}) - (\bar{x}, 0)\| &\le \|(\hat{x}^{(k)}, \hat{y}^{(k)}) - (x^{(k)}, 0)\| + \|(x^{(k)}, 0) - (\bar{x}, 0)\| \\ &\le \beta \operatorname{dist}((x^{(k)}, 0), \operatorname{gph} F) + \|(x^{(k)}, 0) - (\bar{x}, 0)\| \\ &\le (\beta + 1)\|(x^{(k)}, 0) - (\bar{x}, 0)\|\end{aligned}$$

and thus (4.9) holds with $\eta = \beta + 1$ and we can fulfill (4.9) without knowing the solution $\bar{x}$. Further we require that $\mathscr{S}^* F(\hat{x}^{(k)}, \hat{y}^{(k)}) \cap \mathscr{Z}_n^{\mathrm{reg}} \neq \emptyset$ and compute the new iterate as $x^{(k+1)} = \hat{x}^{(k)} - C_L^T \hat{y}^{(k)}$ for some $L \in \mathscr{S}^* F(\hat{x}^{(k)}, \hat{y}^{(k)}) \cap \mathscr{Z}_n^{\mathrm{reg}}$. In fact, in our numerical implementation we will not compute the matrix C_L, but two $n \times n$ matrices A, B such that

$L = \operatorname{rge}(B^T, A^T)$. The next iterate $x^{(k+1)}$ is then obtained by $x^{(k+1)} = \hat{x}^{(k)} + \Delta x^{(k)}$ where $\Delta x^{(k)}$ is a solution of the system $A\Delta x = -B\hat{y}^{(k)}$.

This leads to the following conceptual algorithm.

Algorithm 1 (SCD semismooth* Newton-type method for inclusions)
1. Choose a starting point $x^{(0)}$, set the iteration counter $k := 0$.
2. If $0 \in F(x^{(k)})$, stop the algorithm.
3. **Approximation step:** Compute

$$(\hat{x}^{(k)}, \hat{y}^{(k)}) \in \operatorname{gph} F$$

satisfying (4.9) and such that $\mathscr{S}^* F(\hat{x}^{(k)}, \hat{y}^{(k)}) \cap \mathscr{Z}_n^{\mathrm{reg}} \neq \emptyset$.
4. **Newton step:** Select $n \times n$ matrices $A^{(k)}, B^{(k)}$ with

$$L^{(k)} := \operatorname{rge}\left(B^{(k)^T}, A^{(k)^T}\right) \in \mathscr{S}^* F(\hat{x}^{(k)}, \hat{y}^{(k)}) \cap \mathscr{Z}_n^{\mathrm{reg}},$$

calculate the Newton direction $\Delta x^{(k)}$ as a solution of the linear system

$$A^{(k)}\Delta x = -B^{(k)}\hat{y}^{(k)}$$

and obtain the new iterate via $x^{(k+1)} = \hat{x}^{(k)} + \Delta x^{(k)}$.
5. Set $k := k + 1$ and go to 2.

For this algorithm, local superlinear convergence follows from Proposition 4.3, see also [6, Corollary 5.6].

Theorem 4.4 *Assume that F is SCD semismooth* at $(\bar{x}, 0) \in \operatorname{gph} F$ and SCD regular around $(\bar{x}, 0)$. Then for every $\eta > 0$ there is a neighborhood U of $\bar{x}$ such that for every starting point $x^{(0)} \in U$ Algorithm 1 is well defined and stops after finitely many iterations at a solution of (4.8) or produces a sequence $x^{(k)}$ superlinearly converging to $\bar{x}$ for any choice of $(\hat{x}^{(k)}, \hat{y}^{(k)})$ satisfying (4.9) and any $L^{(k)} \in \mathscr{S}^* F(\hat{x}^{(k)}, \hat{y}^{(k)})$.*

As shown in [6, Corollary 6.4], if F happens to be SCD semismooth* around $(\bar{x}, 0)$, then the assumptions of the above statement are fulfilled whenever F is strongly metrically subregular at all points in a neighborhood of $(\bar{x}, 0)$. Therefore, in particular, these assumptions are satisfied provided F is strongly metrically regular around $(\bar{x}, 0)$, which is used in the test problem discussed in Section 7.

There is an alternative for the computation of the Newton direction $\Delta x^{(k)}$ based on the subspaces from $\mathscr{S} F(\hat{x}^{(k)}, \hat{y}^{(k)})$, cf. [6]:
4. **Newton step:** *Select $n \times n$ matrices $A^{(k)}, B^{(k)}$ with*

$$L^{(k)} := \operatorname{rge}\left(A^{(k)}, B^{(k)}\right) \in \mathscr{S} F(\hat{x}^{(k)}, \hat{y}^{(k)}) \cap \mathscr{Z}_n^{reg},$$

compute a solution p of the linear system

$$B^{(k)} p = -\hat{y}^{(k)}$$

and obtain the new iterate $x^{(k+1)} = \hat{x}^{(k)} + \Delta x^{(k)}$ with Newton direction $\Delta x^{(k)} = A^{(k)} p$.

For the choice between the two approaches for calculating the Newton direction, it is important to consider whether elements from $\mathscr{S}^* F(\hat{x}^{(k)}, \hat{y}^{(k)})$ or from $\mathscr{S} F(\hat{x}^{(k)}, \hat{y}^{(k)})$ are easier to compute.

Note that for an implementation of the Newton step, we need not to know the whole derivative $\mathscr{S}^* F(\hat{x}^{(k)}, \hat{y}^{(k)})$ (or $\mathscr{S} F(\hat{x}^{(k)}, \hat{y}^{(k)})$) but only one element $L^{(k)} \in \mathscr{S}^* F(\hat{x}^{(k)}, \hat{y}^{(k)})$.

At the end of this section we want to clarify the relation between the SCD semismooth* Newton method and other Newton-type methods. Consider first the case where $F : \mathbb{R}^n \to \mathbb{R}^n$ is a single-valued Lipschitzian function. In this case, the approximation step can be simply implemented by $(\hat{x}^{(k)}, \hat{y}^{(k)}) := \big(x^{(k)}, F(x^{(k)})\big)$, because (4.9) is fulfilled due to the Lipschitz continuity of F. As already mentioned in the preceding section, the subspaces $L \in \mathscr{S}^* F\big(x^{(k)}, F(x^{(k)})\big)$ coincide with the graphs of adjoint mappings associated with the matrices in the B-Jacobian of F at $x^{(k)}$. It follows that the Newton direction $\Delta x^{(k)}$ computed in step 4. of Algorithm 1 is also a Newton direction for the classical semismooth Newton method by Qi and Sun [23].

There exist also coderivative-based Newton methods in nonsmooth optimization [16–18, 20]. Mordukhovich and Sarabi [20] introduced a Newton-type algorithm for solving the equation $\nabla\varphi(x) = 0$ for $C^{1,1}$ functions $\varphi : \mathbb{R}^n \to \mathbb{R}$, where in an iterate $x^{(k)}$ the Newton direction $\Delta x^{(k)}$ is obtained by solving the inclusion

$$-\nabla\varphi(x^{(k)}) \in D^*\nabla\varphi\big(x^{(k)}, \nabla\varphi(x^{(k)})\big)(\Delta x). \tag{4.10}$$

Locally superlinear convergence is shown in [20] under the assumptions that $\nabla\varphi$ is semismooth* and the solution $\bar{x}$ is a tilt-stable local minimizer of φ. By taking into account that the limiting coderivative of the gradient mapping $\nabla\varphi$ enjoys some symmetry properties, in particular that $\mathscr{S}\nabla\varphi = \mathscr{S}^*\nabla\varphi$ by [6, Proposition 3.26], it can be shown that the SCD semismooth* Newton direction $\Delta x^{(k)}$ also fulfills (4.10). Thus, for finding a solution of (4.10) one could use, for instance, an appropriate linear system from the SCD semismooth* Newton method. Moreover, the requirement of tilt-stability of $\bar{x}$ is stronger than the corresponding assumption of SCD regularity used in Theorem 4.4.

In order to minimize continuously prox-regular functions $\varphi : \mathbb{R}^n \to \overline{\mathbb{R}}$, it is suggested in [20] to apply the above approach to the *Moreau envelope*

$$e_\lambda\varphi(x) := \min_y \left\{ \frac{1}{2\lambda}\|y - x\|^2 + \varphi(y) \right\}, \ x \in \mathbb{R}^n,$$

where $\lambda > 0$ has to be chosen sufficiently small. This is worked out in more detail in [16] without explicitly computing the Moreau envelope. It turns out that the resulting method is closely related with Algorithm 1 by taking an approximation step

$$\hat{x}^{(k)} := P_\lambda\varphi(x^{(k)}), \quad \hat{y}^{(k)} := \lambda(x^{(k)} - \hat{x}^{(k)}),$$

where

$$P_\lambda\varphi(x) := \arg\min_y \left\{ \frac{1}{2\lambda}\|y - x\|^2 + \varphi(y) \right\}, \ x \in \mathbb{R}^n,$$

denotes the *proximal mapping*. Then the Newton direction $\Delta x^{(k)}$ is computed in [16] as solution of the inclusion

$$-\hat{y}^{(k)} \in D^*\partial\varphi(\hat{x}^{(k)}, \hat{y}^{(k)})(\Delta x). \tag{4.11}$$

Again, it can be shown that the SCD semismooth* Newton direction solves this inclusion and the SCD semismooth* Newton has the same advantages as in the $C^{1,1}$ case, i.e., the Newton direction is easier to compute and Assumption (H2) in [16] (metric regularity of $\partial\varphi$) is more restrictive than SCD regularity.

The inclusions (4.10), (4.11) suggest the possibility of computing the Newton direction as solution of the coderivative-based inclusion

$$-\hat{y}^{(k)} \in D^*F\varphi(\hat{x}^{(k)}, \hat{y}^{(k)})(\Delta x)$$

for general mappings F. However, unless the limiting coderivative D^*F possesses certain symmetry properties, this approach will not result in superlinear convergence as can be easily seen from the differentiable case: If F is single-valued and continuously differentiable, then $D^*F\big(\hat{x}^{(k)}, F(\hat{x}^{(k)})\big)(\Delta x) = \nabla F(\hat{x}^{(k)})^T \Delta x$ yielding $\Delta x^{(k)} = -\big(\nabla F(\hat{x}^{(k)})^T\big)^{-1} F(\hat{x}^{(k)})$, which differs from the classical Newton direction $-\nabla F(\hat{x}^{(k)})^{-1} F(\hat{x}^{(k)})$.

5 Implementation of the semismooth* Newton Method

There is a lot of possibilities how to implement the SCD semismooth* Newton method. Apart from the Newton step, which is not uniquely determined by different choices of subspaces contained in $\mathscr{S}^*F(\hat{x}^{(k)}, \hat{y}^{(k)})$, there is a multitude of possibilities how to perform the approximation step. In this section we will construct an implementable version of the semismooth* Newton method for the numerical solution of GE (1.1) under the assumption that the proximal mapping $P_\lambda q$, defined by

$$P_\lambda q(y) := \arg\min_x \left\{ \frac{1}{2\lambda}\|x - y\|^2 + q(x) \right\}, \ y \in \mathbb{R}^n,$$

can be efficiently evaluated for every $y \in \mathbb{R}^n$ and parameter $\lambda > 0$. Since q is convex, it is well known that for every $\lambda > 0$ the proximal mapping $P_\lambda q$ is single-valued and nonexpansive and $P_\lambda q = (I + \lambda \partial q)^{-1}$, see, e.g. [27, Proposition 12.19].

Given some scaling parameter $\gamma > 0$, we will denote

$$u_\gamma(x) := P_{\frac{1}{\gamma}} q\big(x - \frac{1}{\gamma} f(x)\big) - x.$$

It is well known, that $\bar{x}$ solves (1.1) if and only if $u_\gamma(\bar{x}) = 0$ for some $\gamma > 0$ and in this case we even have $u_\gamma(\bar{x}) = 0$ for all $\gamma > 0$. From the definition of the proximal mapping we obtain that $u_\gamma(x)$ is the unique solution of the uniformly convex optimization problem

$$\min_u \frac{\gamma}{2}\|u\|^2 + \langle f(x), u\rangle + q(x + u).$$

The first-order (necessary and sufficient) optimality condition reads as

$$0 \in \gamma u_\gamma(x) + f(x) + \partial q\big(x + u_\gamma(x)\big). \tag{5.12}$$

Since $P_\lambda q$ is nonexpansive, we obtain the bounds

$$\begin{aligned} &\big\|\big(x+u_\gamma(x)\big)-\big(x'+u_\gamma(x')\big)\big\| \le \big\|(x-x')-\frac{1}{\gamma}\big(f(x)-f(x')\big)\big\| \le \|x-x'\|+\frac{1}{\gamma}\|f(x)-f(x')\|, \\ &\|u_\gamma(x)-u_\gamma(x')\| \le 2\|x-x'\|+\frac{1}{\gamma}\|f(x)-f(x')\|. \end{aligned} \tag{5.13}$$

Our approach is based on an equivalent reformulation of (1.1) in form of the GE

$$0 \in \mathscr{F}(x, d) := \begin{pmatrix} f(x) + \partial q(d) \\ x - d \end{pmatrix} \tag{5.14}$$

in variables $(x, d) \in \mathbb{R}^n \times \mathbb{R}^n$. Clearly, $\bar{x}$ is a solution of (1.1) if and only if $(\bar{x}, \bar{x})$ is a solution of (5.14).

