JOHANNES KEPLER UNIVERSITY LINZ

Institute of Computational Mathematics

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

Helmut Gfrerer
Institute of Computational Mathematics, Johannes Kepler University
Altenberger Str. 69, 4040 Linz, Austria

Jiří V. Outrata
Academy of Sciences of the Czech Republic
Institute of Information Theory and Automation
18208 Prague, Czech Republic
and
Centre for Informatics and Applied Optimization
Federation University of Australia
Ballarat, Vic 3350, Australia

Jan Valdman
Academy of Sciences of the Czech Republic
Institute of Information Theory and Automation
18208 Prague, Czech Republic
and
Institute of Mathematics, Faculty of Science
University of South Bohemia
37005 Ceske Budejovice, Czech Republic

NuMa-Report No. 2020-04 July 2020

A–4040 LINZ, Altenbergerstraße 69, Austria

Technical Reports before 1998:

1995

95-1	Hedwig Brandstetter *Was ist neu in Fortran 90?*	March 1995
95-2	G. Haase, B. Heise, M. Kuhn, U. Langer *Adaptive Domain Decomposition Methods for Finite and Boundary Element Equations.*	August 1995
95-3	Joachim Schöberl *An Automatic Mesh Generator Using Geometric Rules for Two and Three Space Dimensions.*	August 1995

1996

96-1	Ferdinand Kickinger *Automatic Mesh Generation for 3D Objects.*	February 1996
96-2	Mario Goppold, Gundolf Haase, Bodo Heise und Michael Kuhn *Preprocessing in BE/FE Domain Decomposition Methods.*	February 1996
96-3	Bodo Heise *A Mixed Variational Formulation for 3D Magnetostatics and its Finite Element Discretisation.*	February 1996
96-4	Bodo Heise und Michael Jung *Robust Parallel Newton-Multilevel Methods.*	February 1996
96-5	Ferdinand Kickinger *Algebraic Multigrid for Discrete Elliptic Second Order Problems.*	February 1996
96-6	Bodo Heise *A Mixed Variational Formulation for 3D Magnetostatics and its Finite Element Discretisation.*	May 1996
96-7	Michael Kuhn *Benchmarking for Boundary Element Methods.*	June 1996

1997

97-1	Bodo Heise, Michael Kuhn and Ulrich Langer *A Mixed Variational Formulation for 3D Magnetostatics in the Space* $H(rot) \cap H(div)$	February 1997
97-2	Joachim Schöberl *Robust Multigrid Preconditioning for Parameter Dependent Problems I: The Stokes-type Case.*	June 1997
97-3	Ferdinand Kickinger, Sergei V. Nepomnyaschikh, Ralf Pfau, Joachim Schöberl *Numerical Estimates of Inequalities in* $H^{\frac{1}{2}}$.	August 1997
97-4	Joachim Schöberl *Programmbeschreibung NAOMI 2D und Algebraic Multigrid.*	September 1997

From 1998 to 2008 technical reports were published by SFB013. Please see
`http://www.sfb013.uni-linz.ac.at/index.php?id=reports`
From 2004 on reports were also published by RICAM. Please see
`http://www.ricam.oeaw.ac.at/publications/list/`
For a complete list of NuMa reports see
`http://www.numa.uni-linz.ac.at/Publications/List/`

On the application of the semismooth* Newton method to variational inequalities of the second kind

Helmut Gfrerer* Jiří V. Outrata† Jan Valdman‡

Abstract. The paper starts with a concise description of the recently developed 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 exhibiting a local superlinear convergence. Thereafter we suggest several globally convergent hybrid algorithms in which one combines the semismooth* Newton method with selected splitting algorithms for the solution of monotone variational inequalities. Their efficiency is documented by extensive numerical experiments.

Key words. Newton method, semismoothness*, superlinear convergence, global convergence, generalized equation, coderivatives.

AMS Subject classification. 65K10, 65K15, 90C33.

1 Introduction

In [8] the authors have proposed the so-called semismooth* Newton method for the numerical solution of a general inclusion

$$0 \in H(x), \tag{1.1}$$

where $H : \mathbb{R}^n \rightrightarrows \mathbb{R}^{\ltimes}$ is a closed-graph multifunction. The aim of this paper is to work out this Newton method for the numerical solution of the *generalized equation* (GE)

$$0 \in H(x) := f(x) + \partial q(x), \tag{1.2}$$

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 (1.2) is equivalent with the variational inequality (VI):

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

$$\langle f(\bar{x}), x - \bar{x} \rangle + q(x) - q(\bar{x}) \geq 0 \ \text{ for all } \ x \in \mathbb{R}^n. \tag{1.3}$$

The model (1.3) has been introduced in [9] and coined the name *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 (1.2), a number of methods can be used

*Institute of Computational Mathematics, Johannes Kepler University Linz, A-4040 Linz, Austria; helmut.gfrerer@jku.at

†Institute of Information Theory and Automation, Czech Academy of Sciences, 18208 Prague, Czech Republic, and Centre for Informatics and Applied Optimization, Federation University of Australia, POB 663, Ballarat, Vic 3350, Australia, outrata@utia.cas.cz

‡Institute of Information Theory and Automation, Czech Academy of Sciences, 18208 Prague, Czech Republic, and Institute of Mathematics, Faculty of Science, University of South Bohemia, 37005 České Budějovice, Czech Republic, jan.valdman@utia.cas.cz

ranging from nonsmooth optimization methods (applicable when ∇f is symmetric) up to a broad family of splitting methods (usable when H is monotone), cf. [3, Chapter 12]. If GE (1.2) amounts to stationarity condition for a Nash game, then also a simple coordinate-wise optimization technique can be used, cf. [12] and [18]. Concerning the Newton type methods, let us mention, for instance, the possibility to write down GE (1.2) as an equation on a monotone graph, which enables us to apply the Newton procedure from [20]. Note, however, that the subproblems to be solved 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, 23].

As mentioned above, in this paper we will investigate the numerical solution of GE (1.2) via the semismooth* Newton method developed in [8]. This method is based on an important property, which is called semismoothness* and is closely related to the semismoothness property introduced in [14] and [19]. In contrast to the Newton methods by Josephy the multi-valued part of (1.2) is also approximated and, differently to some other Newton-type methods, this approximation is provided by means of the graph of the limiting coderivative of ∂q. To facilitate the computations, we identify a certain linear structure inside the coderivative of the subdifferential mapping ∂q. In this way the computation of the Newton direction reduces to the solution of a linear system of equations. To ensure local superlinear convergence one needs merely metric regularity of the considered GE around the solution.

The plan of the paper is as follows. After the preliminary Section 2 in which we collect the needed notions from modern variational analysis, the semismooth* Newton is described and its convergence is analyzed (Section 3). Thereafter, in Section 4 we develop an implementable version of the method for the solution of GE (1.2) and show its local superlinear convergence under mild assumptions. Section 5 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, under monotonicity assumptions, one combines the semismooth* Newton method with various frequently used splitting methods. The resulting algorithms show better convergence properties than the underlying splitting methods themselves. In fact, using this semismooth* hybrid approach we can solve problems, where the pure splitting methods failed. One possible explanation of this phenomenon consists in the fact that for the convergence of the semismooth* Newton method one needs merely the metric regularity and not monotonicity. Finally, the concluding Section 6 is devoted to the presentation of numerical experiments. It contains a low-dimensional Nash equilibrium which admits both a monotone as well as a non-monotone variant. To its computation we apply the implementation developed in Section 4. Thereafter we report about a rather extensive testing of the heuristic and the hybrid methods by means of a specially constructed family of medium-scale GEs.

The following notations is employed, $\mathscr{B}_\delta(\bar{x})$ is the ball around $\bar{x}$ with radius $\delta, x \xrightarrow{A} \bar{x}$ means convergence within a set A and for a multifunction $\Phi, \operatorname{gph}\Phi := \{(x,y) | y \in \Phi(x)\}$ stands for its graph. Finally, $\|(A \vdots B)\|_F$ signifies the Frobenius norm of the matrix, composed horizontally from the blocks A, B.

2 Preliminaries

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

Definition 2.1. *Let A be a closed set in $\mathbb{R}^n$ and $\bar{x} \in A$. Then*

(i) $T_A(\bar{x}) := \operatorname*{Lim\,sup}_{t \searrow 0} \frac{A-\bar{x}}{t}$ *is the* tangent (contingent, Bouligand) cone *to A at $\bar{x}$ and* $\widehat{N}_A(\bar{x}) := (T_A(\bar{x}))^\circ$ *is the* regular (Fréchet) normal cone *to A at $\bar{x}$.*

(ii) $N_A(\bar{x}) := \operatorname*{Lim\,sup}_{x \xrightarrow{A} \bar{x}} \widehat{N}_A(x)$ *is the* limiting (Mordukhovich) normal cone *to A at $\bar{x}$ and, given a direction $d \in \mathbb{R}^n$,* $N_A(\bar{x}; d) := \operatorname*{Lim\,sup}_{\substack{t \searrow 0 \\ d' \to d}} \widehat{N}_A(\bar{x} + td')$ *is the* directional limiting normal cone *to A at* $\bar{x}$ in direction d.

In this definition "Limsup" stands for the Painlevé-Kuratowski *outer set limit*. If A is convex, then $\widehat{N}_A(\bar{x}) = N_A(\bar{x})$ amounts to the classical normal cone in the sense of convex analysis and we will write $N_A(\bar{x})$. By the definition, the limiting normal cone coincides with the directional limiting normal cone in direction 0, i.e., $N_A(\bar{x}) = N_A(\bar{x}; 0)$, and $N_A(\bar{x}; d) = \emptyset$ whenever $d \notin T_A(\bar{x})$.

The above listed cones enable us to describe the local behavior of set-valued maps via various generalized derivatives. Consider a closed-graph multifunction $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ and the point $(\bar{x}, \bar{y}) \in \operatorname{gph} F$.

Definition 2.2. *(i) The multifunction $\widehat{D}^* F(\bar{x}, \bar{y}) : \mathbb{R}^m \rightrightarrows \mathbb{R}^n$, defined by*

$$\widehat{D}^* F(\bar{x}, \bar{y})(v^*) := \{u^* \in \mathbb{R}^n | (u^*, -v^*) \in \widehat{N}_{\operatorname{gph} F}(\bar{x}, \bar{y})\}, v^* \in \mathbb{R}^m$$

is called the regular (Fréchet) coderivative *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

$$D^* F(\bar{x}, \bar{y})(v^*) := \{u^* \in \mathbb{R}^n | (u^*, -v^*) \in N_{\operatorname{gph} F}(\bar{x}, \bar{y})\}, v^* \in \mathbb{R}^m$$

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

(iii) Given a pair of directions $(u, v) \in \mathbb{R}^n \times \mathbb{R}^m$, the multifunction $D^* F((\bar{x}, \bar{y}); (u, v)) : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$, defined by

$$D^* F((\bar{x}, \bar{y}); (u, v))(v^*) := \{u^* \in \mathbb{R}^n | (u^*, -v^*) \in N_{\operatorname{gph} F}((\bar{x}, \bar{y}); (u, v))\}, v^* \in \mathbb{R}^m$$

is called the directional limiting coderivative *of F at $(\bar{x}, \bar{y})$ in direction (u, v).*

For the properties of the cones $T_A(\bar{x})$, $\widehat{N}_A(\bar{x})$ and $N_A(\bar{x})$ from Definition 2.1 and generalized derivatives (i) and (ii) from Definition 2.2 we refer the interested reader to the monographs [21] and [15]. The directional limiting normal cone and coderivative were introduced by the first author in [6] and various properties of these objects can be found also in [7] and the references therein. Note that $D^* F(\bar{x}, \bar{y}) = D^* F((\bar{x}, \bar{y}); (0, 0))$ and that $\operatorname{dom} D^* F((\bar{x}, \bar{y}); (u, v)) = \emptyset$ whenever $v \notin DF(\bar{x}, \bar{y})(u)$.