Proposition 5.1 (i) *Let $x \in \mathbb{R}^n$, $(d, d^*) \in \operatorname{gph} \partial q$. Then*

$$\begin{aligned}&\mathscr{S}^*\mathscr{F}\big((x,d),(f(x)+d^*,x-d)\big)\\&=\left\{\operatorname{rge}\left(\begin{pmatrix}Y^* & 0\\ 0 & -I\end{pmatrix},\begin{pmatrix}\nabla f(x)^T Y^* & -I\\ X^* & I\end{pmatrix}\right)\,\middle|\,\operatorname{rge}(Y^*,X^*)\in\mathscr{S}^*\partial q(d,d^*)\right\}.\end{aligned} \tag{5.15}$$

(ii) *Let $\bar{x}$ be a solution to (1.1). Then the following statements are equivalent:*

- (a) *H is SCD regular around $(\bar{x}, 0)$.*
- (b) *For every $L \in \mathscr{S}^*\partial q\big(\bar{x}, -f(\bar{x})\big)$ and every $(Y^*, X^*) \in \mathscr{M}(L)$ the matrix $\nabla f(x)^T Y^* + X^*$ is nonsingular.*
- (c) *The mapping $\mathscr{F}$ is SCD regular around $\big((\bar{x},\bar{x}),(0,0)\big)$.*

(iii) *Let $\bar{x}$ be a solution to (1.1). If ∂q is SCD semismooth* at $\big(\bar{x}, -f(\bar{x})\big)$ then $\mathscr{F}$ is SCD semismooth* at $\big((\bar{x},\bar{x}),(0,0)\big)$.*

Proof Ad (i): GE (5.14) can be written down in the form

$$0 \in \mathscr{F}(x,d) = h(x,d) + G(x,d),$$

where $h(x,d) := (f(x), x-d)$ and $G(x,d) := \partial q(d) \times \partial g(x)$ with $g : \mathbb{R}^n \to \mathbb{R}$ given by $g(x) = 0$ for all x. By virtue of [6, Proposition 3.15] we obtain that, at the point $\big((x,d),(f(x)+d^*,x-d)\big) \in \operatorname{gph}\mathscr{F} \subseteq \mathbb{R}^{2n}\times\mathbb{R}^{2n}$, one has

$$\mathscr{S}^*\mathscr{F}\big((x,d),(f(x)+d^*,x-d)\big) = \begin{pmatrix} I & 0\\ \nabla h(x,d)^T & I\end{pmatrix}\mathscr{S}^*G\big((x,d),(d^*,0)\big).$$

Next consider the mapping $\tilde{G} : \mathbb{R}^{2n} \rightrightarrows \mathbb{R}^{2n}$ given by $\tilde{G}(x,d) = \partial\big(g(x)+q(d)\big)$. Since

$$\begin{aligned}\operatorname{gph} G &= \big\{\big((x,d),(d^*,x^*)\big)\,\big|\,(d,d^*)\in\operatorname{gph}\partial q,(x,x^*)\in\operatorname{gph}\partial g\big\}\ \text{and}\\ \operatorname{gph}\tilde{G} &= \big\{\big((x,d),(x^*,d^*)\big)\,\big|\,(d,d^*)\in\operatorname{gph}\partial q,(x,x^*)\in\operatorname{gph}\partial g\big\},\end{aligned}$$

we can employ [6, Proposition 3.14] with $\Phi(x,d,d^*,x^*) := (x,d,x^*,d^*)$ to obtain that

$$\mathscr{S}^*G\big((x,d),(d^*,0)\big) = S_{2n}\nabla\Phi(x,d,d^*,0)^T S_{2n}^T \mathscr{S}^*\tilde{G}\big((x,d),(0,d^*)\big).$$

It remains to compute $\mathscr{S}^*\tilde{G}\big((x,d),(0,d^*)\big)$, which can be done by Theorem 3.6 and Lemma 3.8. We obtain that

$$\mathscr{S}^*\tilde{G}\big((x,d),(0,d^*)\big) = \left\{\operatorname{rge}\left(\begin{pmatrix}I & 0\\ 0 & Y^*\end{pmatrix},\begin{pmatrix}0 & 0\\ 0 & X^*\end{pmatrix}\right)\,\middle|\,\operatorname{rge}(Y^*,X^*)\in\mathscr{S}^*\partial q(d,d^*)\right\}.$$

Putting these ingredients together we may conclude that

$$\begin{aligned}
\mathscr{S}^*\mathscr{F}\big((x,d),(f(x)+d^*,x-d)\big) &= \begin{pmatrix} I & 0 \\ \nabla h(x,d)^T & I \end{pmatrix} S_{2n}\nabla\Phi(x,d,d^*,0)^T S_{2n}^T \mathscr{S}^*\tilde{G}\big((x,d),(0,d^*)\big) \\
&= \begin{pmatrix} 0 & I & 0 & 0 \\ I & 0 & 0 & 0 \\ I & \nabla f(x)^T & I & 0 \\ -I & 0 & 0 & I \end{pmatrix} \mathscr{S}^*\tilde{G}\big((x,d),(0,d^*)\big) \\
&= \left\{ \operatorname{rge}\left[\begin{pmatrix} 0 & I & 0 & 0 \\ I & 0 & 0 & 0 \\ I & \nabla f(x)^T & I & 0 \\ -I & 0 & 0 & I \end{pmatrix}\begin{pmatrix} I & 0 \\ 0 & Y^* \\ 0 & 0 \\ 0 & X^* \end{pmatrix}\right] \Big| \operatorname{rge}(Y^*,X^*) \in \mathscr{S}^*\partial q(d,d^*) \right\} \\
&= \left\{ \operatorname{rge}\left(\begin{pmatrix} 0 & Y^* \\ I & 0 \end{pmatrix}, \begin{pmatrix} I & \nabla f(x)^T Y^* \\ -I & X^* \end{pmatrix}\right) \Big| \operatorname{rge}(Y^*,X^*) \in \mathscr{S}^*\partial q(d,d^*) \right\} \\
&= \left\{ \operatorname{rge}\left(\begin{pmatrix} 0 & Y^* \\ I & 0 \end{pmatrix} S_n, \begin{pmatrix} I & \nabla f(x)^T Y^* \\ -I & X^* \end{pmatrix} S_n\right) \Big| \operatorname{rge}(Y^*,X^*) \in \mathscr{S}^*\partial q(d,d^*) \right\}
\end{aligned}$$

leading to formula (5.15).

Ad (ii): By [6, Proposition 3.15] we have

$$\begin{aligned}
\mathscr{S}^*H(\bar{x},0) &= \begin{pmatrix} I & 0 \\ \nabla f(\bar{x})^T & I \end{pmatrix} \mathscr{S}^*\partial q\big(\bar{x},-f(\bar{x})\big) \\
&= \big\{ \operatorname{rge}(Y^*, \nabla f(\bar{x})^T Y^* + X^*) \,\big|\, \operatorname{rge}(Y^*,X^*) \in \mathscr{S}^*\partial q\big(\bar{x},-f(\bar{x})\big) \big\}
\end{aligned}$$

and the equivalence between (a) and (b) is implied by Proposition 3.4. By (5.15), the mapping $\mathscr{F}$ is SCD regular around $\big((\bar{x},\bar{x}),(0,0)\big)$ if and only if for every pair Y^*, X^* with $\operatorname{rge}(Y^*,X^*) \in \mathscr{S}^*\partial q\big(\bar{x},-f(\bar{x})\big)$ the matrix

$$\begin{pmatrix} \nabla f(x)^T Y^* & -I \\ X^* & I \end{pmatrix} = \begin{pmatrix} \nabla f(x)^T Y^* + X^* & -I \\ 0 & I \end{pmatrix}\begin{pmatrix} I & 0 \\ X^* & I \end{pmatrix}$$

is nonsingular and, by the representation above, this holds if and only if $\nabla f(x)^T Y^* + X^*$ is nonsingular. Hence, (b) is equivalent to (c).

Ad (iii): Let f be Lipschitz continuous with constant l in some ball $\mathscr{B}_r(\bar{x})$ around $\bar{x}$. Consider $\varepsilon > 0$, choose $\delta_q > 0$ such that

$$\begin{aligned}
&|\langle e^*, d-\bar{x}\rangle - \langle e, d^* + f(\bar{x})\rangle| \le \tfrac{\varepsilon}{2\sqrt{2}(l+1)}\|(e,e^*)\|\,\big\|\big(d-\bar{x}, d^*+f(\bar{x})\big)\big\| \\
&\quad \text{for all } (d,d^*) \in \operatorname{gph}\partial q \cap \mathscr{B}_{\delta_q}\big(\bar{x},-f(\bar{x})\big) \text{ and all } (e,e^*) \in L \in \mathscr{S}^*\partial q(d,d^*)
\end{aligned}$$

and then choose $\delta \le \min\{\frac{\delta_q}{1+l}, r\}$ such that

$$\|f(x) - f(\bar{x}) - \nabla f(x)(x-\bar{x})\| \le \frac{\varepsilon}{2\sqrt{2}(l+1)}\|x-\bar{x}\|,\ x \in \mathscr{B}_\delta(\bar{x}).$$

Next consider $\big((x,d),(y_1,y_2)\big) \in \operatorname{gph}\mathscr{F} \cap \mathscr{B}_\delta\big((\bar{x},\bar{x}),(0,0)\big)$, $\big((z_1,z_2),(z_1^*,z_2^*)\big) \in \bar{L} \in \mathscr{S}^*\mathscr{F}\big((x,d),(y_1,y_2)\big)$. Then $y_1 = f(x) + d^*$ with $d^* \in \partial q(d)$, $y_2 = x - d$ and by (5.15) there are $(e,e^*) \in L \in \mathscr{S}^*\partial q(d,d^*)$, $c \in \mathbb{R}^n$ with $\big((z_1,z_2),(z_1^*,z_2^*)\big) = \big((e,-c),(\nabla f(x)^T e - c, e^* + c)\big)$. Then $\|x-\bar{x}\| \le \delta$ and

$$\begin{aligned}
\big\|\big(d-\bar{x}, d^*+f(\bar{x})\big)\big\| &\le \|(d-\bar{x},y_1)\| + \|f(x)-f(\bar{x})\| \\
&\le \big\|\big((x,d),(y_1,y_2)\big) - \big((\bar{x},\bar{x}),(0,0)\big)\big\| + \|f(x)-f(\bar{x})\| \le \delta + l\delta \le \delta_q.
\end{aligned}$$

It follows that

$$
\begin{aligned}
&|\langle (z_1,z_2),(y_1,y_2)\rangle - \langle (z_1^*,z_2^*),(x,d)-(\bar{x},\bar{x})\rangle| \\
&= |\langle e, f(x)+d^*\rangle - \langle c, x-d\rangle - \langle \nabla f(x)^T e - c, x-\bar{x}\rangle - \langle e^*+c, d-\bar{x}\rangle| \\
&\le |\langle e, f(x)-f(\bar{x})-\nabla f(x)(x-\bar{x})\rangle| + |\langle e, d^*+f(\bar{x})\rangle + \langle e^*, d-\bar{x}\rangle| \\
&\le \frac{\varepsilon}{2\sqrt{2}(l+1)}\|e\|\|x-\bar{x}\| + \frac{\varepsilon}{2\sqrt{2}(l+1)}\|(e,e^*)\|\big\|\big(d-\bar{x}, d^*+f(\bar{x})\big)\big\| \\
&\le \frac{\varepsilon}{\sqrt{2}(l+1)}\|(e,e^*)\|\big\|\big(x-\bar{x}, d-\bar{x}, d^*+f(\bar{x})\big)\big\| \\
&\le \frac{\varepsilon}{\sqrt{2}(l+1)}\|(e,e^*)\|\big(\big\|\big(x-\bar{x}, d-\bar{x}, d^*+f(x)\big)\big\| + \|f(x)-f(\bar{x})\|\big) \\
&\le \frac{\varepsilon}{\sqrt{2}}\|(e,e^*)\|\|(x-\bar{x}, d-\bar{x}, d^*+f(x), x-d)\| = \frac{\varepsilon}{\sqrt{2}}\|(e,e^*)\|\big\|\big((x,d),(y_1,y_2)\big)-\big((\bar{x},\bar{x}),(0,0)\big)\big\|.
\end{aligned}
$$

Since $\min_c \|c\|^2 + \|e^*-c\|^2 = \frac{1}{2}\|e^*\|^2$, we obtain $\big\|\big((z_1,z_2),(z_1^*,z_2^*)\big)\big\|^2 \ge \|e\|^2 + \frac{1}{2}\|e^*\|^2 \ge \frac{1}{2}\|(e,e^*)\|^2$ and

$$
\begin{aligned}
&|\langle (z_1,z_2),(y_1,y_2)\rangle - \langle (z_1^*,z_2^*),(x,d)-(\bar{x},\bar{x})\rangle| \\
&\qquad \le \varepsilon\big\|\big((z_1,z_2),(z_1^*,z_2^*)\big)\big\|\big\|\big((x,d),(y_1,y_2)\big)-\big((\bar{x},\bar{x}),(0,0)\big)\big\|.
\end{aligned}
$$

Thus $\mathscr{F}$ is SCD semismooth* at $\big((\bar{x},\bar{x}),(0,0)\big)$. □

We proceed now with the description of the approximation step. Given $(x^{(k)}, d^{(k)})$ and a scaling parameter $\gamma^{(k)}$, we compute $u^{(k)} := u_{\gamma^{(k)}}(x^{(k)})$ and set

$$
\hat{x}^{(k)} = x^{(k)},\ \hat{d}^{(k)} = x^{(k)} + u^{(k)} \quad \text{and} \quad \hat{y}^{(k)} = (\hat{y}_1^{(k)}, \hat{y}_2^{(k)}) = -(\gamma^{(k)} u^{(k)}, u^{(k)}). \tag{5.16}
$$

We observe that

$$
\big((\hat{x}^{(k)}, \hat{d}^{(k)}), (\hat{y}_1^{(k)}, \hat{y}_2^{(k)})\big) \in \operatorname{gph} \mathscr{F},
$$

which follows immediately from the first-order optimality condition (5.12). Note that the result of the approximation step does not depend on the auxiliary variable $d^{(k)}$. In order to apply Theorem 4.4, we have to show the existence of a real $\eta > 0$ such that the estimate

$$
\big\|\big((\hat{x}^{(k)}-\bar{x}, \hat{d}^{(k)}-\bar{x}), \hat{y}^{(k)}\big)\big\| \le \eta \|(x^{(k)}-\bar{x}, d^{(k)}-\bar{x})\|, \tag{5.17}
$$

corresponding to (4.9), holds for all $(x^{(k)}, d^{(k)})$ with $x^{(k)}$ close to $\bar{x}$. By virtue of (5.16) the left-hand side of (5.17) amounts to

$$
\begin{aligned}
\big\|\big((\hat{x}^{(k)}-\bar{x}, \hat{x}^{(k)}+u^{(k)}-\bar{x}), (-\gamma^{(k)}u^{(k)}, -u^{(k)})\big)\big\| &\le \|(\hat{x}^{(k)}-\bar{x}, \hat{x}^{(k)}-\bar{x}, 0, 0)\| + \|(0, u^{(k)}, -\gamma^{(k)}u^{(k)}, -u^{(k)})\| \\
&\le 2\|\hat{x}^{(k)}-\bar{x}\| + (2+\gamma^{(k)})\|u^{(k)}\|. \qquad (5.18)
\end{aligned}
$$

Since $u_{\gamma^{(k)}}(\bar{x}) = 0$, we obtain from (5.13) the bounds

$$
\begin{aligned}
\|\hat{d}^{(k)} - \bar{x}\| &\le \|x^{(k)} - \bar{x}\| + \frac{1}{\gamma^{(k)}}\|f(x^{(k)}) - f(\bar{x})\| \\
\|u^{(k)}\| &\le 2\|x^{(k)} - \bar{x}\| + \frac{1}{\gamma^{(k)}}\|f(x^{(k)}) - f(\bar{x})\|.
\end{aligned}
$$

The latter estimate, together with (5.18), implies

$$
\begin{aligned}
\big\|\big((\hat{x}^{(k)}-\bar{x}, \hat{d}^{(k)}-\bar{x}), \hat{y}^{(k)}\big)\big\| &\le \Big(2 + (2+\gamma^{(k)})\big(2 + \frac{l}{\gamma^{(k)}}\big)\Big)\|x^{(k)} - \bar{x}\| \\
&\le \Big(2 + (2+\gamma^{(k)})\big(2+\frac{l}{\gamma^{(k)}}\big)\Big)\|(x^{(k)}-\bar{x}, d^{(k)}-\bar{x})\|,
\end{aligned} \tag{5.19}
$$

where l is the Lipschitz constant of f in a neighborhood of $\bar{x}$. Thus, the desired inequality (5.17) holds, as long as $\gamma^{(k)}$ remains bounded and bounded away from zero.

Next we describe the Newton step. According to Algorithm 1 and (5.15), we have to compute a pair $Y^{*(k)}, X^{*(k)}$ with rge $(Y^{*(k)}, X^{*(k)}) \in \mathscr{S}^* \partial q(\hat{d}^{(k)}, \hat{d}^{*(k)})$ and then solve the linear system

$$\begin{pmatrix} {Y^{*(k)}}^T \nabla f(x^{(k)}) & {X^{*(k)}}^T \\ -I & I \end{pmatrix} \begin{pmatrix} \Delta x^{(k)} \\ \Delta d^{(k)} \end{pmatrix} = - \begin{pmatrix} {Y^{*(k)}}^T & 0 \\ 0 & -I \end{pmatrix} \begin{pmatrix} \hat{y}_1^{(k)} \\ \hat{y}_2^{(k)} \end{pmatrix}$$

Simple algebraic transformations yield

$$({Y^{*(k)}}^T \nabla f(x^{(k)}) + {X^{*(k)}}^T) \Delta x^{(k)} = -({Y^{*(k)}}^T \hat{y}_1^{(k)} + {X^{*(k)}}^T \hat{y}_2^{(k)}) \tag{5.20}$$

and $\Delta d^{(k)} = \hat{y}_2^{(k)} + \Delta x^{(k)}$. Using (5.16) the system (5.20) amounts to

$$({Y^{*(k)}}^T \nabla f(x) + {X^{*(k)}}^T) \Delta x^{(k)} = (\gamma^{(k)} {Y^{*(k)}}^T + {X^{*(k)}}^T) u^{(k)}. \tag{5.21}$$

Having computed the Newton direction, the new iterate is given by $x^{(k+1)} = x^{(k)} + \Delta x^{(k)} = d^{(k+1)}$. We summarize our considerations in the following algorithm, where the auxiliary variable $d^{(k)}$ is omitted.

Algorithm 2 (semismooth* Newton Method for VI of the second kind (1.1))

1. Choose starting point $x^{(0)}$ and set the iteration counter $k := 0$.
2. If $0 \in H(x^{(k)})$ stop the algorithm.
3. Select a parameter $\gamma^{(k)} > 0$, compute $u^{(k)} := u_{\gamma^{(k)}}(x^{(k)})$ and set $\hat{d}^{(k)} := x^{(k)} + u^{(k)}$, $\hat{d}^{*(k)} := -\gamma^{(k)} u^{(k)} - f(x^{(k)})$.
4. Select $(X^{*(k)}, Y^{*(k)})$ with rge $(Y^{*(k)}, X^{*(k)}) \in \mathscr{S}^* \partial q\big((\hat{d}^{(k)}, \hat{d}^{*(k)})\big)$, compute the Newton direction $\Delta x^{(k)}$ from (5.21) and set $x^{(k+1)} = x^{(k)} + \Delta x^{(k)}$.
5. Increase the iteration counter $k := k + 1$ and go to Step 2.

Combining Theorem 4.4 with Proposition 5.1 we obtain the following convergence result.

Theorem 5.2 *Let $\bar{x} \in H^{-1}(0)$ be a solution of (1.1) and assume that ∂q is SCD semismooth* at $(\bar{x}, -f(\bar{x}))$. Further suppose that H is SCD regular around $(\bar{x}, 0)$. Then for every pair $\underline{\gamma}, \bar{\gamma}$ with $0 < \underline{\gamma} \leq \bar{\gamma}$ there exists a neighborhood U of $\bar{x}$ such that for every starting point $x^{(0)} \in U$ Algorithm 2 produces a sequence $x^{(k)}$ converging superlinearly to $\bar{x}$, provided we choose in every iteration step $\gamma^{(k)} \in [\underline{\gamma}, \bar{\gamma}]$.*

Let us now compare the assumptions of Theorem 5.2 with those required for the Josephy-Newton method [12, 13], in which one computes the next iterate $x^{(k+1)}$ as a solution of the linearized GE

$$0 \in f(x^{(k)}) + \nabla f(x^{(k)})(x - x^{(k)}) + \partial q(x).$$

Apart from semismoothness*, the assumed SCD regularity is weaker than the metric regularity often imposed on (1.1) to ensure that the Josephy-Newton method is well defined. We refer to [5, Example 5.13] for an example, where the assumptions of Theorem 5.2 are fulfilled, the semismooth* Newton method works well, but the Jospehy-Newton method collapses. On the other hand, if (1.1) models the KKT conditions of a mathematical program, see [24], Bonnans [2] showed that the strong metric subregularity of H at $\big((\bar{x}, \bar{\lambda}), (0, 0)\big)$ is sufficient for local superlinear convergence of the Josephy-Newton method, where $\bar{x}$ denotes a local minimizer for the mathematical program and $\bar{\lambda}$ is the corresponding unique

Lagrange multiplier. The strong metric subregularity of H at the solution is weaker than the SCD regularity but, as shown in the next example, this weakening does not work in the SCD semismooth* Newton method. Consider the mathematical program

$$\min_{x\in\mathbb{R}^2} \frac{1}{2}x_1^2 \quad \text{s.t.} \quad x \in C := \{x \in \mathbb{R}^2 \mid -x_1 + x_2 \le 0, -x_1 - x_2 \le 0\}.$$

The first order optimality conditions in multiplier-free form read as

$$(0,0) \in H(x) := (x_1, 0) + \partial\delta_C(x) = (x_1, 0) + N_C(x). \tag{5.22}$$

The mapping H is a polyhedral mapping and therefore metrically subregular at any point of its graph. Further, $H^{-1}(0) = \{0\}$ and consequently H is strongly metrically subregular at $(0, 0)$. Obviously, the Newton-Josephy method reaches the solution in one step from any starting point $x^{(0)} \in \mathbb{R}^2$. On the other hand, for any $x = (x_1, x_2) \in \operatorname{int} C$ we have $H(x) = \{(x_1, 0)\}$ and

$$\mathscr{S}^*H\big((x_1, x_2), (x_1, 0)\big)=\{\bar{L}\} \quad \text{with} \quad \bar{L}=\{(p_1, p_2, p_1, 0) \in \mathbb{R}^4 \mid (p_1, p_2)\in\mathbb{R}^2\} \notin \mathscr{Z}_2^{\text{reg}}.$$

Consequently $\bar{L} \in \mathscr{S}^*H\big((0, 0), (0, 0)\big)$, implying that H is not SCD regular at the solution. Furthermore, the Newton step is not defined at iterates $(\hat{x}^{(k)}, \hat{y}^{(k)})$ with $\hat{x}^{(k)} \in \operatorname{int} C$. This shows that in this example the assumption of SCD regularity is indispensable. A similar situation arises when, instead of (5.22), we consider the KKT conditions of the problem.

6 Globalization

In the preceding section we showed locally superlinear convergence of our implementation of the SCD semismooth* Newton method. However, we do not only want fast local convergence but also convergence from arbitrary starting points. To this end we consider a non-monotone line-search heuristic as well as hybrid approaches, which combine this heuristic with some globally convergent method.

To perform the line search, we need some merit function. Similar to the damped Newton method for solving smooth equations, we use some kind of residual. Here we define the residual by means of the approximation step, i.e., given x and $\gamma > 0$, we use

$$r_\gamma(x) := \|(\hat{y}_1^{(k)}, \hat{y}_2^{(k)})\| = \big\|\big(\gamma u_\gamma(x), u_\gamma(x)\big)\big\| = \sqrt{1+\gamma^2}\|u_\gamma(x)\| \tag{6.23}$$

as motivated by (5.16). Note that every evaluation of the residual function $r_\gamma(x)$ requires the computation of $u_\gamma(x)$.

Our globalization approaches are intended mainly for the case when the variational inequality (1.1) does not correspond to the solution of some nonsmooth optimization problem. For the solution of optimization problems, namely, there exist more efficient globalization strategies based on merit functions derived from the objective, and this case will be treated in a forthcoming paper.