Recall that a set-valued mapping $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is said to be *metrically regular* around a point $(\bar{x}, \bar{y}) \in \operatorname{gph} F$, if the graph of F is locally closed at $(\bar{x}, \bar{y})$, and there is a constant $\kappa \geq 0$ along with neighborhoods of U of $\bar{x}$ and V of $\bar{y}$ such that

$$\operatorname{dist}(x, F^{-1}(y)) \leq \kappa \operatorname{dist}(y, F(x)) \text{ for all } (x, y) \in U \times V.$$

The infimum of κ over all such combinations of κ, U and V is called the *regularity modulus* for F at $(\bar{x}, \bar{y})$ and denoted by $\operatorname{reg} F(\bar{x}, \bar{y})$. The following statement follows from [21, Theorem 9.43]

Theorem 2.3. *A mapping* $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ *is metrically regular at* $(\bar{x},\bar{y}) \in \operatorname{gph} F$*, if and only if* $\operatorname{gph} F$ *is locally closed at* $(\bar{x},\bar{y})$ *and*

$$0 \in D^*F(\bar{x},\bar{y})(y^*) \;\Rightarrow\; y^* = 0. \tag{2.4}$$

Further, in this case one has

$$\operatorname{reg} F(\bar{x},\bar{y}) = 1/\min\{\operatorname{dist}(0, D^*F(\bar{x},\bar{y})(y^*)) \mid \|y^*\| = 1\}. \tag{2.5}$$

In the construction of the announced globally convergent hybrid algorithms we employ the notion of monotonicity.

Definition 2.4. *A mapping* $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ *is said to be* monotone *if it has the property that*

$$\langle y_2 - y_1, x_2 - x_1 \rangle \geq 0 \;\text{ for all } (x_1,y_1),(x_2,y_2) \in \operatorname{gph} F.$$

If, in addition, there is some $\mu > 0$ *such that*

$$\langle y_2 - y_1, x_2 - x_1 \rangle \geq \mu \|x_2 - x_1\|^2 \;\text{ for all } (x_1,y_1),(x_2,y_2) \in \operatorname{gph} F,$$

the mapping is called strongly monotone.

Recall that a monotone mapping $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is maximal monotone if no enlargement of its graph is possible in $\mathbb{R}^n \times \mathbb{R}^n$ without destroying monotonicity. Given a maximal monotone mapping F and a positive real λ, the mapping $(I + \lambda F)^{-1}$ is called the *resolvent* of F. It is well known that this mapping is single-valued and Lipschitz on the whole $\mathbb{R}^n$, see, e.g., [21, Theorem 12.12].

3 On the semismooth* Newton method

In this section we describe the semismooth* Newton method as introduced in [8]. Consider the inclusion

$$0 \in F(x), \tag{3.6}$$

where $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is a set-valued mapping with closed graph. The semismoothness* property of F can be defined as follows.

Definition 3.1. *A set-valued mapping* $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ *is called* semismooth* *at a point* $(\bar{x},\bar{y}) \in \operatorname{gph} F$*, if for all* $(u,v) \in \mathbb{R}^n \times \mathbb{R}^m$ *we have*

$$\langle u^*, u \rangle = \langle v^*, v \rangle \;\forall (v^*, u^*) \in \operatorname{gph} D^*F((\bar{x},\bar{y});(u,v)). \tag{3.7}$$

In some situations it is convenient to make use of equivalent characterizations in terms of standard (regular and limiting) coderivatives, respectively.

Proposition 3.2 ([8, Corollary 3.3])**.** *Let* $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ *and* $(\bar{x},\bar{y}) \in \operatorname{gph} F$ *be given. Then the following three statements are equivalent.*

(i) F *is semismooth** *at* $(\bar{x},\bar{y})$.

(ii) For every $\varepsilon > 0$ *there is some* $\delta > 0$ *such that*

$$|\langle x^*, x - \bar{x} \rangle - \langle y^*, y - \bar{y} \rangle| \leq \varepsilon \|(x,y) - (\bar{x},\bar{y})\| \|(x^*,y^*)\| \\ \forall (x,y) \in \mathscr{B}_\delta(\bar{x},\bar{y}) \;\forall (y^*,x^*) \in \operatorname{gph} \widehat{D}^*F(x,y). \tag{3.8}$$

(iii) For every $\varepsilon > 0$ there is some $\delta > 0$ such that

$$|\langle x^*, x - \bar{x}\rangle - \langle y^*, y - \bar{y}\rangle| \le \varepsilon \|(x,y) - (\bar{x},\bar{y})\| \|(x^*,y^*)\| \quad \forall (x,y) \in \mathscr{B}_\delta(\bar{x},\bar{y})\ \forall (y^*,x^*) \in \operatorname{gph} D^*F(x,y). \tag{3.9}$$

The idea behind the semismooth* Newton method for solving (3.6) is as follows. If F is semismooth* at $(\bar{x},0)$ and we are given some point $(x,y) \in \operatorname{gph} F$ close to $(\bar{x},0)$, then for every $(y^*,x^*) \in \operatorname{gph} D^*F(x,y)$ there holds

$$\langle x^*, x - \bar{x}\rangle = \langle y^*, y - 0\rangle + o(\|(x,y) - (\bar{x},\bar{y})\| \|(x^*,y^*)\|)$$

by the definition of the semismoothness* property. We choose now n pairs $(v_i^*, u_i^*) \in \operatorname{gph} D^*F(x,y)$, $i = 1,\dots,n$, compute a solution Δx of the system

$$\langle x_i^*, \Delta x\rangle = -\langle y_i^*, y\rangle,\ i = 1,\dots,n \tag{3.10}$$

and expect that $\|(x + \Delta x) - \bar{x}\| = o(\|x - \bar{x}\|$.

In order to work out this basic idea, we introduce the following notation. Given $(x,y) \in \operatorname{gph} F$, we denote by $\mathscr{A}F(x,y)$ the collection of all pairs of $n \times n$ matrices (A,B), such that there are n elements $(y_i^*, x_i^*) \in \operatorname{gph} D^*F(x,y)$, $i = 1,\dots,n$, and the i-th row of A and B are ${x_i^*}^T$ and ${y_i^*}^T$, respectively. Thus, the system (3.10) is of the form

$$A\Delta x = -By$$

with $(A,B) \in \mathscr{A}F(x,y)$. This system should have a unique solution and this leads us to the definition

$$\mathscr{A}_{\text{reg}}F(x,y) := \{(A,B) \in \mathscr{A}F(x,y) \mid A \text{ non-singular}\}.$$

The set $\mathscr{A}_{\text{reg}}F(x,y)$ is nonempty if, e.g., the mapping F is strongly metrically regular around (x,y), cf. [8, Theorem 4.1]. However, strong metric regularity is only a sufficient condition, for the problem (1.2) we will show below that the weaker assumption of metric regularity is also sufficient.

For the local convergence analysis of the semismooth* Newton method, the following result plays a central role.

Proposition 3.3 ([8, Proposition 4.3]). *Assume that the mapping $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is semismooth* at $(\bar{x},0) \in \operatorname{gph} F$. Then for every $\varepsilon > 0$ there is some $\delta > 0$ such that for every $(x,y) \in \operatorname{gph} F \cap \mathscr{B}_\delta(\bar{x},0)$ and every pair $(A,B) \in \mathscr{A}_{\text{reg}}F(x,y)$ one has*

$$\|(x - A^{-1}By) - \bar{x}\| \le \varepsilon \|A^{-1}\| \|(A \vdots B)\|_F \|(x,y) - (\bar{x},0)\|. \tag{3.11}$$

As a byproduct of this statement we obtain the following corollary which is not directly related with the semismooth* Newton method.

Corollary 3.4. *Assume that the mapping $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is semismooth* at $(\bar{x},0) \in \operatorname{gph} F$ and assume that there are positive reals $\bar{\delta}$ and $\bar{\kappa}$ such that for every $(x,y) \in \operatorname{gph} F \cap \mathscr{B}_{\bar{\delta}}(\bar{x},0)$ there are matrices $(A,B) \in \mathscr{A}_{\text{reg}}F(x,y)$ such that*

$$\|A^{-1}\| \|(A \vdots B)\|_F \le \kappa.$$

Then $\bar{x}$ is an isolated solution of the inclusion $0 \in F(x)$.

Proof. By contraposition. Assume that $\bar{x}$ is not an isolated solution and find $0 < \delta < \bar{\delta}$ such that (3.11) holds with $\varepsilon = 1/(2\kappa)$. Since $\bar{x}$ is not an isolated solution, there exists another solution $\tilde{x} \neq \bar{x}$ in $\mathscr{B}_\delta(\bar{x})$ and by picking suitable matrices $(A,B) \in \mathscr{A}_{\text{reg}}F(\tilde{x},0)$ we obtain the contradiction

$$\|\tilde{x}-\bar{x}\| = \|(\tilde{x}-A^{-1}B0)-\bar{x}\| \leq \frac{1}{2\kappa}\|A^{-1}\|\|(A\,\vdots\,B)\|_F\|(\tilde{x},0)-(\bar{x},0)\| \leq \frac{1}{2}\|\tilde{x}-\bar{x}\|.$$

□

We are now in the position to describe the iteration step of the semismooth* Newton method. Assume we are given some iterate $x^{(k)}$. 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 $\bar{x}$. Thus we perform first some step which yields $(\hat{x}^{(k)},\hat{y}^{(k)}) \in \operatorname{gph} F$ as an approximate projection of $(x^{(k)},0)$ on $\operatorname{gph} F$. Further we require that $\mathscr{A}_{\text{reg}}F(\hat{x}^{(k)},\hat{y}^{(k)}) \neq \emptyset$ and compute the new iterate as $x^{(k+1)} = \hat{x}^{(k)} - A^{-1}B\hat{y}^{(k)}$ for some $(A,B) \in \mathscr{A}_{\text{reg}}F(\hat{x}^{(k)},\hat{y}^{(k)})$. This leads to the following conceptual algorithm.

Algorithm 1 (semismooth* Newton-type method for generalized equations)**.**
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*

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

close to $(x^{(k)},0)$ such that $\mathscr{A}_{\text{reg}}F(\hat{x}^{(k)},\hat{y}^{(k)}) \neq \emptyset$.
4. **Newton step:** *Select*

$$(A,B) \in \mathscr{A}_{\text{reg}}F(\hat{x}^{(k)},\hat{y}^{(k)})$$

and compute the new iterate

$$x^{(k+1)} = \hat{x}^{(k)} - A^{-1}B\hat{y}^{(k)}.$$

5. Set $k := k+1$ and go to 2.

Now let us consider convergence properties of Algorithm 1. Given two reals $L,\kappa > 0$ and a solution $\bar{x}$ of (3.6), we denote

$$\mathscr{G}^{L,\kappa}_{F,\bar{x}}(x) := \{(\hat{x},\hat{y},A,B) \mid \|(\hat{x}-\bar{x},\hat{y})\| \leq L\|x-\bar{x}\|,\ (A,B) \in \mathscr{A}_{\text{reg}}F(\hat{x},\hat{y}), \|A^{-1}\|\|(A\,\vdots\,B)\|_F \leq \kappa\}.$$

Theorem 3.5 ([8, Theorem 4.4])**.** *Assume that F is semismooth* at $(\bar{x},0) \in \operatorname{gph} F$ and assume that there are $L,\kappa > 0$ such that for every $x \notin F^{-1}(0)$ sufficiently close to $\bar{x}$ we have $\mathscr{G}^{L,\kappa}_{F,\bar{x}}(x) \neq \emptyset$. Then there exists some $\delta > 0$ such that for every starting point $x^{(0)} \in \mathscr{B}_\delta(\bar{x})$ Algorithm 1 either stops after finitely many iterations at a solution or produces a sequence $x^{(k)}$ which converges superlinearly to $\bar{x}$, provided we choose in every iteration $(\hat{x}^{(k)},\hat{y}^{(k)},A,B) \in \mathscr{G}^{L,\kappa}_{F,\bar{x}}(x^{(k)})$.*

According to Theorem 3.5, the outcome $(\hat{x}^{(k)},\hat{y}^{(k)}) \in \operatorname{gph} F$ from the approximation step has to fulfill the inequality

$$\|(\hat{x}^{(k)},\hat{y}^{(k)}) - (\bar{x},0)\| \leq L\|x^{(k)}-\bar{x}\|. \tag{3.12}$$

It is easy to show that this estimate holds true if

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

i.e., $(\hat{x}^{(k)},\hat{y}^{(k)})$ is some approximate projection of $(x^{(k)},0)$ on $\mathrm{gph}\,F$. In fact, it suffices when the deviation of $(\hat{x}^{(k)},\hat{y}^{(k)})$ from the exact projection is proportional to the distance $\mathrm{dist}((x^{(k)},0),\mathrm{gph}\,F)$. So the approximation of the projection can be rather crude.