6.1 A Non-monotone Line-search Heuristic

In general, we replace the full Newton step 4. in Algorithm 2 by a damped step of the form

$$x^{(k+1)} = \hat{x}^{(k)} + \alpha^{(k)}\triangle x^{(k)},$$

where $\alpha^{(k)} \in (0, 1]$ is chosen so that the line search condition

$$r_{\gamma^{(k)}}(\hat{x}^{(k)} + \alpha^{(k)}\triangle x^{(k)}) \le (1 + \delta^{(k)} - \nu\alpha^{(k)})r_{\gamma^{(k)}}(\hat{x}^{(k)}) \tag{6.24}$$

is fulfilled, where $\nu \in (0, 1)$ and $\delta^{(k)}$ is a given sequence of positive numbers converging to 0.

Obviously, the step size $\alpha^{(k)}$ exists since the residual function $r_\gamma(x)$ is continuous. However, it is not guaranteed that the residual is decreasing, i.e., that $r_{\gamma^{(k)}}(x^{(k+1)}) < r_{\gamma^{(k)}}(\hat{x}^{(k)})$.

The computation of $\alpha^{(k)}$ can be done in the usual way. For instance, as $\alpha^{(k)}$ we can choose the first element of a sequence (β_j), which fulfills $\beta_0 = 1$ and converges monotonically to zero, such that the line search condition (6.24) is fulfilled.

For $\gamma^{(k)}$ we suggest a choice with $\gamma^{(k)} \approx \|\nabla f(x^{(k)})\|$. Since the spectral norm $\|\nabla f(x^{(k)})\|$ is difficult to compute, we use an easy computable norm instead, e.g., the maximum absolute column sum norm $\|\nabla f(x^{(k)})\|_1$.

Although we are not able to show convergence properties for this heuristic, it showed good convergence properties in practice.

6.2 Globally Convergent Hybrid Approaches

In this subsection, we suggest a combination of the semismooth* Newton method with some existing globally convergent method which exhibits both global convergence and local superlinear convergence. For this purpose we assume that we dispose with a method, which is globally convergent in the sense that it generates from an arbitrary starting point $x^{(0)}$ a sequence $x^{(k)}$ such that at least one accumulation point of $x^{(k)}$ is a solution to the GE (1.1). We suppose that this method is formally given by some mapping $\mathscr{T} : \mathbb{R}^n \to \mathbb{R}^n$, which computes from the iterate $x^{(k)}$ the next iterate by

$$x^{(k+1)} = \mathscr{T}(x^{(k)}).$$

Of course, $\mathscr{T}$ must depend on the problem (1.1) which we want to solve and will presumably depend also on some additional parameters which control the behavior of the method. In our notation, these dependencies are neglected to a large extent.

Consider the following well-known examples for such a mapping $\mathscr{T}$, which have the required convergence properties at least under some suitable monotonicity assumptions on (1.1).

1. For the forward-backward splitting method, the mapping $\mathscr{T}$ is given by

$$\mathscr{T}_\lambda^{\mathrm{FB}}(x) = (I + \lambda \partial q)^{-1}(I - \lambda f)(x),$$

   where $\lambda > 0$ is a suitable parameter. Note that $\mathscr{T}_\lambda^{\mathrm{FB}}(x) = x + u_{1/\lambda}(x)$.
2. For the Douglas-Rachford splitting method we have

$$\mathscr{T}_\lambda^{\mathrm{DR}}(x) = (I+\lambda f)^{-1}\Big((I+\lambda\partial q)^{-1}(I-\lambda f)+\lambda f\Big)(x) = (I+\lambda f)^{-1}(\mathscr{T}_\lambda^{\mathrm{FB}}+\lambda f)(x),$$

   where $\lambda > 0$ is again some parameter.
3. A third method is given by the hybrid projection-proximal point algorithm due to Solodov and Svaiter [28]. Let x and a suitable parameter $\mu > 0$ be given and consider $\hat{x} = \mathscr{T}_{1/\mu}^{\mathrm{FB}}(x)$, i.e. $\hat{x} - x = u_\mu(x)$. Then $0 \in \mu(\hat{x} - x) + f(x) + \partial q(\hat{x})$ and consequently

$$v := -\mu(\hat{x} - x) + f(\hat{x}) - f(x) \in H(\hat{x}) \quad \text{and} \quad 0 = v + \mu(\hat{x} - x) + \big(f(x) - f(\hat{x})\big). \tag{6.25}$$

Then, in the hybrid projection-proximal point algorithm the mapping $\mathscr{T}$ is given by the projection of x on the hyperplane $\{z \mid \langle v, z - \hat{x}\rangle = 0\}$, i.e.,

$$\mathscr{T}_\mu^{\mathrm{PM}}(x) = x - \frac{\langle v, \hat{x} - x\rangle}{\|v\|^2} v.$$

Note that in principle we could also use other methods which depend not only on the last iterate like the golden ratio algorithm [19], but for ease of presentation these methods are not considered.

Algorithm 3 (Globally convergent hybrid semismooth* Newton method for VI of the second kind)
Input: A method for solving (1.1) given by the iteration operator $\mathscr{T} : \mathbb{R}^n \to \mathbb{R}^n$, a starting point $x^{(0)}$, line search parameter $0 < \nu < 1$, a sequence $\delta^{(k)} \in (0, 1)$, a sequence $\beta_j \downarrow 0$ with $\beta_0 = 1$ and a stopping tolerance $\varepsilon_{tol} > 0$.
1. Choose $\gamma^{(0)}$, set $r_N^{(0)} := r_{\gamma^{(0)}}(x^{(0)})$ and set the counters $k := 0, l := 0$.
2. If $r_{\gamma^{(k)}}(x^{(k)}) \le \varepsilon_{tol}$ stop the algorithm.
3. Perform the approximation step as in Algorithm 2 and compute the Newton direction $\Delta x^{(k)}$ by solving (5.21). Try to determine the step size $\alpha^{(k)}$ as the first element from the sequence β_j satisfying $\beta_j > \delta^{(l)}$ and

$$r_{\gamma^{(k)}}(x^{(k)} + \beta_j \Delta x^{(k)}) \le (1 - \nu\beta_j) r_N^{(l)}.$$

4. If both $\Delta x^{(k)}$ and $\alpha^{(k)}$ exist, set $x^{(k+1)} = x^{(k)} + \alpha^{(k)} \Delta x^{(k)}$, $r_N^{(l+1)} = r_{\gamma^{(k)}}(x^{(k+1)})$ and increase $l := l + 1$.
5. Otherwise, if the Newton direction $\Delta x^{(k)}$ or the step length $\alpha^{(k)}$ does not exist, compute $x^{(k+1)} = \mathscr{T}(x^{(k)})$.
6. Update $\gamma^{(k+1)}$ and increase the iteration counter $k := k + 1$ and go to Step 2.

In what follows, we denote by k_l the subsequence of iterations where the new iterate x^{k+1} is computed by damped Newton Step 4, i.e.,

$$x^{(k_l)} = x^{(k_l - 1)} + \alpha^{(k_l - 1)} \Delta x^{(k_l - 1)}, \; r_N^{(l)} = r_{\gamma^{(k_l - 1)}}(x^{(k_l)}).$$

Theorem 6.1 *Assume that the GE (1.1) has at least one solution and assume that the solution method given by the iteration mapping $\mathscr{T} : \mathbb{R}^n \to \mathbb{R}^n$ has the property that for every starting point $y^{(0)} \in \mathbb{R}^n$ the sequence $y^{(k)}$, given by the recursion $y^{(k+1)} = \mathscr{T}(y^{(k)})$, has at least one accumulation point which is a solution to the GE (1.1). Then for every starting point $x^{(0)}$ the sequence $x^{(k)}$, produced by Algorithm 3 with $\varepsilon_{tol} = 0$ and $\sum_{k=0}^{\infty} \delta^{(k)} = \infty$, has the following properties.*

(i) *If the Newton step is accepted only finitely many times in step 4, then the sequence $x^{(k)}$ has at least one accumulation point that solves (1.1).*
(ii) *If the Newton step is accepted infinitely many times in step 4, then every accumulation point of the subsequence $x^{(k_l)}$ is a solution to (1.1).*
(iii) *If there exists an accumulation point $\bar{x}$ of the sequence $x^{(k)}$ which solves (1.1), the mapping H is SCD regular around $(\bar{x}, 0)$ and ∂q is SCD semismooth* at $\big(\bar{x}, -f(\bar{x})\big)$, then the sequence $x^{(k)}$ converges superlinearly to $\bar{x}$ and the Newton step in step 3 is*

accepted with step length $\alpha^{(k)} = 1$ *for all* k *sufficiently large, provided the sequence* $\gamma^{(k)}$ *satisfies*

$$0 < \underline{\gamma} \leq \gamma^{(k)} \leq \bar{\gamma} \ \forall k$$

for some positive reals $\underline{\gamma}, \bar{\gamma}$.

Proof The first statement is an immediate consequence of our assumption on $\mathscr{T}$. In order to show the second statement, observe that the sequence $r_N^{(l)}$ satisfies $r_N^{(l+1)} \leq (1-\nu\delta^{(l)})r_N^{(l)}$ implying

$$\lim_{l\to\infty} \ln(r_N^{(l+1)}) - \ln(r_N^{(0)}) \leq \lim_{l\to\infty} \sum_{i=0}^{l} \ln(1-\nu\delta^{(i)}) \leq -\lim_{l\to\infty} \sum_{i=0}^{l} \nu\delta^{(i)} = -\infty.$$

Thus $\lim_{l\to\infty} r_N^{(l)} = \lim_{l\to\infty} \sqrt{1+\gamma^{(k_l-1)^2}}\|u_{\gamma^{(k_l-1)}}(x^{(k_l)})\| = 0$ and we can conclude that

$$\lim_{l\to\infty} \|u_{\gamma^{(k_l-1)}}(x^{(k_l)})\| = \lim_{l\to\infty} \gamma^{(k_l-1)}\|u_{\gamma^{(k_l-1)}}(x^{(k_l)})\| = 0.$$

Together with the inclusion

$$0 \in \gamma^{(k_l-1)}u_{\gamma^{(k_l-1)}}(x^{(k_l)}) + f(x^{(k_l)}) + \partial q\big(x^{(k_l)} + u_{\gamma^{(k_l-1)}}(x^{(k_l)})\big),$$

the continuity of f and the closedness of gph ∂q, it follows that $0 \in f(\bar{x}) + \partial q(\bar{x})$ holds for every accumulation point $\bar{x}$ of the subsequence $x^{(k_l)}$. This proves our second assertion.

Finally, we want to show (iii). Assume that $\bar{x}$ is an accumulation point of the sequence $x^{(k)}$ such that the mapping H is SCD regular around $(\bar{x}, 0)$ and ∂q is SCD semismooth* at $\big(\bar{x}, -f(\bar{x})\big)$. By Proposition 5.1 the mapping $\mathscr{F}$ is SCD regular and SCD semismooth* at $\big((\bar{x},\bar{x}),(0,0)\big)$. By invoking [6, Theorem 6.2], the mapping $\mathscr{F}$ is strongly metrically subregular at $\big((\bar{x},\bar{x}),(0,0)\big)$ and, moreover, there are some $\kappa > 0$ and some neighborhoods U of $(\bar{x},\bar{x})$ and V of $(0,0)$ such that

$$\|(x,d) - (\bar{x},\bar{x})\| \leq \kappa \operatorname{dist}((0,0), \mathscr{F}(x,d)) \ \forall (x,d) \in U, \tag{6.26}$$

$$L \in \mathscr{Z}_{2n}^{\mathrm{reg}} \text{ and } \|C_L\| \leq \kappa \ \forall L \in \mathscr{S}^*\mathscr{F}\big((x,d),(y_1,y_2)\big) \ \forall \big((x,d),(y_1,y_2)\big) \in U \times V. \tag{6.27}$$

Thus, whenever $\big((\hat{x}^{(k)},\hat{d}^{(k)}),\hat{y}^{(k)}\big) \in U \times V$, the Newton direction $(\Delta x^{(k)}, \Delta d^{(k)})$ exists and satisfies $\|(\Delta x^{(k)}, \Delta d^{(k)})\| \leq \kappa\|\hat{y}^{(k)}\|$.

By Proposition 4.2 and (6.27), for every $\varepsilon > 0$ there is some $\delta > 0$ such that

$$\|\hat{x}^{(k)} + \Delta x^{(k)} - \bar{x}\| \leq \left\| \begin{pmatrix} \hat{x}^{(k)} + \Delta x^{(k)} - \bar{x} \\ \hat{d}^{(k)} + \Delta d^{(k)} - \bar{x} \end{pmatrix} \right\| \leq \varepsilon\sqrt{2n(1+\kappa^2)} \left\| \big((\hat{x}^{(k)} - \bar{x}, \hat{d}^{(k)} - \bar{x}), \hat{y}^{(k)}\big) \right\|$$

whenever $\big((\hat{x}^{(k)},\hat{d}^{(k)}),\hat{y}^{(k)}\big) \in \mathscr{B}_\delta\big((\bar{x},\bar{x}),(0,0)\big)$. Therefore, we can find some $\delta' \in (0,1]$ such that $\mathscr{B}_{\delta'}\big((\bar{x},\bar{x}),(0,0)\big) \subset U \times V$ and

$$\|\hat{x}^{(k)} + \Delta x^{(k)} - \bar{x}\| \leq \min\left\{ \frac{1-\nu}{c_1 c_2 \kappa\sqrt{1+\bar{\gamma}^2}}, \frac{1}{2c_2} \right\} \left\| \big((\hat{x}^{(k)} - \bar{x}, \hat{d}^{(k)} - \bar{x}), \hat{y}^{(k)}\big) \right\|$$

for $\big((\hat{x}^{(k)},\hat{d}^{(k)}),\hat{y}^{(k)}\big) \in \mathscr{B}_{\delta'}\big((\bar{x},\bar{x}),(0,0)\big)$, where $c_1 := 2 + \frac{l}{\underline{\gamma}}$, $c_2 := 2 + (2+\bar{\gamma})c_1$ and l is some Lipschitz constant of f in $\mathscr{B}_1(\bar{x})$. From (5.19) we deduce $\left\| \big((\hat{x}^{(k)} - \bar{x}, \hat{d}^{(k)} - \bar{x}), \hat{y}^{(k)}\big) \right\| \leq c_2\|x^{(k)} - \bar{x}\|$ that yields

$$\|\hat{x}^{(k)} + \Delta x^{(k)} - \bar{x}\| \leq \min\left\{ \frac{1-\nu}{c_1\kappa\sqrt{1+\bar{\gamma}^2}}, \frac{1}{2} \right\} \|\hat{x}^{(k)} - \bar{x}\| \tag{6.28}$$

for $x^{(k)} \in \mathscr{B}_{\bar\delta}(\bar x)$ with $\bar\delta := \delta'/c_2$. We now claim that for every iterate $x^{(k)} \in \mathscr{B}_{\bar\delta}(\bar x)$ the Newton step with step size $\alpha^{(k)} = 1$ is accepted. If $x^{(k)} \in \mathscr{B}_{\bar\delta}(\bar x)$ then $\big((\hat x^{(k)}, \hat d^{(k)}), \hat y^{(k)}\big) \in \mathscr{B}_{\delta'}\big((\bar x, \bar x), (0,0)\big) \subset U \times V$ and from (6.26) we obtain

$$\|x^{(k)} - \bar x\| \le \|(\hat x^{(k)}, \hat d^{(k)}) - (\bar x, \bar x)\| \le \kappa \operatorname{dist}((0,0), \mathscr{F}(\hat x^{(k)}, \hat d^{(k)})) \le \kappa \|\hat y^{(k)}\| \le \kappa\sqrt{1+\bar\gamma^2}\,\|u^{(k)}\|.$$

Since $u_{\gamma^{(k)}}(\bar x) = 0$, we obtain from (5.13) and (6.28) that

$$\begin{aligned}\|u_{\gamma^{(k)}}(x^{(k)} + \Delta x^{(k)})\| &\le c_1 \|x^{(k)} + \Delta x^{(k)} - \bar x\| \le \frac{1-\nu}{\kappa\sqrt{1+\bar\gamma^2}} \|x^{(k)} - \bar x\| \le (1-\nu)\|u^{(k)}\| \\ &= (1-\nu)\|u_{\gamma^{(k)}}(x^{(k)})\|,\end{aligned}$$

showing

$$\begin{aligned} r_{\gamma^{(k)}}(x^{(k)} + \Delta x^{(k)}) &= \sqrt{1+\gamma^{(k)^2}}\,\|u_{\gamma^{(k)}}(x^{(k)} + \Delta x^{(k)})\| \le (1-\nu)\sqrt{1+\gamma^{(k)^2}}\,\|u_{\gamma^{(k)}}(x^{(k)})\| \\ &= (1-\nu) r_{\gamma^{(k)}}(x^{(k)}). \end{aligned}$$

From this we conclude that the step size $\alpha^{(k)} = 1$ is accepted. Now let $\bar k$ denote the first index such that $x^{(\bar k)}$ enters the ball $\mathscr{B}_{\bar\delta}$. Then for all $k \ge \bar k$ we have

$$x^{(k+1)} = x^{(k)} + \Delta x^{(k)}, \quad \|x^{(k+1)} - \bar x\| \le \frac{1}{2}\|x^{(k)} - \bar x\|$$

and superlinear convergence follows from Theorem 5.2. □

One of the reviewers asked us to work out the iteration $x^{(k+1)} = \mathscr{T}(x^{(k)})$ in step 5. of Algorithm 3 in more detail and we will do this now for the hybrid projection-proximal method $\mathscr{T}_\mu^{\mathrm{PM}}$. For the rest of this section, assume that H is maximally monotone. To ensure convergence of the method, the parameter μ must be properly chosen. In the following implementation of the hybrid projection-proximal method we do not use a constant value of μ, but adapt it properly in each iteration so that one step of the hybrid projection-proximal method is formally given by

$$(x^{(k+1)}, \mu^{(k+1)}) = \mathscr{T}^{\mathrm{PM}}(x^{(k)}, \mu^{(k)}).$$

Algorithm 4 (One step of the hybrid projection-proximal method)
Input: Previous iterate $(x^{(k)}, \mu^{(k)})$, constants $\hat\mu > 0$, $0 < \alpha_1 < \alpha_2 < 1$, $0 < \xi_2 < 1 < \xi_1$.
Output: Next iterate $(x^{(k+1)}, \mu^{(k+1)}) = \mathscr{T}^{\mathrm{PM}}(x^{(k)}, \mu^{(k)})$

1. Set $\mu := \mu^{(k)}$.
2. Compute $\hat x^{(k)} := \mathscr{T}^{\mathrm{FB}}_{1/\mu}(x^{(k)})$, $v^{(k)} := -\mu(\hat x^{(k)} - x^{(k)}) + f(\hat x^{(k)}) - f(x^{(k)})$.
3. If $\hat x^{(k)} == x^{(k)}$ then stop: $x^{(k)}$ solves (1.1).
4. If $\langle v^{(k)}, x^{(k)} - \hat x^{(k)}\rangle \le \alpha_1 \|v^{(k)}\|\|x^{(k)} - \hat x^{(k)}\|$ then set $\mu := \xi_1\mu$ and go to step 2.
5. Set $x^{(k+1)} := x^{(k)} - \frac{\langle v^{(k)}, \hat x^{(k)} - x^{(k)}\rangle}{\|v^{(k)}\|^2} v^{(k)}$.
6. If $\mu == \mu^{(k)}$ and $\langle v^{(k)}, x^{(k)} - \hat x^{(k)}\rangle > \alpha_2 \|v^{(k)}\|\|x^{(k)} - \hat x^{(k)}\|$ then set $\mu^{(k+1)} := \max\{\xi_2\mu, \hat\mu\}$ else set $\mu^{(k+1)} := \mu$.

For our convergence analysis we assume that the algorithm does not stop prematurely in step 3. with a solution of the GE (1.1).

Theorem 6.2 *Assume that the GE (1.1) has at least one solution and that H is monotone. Let $x^{(0)} \in \mathbb{R}^n$, $\mu^{(0)} > 0$ be given and consider a sequence $(x^{(k)}, \mu^{(k)})$ produced by Algorithm 4, where we assume $\mu^{(0)} \geq \hat{\mu}$. Then the sequence $(x^{(k)}, \mu^{(k)})$ is well defined and bounded and every accumulation point of $x^{(k)}$ is a solution of (1.1).*

Proof Consider $\bar{x} \in H^{-1}(0)$ and let $\rho_1 := \|x^{(0)} - \bar{x}\|$. The mapping $x \mapsto u_{\hat{\mu}}(x)$ is continuous and thus $\rho_2 := \max\{\|u_{\hat{\mu}}(x)\| \mid x \in \mathscr{B}_{\rho_1}(\bar{x})\} < \infty$. Set $\rho := \rho_1 + \rho_2$ and $l := \max\{\|\nabla f(x)\| \mid x \in \mathscr{B}_{\rho}(\bar{x})\}$. Note that f is Lipschitz continuous on $\mathscr{B}_{\rho}(\bar{x})$ with constant l. Next consider arbitrary $x \in \mathscr{B}_{\rho_1}(\bar{x})$, $\mu \geq \hat{\mu}$ with $\mu > \bar{\mu} := l(1+\alpha_1)/(1-\alpha_1)$ and the related quantities $\hat{x} - x = u_\mu(x)$, $v := -\mu(\hat{x} - x) + f(\hat{x}) - f(x)$. Assuming $x \neq \hat{x}$, we claim that

$$\langle v, x - \hat{x}\rangle > \alpha_1 \|v\| \|x - \hat{x}\|.$$

Indeed, the mapping $\mu \mapsto \|u_\mu(x)\|$ is non-increasing, cf. [1, Theorem 10.9], and therefore $\|x - \hat{x}\| = \|u_\mu(x)\| \leq \|u_{\hat{\mu}}(x)\| \leq \rho_2$. Hence $x, \hat{x} \in \mathscr{B}_{\rho}(\bar{x})$ and, since condition $\mu > l(1+\alpha_1)/(1-\alpha_1)$ is equivalent to $\mu - l > \alpha_1(\mu + l)$, we obtain that

$$\begin{aligned}\langle v, x - \hat{x}\rangle &= \mu\|x - \hat{x}\|^2 + \langle f(\hat{x}) - f(x), x - \hat{x}\rangle \geq \mu\|x - \hat{x}\|^2 - l\|\hat{x} - x\|\|x - \hat{x}\| = (\mu - l)\|x - \hat{x}\|^2 \\ &> \alpha_1(\mu + l)\|x - \hat{x}\|^2 = \alpha_1(\mu\|x - \hat{x}\| + l\|x - \hat{x}\|)\|x - \hat{x}\| \geq \alpha_1\|v\|\|x - \hat{x}\|\end{aligned}$$

verifying our claim. By our assumptions, we have $x^{(0)} \in \mathscr{B}_{\rho_1}(\bar{x})$, $\mu^{(0)} \geq \hat{\mu}$ and consider any iterate $(x^{(k)}, \mu^{(k)}) \in \mathscr{B}_{\rho_1}(\bar{x}) \times [\hat{\mu}, \infty)$. As soon as the value of μ in the loop between steps 2. and 4. of Algorithm 4 exceeds $\bar{\mu}$, we have

$$\langle v^{(k)}, x^{(k)} - \hat{x}^{(k)}\rangle > \alpha_1 \|v^{(k)}\| \|x^{(k)} - \hat{x}^{(k)}\|$$

by our claim, implying that the loop is terminated with a final value of $\mu \leq \max\{\mu^{(k)}, \xi_1\bar{\mu}\}$ and consequently $x^{(k+1)}$ and $\mu^{(k+1)} \in [\hat{\mu}, \max\{\mu^{(k)}, \xi_1\bar{\mu}\}]$ are well defined. Since $\langle v^{(k)}, x^{(k)} - \hat{x}^{(k)}\rangle > 0$, we can apply [28, Lemma 2.1] to obtain

$$\|x^{(k+1)} - \bar{x}\|^2 \leq \|x^{(k)} - \bar{x}\|^2 - \left(\frac{\langle v^{(k)}, x^{(k)} - \hat{x}^{(k)}\rangle}{\|v^{(k)}\|}\right)^2 \leq \|x^{(k)} - \bar{x}\|^2 - \alpha_1^2\|\hat{x}^{(k)} - x^{(k)}\|^2, \tag{6.29}$$

showing $x^{(k+1)} \in \mathscr{B}_{\rho_1}(\bar{x})$. Thus, an inductive argument yields that all iterates $(x^{(k)}, \mu^{(k)})$ belong to the bounded set $\mathscr{B}_{\rho_1}(\bar{x}) \times [\hat{\mu}, \max\{\mu^{(0)}, \xi_1\bar{\mu}\}]$. Summing up inequality (6.29) repeatedly, we obtain

$$0 \leq \|x^{(i+1)} - \bar{x}\|^2 \leq \|x^{(0)} - \bar{x}\|^2 - \alpha_1^2 \sum_{k=0}^{i} \|\hat{x}^{(k)} - x^{(k)}\|^2 \ \forall i$$

and we can conclude that

$$\lim_{k\to\infty} \|\hat{x}^{(k)} - x^{(k)}\| = \lim_{k\to\infty} \|u_{\mu^{(k)}}(x^{(k)})\| = 0.$$

Now let $\tilde{x}$ be an accumulation point of the sequence $x^{(k)}$. Then we can find a subsequence k_i with $x^{(k_i)} \to \tilde{x}$ and such that $\mu^{(k_i)}$ converges to some $\mu \in [\hat{\mu}, \max\{\mu^{(0)}, \xi_1\bar{\mu}\}]$. By [27, Theorem 1.25] we obtain that $u_\mu(\tilde{x}) = 0$ and therefore $\tilde{x}$ solves (1.1). □

Remark 6.3 Note that Theorem 6.2 does not require that f is globally Lipschitz continuous on $\mathbb{R}^n$. We only need that f is strictly continuous, cf. [27, Definition 9.1], so that f is Lipschitz continuous on convex compact sets.