In the computation of matrices A, B needed in the Newton step we will make use of the following result which is interesting also for its own sake.

Theorem 3.6. *Let $q:\mathbb{R}^n\to\bar{\mathbb{R}}$ be proper convex and lsc. Then for every $(x,x^*)\in\mathrm{gph}\,\partial q$ there is a positive semidefinite matrix G with $\|G\|\le 1$ such that*

$$\mathrm{rge}\,(I-G,G):=\{\big((I-G)v^*,Gv^*\big)\mid v^*\in\mathbb{R}^n\}\subset\mathrm{gph}\,D^*(\partial q)(x,x^*). \tag{3.13}$$

Proof. Consider the Moreau envelope function

$$e_1q(y):=\inf_x\big(q(x)+\frac{1}{2}\|x-y\|^2\big). \tag{3.14}$$

By [21, Exercise 12.23], e_1q is continuously differentiable on $\mathbb{R}^n$ and ∇e_1q is a maximal monotone, single valued mapping, which is Lipschitz continuous with constant 1, and

$$\nabla e_1q=(I+(\partial q)^{-1})^{-1}.$$

Thus

$$y^*=\nabla e_1q(y)\ \Leftrightarrow\ y\in(I+(\partial q)^{-1})(y^*)\ \Leftrightarrow\ y^*\in\partial q(y-y^*)$$

and, consequently, $(z,z^*)\in\mathrm{gph}\,\partial q\Leftrightarrow z^*=\nabla e_1q(z+z^*)$. Next, consider an element G from the B-subdifferential $\overline{\nabla}(\nabla e_1q)(x+x^*)$ together with sequences $y_k\to x+x^*$ and $G_k\to G$ with $G_k=\nabla(\nabla e_1q)(y_k)$. From the monotonicity of ∇e_1q it follows that G_k is positive semidefinite and hence so is G as well. Further, by [21, Theorem 13.52], G is symmetric and from the Lipschitz continuity of ∇e_1q we deduce that $\|G\|\le 1$. Since G belongs to the B-subdifferential of ∇e_1q at $x+x^*$, we have

$$(v^*,G^Tv^*)=(v^*,Gv^*)\in D^*(\nabla e_1q)(x+x^*,x^*)=D^*\Big((I+(\partial q)^{-1})^{-1}\Big)(x+x^*,x^*)\ \forall v^*\in\mathbb{R}^n.$$

Taking into account some elementary calculus rules for coderivatives, we conclude that

$$\begin{aligned}(v^*,Gv^*)\in D^*\Big((I+(\partial q)^{-1})^{-1}\Big)(x+x^*,x^*)&\Leftrightarrow(-Gv^*,-v^*)\in D^*(I+(\partial q)^{-1})(x^*,x+x^*)\\ \Leftrightarrow(-Gv^*,-v^*+Gv^*)\in D^*(\partial q)^{-1}(x^*,x)&\Leftrightarrow(v^*-Gv^*,Gv^*)\in D^*(\partial q)(x,x^*),\end{aligned}$$

and the assertion of the theorem follows. □

4 Implementation of the semismooth* Newton method

There are a lot of possibilities how to implement the semismooth* Newton method. Apart from the Newton step, which is not uniquely determined by different selections of the coderivatives, 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 (1.2). We restrict ourselves to the case where the approximation step is performed by means of the mapping u_γ defined as

$$u_\gamma(x):=\arg\min_u\big(\frac{1}{2}\gamma\|u\|^2+\langle f(x),u\rangle+q(x+u)\big), \tag{4.15}$$

where $\gamma > 0$ is some scaling parameter. Note that u_γ is clearly single-valued due to the strong convexity of the objective. The first-order (necessary and sufficient) optimality condition reads as

$$0 \in \gamma u_\gamma(x) + f(x) + \partial q(x + u_\gamma(x)), \tag{4.16}$$

which can be equivalently written as

$$\gamma x - f(x) \in (\gamma I + \partial q)(x + u_\gamma(x)).$$

Let us premultiply this inclusion by $\lambda := 1/\gamma$. One obtains that

$$x - \lambda f(x) \in (I + \lambda \partial q)(x + u_\gamma(x)),$$

which yields the equality

$$x + u_\gamma(x) = (I + \lambda \partial q)^{-1}(x - \lambda f(x)), \tag{4.17}$$

because the resolvent $(I + \lambda \partial q)^{-1}$ is single-valued due the maximal monotonicity of ∂q. Since this resolvent is also nonexpansive, cf. [21, Theorem 12.12], for arbitrary two points $x, x' \in \mathbb{R}^n$ we obtain the bounds

$$\|(x + u_\gamma(x)) - (x' + u_\gamma(x')\| \le \|(x - x') - \lambda(f(x) - f(x'))\| \le \|x - x'\| + \frac{1}{\gamma}\|f(x) - f(x')\| \tag{4.18}$$

$$\|u_\gamma(x) - u_\gamma(x')\| \le 2\|x - x'\| + \frac{1}{\gamma}\|f(x) - f(x')\|. \tag{4.19}$$

They will be used in the estimates below.

Remark 4.1. *Equation* (4.17) *tells us, that $x + u_\gamma(x)$ is the outcome of one step of the so-called* forward-backward splitting method*, see, e.g., [13].*

Our approach is based on an equivalent reformulation of (1.2) in form of the GE

$$0 \in \mathscr{F}(x, d) := \begin{pmatrix} f(x) + \partial q(d) \\ x - d \end{pmatrix} \tag{4.20}$$

in variables $(x, d) \in \mathbb{R}^n \times \mathbb{R}^n$. Clearly, $\bar{x}$ is a solution of (1.2) if and only if $(\bar{x}, \bar{x})$ is a solution of (4.20). Further, it is easy to see that $\mathscr{F}$ is semismooth* at $((\bar{x}, \bar{x}), (0, 0))$ if and only if ∂q is semismooth* at $(\bar{x}, -f(\bar{x}))$.

We start 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

$$\hat{x}^{(k)} = x^{(k)},\ \hat{d}^{(k)} = x^{(k)} + u^{(k)} \quad \text{and} \quad \hat{y}^{(k)} = (\hat{y}_1^{(k)}, \hat{y}_2^{(k)}) = (-\gamma^{(k)} u^{(k)}, u^{(k)}). \tag{4.21}$$

We observe that

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

which follows immediately from the first-order optimality condition (4.16). Note that the outcome of the approximation step does not depend on the auxiliary variable $d^{(k)}$. In order to apply Theorem 3.5, we shall show the existence of a real $L > 0$ such that the estimate

$$\|((\hat{x}^{(k)} - \bar{x}, \hat{d}^{(k)} - \bar{x}), \hat{y}^{(k)}\| \le L\|(x^{(k)} - \bar{x}, d^{(k)} - \bar{x})\|, \tag{4.22}$$

corresponding to (3.12), holds for all $(x^{(k)},d^{(k)})$ with $x^{(k)}$ close to $\bar{x}$. We observe that the left-hand side of (4.22) amounts to

$$\begin{aligned}\|((\hat{x}^{(k)}-\bar{x},\hat{x}^{(k)}+u^{(k)}-\bar{x}),(-\gamma^{(k)}u^{(k)},u^{(k)})\| &\le \|(\hat{x}^{(k)}-\bar{x},\hat{x}^{(k)}-\bar{x},0,0)\|+\|(0,u^{(k)},-\gamma^{(k)}u^{(k)},u^{(k)})\|\\ &\le 2\|\hat{x}^{(k)}-\bar{x}\|+(2+\gamma^{(k)})\|u^{(k)}\|. \qquad (4.23)\end{aligned}$$

Since $u_{\gamma^{(k)}}(\bar{x})=0$, we obtain from (4.19) the bounds

$$\|\hat{d}^{(k)}-\bar{x}\| \le \|x^{(k)}-\bar{x}\|+\frac{1}{\gamma^{(k)}}\|f(x^{(k)})-f(\bar{x})\| \qquad (4.24)$$

$$\|u^{(k)}\| \le 2\|x^{(k)}-\bar{x}\|+\frac{1}{\gamma^{(k)}}\|f(x^{(k)})-f(\bar{x})\|. \qquad (4.25)$$

The latter estimate, together with (4.23), imply

$$\begin{aligned}\|((\hat{x}^{(k)}-\bar{x},\hat{d}^{(k)}-\bar{x}),\hat{y}^{(k)}\| &\le \Big(2+(2+\gamma^{(k)})\big(2+\frac{l}{\gamma^{(k)}}\big)\Big)\|x^{(k)}-\bar{x}\|\\ &\le \Big(2+(2+\gamma^{(k)})\big(2+\frac{l}{\gamma^{(k)}}\big)\Big)\|(x^{(k)}-\bar{x},d^{(k)}-\bar{x})\|, \qquad (4.26)\end{aligned}$$

where l is the Lipschitz constant of f on a neighborhood of $\bar{x}$. Thus the desired inequality (4.22) holds, as long as $\gamma^{(k)}$ remains bounded and bounded away from 0.

Let us now consider the Newton step. By calculus of coderivatives we have for any $s,s^*\in\mathbb{R}^n$ the equality

$$D^*\mathscr{F}((\hat{x}^{(k)},\hat{d}^{(k)}),\hat{y}^{(k)})\begin{pmatrix} s\\ s^*\end{pmatrix}=\begin{pmatrix}\nabla f(\hat{x}^{(k)})^T s+s^*\\ D^*(\partial q)(\hat{d}^{(k)},\hat{d}^{*(k)})(s)-s^*\end{pmatrix},$$

where $\hat{d}^{*(k)}:=\hat{y}_1^{(k)}-f(\hat{x}^{(k)})\in\partial q(\hat{d}^{(k)})$. Assume that, according to Theorem 3.6, we have a symmetric, positive definite matrix $G^{(k)}$ satisfying $\|G^{(k)}\|\le 1$ and

$$\text{rge }(I-G^{(k)},G^{(k)})\subseteq \text{gph}\,D^*(\partial q)(\hat{d}^k,\hat{d}^{*(k)}) \qquad (4.27)$$

at our disposal. We now choose

$$v_i^*:=\begin{pmatrix}(I-G^{(k)})e_i\\ 0\end{pmatrix},\ u_i^*=\begin{pmatrix}\nabla f(\hat{x}^{(k)})^T(I-G^{(k)})e_i\\ G^{(k)}e_i\end{pmatrix},\ i=1,\dots,n,$$

$$v_i^*:=\begin{pmatrix}0\\ e_{i-n}\end{pmatrix},\ u_i^*=\begin{pmatrix}e_{i-n}\\ -e_{i-n}\end{pmatrix},\ i=n+1,\dots,2n,$$

so that $(A,B)\in\mathscr{A}\mathscr{F}((\hat{x}^{(k)},\hat{d}^{(k)}),\hat{y}^{(k)})$, where

$$A=\begin{pmatrix}(I-G^{(k)})\nabla f(\hat{x}^{(k)}) & G^{(k)}\\ I & -I\end{pmatrix},\ B=\begin{pmatrix}I-G^{(k)} & 0\\ 0 & I\end{pmatrix}. \qquad (4.28)$$

Elementary calculations show that

$$A^{-1}=\begin{pmatrix}C^{-1} & G^{(k)}C^{-1}\\ C^{-1} & -(I-G^{(k)})\nabla f(\hat{x}^{(k)})C^{-1}\end{pmatrix},$$

provided the matrix $C:=(I-G^{(k)})\nabla f(\hat{x}^{(k)})+G^{(k)}$ is nonsingular. In this case, since $G^{(k)}$ is positive semidefinite and $\|G^{(k)}\|\le 1$, the matrices A,B given by (4.28) fulfill a bound of the form

$$\|A^{-1}\|\|(A\vdots B)\|_F\le \|\big((I-G^{(k)})\nabla f(\hat{x}^{(k)})+G^{(k)}\big)^{-1}\|\big(C_1+C_2\|\nabla f(\hat{x}^{(k)})\|\big)^2 \qquad (4.29)$$

with constants $C_1,C_2>0$.