7 Numerical Experiments

Based on the general results from [4], the authors in [22] considered an evolutionary Cournot-Nash equilibrium, where in the course of time the players (producers) adjust their productions to respond adequately to changing external parameters. Following [4], however, each change of production is generally associated with some expenses, called *costs of change*. In this way one obtains a generalized equation (1.1) which has to be solved repeatedly in each selected time step.

In this paper we make the model from [22] more involved by admitting multiple commodities and more realistic production constraints. As the solver of the respective generalized equation (1.1), the SCD semismooth* Newton method (Algorithm 2) will be employed. The new model is described as follows: Let n, m be the number of players and the number of produced commodities, respectively. Further, let $x = (x^1, \dots, x^n) \in (\mathbb{R}^m)^n$ be the cumulative vector of productions, where

$$x^i = (x^i_1, x^i_2, \dots, x^i_m) \in \mathbb{R}^m_+, \quad i = 1, 2, \dots, n$$

stands for the *production portfolio* of the i-th player. With each player we associate

- the mapping $c^i : \mathbb{R}^m_+ \to \mathbb{R}$ which assigns x^i the respective production cost;
- the linear system of inequalities $\Xi^i x^i \le \zeta^i$ with a $p^i \times m$ matrix Ξ^i and a vector $\zeta^i \in \mathbb{R}^{p^i}$ which specifies the set of *feasible productions* $\Omega^i = \{x^i \in \mathbb{R}^m \mid \Xi^i x^i \le \zeta^i\} \subseteq \mathbb{R}^m_+$, and
- the cost of change $z^i : \mathbb{R}^m \to \mathbb{R}$ which assigns each change of the production portfolio $\triangle x^i \in \mathbb{R}^m$ the corresponding cost.

Clearly, the vector $t = (t_1, t_2, \dots, t_m)$ with $t_j = \sum_{i=1}^n x^i_j$, $j = 1, \dots, m$, provides the overall amounts of single commodities which are available on the market in the considered time period. The price of the j-th commodity is given via the respective *inverse demand function* $\pi_j : \mathbb{R}_+ \to \mathbb{R}_+$ assigning each value t_j the corresponding price, at which the consumers are willing to buy.

Putting everything together, one arrives at the GE (1.1), where

$$f(x) = \begin{pmatrix} f^1(x) \\ \vdots \\ f^n(x) \end{pmatrix} \quad \text{with} \quad f^i(x) = \nabla c^i(x^i) - \begin{pmatrix} \pi_1(t_1) \\ \vdots \\ \pi_m(t_m) \end{pmatrix} - \begin{pmatrix} x^i_1 \nabla \pi_1(t_1) \\ \vdots \\ x^i_m \nabla \pi_m(t_m) \end{pmatrix}$$

and $q(x) = \sum_{i=1}^n \big(z^i(x^i) + \delta_{\Omega^i}(x^i)\big)$, $i = 1, 2, \dots, n$. Concerning functions c^i, $i = 1, \dots, n$, and π_j, $j = 1, \dots, m$, we use functions of the same type as in [21], i.e.,

$$c^i(x^i) = \sum_{j=1}^m \left(b^i_j x^i_j + \frac{\delta^i_j}{\delta^i_j + 1} K^i_j{}^{-\frac{1}{\delta^i_j}} |x^i_j|^{\frac{\delta^i_j + 1}{\delta^i_j}} \right), \ i = 1, \dots, n, \tag{7.30}$$

with positive parameters b^i_j, δ^i_j and K^i_j, and

$$\pi_j(t_j) = (1000n)^{\frac{1}{\gamma_j}} t_j^{-\frac{1}{\gamma_j}}, \ j = 1, \dots, m, \tag{7.31}$$

with positive parameters γ_j.

The functions z^i are modeled in the form

$$z^i(\triangle x^i) = z^i(x^i - a^i) = \sum_{j=1}^{m} \beta_j^i |x_j^i - a_j^i|, \ i = 1, \ldots, n, \tag{7.32}$$

where $a^i \in \Omega^i$ signifies the "previous" production portfolio of the i-th player and the weights β_j^i are positive reals indicating the costs of a "unit" change of production of the j-th commodity by the i-th player.

On the basis of [22] and [21] it can be shown that for each fixed choice of the parameters in (7.30), (7.31) and (7.32) the mapping $H(x) = f(x) + \partial q(x)$ is strictly monotone and the respective GE (1.1) has a unique solution $\bar{x}$ such that H is strongly metrically regular around $(\bar{x}, 0)$. From Theorem 3.6 and [6, Proposition 3.15] it follows that H is an SCD mapping whenever f is continuously differentiable near $\bar{x}$. Consequently, since $\operatorname{gph} \partial q$ is a polyhedral mapping, we infer from [5, Propositions 3.5, 3.6, 3.7] that in such a situation H is SCD semismooth* and so the conceptual Algorithm 1 may be used. However, when implementing Algorithm 2, one has to be careful because the mapping f does not meet the requirement of continuous differentiability on $\mathbb{R}^n$. Therefore we replace π_j by the twice continuously differentiable functions

$$\hat{\pi}_j(t_j) := \begin{cases} \pi_j(t_j) & \text{if } t_j > \varepsilon_1 \\ \pi_j(\varepsilon_1) + \pi_j'(\varepsilon_1)(t_j - \varepsilon_1) + \frac{1}{2}\pi_j''(\varepsilon_1)(t_j - \varepsilon_1)^2 & \text{if } t_j \le \varepsilon_1 \end{cases}$$

and, in the definition of $c^i(x^i)$, we replace the term $|x_j^i|$ by $\sqrt{(x_j^i)^2 + \varepsilon_2^2}$ whenever $\delta_j^i < 1$ (in our implementation we used $\varepsilon_1 := 10^{-1}, \varepsilon_2 := 10^{-10}$). Since the functions c^i are convex, one could alternatively incorporate them in q without smoothing instead of treating them as part of f.

Next we describe the approximation step of Algorithm 2, where $x^{(k)} = \big((x^1)^{(k)}, \ldots, (x^n)^{(k)}\big)$ stands for the k-th iterate. For a given scaling parameter $\gamma^{(k)} > 0$ and $i = 1, 2, \ldots, n$, we compute consecutively the (unique) solutions $(u^i)^{(k)}$, $i = 1, \ldots, n$, of the strictly convex optimization problems

$$\min_{u^i \in \mathbb{R}^m} \frac{\gamma^{(k)}}{2} \|u^i\|^2 + \langle f^i(x^{(k)}), u^i \rangle + q^i\big((x^i)^{(k)} + u^i\big), \tag{7.33}$$

obtaining thus the vector $u^{(k)} = \big((u^1)^{(k)}, \ldots, (u^n)^{(k)}\big) \in (\mathbb{R}^m)^n$. Due to the specific structure of the functions q^i, problem (7.33) can be replaced by the standard quadratic program

$$\begin{array}{ll} \min\limits_{(u^i, v^i) \in \mathbb{R}^m \times \mathbb{R}^m} & \frac{\gamma^{(k)}}{2} \|u^i\|^2 + \langle f^i(x^{(k)}), u^i \rangle + \sum_{j=1}^m \beta_j^i v_j^i \\ \text{subject to} & \Xi^i\big((x^i)^{(k)} + u^i\big) \le \zeta^i \\ & \left.\begin{array}{l} v_j^i \ge \quad (x_j^i)^{(k)} + u_j^i - a_j^i \\ v_j^i \ge -\big((x_j^i)^{(k)} + u_j^i - a_j^i\big) \end{array}\right\} j = 1, \ldots, m. \end{array}$$

Clearly, the u-component of the solution amounts exactly to the (unique) solution of (7.33). The outcome of the projection step is then given by the update (5.16), i.e.,

$$\hat{x}^{(k)} = x^{(k)}, \ \hat{d}^{(k)} = x^{(k)} + u^{(k)} = \big((\hat{d}^1)^{(k)}, \ldots, (\hat{d}^n)^{(k)}\big) \quad \text{and} \quad \hat{y}^{(k)} = -(\gamma^{(k)} u^{(k)}, u^{(k)}).$$

In the Newton step, we make use of the following theorem.

Theorem 7.1 *Let $g : \mathbb{R}^m \to \overline{\mathbb{R}}$ be given by $g(x) = \sum_{j=1}^m \beta_j |x_j - a_j| + \delta_\Omega(x)$, where $\beta_j \geq 0$, $a_j \in \mathbb{R}$, $j = 1, \ldots, m$ and $\Omega = \{x \in \mathbb{R}^n \mid \langle \xi_l, x\rangle \leq \zeta_l,\ l = 1, \ldots, p\}$ is a convex polyhedral set given by the vectors $\xi_l \in \mathbb{R}^m$ and scalars $\zeta_l \in \mathbb{R}$, $l = 1, \ldots, p$. Then for every $(x, x^*) \in \operatorname{gph} \partial g$ there holds*

$$W(x) \times W(x)^\perp \in \mathscr{S}\partial g(x, x^*) = \mathscr{S}^*\partial g(x, x^*),$$

where $W(x) := \{w \in \mathbb{R}^m \mid w_i = 0,\ i \in J_0(x), \langle \xi_l, w\rangle = 0, l \in L(x)\}$ with $J_0(x) := \{j \mid \beta_j > 0,\ x_j = a_j\}$ and $L(x) = \{l \mid \langle \xi_l, x\rangle = \zeta_l\}$.

Proof By standard calculus rules of convex analysis, for every $x \in \Omega$ we have

$$\partial g(x) = N_\Omega(x) + \sum_{j:\beta_j>0} \beta_j \partial |x_j - a_j| = \left\{ \sum_{l\in L(x)} \xi_l \mu_l + \sum_{j\in J_+(x)} \beta_j e_j - \sum_{j\in J_-(x)} \beta_j e_j \right.$$
$$\left. + \sum_{j\in J_0(x)1}^m \beta_j \tau_j e_j \mid \mu_l \geq 0,\ l \in L(x), \tau_j \in [-1, 1],\ j \in J_0(x)\right\},$$

where $J_+(x) := \{j \mid \beta_j > 0,\ x_j > a_j\}$, $J_-(x) := \{j \mid \beta_j > 0,\ x_j < a_j\}$ and e_j denotes the j-th unit vector. For every partition J_0, J_+, J_- of $\{j \in \{1, \ldots, m\} \mid \beta_j > 0\}$ and every index set $L \subseteq \{1, \ldots, p\}$ let

$$D_{J_0,J_+,J_-,L} := \left\{ x \,\middle|\, \begin{array}{l} x_j = a_j,\ j \in J_0,\ x_j \geq a_j,\ j \in J_+,\ x_j \leq a_j,\ j \in J_- \\ \langle \xi_l, x\rangle = \zeta_l,\ l \in L,\ \langle \xi_l, x\rangle \leq \zeta_l,\ l \notin L \end{array} \right\},$$

$$\tilde{D}_{J_0,J_+,J_-,L} = \left\{ \sum_{j\in J_0} \beta_j \tau_j e_j + \sum_{j\in J_+} \beta_j e_j - \sum_{j\in J_-} \beta_j e_j + \sum_{l\in L} \xi_l \mu_l \,\middle|\, \tau_j \in [-1, +1],\ j \in J_0,\ \mu_l \geq 0,\ l \in L \right\},$$

$$E_{J_0,J_+,J_-,L} := D_{J_0,J_+,J_-,L} \times \tilde{D}_{J_0,J_+,J_-,L}.$$

Further, we denote by $\mathscr{I}$ the collection of all those index sets (J_0, J_+, J_-, L) such that

$$\operatorname{ri} D_{J_0,J_+,J_-,L} = \left\{ x \,\middle|\, \begin{array}{l} x_j = a_j,\ j \in J_0,\ x_j > a_j,\ j \in J_+,\ x_j < a_j,\ j \in J_- \\ \langle \xi_l, x\rangle = \zeta_l,\ l \in L,\ \langle \xi_l, x\rangle < \zeta_l,\ l \notin L \end{array} \right\} \neq \emptyset.$$

It follows that for every $(J_0, J_+, J_-, L) \in \mathscr{I}$ and every $x \in D_{J_0,J_+,J_-,L}$ we have $x \in \Omega$ and $\tilde{D}_{J_0,J_+,J_-,L} \subseteq \partial g(x)$. Further, for every $x \in \Omega$ there holds $\big(J_0(x), J_+(x), J_-(x), L(x)\big) \in \mathscr{I}$ and $\tilde{D}_{J_0(x),J_+(x),J_-(x),L(x)} = \partial g(x)$ implying

$$\operatorname{gph} \partial g = \bigcup_{(J_0,J_+,J_-,L)\in\mathscr{I}} E_{J_0,J_+,J_-,L}.$$

We now claim that for any two elements $(J_0, J_+, J_-, L) \neq (J_0', J_+', J_-', L') \in \mathscr{I}$ we have $E_{J_0',J_+',J_-',L'} \cap \operatorname{ri} E_{J_0,J_+,J_-,L} = \emptyset$. Note that $\operatorname{ri} E_{J_0,J_+,J_-,L} = \operatorname{ri} D_{J_0,J_+,J_-,L} \times \operatorname{ri} \tilde{D}_{J_0,J_+,J_-,L}$ and that

$$\operatorname{ri} \tilde{D}_{J_0,J_+,J_-,L} = \Big\{ \sum_{j\in J_0} \beta_j \tau_j e_j + \sum_{j\in J_+} \beta_j e_j - \sum_{j\in J_-} \beta_j e_j + \sum_{l\in L} \xi_l \mu_l \mid \tau_j \in (-1, +1),\ j \in J_0,\ \mu_l > 0,\ l \in L \Big\}$$

by [26, Theorem 6.6.]. Assuming that this claim does not hold for some $(J_0, J_+, J_-, L) \neq (J_0', J_+', J_-', L') \in \mathscr{I}$, there are reals $\mu_l > 0, l \in L$, $\mu_l' \geq 0, l \in L'$, $\tau_j \in (-1, 1)$, $j \in J_0$, $\tau_j' \in [-1, 1]$, $j \in J_0'$ such that

$$\sum_{j\in J_0} \beta_j \tau_j e_j + \sum_{j\in J_+} \beta_j e_j - \sum_{j\in J_-} \beta_j e_j + \sum_{l\in L} \xi_l \mu_l = \sum_{j\in J_0'} \beta_j \tau_j' e_j + \sum_{j\in J_+'} \beta_j e_j - \sum_{j\in J_-'} \beta_j e_j + \sum_{l\in L'} \xi_l \mu_l' \tag{7.34}$$

and some $x \in D_{J_0',J_+',J_-',L'} \cap \operatorname{ri} D_{J_0,J_+,J_-,L}$ implying $J_0' \subseteq J_0$ and $L' \subseteq L$, where equality can not simultaneously hold in both inclusions. Choosing $x' \in \operatorname{ri} D_{J_0',J_+',J_-',L'}$ and setting $u = x' - x$, we obtain

$$\begin{aligned} u_j &= 0,\ j \in J_0',\ u_j > 0,\ j \in (J_0 \setminus J_0') \cap J_+',\ u_j < 0,\ j \in (J_0 \setminus J_0') \cap J_-',\ \langle \xi_l, u\rangle \\ &= 0,\ l \in L',\ \langle \xi_l, u\rangle < 0,\ l \in L \setminus L'. \end{aligned}$$

Rearranging (7.34) yields

$$\sum_{j\in(J_0\setminus J_0')\cap J_+'} \beta_j(\tau_j-1)e_j + \sum_{j\in(J_0\setminus J_0')\cap J_-'} \beta_j(\xi_j+1)e_j + \sum_{l\in L\setminus L'} \mu_l\xi_l = \sum_{j\in J_0'} \beta_j(\tau_j'-\tau_j)e_j + \sum_{l\in L'}(\mu_l'-\mu_l)\xi_l$$

and by multiplying this equation with u we obtain the contradiction

$$\begin{aligned} 0 > &\sum_{j\in(J_0\setminus J_0')\cap J_+'} \beta_j(\tau_j - 1)u_j + \sum_{j\in(J_0\setminus J_0')\cap J_-'} \beta_j(\tau_j + 1)u_j + \sum_{l\in L\setminus L'} \mu_l\langle \xi_l, u\rangle \\ &= \sum_{j\in J_0'} \beta_j(\tau_j' - \tau_j)u_j + \sum_{l\in L'}(\mu_l' - \mu_l)\langle \xi_l, u\rangle = 0. \end{aligned}$$

Hence, our claim holds true and we may conclude that for every $(J_0, J_+, J_-, L) \in \mathscr{I}$ and every $(z, z^*) \in \operatorname{ri} E_{J_0,J_+,J_-,L}$ we have

$$\begin{aligned} T_{\operatorname{gph}\partial g}(z, z^*) &= T_{E_{J_0,J_+,J_-,L}}(z, z^*) = T_{D_{J_0,J_+,J_-,L}}(z) \times T_{\tilde{D}_{J_0,J_+,J_-,L}}(z^*) \\ &= \{w \mid w_j = 0,\ j \in J_0,\ \langle \xi_l, w\rangle = 0,\ l \in L\} \times \{\sum_{j\in J_0} \beta_j\sigma_j e_j \\ &\quad + \sum_{l\in L} \xi_l\nu_l \mid \sigma_j \in \mathbb{R},\ j \in J_0,\ \nu_l \in \mathbb{R},\ l \in L\} = W(z) \times W(z)^{\perp}, \end{aligned}$$

where the last equality follows from $J_0 = J_0(z)$ and $L = L(z)$. Now consider $(x, x^*) \in \operatorname{gph}\partial g$. Then $\big(J_0(x), J_+(x), J_-(x), L(x)\big) \in \mathscr{I}$ and $x \in \operatorname{ri} D_{I_0(x),I_+(x),I_-(x),J(x)}$. Selecting $z^* \in \operatorname{ri} \tilde{D}_{J_0(x),J_+(x),J_-(x),L(x)}$, for all $\alpha \in (0, 1]$ we have $x_\alpha^* := (1-\alpha)x^* + \alpha z^* \in \operatorname{ri} \tilde{D}_{J_0(x),J_+(x),J_-(x),L(x)}$ implying $T_{\operatorname{gph}\partial g}(x, x_\alpha^*) = W(x) \times W(x)^{\perp}$. Now the assertion follows from the definition of $\mathscr{S}\partial g(x, x^*)$ together with Theorem 3.6. □

Let $\hat{d}^{*(k)} := -\gamma^{(k)}u^{(k)} - f(x^{(k)})$. By Lemma 3.8 and consecutive application of Theorem 7.1 with $g = q^i$ we obtain

$$\prod_{i=1}^{n} (W^i)^{(k)} \times \prod_{i=1}^{n} (W^i)^{(k)^{\perp}} \in \mathscr{S}^*\partial q(\hat{d}^{(k)}, \hat{d}^{*(k)}),$$

where for each $i = 1, \ldots, n$ the subspace $(W^i)^{(k)} \subset \mathbb{R}^m$ is given by

$$(W^i)^{(k)} := \{w \mid \langle \xi_l^i, w\rangle = 0, l \in (L^i)^{(k)},\ w_j = 0,\ j \in (J_0^i)^{(k)}\}$$

with $(J_0^i)^{(k)} := \{j \in \{1, \ldots, m\} \mid (\hat{d}_j^i)^{(k)} = a_j^i\}$, $(L^i)^{(k)} := \{l \in \{1, \ldots, p^i\} \mid \langle \xi_l^i, (\hat{d}^i)^{(k)}\rangle = \zeta_l^i\}$, and the vectors ξ_l^i, $l = 1, \ldots, p^i$, given by the l-th row of the matrix Ξ^i.

The required matrices $Y^{(k)} = \operatorname{diag}\big((Y^1)^{(k)}, \ldots, (Y^n)^{(k)}\big)$ and $X^{(k)} = \operatorname{diag}\big((X^1)^{(k)}, \ldots, (X^n)^{(k)}\big)$ are block diagonal matrices, where the diagonal $m \times m$ blocks can be computed as

$$(Y^i)^{(k)} = Q_2^i \times {Q_2^i}^T, \quad (X^i)^{(k)} = Q_1^i \times {Q_1^i}^T$$

Table 1 Input parameters b_j^i, δ_j^i, K_j^i of production costs and market elasticities γ_j

	$i=1$			$i=2$			$i=3$			$i=4$			$i=5$			
	b_j^i	δ_j^i	K_j^i	b_j^i	δ_j^i	K_j^i	b_j^i	δ_j^i	K_j^i	b_j^i	δ_j^i	K_j^i	b_j^i	δ_j^i	K_j^i	γ_j
j=1	9.0	1.2	5.0	7.0	1.1	5.0	3.0	1.0	5.0	4.0	0.9	5.0	2.0	0.8	5.0	1.0
j=2	9.0	1.2	5.0	7.0	1.1	5.0	3.0	1.0	5.0	4.0	0.9	5.0	2.0	0.8	5.0	0.9
j=3	9.0	1.2	5.0	7.0	1.1	5.0	3.0	1.0	5.0	4.0	0.9	5.0	2.0	0.8	5.0	0.8

and the columns of Q_2^i and Q_1^i are orthonormal bases for the subspaces $(W^i)^{(k)}$ and $(W^i)^{(k)\perp}$, respectively. The matrices Q_1^i and Q_2^i can be computed, e.g., via a QR-factorization with column pivoting for the matrix with columns $\xi_l^i/\|\xi_l^i\|$, $l \in (L^i)^{(k)}$, and e_j, $j \in (J_0^i)^{(k)}$, see, e.g., [9, Section 2.2.5.3].

Concerning the numerical tests[1], we consider first an academic example with $n = 5$ (the number of firms) and $m = 3$ (the number of comodities). The parameters b_j^i, δ_j^i, K_j^i of production cost functions together with the market elasticities γ_j arising in the inverse demand functions are displayed in Table 1. In the constraints $\Xi^i x^i \le \zeta^i$, defining the sets of feasible productions, we assume that matrices Ξ^i have only one row (i.e., $p^i = 1$). The respective data are listed in Table 2 together with the weights β_j^i specifying the costs of change and the "previous" productions a_j^i. Finally, Table 3 presents the initial iterations of productions $(x_j^i)^{(0)}$ and the calculated results, including both the equilibrium productions (x_j^i) as well as the corresponding costs of change (z_j^i).

The results shown in Table 3 have been achieved in 6 iterations of the Algorithm 2 and the final residual amounts to 2.7×10^{-12}. Note that the third firm exhausts its maximum production capacity (so it holds $\Xi_1^3 x_1^3 + \Xi_2^3 x_2^3 + \Xi_3^3 x_3^3 = \zeta^3$), whereas the other firms do not. We also observe the prohibitive influence of the high value of β_3^1, due, as expected, to $x_3^1 = a_3^1 = 47.8$.

Next, to demonstrate the computational efficiency of the SCD semismooth* Newton method, we increase substantially the values of n and m. In dependence of n and m, we generated test problems by drawing the data independently from the uniform distributions with the following parameters:

$$\left.\begin{array}{l} b_j^i \sim \mathscr{U}(2,20),\ \delta_j^i \sim \mathscr{U}(0.5,2),\ K_j^i \sim \mathscr{U}(0.1,10) \\ \Xi_{lj}^i \sim \mathscr{U}(0,1),\ \beta_j^i \sim \mathscr{U}(1,10),\ a_j^i \sim \mathscr{U}(20,50), \\ \gamma_j \sim \mathscr{U}(1,2) \end{array}\right\}, \quad i=1,\dots,n,\ j=1,\dots,m,\ l=1,\dots,p^i.$$

Here, the numbers $p^i, i = 1,\dots,n$, are obtained by rounding numbers drawn independently from $\mathscr{U}(1, 1.5m+1)$. Furthermore, we set $\zeta^i := \Xi^i z^i$, where for each $i = 1,\dots,n$ the elements z_j^i, $j = 1,\dots,m$ are drawn from $\mathscr{U}(1,15)$. For each pair (n,m) belonging to the set

[1] All codes can be found on https://www.numa.uni-linz.ac.at/~gfrerer/Software/Cournot_Nash/

Table 2 Input parameters Ξ_j^i, ζ^i defining feasible productions, parameters β_j^i of costs of change and previous productions a_j^i

	$i=1$			$i=2$			$i=3$			$i=4$			$i=5$		
	Ξ_j^i	β_j^i	a_j^i	Ξ_j^i	β_j^i	a_j^i	Ξ_j^i	β_j^i	a_j^i	Ξ_j^i	β_j^i	a_j^i	Ξ_j^i	β_j^i	a_j^i
j=1	1.0	0.5	47.8	1.0	1.0	51.1	1.0	2.0	51.3	1.0	0.0	48.5	1.0	0.0	43.5
j=2	1.0	0.5	47.8	1.0	1.0	51.1	1.0	2.0	51.3	1.0	0.0	48.5	1.0	0.0	43.5
j=3	1.0	20.0	47.8	1.0	1.0	51.1	1.0	2.0	51.3	1.0	0.0	48.5	1.0	0.0	43.5
ζ^i	200			250			100			200			200		

$\{(5, 200), (25, 40), (200, 5)\}$ we generated 50 test problems and solved them as well with the heuristic from Section 6.1 as with the globalized semismooth* Newton method of Algorithm 3 with $\mathscr{T} = \mathscr{T}_\gamma^{\mathrm{PM}}$. As a stopping criterion we used $r_{\gamma^{(k)}}(x^{(k)}) \leq 10^{-12} r_{\gamma^{(0)}}(x^{(0)})$ and as a starting point we chose the vector $(5, 5, \ldots, 5)$. Both methods have been successful in all 150 test problems. In Table 4 we report for each scenario the mean value of the iterations needed, the standard deviation and the maximum iteration number.