Proposition 4.2. *Assume that*

$$0 \in \nabla f(\hat{x}^{(k)})^T s + D^*(\partial q)(\hat{d}^k, \hat{d}^{*(k)})(s) \ \Rightarrow s = 0 \tag{4.30}$$

Then $(I - G^{(k)})\nabla f(\hat{x}^{(k)}) + G^{(k)}$ *is non-singular and*

$$\|((I - G^{(k)})\nabla f(\hat{x}^{(k)}) + G^{(k)})^{-1}\| \leq 1 + \frac{1}{\mu}\|\nabla f(\hat{x}^{(k)}) - I\|,$$

where

$$\mu = \min_{\|s\|=1} \operatorname{dist}(0, \nabla f(\hat{x}^{(k)})^T s + D^*(\partial q)(\hat{d}^k, \hat{d}^{*(k)})(s)). \tag{4.31}$$

Proof. By the definition of μ we have for every $u \in \mathbb{R}^n$ the estimate

$$\begin{aligned}\|\nabla f(\hat{x}^{(k)})^T(I - G^{(k)})u + G^{(k)}u\| &\geq \operatorname{dist}(0, \nabla f(\hat{x}^{(k)})^T(I - G^{(k)})u + D^*(\partial q)(\hat{d}^k, \hat{d}^{*(k)})((I - G^{(k)})u)) \\ &\geq \mu\|(I - G^{(k)})u\|\end{aligned}$$

implying

$$\|\nabla f(\hat{x}^{(k)})^T(I - G^{(k)})u + G^{(k)}u\| \geq \frac{\mu}{\mu + \|\nabla f(\hat{x}^{(k)}) - I\|}\|u\|$$

whenever $\|u\| \leq \|(I - G^{(k)})u\|(\mu + \|\nabla f(\hat{x}^{(k)}) - I\|)$. On the other hand, if $\|u\| > \|(I - G^{(k)})u\|(\mu + \|\nabla f(\hat{x}^{(k)}) - I\|)$, then

$$\begin{aligned}\|\nabla f(\hat{x}^{(k)})^T(I - G^{(k)})u + G^{(k)}u\| &= \|u + (\nabla f(\hat{x}^{(k)})^T - I)(I - G^{(k)})u\| \\ &\geq \|u\| - \|\nabla f(\hat{x}^{(k)})^T - I\|\|(I - G^{(k)})u\| > \|u\| - \frac{\|\nabla f(\hat{x}^{(k)})^T - I\|}{\mu + \|\nabla f(\hat{x}^{(k)}) - I\|}\|u\| \\ &= \frac{\mu}{\mu + \|\nabla f(\hat{x}^{(k)}) - I\|}\|u\|,\end{aligned}$$

where we have taken into account that $\|\nabla f(\hat{x}^{(k)})^T - I\| = \|\nabla f(\hat{x}^{(k)}) - I\|$. Hence,

$$\|(\nabla f(\hat{x}^{(k)})^T(I - G^{(k)}) + G^{(k)})u\| \geq \frac{\mu}{\mu + \|\nabla f(\hat{x}^{(k)}) - I\|}\|u\| \quad \forall u$$

and

$$\|((I - G^{(k)})\nabla f(\hat{x}^{(k)}) + G^{(k)})^{-1}\| = \|(\nabla f(\hat{x}^{(k)})^T(I - G^{(k)}) + G^{(k)})^{-1}\| \leq \frac{\mu + \|\nabla f(\hat{x}^{(k)}) - I)\|}{\mu}$$

follows. □

In order to actually perform the Newton step, we denote by $(\Delta x^{(k)}, \Delta d^{(k)})$ the solution of the linear system

$$A\begin{pmatrix}\Delta x \\ \Delta d\end{pmatrix} = \begin{pmatrix}(I - G^{(k)})\nabla f(\hat{x}^{(k)}) & G^{(k)} \\ I & -I\end{pmatrix}\begin{pmatrix}\Delta x \\ \Delta d\end{pmatrix} = -B\hat{y}^{(k)} = \begin{pmatrix}\gamma^{(k)}(I - G^{(k)})u^{(k)} \\ u^{(k)}\end{pmatrix}$$

and set $x^{(k+1)} := \hat{x}^{(k)} + \Delta x^{(k)}$. Note that the variable Δd can be easily eliminated from the system above yielding

$$\begin{aligned}&((I - G^{(k)})\nabla f(\hat{x}^{(k)}) + G^{(k)})\Delta x^{(k)} = (\gamma^{(k)}(I - G^{(k)}) + G^{(k)})u^{(k)}, \\ &x^{(k+1)} = d^{(k+1)} = x^{(k)} + \Delta x^{(k)}.\end{aligned} \tag{4.32}$$

We now present sufficient conditions for the fulfilment of condition (4.30). Note that by Theorem 2.3, condition (4.30) is fulfilled if and only if the mapping $H_{u^{(k)}} : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$, defined by $H_{u^{(k)}}(x) := f(x) + \partial q(x + u^{(k)})$, is metrically regular around $(x^{(k)}, \hat{y}_1^{(k)})$.

Lemma 4.3. *Assume that H is metrically regular around $(\bar{x}, 0) \in \operatorname{gph} H$ and let two positive real numbers $\underline{\gamma} \leq \bar{\gamma}$ be given. Then for every $\kappa' > \operatorname{reg} H(\bar{x}, 0)$ there exists some positive radius ρ' such that for every $x^{(k)} \in \mathscr{B}_{\rho'}(\bar{x})$ and every $\gamma^{(k)} \in [\underline{\gamma}, \bar{\gamma}]$ one has*

$$\operatorname{dist}(0, \nabla f(\hat{x}^{(k)})^T s + D^*(\partial q)(\hat{d}^{(k)}, \hat{d}^{*(k)})(s)) \geq \Big(\frac{1}{\kappa'} - \|\nabla f(\hat{x}^{(k)}) - \nabla f(\hat{d}^{(k)})\|\Big)\|s\| \geq \frac{1}{2\kappa'}\|s\| \ \forall s$$

and consequently

$$\|((I - G^{(k)})\nabla f(\hat{x}^{(k)}) + G^{(k)})^{-1}\| \leq 1 + 2\kappa'\|\nabla f(\hat{x}^{(k)}) - I\|$$

and

$$\|\Delta x^{(k)}\| \leq \Big(1 + 2\kappa'\|\nabla f(\hat{x}^{(k)}) - I\|\Big) \max\{1, \gamma^{(k)}\}\|u^{(k)}\|.$$

Proof. Let $\kappa' > \operatorname{reg} H(\bar{x}, 0)$ be arbitrarily fixed. We can find a positive radius $\rho > 0$ such that

$$\operatorname{dist}(x, H^{-1}(y)) \leq \kappa' \operatorname{dist}(y, H(x)) \ \forall (x, y) \in \mathscr{B}_\rho(\bar{x}) \times \mathscr{B}_\rho(0).$$

Hence, by Theorem 2.3, for every $(x, y) \in \operatorname{gph} H \cap (\operatorname{int} \mathscr{B}_\rho(\bar{x}) \times \operatorname{int} \mathscr{B}_\rho(0))$ we have

$$\operatorname{dist}(0, \nabla f(x)^T s + D^*(\partial q)(x, y - f(x))(s)) \geq \frac{1}{\kappa'}\|s\| \ \forall s.$$

We can choose ρ small enough such that

$$\|\nabla f(x) - \nabla f(d)\| \leq \frac{1}{2\kappa'} \ \forall x, d \in \mathscr{B}_\rho(\bar{x}).$$

Let l denote the Lipschitz constant of f on $\mathscr{B}_\rho(\bar{x})$ and choose $\rho' > 0$ such that

$$(\gamma + l)(2 + \frac{l}{\gamma})\rho' < \rho, \quad (1 + \frac{l}{\gamma})\rho' < \rho \quad \forall \gamma \in [\underline{\gamma}, \bar{\gamma}].$$

Consider $(x^{(k)}, d^{(k)}) \in \mathscr{B}_{\rho'}(\bar{x}) \times \mathbb{R}^n$ and $\gamma^{(k)} \in [\underline{\gamma}, \bar{\gamma}]$. By (4.24) we have $\|\hat{d}^{(k)} - \bar{x}\| \leq (1 + \frac{l}{\gamma^{(k)}})\rho' < \rho$. Further, $f(\hat{d}^{(k)}) + \hat{d}^{*(k)} = f(\hat{d}^{(k)}) - \gamma^{(k)} u^{(k)} - f(\hat{x}^{(k)}) \in H(\hat{d}^{(k)})$ and

$$\|f(\hat{d}^{(k)}) + \hat{d}^{*(k)}\| \leq (l + \gamma^{(k)})\|u^{(k)}\| \leq (l + \gamma^{(k)})(2 + \frac{l}{\gamma^{(k)}})\rho' < \rho,$$

where we have used (4.25). Thus

$$\operatorname{dist}(0, \nabla f(\hat{d}^{(k)})^T s + D^*(\partial q)(\hat{d}^{(k)}, \hat{d}^{*(k)})(s)) \geq \frac{1}{\kappa'}\|s\| \quad \forall s$$

implying

$$\operatorname{dist}(0, \nabla f(\hat{x}^{(k)})^T s + D^*(\partial q)(\hat{d}^{(k)}, \hat{d}^{*(k)})(s)) \geq \Big(\frac{1}{\kappa'} - \|\nabla f(\hat{x}^{(k)}) - \nabla f(\hat{d}^{(k)})\|\Big)\|s\| \geq \frac{1}{2\kappa'}\|s\| \quad \forall s.$$

Hence we can apply Proposition 4.2 to obtain

$$\|((I - G^{(k)})\nabla f(\hat{x}^{(k)}) + G^{(k)})^{-1}\| \leq 1 + 2\kappa'\|\nabla f(\hat{x}^{(k)}) - I\|.$$

The estimate for $\|\Delta x^{(k)}\|$ follows from (4.32) by taking into account that $\|(\gamma^{(k)}(I - G^{(k)}) + G^{(k)})u^{(k)}\| \leq \max\{1, \gamma^{(k)}\}\|u^{(k)}\|$ because $G^{(k)}$ is symmetric and positive semidefinite with $\|G^{(k)}\| \leq 1$. $\square$

Under an additional condition we can give an estimate for the constant μ defined by (4.31) and for the length of the Newton direction $\|\Delta x^{(k)}\|$.