For each of the 3 scenarios, we have a problem with $nm = 1000$ unknowns. The time-consuming parts of the semismooth* Newton method are the approximation step and the Newton step: In the approximation step, we have to solve n quadratic problems with $2m$ variables, while in the Newton step, we must solve a linear system with nm variables. Thus, in case where $(n, m) = (5, 200)$ the approximation step is more time-consuming than the Newton step, whereas in case where $(n, m) = (200, 5)$ the approximation step is much cheaper than the Newton step. We can see that the numbers of iterations needed are fairly small. Note that the given iteration numbers essentially reflect the *global* convergence behavior: The majority of iterations are needed to get sufficiently close to the solution and then, by superlinear convergence of the semismooth* Newton method, only 3–6 iterations more are required to approximate the solution with the desired accuracy. In Fig. 1 we show the residuals $r_{\gamma^{(k)}}(x^{(k)})$ given by (6.23) for a test problem with $(n, m) = (5, 200)$ for both the Algorithm 3 and the heuristic of Section 6.1. The Algorithm 3 needed 16 iterations to reduce the initial residual of 840.7 to 6.0 and the method stopped after 6 additional iterations with a residual of 4×10^{-12}. Similarly, for the heuristic we obtained at the 15-th iterate a residual of 8.5 and the method stopped after 21 iterations with a final residual of 5.7×10^{-12}.

We now compare the semismooth* Newton method with several first-order splitting methods, namely the Forward-Backward splitting method FB, the golden ratio algorithm

Table 3 The initial productions $(x_j^i)^{(0)}$, the calculated equilibrium productions x_j^i, and the corresponding costs of change z_j^i

	$i=1$			$i=2$			$i=3$			$i=4$			$i=5$		
	$(x_j^i)^{(0)}$	x_j^i	z_j^i	$(x_j^i)^{(0)}$	x_j^i	z_j^i	$(x_j^i)^{(0)}$	x_j^i	z_j^i	$(x_j^i)^{(0)}$	x_j^i	z_j^i	$(x_j^i)^{(0)}$	x_j^i	z_j^i
j=1	45.0	54.4	3.3	45.0	54.6	3.5	45.0	20.6	61.4	45.0	50.8	0.0	45.0	45.3	0.0
j=2	45.0	67.9	10.0	45.0	66.2	15.0	45.0	30.6	41.5	45.0	58.2	0.0	45.0	50.6	0.0
j=3	45.0	47.8	0.0	45.0	85.0	33.8	45.0	48.8	5.0	45.0	70.7	0.0	45.0	60.0	0.0

Table 4 Statistics of iteration numbers for 50 test problems per scenario

	Hybrid method			Heuristic		
(n, m)	mean value	std. dev.	max. iteration #	mean value	std. dev.	max. iteration #
(5,200)	20.2	9.6	46	20.4	4.7	39
(25,40)	28.9	10.3	52	28.2	7.6	50
(200,5)	32.4	13.8	76	27.9	9.3	75

aGRAAL [19], the Douglas-Rachford splitting algorithm DR and the hybrid projection-proximal point algorithm PM [28]. We performed this comparison only for the scenario with $(n, m) = (200, 5)$, where one evaluation of the proximal mapping is relatively cheap, that is, we have to solve 200 quadratic programs with 10 variables. We generated 3 test problems and computed with the semismooth* Newton method a fairly accurate approximation $\tilde{x}$ of the exact solution: For each of the 3 test problems, the final residual was less than 2.4×10^{-12}. Using this approximate solution $\tilde{x}$, we computed for the aforementioned methods the relative error of the iterates $x^{(k)}$ defined as $\max\{\frac{|x_i^{(k)} - \tilde{x}_i|}{\max\{1, |\tilde{x}_i|\}} \mid i = 1, \ldots, nm\}$. In Fig. 2 we plot this relative error against the CPU time needed to calculate $x^{(k)}$. We set for the first-order methods as a time limit five times the time needed for the semismooth* Newton method to converge.

We can see that only for the first test problem the FB method was able to produce an approximate solution with high accuracy within the time limit. For the FB method, the final relative error was less than 10^{-5}, the other methods terminated with a relative error in the range between 4% and 7%. For the second test problem, the relative accuracy of the final iterate for the FB-method was about 8%, while we were unable to get even one significant digit with the other methods. For the third test problem, the relative error for all first-order methods was about 100%.

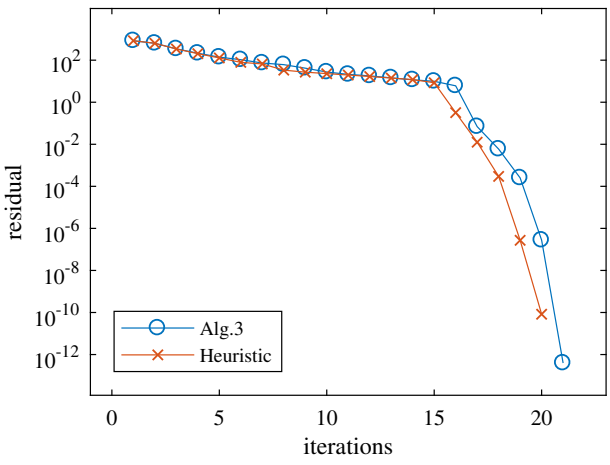

Fig. 1 Comparison of Algorithm 3 with heuristic

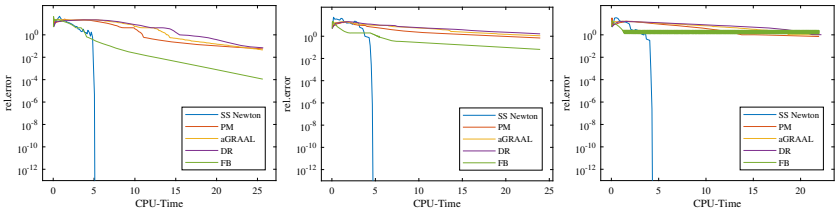

Fig. 2 Comparison of the semismooth* Newton method with several first-order methods

8 Conclusion

The semismooth* Newton method from [5] and its SCD variant from [6] provide us with a powerful tool for numerical solution of a broad class of problems governed by GEs. When faced with a concrete problem of this sort, one has to employ appropriate results of variational analysis in order to implement the approximation step and the Newton step in an efficient way. In this paper, we suggest an implementation of the SCD semismooth* Newton method for the case of variational inequalities of the second kind, which is a useful modeling framework for a number of practical problems. In particular, in this way one can model Nash games with convex, possibly nonsmooth costs, frequently arising, e.g., in economics and biology. Without substantial changes, this implementation can also be adopted for the case of the so-called hemivariational inequalities, cf. [10], which are frequently used in various models in nonsmooth mechanics. This could be a topic for future research.

Acknowledgements The authors are deeply indebted to the reviewers for careful reading and important suggestions and remarks.

Funding The research of the first author was supported by the Austrian Science Fund (FWF) under grant P29190-N32. The research of the second author was supported by the Grant Agency of the Czech Republic, Project 21-06569K, and the Australian Research Council, Project DP160100854. The research of the third author was supported by the Grant Agency of the Czech Republic, Project 21-06569K.

Data Availability Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

Declarations

Conflict of Interests The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Beck, A.: First-order Methods in Optimization MOS-SIAM Series on Optimization, vol. 25. SIAM, MOS, Philadelphia (2017)
2. Bonnans, J.F.: Local analysis of Newton-type methods for variational inequalities and nonlinear programming. Appl. Math. Optim. **29**, 161–186 (1994)
3. Facchinei, F., Pang, J.-S.: Finite-Dimensional Variational Inequalities and Complementarity Problems, Vol. I+II. Springer, New York (2003)
4. Flåm, S.D.: Games and cost of change. Ann. Oper. Res. **301**, 107—119 (2021)
5. Gfrerer, H., Outrata, J.V.: On a semismooth* Newton method for solving generalized equations. SIAM J. Optim. **31**, 489–517 (2021)
6. Gfrerer, H., Outrata, J.V.: On (local) analysis of multifunctions via subspaces contained in graphs of generalized derivatives, J. Math. Annal. Appl. https://doi.org/10.1016/j.jmaa.2021.125895 (2022)

7. Gfrerer, H., Ye, J.J.: New constraint qualifications for mathematical programs with equilibrium constraints via variational analysis. SIAM J. Optim. **27**, 842–865 (2017)
8. Glowinski, R.: Numerical Methods for Nonlinear Variational Problems. Springer, New York (1984)
9. Gill, P.E., Murray, W., Wright, M.H.: Practical Optimization. Academic Press, London (1981)
10. Haslinger, J., Miettinen, M., Panagiotopoulos, P.D.: Finite Element Method for Hemivariational Inequalities: Theory, Methods and Applications. Kluwer Academic Publishers, Boston (1999)
11. Ito, K., Kunisch, K.: On a semi-smooth Newton method and its globalization. Math. Program. **118**, 347–370 (2009)
12. Josephy, N.H.: Newton's method for generalized equations and the PIES energy model, Ph.D. Dissertation, Department of Industrial Engineering University of Wisconsin-Madison (1979)
13. Josephy, N.H.: Quasi-Newton Method for Generalized Equations. Technical Summary Report, vol. 1966. Mathematics Research Center, University of Wisconsin-Madison (1979)
14. Jourani, A.: Radiality and semismoothness. Control. Cybern. **36**, 669–680 (2007)
15. Kanzow, C., Schwartz, A.: Spieltheorie. Springer, Cham (2018)
16. Khanh, P.D., Mordukhovich, B.S., Phat, V.T.: A generalized Newton method for subgradient systems, to appear in Math. Oper. Res., arXiv:2009.10551
17. Khanh, P.D., Mordukhovich, B.S., Phat, V.T., Tran, D.B.: Generalized damped Newton algorithms in nonsmooth optimization via second-order subdifferentials, to appear in J. Global Optim., arXiv:2101.10555
18. Khanh, P.D., Mordukhovich, B.S., Phat, V.T., Tran, D.B.: Globally convergent coderivative-based generalized Newton methods in nonsmooth optimization, to appear in Math Program, arXiv:2109.02093 (2022)
19. Malitsky, Y.: Golden ratio algorithms for variational inequalities. Math. Program. **184**, 383–410 (2020)
20. Mordukhovich, B.S., Sarabi, M.E.: Generalized Newton algorithms for tilt-stable minimizers in nonsmooth optimization. SIAM J. Optim. **31**, 1184–1214 (2021)
21. Murphy, F.H., Sherali, H.D., Soyster, A.L.: A mathematical programming approach for determining oligopolistic market equilibrium. Math. Program. **24**, 2–106 (1982)
22. Outrata, J.V., Valdman, J.: On computation of optimal strategies in oligopolistic markets respecting the cost of change. Math. Meth. Oper. Res. **92**, 489–509 (2020)
23. Qi, L., Sun, J.: A nonsmooth version of Newton's method. Math. Program. **58**, 353–367 (1993)
24. Robinson, S.M.: Newton's method for a class of nonsmooth functions. Set-Valued Anal **2**, 291–305 (1994)
25. Robinson, S.M.: A point-of-attraction result for Newton's method with point-based approximations. Optimization **60**, 89–99 (2011)
26. Rockafellar, R.T.: Convex Analysis. Princeton, New Jersey (1970)
27. Rockafellar, R.T., Wets, R.J.-B.: Variational Analysis. Springer, Berlin (1998)
28. Solodov, M.V., Svaiter, B.F.: A hybrid projection–proximal point algorithm. J. Conv. Anal. **6**, 59–70 (1999)
29. Xiao, X., Li, Y., Wen, Z., Zhang, L.: A regularized semi-smooth newton method with projection steps for composite convex programs. J. Sci. Comput. **76**, 364–389 (2018)