Lemma 4.4. *Assume that f is monotone. Given $x \in \mathbb{R}^n$ and $d \in \operatorname{dom}\partial q$, consider the numbers*

$$\mu_f(x) := \min\{\langle \nabla f(x)u, u\rangle \mid \|u\| = 1\}, \tag{4.33}$$

$$\mu_q(d) := \liminf_{\rho\downarrow 0}\{\frac{\langle d_1^* - d_2^*, d_1 - d_2\rangle}{\|d_1 - d_2\|^2} \mid d_i \in \mathscr{B}_\rho(d), (d_i, d_i^*) \in \operatorname{gph}\partial q,\ i = 1,2, d_1 \neq d_2\}. \tag{4.34}$$

Then for every $d^* \in \partial q(d)$ we have

$$\min_{\|s\|=1} \operatorname{dist}(0, \nabla f(x)^T s + D^*(\partial q)(d, d^*)(s)) \geq \mu_f(x) + \mu_q(d).$$

Proof. If $\mu_f(x) + \mu_q(d) = 0$ the assertion trivially holds true. Hence we may assume $\mu_f(x) + \mu_q(d) > 0$. Consider $0 < \varepsilon < \mu_f + \mu_q$ and pick some $\rho > 0$ such that

$$\inf_{x' \in \mathscr{B}_\rho(x)} \min\{\langle \nabla f(x')u, u\rangle \mid \|u\| = 1\} \geq \mu_f(x) - \frac{\varepsilon}{2},$$

$$\inf\{\frac{\langle d_1^* - d_2^*, d_1 - d_2\rangle}{\|d_1 - d_2\|^2} \mid d_i \in \mathscr{B}_\rho(d), (d_i, d_i^*) \in \operatorname{gph}\partial q,\ i = 1,2,\ d_1 \neq d_2\} \geq \mu_q(d) - \frac{\varepsilon}{2}.$$

Utilizing [21, Exercise 12.45] we see that the mapping $H^\rho_{d-x} : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$, defined by $H^\rho_{d-x}(x') := f(x') + \partial q(x' + (d - x)) + N_{\mathscr{B}_\rho(x)}(x')$ is maximally monotone. Further, by construction the mapping H^ρ_{d-x} is strongly monotone with constant $\mu_f(x) + \mu_g(d) - \varepsilon$ and therefore its inverse is single valued and Lipschitzian on $\mathbb{R}^n$ with constant $1/(\mu_f(x) + \mu_g(d) - \varepsilon)$. But this implies that H^ρ_{d-x} is (strongly) metrically regular around every point (x', x'^*) of its graph and by [21, Theorem 9.43] we obtain

$$\min_{\|s\|=1} \operatorname{dist}(0, D^* H^\rho_{d-x}(x', x'^*)(s)) \geq \mu_f(x) + \mu_g(d) - \varepsilon.$$

Taking into account that $N_{\mathscr{B}_\rho(x)}(x) = \{0\}$, for every $d^* \in \partial q(d)$ we have $D^* H^\rho_{d-x}(x, f(x) + d^*)(s) = \nabla f(x)^T s + D^*(\partial q)(d, d^*)(s)$ and since we can choose $\varepsilon > 0$ arbitrarily small, the assertion follows. □

Corollary 4.5. *Assume that f is monotone and assume that either f or ∂q is strongly monotone. Then for every iterate $x^{(k)} \in \mathbb{R}^n$ and every scaling parameter $\gamma^{(k)} > 0$ the new iterate $x^{(k+1)}$ given by* (4.21) *and* (4.32) *is well defined. Moreover,*

$$\|\Delta x^{(k)}\| \leq \Big(1 + \frac{1}{\mu_f + \mu_q}\|\nabla f(\hat{x}^{(k)}) - I\|\Big) \max\{1, \gamma^{(k)}\}\|u^{(k)}\|,$$

where $\mu_f := \inf_{x \in \mathbb{R}^n} \mu_f(x)$, $\mu_q := \inf_{d \in \operatorname{dom}\partial q} \mu_q(d)$.

Proof. Note that $\mu_f = \inf_{x_1 \neq x_2} \frac{\langle f(x_1) - f(x_2), x_1 - x_2\rangle}{\|x_1 - x_2\|^2}$ and

$$\mu_q \geq \inf\{\frac{\langle d_1^* - d_2^*, d_1 - d_2\rangle}{\|d_1 - d_2\|^2} \mid (d_i, d_i^*) \in \operatorname{gph}\partial q,\ i = 1,2,\ d_1 \neq d_2\}.$$

Thus $\mu_f + \mu_q > 0$ and the assertion follows from Lemma 4.4, Proposition 4.2 and formula (4.32) □

We now prove locally superlinear convergence of the semismooth* Newton method. We restrict ourselves to the special case when $\gamma^{(k)}$ is constant.

Theorem 4.6. *Assume that the mapping $H = f + \partial q$ is metrically regular at $(\bar{x}, 0) \in \operatorname{gph} H$ and assume that ∂q is semismooth* at $(\bar{x}, -f(\bar{x}))$. Then for every positive number γ there exists a neighborhood U of $\bar{x}$ such that for every starting point $x^{(0)} \in U$ the semismooth* Newton method of Algorithm 1 with $\gamma^{(k)} = \gamma$ for all k, with approximation step* (4.21) *and with Newton step given by* (4.32)*, converges superlinearly to $\bar{x}$.*

Proof. Fix $\kappa' > \operatorname{reg} H(\bar{x}, 0)$, set $\underline{\gamma} = \bar{\gamma} = \gamma$ and determine $\rho' > 0$ according to Lemma 4.3. Denoting by l the Lipschitz constant of f on $\mathscr{B}_{\rho'}(\bar{x})$ and taking into account (4.26) and (4.29), we can conclude that

$$\mathscr{G}^{L,\kappa}_{\mathscr{F},(\bar{x},\bar{x})}(x,d) \neq \emptyset \ \forall (x,d) \in \mathscr{B}_{\rho'}(\bar{x}) \times \mathbb{R}^n,$$

where $L := 2 + (2+\gamma)\big(2 + \frac{l}{\gamma}\big)$ and $\kappa = (1 + 2\kappa'(l+1))(C_1 + C_2 l)^2$. Since we choose in every step $((\hat{x}^{(k)}, \hat{d}^{(k)}), \hat{y}^{(k)}, A, B) \in \mathscr{G}^{L,\kappa}_{\mathscr{F},(\bar{x},\bar{x})}(x^{(k)}, d^{(k)})$, the assertion follows from Theorem 3.5.

□

5 Globalization

In the last section we showed locally superlinear convergence of our implementation of the 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 like the forward-backward (FB) splitting method, the Douglas-Rachford (DR) splitting method and some hyperplane projection method, respectively, in order to ensure global convergence.

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

$$r_\gamma(x) := \|(-\gamma u_\gamma(x), u_\gamma(x))\| = \sqrt{1+\gamma^2}\|u_\gamma(x)\| \tag{5.35}$$

as motivated by (4.21).

5.1 A non-monotone line-search heuristic

In general, we replace the full Newton step 4. in Algorithm 1 by a damped step of the form

$$x^{k+1} = \hat{x}^{(k)} + \alpha^{(k)} \triangle x^{(k)} \text{ with } \triangle x^{(k)} := -A^{-1} B \hat{y}^{(k)},$$

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

$$r_{\gamma^{(k)}}(\hat{x}^{(k)} + \alpha^{(k)} s^{(k)}) \le (1 + \delta^{(k)} - \mu \alpha^{(k)}) r_{\gamma^{(k)}}(\hat{x}^{(k)}) \tag{5.36}$$

is fulfilled, where $\mu \in (0, 1)$ 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, e.g., we can choose the first element of a sequence (β_j), which has $\beta_0 = 1$ and converges monotonically to zero, such that the line search condition (5.36) is fulfilled.

Algorithm 2 (Globalized semismooth* Newton heuristic for VI of the second kind)**.**
Input: starting point $x^{(0)}$*, line search parameter* $0 < \nu < 1$*, a sequence* $\delta^{(k)} \downarrow 0$*, a sequence* $\beta_j \downarrow 0$ *with* $\beta_0 = 1$ *and a stopping tolerance* $\varepsilon_{tol} > 0$*.*
1. Choose $\gamma^{(0)}$ *and set the iteration counter* $k := 0$*.*
2. If $r_{\gamma^{(k)}}(x^{(k)}) \le \varepsilon_{tol}$*, stop the algorithm.*
3. Compute $G^{(k)}$ *fulfilling* (4.27) *and the Newton direction* $\Delta x^{(k)}$ *by solving* (4.32)*.*
4. Determine the step size $\alpha^{(k)}$ *as the first element from the sequence* β_j *satisfying*

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

5. Set $x^{(k+1)} = x^{(k)} + \alpha^{(k)} \Delta x^{(k)}$ *and update* $\gamma^{(k+1)}$*.*
6. Increase the iteration counter $k := k+1$ *and go to Step 2.*

Note that every evaluation of the residual function $r_\gamma(x)$ requires the computation of $u_\gamma(x)$, i.e., essentially one step of the FB splitting method. For $\gamma^{(k)}$ we suggest a choice $\gamma^{(k)} \approx \|\nabla f(x^{(k)}\|$. 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 $\|\nabla f(x^{(k)})\|_1$.

Algorithm 2 is a heuristic and we are not able to show convergence properties. Nevertheless it showed good convergence properties in practice and therefore we incorporate its principles in other algorithms to improve their performance.

5.2 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. Assume that the used globally convergent method is formally given by some mapping $\mathscr{T} : \mathbb{R}^n \to \mathbb{R}^n$, which computes from some iterate $x^{(k)}$ the next iterate by

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

Of course, $\mathscr{T}$ must depend on the problem (1.2) which we want to solve and will presumably depend also on some additional parameters which control the behavior of the method. In our notation we neglect to a large extent these dependencies.

Consider the following well-known examples for such a mapping $\mathscr{T}$.

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

$$\mathscr{T}_\lambda^{\mathrm{FB}}(x) = (I + \lambda \partial q)^{-1}(I - \lambda f)(x), \tag{5.37}$$

where $\lambda > 0$ is a suitable prarameter. Note that $\mathscr{T}_\lambda^{\mathrm{FB}}(x) = x + u_{1/\lambda}(x)$.

2. For the Douglas-Rachford splitting method we have

$$\mathscr{T}_\lambda^{\mathrm{DR}}(x) = (I + \lambda f)^{-1}\Big((I + \lambda \partial q)^{-1}(I - \lambda f) + \lambda f\Big)(x) = (I + \lambda f)^{-1}(\mathscr{T}_\lambda^{\mathrm{FB}} + \lambda f)(x), \tag{5.38}$$

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 [22]. Let x and $\gamma > 0$ be given and consider $\hat{x} = \mathscr{T}_{1/\gamma}^{\mathrm{FB}}(x)$, i.e. $\hat{x} - x = u_\gamma(x)$. Then $0 \in \gamma(\hat{x} - x) + f(x) + \partial q(\hat{x})$ and consequently

$$0 \in v + \gamma(\hat{x} - x) + (f(x) - f(\hat{x})), \tag{5.39}$$

where $v := -\gamma(\hat{x}-x)+f(\hat{x})-f(x) \in H(\hat{x})$. 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.,

$$\mathscr{T}_\gamma^{\mathrm{PM}}(x) = x - \frac{\langle v, \hat{x}-x\rangle}{\|v\|^2} v. \tag{5.40}$$

Algorithm 3 (Globally convergent hybrid semismooth* Newton method for VI of the second kind)**.**
Input: A method for solving (1.2) *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. Compute $G^{(k)}$ *fulfilling* (4.27) *and the Newton direction* $\Delta x^{(k)}$ *by solving* (4.32)*. 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.*

Recall that the Newton direction $\Delta x^{(k)}$ exists, whenever condition (4.30) is fulfilled. In particular, by Corollary 4.5 this holds if f is monotone and either f or ∂q is strongly monotone.

In what follows we denote by k_l the subsequence of iterations where the new iterate x^{k+1} is computed in the 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 5.1. *Assume that the GE* (1.2) *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* (1.2)*. 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.*

1. *If the Newton iterate is accepted only finitely many times in step 4, then the sequence* $x^{(k)}$ *has at least one accumulation point which solves* (1.2)*. Further, if the sequence* $\gamma^{(k)}$ *is bounded and bounded away from* 0*, for every accumulation point* $\bar{x}$ *of the sequence* $x^{(k)}$ *which is a solution to* (1.2)*, the mapping* H *is not metrically regular around* $(\bar{x}, 0)$.

2. *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* (1.2)*.*

3. *If there exists an accumulation point* $\bar{x}$ *of the sequence* $x^{(k)}$ *which solves* (1.2) *and where the mapping* H *is metrically regular and semismooth* at* $(\bar{x}, 0)$*, then the sequence* $x^{(k)}$ *converges superlinearly to* $\bar{x}$ *and the Newton step in step 4 is accepted with step length* $\alpha^{(k)} = 1$ *for all* k *sufficiently large, provided the sequence* $\gamma^{(k)}$ *satisfies*

$$0 < \underline{\gamma} \le \gamma^{(k)} \le \bar{\gamma}\ \forall k \tag{5.41}$$

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

Proof. The first statement is an immediate consequence of our assumption on $\mathscr{T}$ and the third assertion. 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

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

the continuity of f and the closedness of $\operatorname{gph}\partial q$, 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 shows our second assertion.

Finally, assume that $\bar{x}$ is an accumulation point of the sequence $x^{(k)}$ such that $0 \in H(\bar{x})$ and H is both metrically regular and semismooth* at $(\bar{x}, 0)$ and assume that (5.41) holds. Fixing $\kappa' > \operatorname{reg} H(\bar{x}, 0)$, by Lemma 4.3 we can find a positive radius $\rho' > 0$ such that for all $(x^{(k)}, d^{(k)}) \in \mathscr{B}_{\rho'}(\bar{x}) \times \mathbb{R}^n$ the Newton direction $\Delta x^{(k)}$ exists and the matrices A, B given by (4.28) fulfill by virtue of (4.29) the inequality

$$\|A^{-1}\| \|(A \vdots B)\|_F \leq (1+2\kappa'(l+1))(C_1 + C_2 \|\nabla f(\hat{x}^{(k)})\|)^2 \leq (1+2\kappa'(l+1))(C_1 + C_2 l)^2,$$

where l denotes the Lipschitz constant of f in $\mathscr{B}_{\rho'}(\bar{x})$. By Corollary 3.4 the solution $\bar{x}$ is isolated and we can choose ρ' possibly smaller such that $\operatorname{dist}(x, H^{-1}(0)) = \|x - \bar{x}\| \; \forall x \in \mathscr{B}_{\rho'}(\bar{x})$.

By Proposition 3.3 together with the first equation in (4.26), for every $\varepsilon > 0$ there is some $\delta > 0$ such that

$$\begin{aligned} \|x^{(k)} + \Delta x^{(k)} - \bar{x}\| &\leq \left\| \begin{pmatrix} \hat{x}^{(k)} + \Delta x^{(k)} - \bar{x} \\ \hat{d}^{(k)} + \Delta d^{(k)} - \bar{x} \end{pmatrix} \right\| \leq \varepsilon \|A^{-1}\| \|(A \vdots B)\|_F \|(\hat{x}^{(k)} - \bar{x}, \hat{d}^{(k)} - \bar{x}, \hat{y}_1^{(k)}, \hat{y}_2^{(k)})\| \\ &\leq \varepsilon (1+2\kappa'(l+1))(C_1 + C_2 l)^2 \Big(2 + (2+\gamma^{(k)})\big(2 + \frac{l}{\gamma^{(k)}}\big)\Big) \|x^{(k)} - \bar{x}\| \end{aligned} \quad (5.42)$$

whenever $x^{(k)} \in \mathscr{B}_\delta(\bar{x})$. In particular, we can find some $0 < \bar{\delta} < \rho'/(1 + \frac{l}{\underline{\gamma}})$ such that

$$\|x^{(k)} + \Delta x^{(k)} - \bar{x}\| \leq \frac{1-\nu}{(2+\frac{l}{\underline{\gamma}})\big(1+\kappa'(\bar{\gamma}+l)\big)} \|x^{(k)} - \bar{x}\| < \frac{1}{2} \|x^{(k)} - \bar{x}\|$$

whenever $x^{(k)} \in \mathscr{B}_{\bar{\delta}}(\bar{x})$. 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. Indeed, consider $x^{(k)} \in \mathscr{B}_{\bar{\delta}}(\bar{x})$. Then, by (4.24) we obtain

$$\|x^{(k)} + u^{(k)} - \bar{x}\| = \|\hat{d}^{(k)} - \bar{x}\| \leq (1 + \frac{l}{\underline{\gamma}}) \|x^{(k)} - \bar{x}\| < \rho'$$

and by the definition of $u^{(k)}$ we have

$$-\gamma^{(k)} u^{(k)} + f(x^{(k)} + u^{(k)}) - f(x^{(k)}) \in H(x^{(k)} + u^{(k)}).$$

Due to metric regularity we conclude

$$\begin{aligned}\operatorname{dist}(x^{(k)}+u^{(k)},H^{-1}(0)) = \|x^{(k)}+u^{(k)}-\bar{x}\| &\le \kappa'\operatorname{dist}(0,H(x^{(k)}+u^{(k)}))\\ &\le \kappa'\|-\gamma^{(k)}u^{(k)}+f(x^{(k)}+u^{(k)})-f(x^{(k)})\| \le \kappa'(\bar{\gamma}+l)\|u^{(k)}\|\end{aligned}$$

implying

$$\|x^{(k)}-\bar{x}\| \le \big(1+\kappa'(\bar{\gamma}+l)\big)\|u^{(k)}\|.$$

Since $u_{\gamma^{(k)}}(\bar{x})=0$, we obtain from (4.19)

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

showing

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

From this we conclude that the step size $\alpha^{(k)}=1$ is accepted and thus our claim holds true. 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

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

establishing convergence of the sequence $x^{(k)}$ to $\bar{x}$. The superlinear speed of convergence is a consequence of (5.42). □

In the following subsections we discuss some implementation details and alternatives for the three different mappings $\mathscr{T}$ introduced at the beginning of this section.

5.2.1 Douglas-Rachford splitting method

In order that $\mathscr{T}_\lambda^{\mathrm{DR}}$ meets the assumptions of Theorem 5.1 it is sufficient that f is monotone, see, e.g., [13, Corollary 2]. We now discuss a variant of Algorithm 3 which seems to be slightly more efficient. Consider the sequences $x^{(k)}$ and $v^{(k)}$ generated by

$$x^{(k+1)} = \mathscr{T}_\lambda^{\mathrm{DR}}(x^{(k)}),\ v^{(k)} = (I+\lambda f)(x^{(k)}).$$

Then it is well known, see, e.g., [13], that $v^{(k+1)} = \mathscr{G}_\lambda(v^{(k)})$, where

$$\mathscr{G}_\lambda(v) := \Big(J_{\partial q}^\lambda\big(2J_f^\lambda - I\big)+I-J_f^\lambda\Big)(v)$$

with resolvents $J_{\partial q}^\lambda := (I+\lambda\partial q)^{-1}$, $J_f^\lambda := (I+\lambda f)^{-1}$. From (5.38) we obtain

$$v^{(k+1)} = \mathscr{T}_\lambda^{\mathrm{FB}}(x^{(k)})+\lambda f(x^{(k)}) = x^{(k)}+u_{\frac{1}{\lambda}}(x^{(k)})+\lambda f(x^{(k)}) = v^{(k)}+u_{\frac{1}{\lambda}}(x^{(k)}) \tag{5.43}$$

implying $u_{\frac{1}{\lambda}}(x^{(k)}) = v^{(k+1)} - v^{(k)}$. Thus, $u_{\frac{1}{\lambda}}(x^{(k+1)}) = v^{(k+2)} - v^{(k+1)} = \mathscr{G}_\lambda(v^{(k+1)}) - \mathscr{G}_\lambda(v^{(k)})$ and by [13, relation (17)] we obtain

$$\begin{aligned}
\|u_{\frac{1}{\lambda}}(x^{(k+1)})\|^2 &= \|\mathscr{G}_\lambda(v^{(k+1)}) - \mathscr{G}_\lambda(v^{(k)})\|^2 \\
&\le \langle \mathscr{G}_\lambda(v^{(k+1)}) - \mathscr{G}_\lambda(v^{(k)}), v^{(k+1)} - v^{(k)} \rangle \\
&\qquad - \langle (I - J_f^\lambda)(v^{(k+1)}) - (I - J_f^\lambda)(v^{(k)}), J_f^\lambda(v^{(k+1)}) - J_f^\lambda(v^{(k)}) \rangle \\
&= \langle u_{\frac{1}{\lambda}}(x^{(k+1)}), u_{\frac{1}{\lambda}}(x^{(k)}) \rangle - \lambda \langle f(x^{(k+1)}) - f(x^{(k)}), x^{(k+1)} - x^{(k)} \rangle.
\end{aligned} \tag{5.44}$$

For the following analysis we require that f is even strongly monotone, i.e.,

$$\mu_f := \inf_{x_1 \ne x_2} \frac{\langle f(x_1) - f(x_2), x_1 - x_2 \rangle}{\|x_1 - x_2\|^2} > 0. \tag{5.45}$$

Recall that $\mu_f = \inf_{x \in \mathbb{R}^n} \mu_f(x)$ with $\mu_f(x)$ given by (4.33). Then we obtain from (5.44) that

$$\|u_{\frac{1}{\lambda}}(x^{(k+1)})\|^2 \le \|u_{\frac{1}{\lambda}}(x^{(k+1)})\| \|u_{\frac{1}{\lambda}}(x^{(k+1)})\| - \mu_f \|x^{(k+1)} - x^{(k)}\|^2$$

and thus $\|u_{\frac{1}{\lambda}}(x^{(k+1)})\| < \|u_{\frac{1}{\lambda}}(x^{(k)})\|$, i.e., the residual $r_{\frac{1}{\lambda}}(x^{(k)})$ is strictly decreasing. The basic idea is now, to perform alternately a step of the Douglas-Rachford splitting method with parameter $\lambda = \frac{1}{\gamma}$ and then a semismooth* Newton step with line search with parameter γ, where in the line search we possibly sacrifice a part of the reduction in the residual gained in the Douglas-Rachford splitting step. This procedure is mainly motivated by the excellent performance of the non-monotone line search heuristic of Algorithm 2, where we now have a possibility to control the increment in the residual in order to guarantee convergence.

Algorithm 4 (Globally convergent hybrid semismooth* Newton - Douglas Rachford method for VI of the second kind)**.**
Input: A starting point $x^{(0)}$, a parameter $\gamma > 0$, line search parameters $0 < \nu < 1$, $0 < \xi < 1$, a sequence $\beta_j \downarrow 0$ with $\beta_0 = 1$ and a stopping tolerance $\varepsilon_{tol} > 0$.
1. Compute $u^{(0)} := u_\gamma(x^{(0)})$ and set the counter $k := 0$.
2. If $r_\gamma(x^{(2k)}) \le \varepsilon_{tol}$, stop the algorithm.
3. Compute $x^{2k+1} = \mathscr{T}^{\mathrm{DR}}_{1/\gamma}(x^{(2k)})$ and $u^{(2k+1)} := u_\gamma(x^{(2k+1)})$.
4. Compute $G^{(2k+1)}$ fulfilling (4.27) and the Newton direction $\Delta x^{(2k+1)}$ by solving (4.32). Determine the step size $\alpha^{(2k+1)}$ as the first element from the sequence β_j satisfying

$$\|u_\gamma(x^{(2k+1)} + \beta_j \Delta x^{(2k+1)})\| \le (1 - \nu\beta_j)\big(\xi \|u^{(2k)}\| + (1-\xi)\|u^{(2k+1)}\|\big).$$

5. Set $x^{(2k+2)} = x^{(2k+1)} + \alpha^{(2k+1)} \Delta x^{(2k+1)}$ and $u^{(2k+2)} := u_\gamma(x^{(2k+2)})$.
6. Increase the counter $k := k+1$ and go to Step 2.

Theorem 5.2. *If f is strongly monotone then Algorithm 4 is well defined. If, in addition, f is Lipschitzian on the set $S^{(0)} := \{x \mid \|u_\gamma(x)\| \le \|u_\gamma(x^{(0)})\|\}$, then the sequence $u^{(2k)}$ converges at least Q-linearly to 0 and the sequence $x^{(j)}$ converges at least R-linearly to the unique solution $\bar{x}$ of (1.2). If H is semismooth* at $(\bar{x}, 0)$, then convergence of the sequence $x^{(2k)}$ is Q-superlinear.*

Proof. Since f is strongly monotone, for every k the Newton direction $\Delta x^{(2k+1)}$ is well defined by Corollary 4.5. Further, from (5.44) it follows that $\|u^{(2k+1} \| < \|u^{(2k)}\|$ implying $\|u^{(2k+1)}\| < \xi \|u^{(2k)}\| +$

$(1-\xi)\|u^{(2k+1)}\| < \|u^{(2k}\|$. Since the function $\phi^{(2k+1)}(\alpha) := \|u_\gamma(x^{(2k+1)} + \alpha\Delta x^{(2k+1)})\|$ is continuous by virtue of (4.19) and $\phi^{(2k+1)}(0) = \|u^{(2k+1)}\|$, we conclude that $\phi^{(2k+1)}(\alpha) < (1-\nu\alpha)(\xi\|u^{(2k)}\| + (1-\xi)\|u^{(2k+1)}\|)$ for all sufficiently small $\alpha > 0$. Thus, also the step size $\alpha^{(2k+1)}$ is well defined and hence so is the whole algorithm. Note that

$$\|u^{(2k+2)}\| = \phi^{(2k+1)}(\alpha^{(2k+1)}) < \xi\|u^{(2k)}\| + (1-\xi)\|u^{(2k+1)}\| < \|u^{(2k)}\| \tag{5.46}$$

and therefore $x^{(j)} \in S^{(0)}$ $\forall j > 0$. Denoting by l the Lipschitz constant of f on $S^{(0)}$, we conclude that $\|\nabla f(x^{(2k+1)})\| \leq l$ $\forall k$ because of $x^{(2k+1)} \in \operatorname{int} S^{(0)}$. From (5.43) we deduce

$$(I + \frac{1}{\gamma} f)(x^{(2k+1)}) - (I + \frac{1}{\gamma} f)(x^{(2k)}) = u^{(2k)}$$

implying

$$\|u^{(2k)}\| > \|x^{(2k+1)} - x^{(2k)}\| \geq \frac{\gamma}{\gamma + l}\|u^{(2k)}\|.$$

Using (5.44) we obtain

$$\|u^{(2k+1)}\|^2 \leq \langle u^{(2k+1)}, u^{(2k)}\rangle - \frac{\mu_f}{\gamma}\|x^{(2k+1)} - x^{(2k)}\|^2 \leq \|u^{(2k+1)}\|\|u^{(2k)}\| - \frac{\mu_f \gamma}{(\gamma+l)^2}\|u^{(2k)}\|^2$$

and consequently

$$\|u^{(2k+1)}\| < \tau\|u^{(2k)}\| \quad \text{with} \quad \tau := 1 - \frac{\mu_f\gamma}{(\gamma+l)^2} < 1.$$

Combining this estimate with (5.46) we obtain

$$\|u^{(2k+2)}\| < \bar{\tau}\|u^{(2k)}\| \quad \text{with} \quad \bar{\tau} := \xi + \tau(1-\xi)) < 1$$

and Q-linear convergence of the sequence $u^{(2k)}$ is established. By Corollary 4.5 we obtain

$$\|\Delta x^{(2k+1)}\| \leq \Big(1 + \frac{1}{\mu_f}\|\nabla f(x^{(2k+1)}) - I\|\Big)\max\{1,\gamma\}\|u^{(2k+1)}\| \leq \Big(1 + \frac{l+1}{\mu_f}\Big)\max\{1,\gamma\}\|u^{(2k+1)}\|$$

and together with $\alpha^{(2k+1)} \leq 1$ we have

$$\|x^{(2k+2)} - x^{(2k)}\| \leq \|x^{(2k+2)} - x^{(2k+1)}\| + \|x^{(2k+1)} - x^{(2k)}\| \leq \|\Delta x^{(2k+1)}\| + \|u^{(2k)}\| \leq C\|u^{(2k)}\|$$

with $C := \tau\big(1 + \frac{l+1}{\mu_f}\big)\max\{1,\gamma\} + 1$. This implies

$$\|x^{(2j)} - x^{(2k)}\| \leq C\sum_{i=k}^{j-1}\|u^{(2i)}\| \leq \frac{C}{1-\bar{\tau}}\|u^{(2k)}\|$$

for all $0 < k < j$. Thus $x^{(2k)}$ is a Cauchy sequence and therefore convergent to some $\tilde{x}$. By continuity of $u_\gamma(\cdot)$ we have $u_\gamma(\tilde{x}) = \lim_{k\to\infty} u_\gamma(x^{(2k)}) = \lim_{k\to\infty} u^{(2k)} = 0$ and hence $0 \in H(\tilde{x})$. But by strong monotonicity of H the solution of (1.2) is unique and $\tilde{x} = \bar{x}$ follows. Further we have

$$\|x^{(2k)} - \bar{x}\| \leq \frac{C}{1-\bar{\tau}}\bar{\tau}^k\|u^{(0)}\|,$$

$$\|x^{(2k+1)} - \bar{x}\| \leq \|x^{(2k)} - \bar{x}\| + \|x^{(2k+1)} - x^{(2k)}\| \leq \|x^{(2k)} - \bar{x}\| + \|u^{(2k)}\| \leq \Big(1 + \frac{C}{1-\bar{\tau}}\Big)\bar{\tau}^k\|u^{(0)}\|$$

and R-linear convergence of $x^{(j)}$ to $\bar{x}$ with convergence factor $\sqrt{\tau}$ follows.

There remains to show the superlinear convergence of the sequence $x^{(2k)}$. Strong monotonicity of f implies that H is a maximal strongly monotone mapping and hence it is (strongly) metrically regular around $(\bar{x},0)$. Using similar argument as in the proof of Theorem 5.1 we can conclude that $\alpha^{(k)}=1$ for all k sufficiently large and $\lim_{k\to\infty}\|x^{(2k+2)}-\bar{x}\|/\|x^{(2k+1)}-\bar{x}\|=0$. Further,

$$\frac{\|x^{(2k+1)}-\bar{x}\|}{\|x^{(2k)}-\bar{x}\|}\leq 1+\frac{\|u^{(2k)}\|}{\|x^{(2k)}-\bar{x}\|}\leq 3+\frac{l}{\gamma}$$

by (4.25) and $\lim_{k\to\infty}\|x^{(2k+2)}-\bar{x}\|/\|x^{(2k)}-\bar{x}\|=0$ follows. This completes the proof. □

Remark 5.3. *The factor τ appearing in the proof of Theorem 5.2 is the smallest when $\gamma=l$, the Lipschitz constant of f. This is in accordance with our practical experience with the heuristic Algorithm 2 that a choice $\gamma^{(k)}\approx\|\nabla f(x^{(k)})\|$ yields good results.*

Remark 5.4. *The requirement that f is Lipschitzian on $S^{(0)}$ is, e.g., fulfilled if $S^{(0)}$ is bounded. In particular, since $x+u_\gamma(x)\in\operatorname{dom}\partial q$, this is the case when $\operatorname{dom}\partial q$ is bounded.*

5.2.2 Forward-backward splitting method

Most of the research on forward-backward splitting methods has relied on assumptions of strong monotonicity, cf. [5]. E.g., when f is Lipschitzian on $\operatorname{dom}\partial q$, and either f or ∂q is strongly monotone, then $\mathcal{T}_\lambda^{\rm FB}$ fulfills the requirements of Theorem 5.1 provided λ is chosen sufficiently small, see, e.g., [2]. Using [2, Theorem 2.4], it is not difficult to show, that for the sequence $x^{(k+1)}=\mathcal{T}_\lambda^{\rm FB}(x^{(k)})$ we have

$$\|u_{1/\lambda}(x^{(k+1)})\|<\tau\|u_{1/\lambda}(x^{(k)})\|$$

with some factor $\tau<1$ for λ small enough. We could proceed in the same way as we have used for the Douglas-Rachford splitting method, but we omit to do this for the following reason. When we are forced to choose λ very small, in particular when $\frac{1}{\lambda}$ is much larger than the Lipschitz constant of f, our numerical experiments do not show a favourable behaviour compared with our semismooth* Newton approaches based on $\mathcal{T}_\lambda^{\rm DR}$ and $T_\lambda^{\rm PM}$. On the other hand, if we are allowed to choose λ comparatively large, then the pure forward-backward method shows a good convergence behaviour and we need not to use the semismooth* Newton method at all.

5.2.3 Hybrid projection-proximal point algorithm

When using $\mathcal{T}_\gamma^{\rm PM}$, we only need monotonicity of H, monotonicity of f is not required.

Consider a sequence $x^{(k+1)}=\mathcal{T}_{\gamma^{(k)}}^{\rm PM}(x^{(k)})$. It follows from [22, Theorem 2.2] that the following two conditions are sufficient in order to meet the assumptions of Theorem 5.1:

1. $\gamma^{(k)}>0\ \forall k$ and $\sum_{k=0}^\infty(\gamma^{(k)})^{-2}=\infty$.
2. There is some $\sigma\in[0,1)$ such that

$$\|f(x^{(k)}+u^{(k)})-f(x^{(k)})\|\leq\sigma\max\{\|-\gamma^{(k)}u^{(k)}+f(x^{(k)}+u^{(k)})-f(x^{(k)})\|,\gamma^{(k)}\|u^{(k)}\|\}\quad\forall k,\tag{5.47}$$

where $u^{(k)}:=u_{\gamma^{(k)}}(x^{(k)})$.

We now demonstrate that both conditions can be fulfilled by setting $\gamma^{(k)} = \hat{\gamma}\ \forall k$ for any $\hat{\gamma} \geq \hat{l}/\sigma$ with

$$\hat{l} := \max\{\|\nabla f(x)\| \mid x \in \mathscr{B}_{2\|x^{(0)}-\bar{x}\|}(\bar{x})\},$$

where $\bar{x}$ denotes any solution of (1.2) and $\sigma \in (0,1)$ is arbitrarily fixed. It follows that f is Lipschitzian on $\mathscr{B}_{2\|x^{(0)}-\bar{x}\|}(\bar{x})$ with constant $\hat{l}$. Of course, condition 1. is trivially fulfilled and there remains to show the second one. Consider any iterate $x^{(k)} \in \mathscr{B}_{\|x^{(0)}-\bar{x}\|}(\bar{x})$. By (4.18) we obtain

$$\|x^{(k)} + u^{(k)} - \bar{x}\| \leq \|x^{(k)} - \bar{x}\| + \frac{\|f(x^{(k)}) - f(\bar{x})\|}{\hat{\gamma}} \leq (1+\sigma)\|x^{(k)} - \bar{x}\|$$

and therefore

$$\|f(x^{(k)} + u^{(k)}) - f(x^{(k)})\| \leq \hat{l}\|u^{(k)}\| \leq \sigma\hat{\gamma}\|u^{(k)}\|.$$

By [22, Lemma 2.1] we have $\|x^{(k+1)} - \bar{x}\| \leq \|x^{(k)} - \bar{x}\|$ and our claim follows by induction.

In practice we choose $\gamma^{(k)}$ not constant in every iteration, but we try it to adjust it to a local Lipschitz constant of f near $x^{(k)}$. E.g., we can choose $\gamma^{(k)}$ as the first element of a sequence $\chi_j \uparrow \infty$ such that $\chi_j \geq \|\nabla f(x^{(k)})\|$ and inequality (5.47) holds.

6 Numerical experiments

All variants of the semismooth* Newton method presented in the preceding sections have been extensively tested by means of a wide range of examples. In this section we will show first the behavior of Algorithm 1 via a low-dimensional example with an economic background. Thereafter we will illustrate the efficiency of the family of methods, presented in Section 5, by means of an artificially constructed set of problems having a variable scale.

6.1 An economic equilibrium

In [18] the authors considered an evolution process in an oligopolistic market, where the players (firms) adapt their strategies (productions) according to changing external parameters (input prices etc.). In their decisions, however, they must take into account that each change of production may be associated with some costs, see [4]. As derived in [18], the respective Cournot-Nash equilibrium at some time instant is governed by GE (1.2) with

$$q = \tilde{q} + \delta_A, \qquad A = \prod_{i=1}^{n} A_i, \qquad \tilde{q}(x) = \sum_{i=1}^{n} \tilde{q}_i(x_i)$$

where n denotes the number of players. The strategy sets A_i are nonempty and compact intervals $[b_i, d_i]$ and $\tilde{q}_i(x_i) = \beta_i|x_i - a_i|$ for some non-negative reals β_i and parameters $a_i \in A_i, i = 1, 2, \ldots, n$. Mapping f is continuously differentiable on an open set containing $\operatorname{dom}\partial q = A$ and its description can be found in, e.g., [16] and [17]; see also [18], where the values of $a = (a_1, a_2, \ldots, a_n)$ and $\beta = (\beta_1, \beta_2, \ldots, \beta_n)$ are specified. The implementation of the semismooth* Newton method described in Section 4 has been first applied to the problem formulation from [18], where all production cost functions are convex and f is strongly monotone on A. Thereafter we have replaced the production cost function of the first player by a (more realistic) concave one. As a consequence, mapping f has lost its monotonicity on A and the respective GE (1.2) might have possibly multiple solutions with not all of them being necessarily Cournot-Nash equilibria. It has turned out, however, that, in our example, the Jacobian of

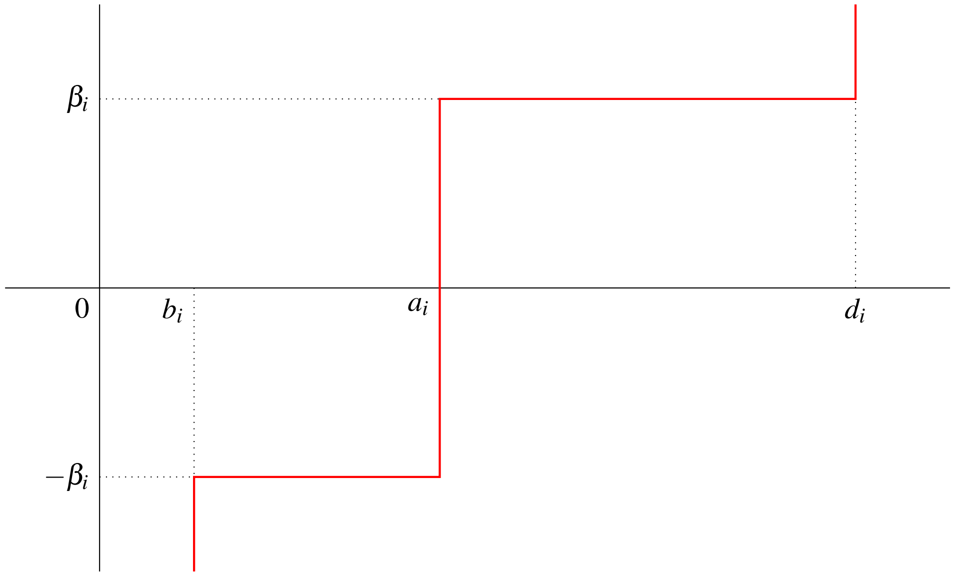

Figure 1: $\operatorname{gph}\partial q_i$ with $A_i = [b_i, d_i]$.

f is positive definite at the obtained solution. This implies in particular that this point is a Cournot-Nash equilibrium and the respective multifunction H is metrically regular there.

It is easy to see that in both cases the resulting multifunction $\mathscr{F}$ is semismooth* at any point of its graph. Since q is a separable function, one has that $\partial q(x) = \prod_{i=1}^{n} \partial q_i(x_i)$ and the sets $\operatorname{gph}\partial q_i$ attain the form depicted in Figure 1.

Next we provide a simple formula for the computation of the matrix G needed in the Newton step. This formula, however, will be given for a more general situation considered in the family of test examples discussed in Subsection 6.2. Assume that each $\operatorname{gph}\partial q_i$ is a polygonal line in $\mathbb{R}^2$ connecting the given points

$$(\xi_1^i, -\infty), (\xi_1^i, \eta_1^i), (\xi_2^i, \eta_2^i), \ldots, (\xi_{2m_i-1}^i, \eta_{2m_i-1}^i), (\xi_{2m_i}^i, \eta_{2m_i}^i), (\xi_{2m_i}^i, \infty), \quad i = 1, 2, \ldots, n, \tag{6.48}$$

for some integer $m_i \geq 1$. Further suppose that

$$\Delta\xi_j^i := \xi_{j+1}^i - \xi_j^i \begin{cases} > 0 & \text{if } j \text{ is odd} \\ = 0 & \text{if } j \text{ is even,} \end{cases}$$

$$\Delta\eta_j^i := \eta_{j+1}^i - \eta_j^i \begin{cases} \geq 0 & \text{if } j \text{ is odd} \\ > 0 & \text{if } j \text{ is even.} \end{cases}$$

Observe that $\Delta\xi_j^i + \Delta\eta_j^i > 0$ holds for all i and all $j = 1, \ldots, 2m_i - 1$. It follows that a polygonal line given this way is monotone increasing and, consequently, q_i is a convex piecewise linear-quadratic function with $\operatorname{dom} q_i = [\xi_1^i, \xi_{2m_i}^i]$. Clearly, in the example depicted in Figure 1 one has

$$m_i = 2, \xi_1^i = b_i, \xi_2^i = \xi_3^i = a_i, \xi_4^i = d_i, \eta_1^i = \eta_2^i = -\beta_i, \eta_3^i = \eta_4^i = \beta_i.$$

With this problem structure it is not difficult to compute the quantity $u_\gamma(x)$ as follows. For each $i = 1, \ldots, n$ we denote by u_i and f_i the ith component of $u_\gamma(x)$ and $f(x)$, respectively, i.e., u_i solves the inclusion

$$0 \in \gamma u_i + f_i + \partial q_i(x_i + u_i) = \gamma(x_i + u_i) + (f_i - \gamma x_i) + \partial q_i(x_i + u_i)$$

Let

$$j_i = \min\{j \in \{1,\dots,2m_i\} \mid \gamma x_i - f_i < \gamma \xi^i_j + \eta^i_j\} \quad (\infty, \text{if } \gamma x_i - f_i \geq \gamma \xi^i_{2m_i} + \eta^i_{2m_i}).$$

Then

$$u_i + x_i = \begin{cases} \xi^i_1 & \text{if } j_i = 1, \\ \xi^i_{j_i-1} + t_i \Delta\xi^i_{j_i-1} & \text{if } 1 < j_i \leq 2m_i, \text{ with } t_i = \dfrac{\gamma x_i - f_i - (\gamma \xi^i_{j_i-1} + \eta^i_{j_i-1})}{\gamma \Delta\xi^i_{j_i-1} + \Delta\eta^i_{j_i-1}}. \\ \xi^i_{2m_i} & \text{if } j_i = \infty, \end{cases}$$

The matrix G, needed in the Newton step, can be computed as follows.

Proposition 6.1. *Let $(x,x^*) \in \operatorname{gph} \partial q$ and G be $n \times n$ diagonal matrix with entries*

$$G_{ii} = \begin{cases} 1 & \text{if } x_i \in \mathscr{A}_i := \{\xi^i_1\} \cup \{\xi^i_{2m_i}\} \cup \{\xi^i_j \,|\, \Delta\xi^i_j = 0\} \\ \frac{\Delta\eta^i_j}{\Delta\xi^i_j + \Delta\eta^i_j} & \text{if } x_i \in [\xi^i_j, \xi^i_{j+1}] \setminus \mathscr{A}_i \text{ and } j \in \{1,\dots,2m_i\} \text{ is odd} \end{cases} \tag{6.49}$$

for $i = 1,2,\dots,n$. Then G fulfills the conditions stated in Theorem 3.6.

Proof. Observe first that the set

$$\{(x_i, x^*_i) \in \operatorname{gph} \partial q_i \,|\, x_i \in \mathscr{A}_i\}$$

comprises all points of $\operatorname{gph} \partial q_i$ lying in its vectical line segments and

$$\bigcup_{\substack{j=1 \\ j \text{ odd}}}^{2m_i} [\xi^i_j, \xi^i_{j+1}] \setminus \mathscr{A}_i = \{x_i \in \operatorname{dom} \partial q_i \,|\, \partial q(x_i) \text{ is a singleton}\}.$$

Next let us notice that $\operatorname{gph} D^*(\partial q)(x,x^*) = \prod_{i=1}^{n} \operatorname{gph} D^*(\partial q_i)(x_i, x^*_i)$, where

$$\operatorname{gph} D^*(\partial q_i)(x_i, x^*_i) = \{(u,v) | (v,-u) \in N_{\operatorname{gph} \partial q_i}(x_i, x^*_i)\}$$

and

$$N_{\operatorname{gph} \partial q_i}(x_i, x^*_i) \supset \begin{cases} \{0\} \times \mathbb{R} & \text{provided } x_i \in \mathscr{A}_i \\ \mathbb{R}(\Delta\eta^i_j, -\Delta\xi^i_j) & \text{provided } x_i \in [\xi^i_j, \xi^i_{j+1}] \setminus \mathscr{A}_i \text{ and } j \in \{1,\dots,2m_i - 1\} \text{ is odd.} \end{cases}$$

From this analysis it follows that, in order to fulfill inclusion (3.13), it suffices to construct G as a diagonal matrix, where $G_{ii} = 1$ provided $x_i \in \mathscr{A}_i$. Otherwise, if $x_i \in [\xi^i_j, \xi^i_{j+1}] \setminus \mathscr{A}_i$ for some odd $j \in \{1,\dots,2m_i - 1\}$, then we put G_{ii} as the (unique) solutions of the equation

$$\frac{\Delta\eta^i_j}{\Delta\xi^i_j} = \frac{G_{ii}}{1 - G_{ii}}.$$

The above equation is well-posed because $\Delta\xi^i_j > 0$ for j odd and leads to the second line in formula (6.49). By construction, all elements G_{ii} belong to $[0,1]$, which implies that G is positive semidefinite and $\|G\| \leq 1$. Thus, the proof is complete. □

	i	1	2	3	4	5
convex c_1	strategies	49.411	51.140	54.236	48.054	43.095
from [18]	objectives	-377.239	-459.943	-639.952	-503.445	-507.100
	value $q_i(x_i)$	0.800	0	5.831	0	0
concave c_1	strategies	97.191	51.140	51.320	43.254	39.470
from (6.51)	objectives	-427.320	-325.770	-491.385	-367.618	-387.836
	value $q_i(x_i)$	24.691	0	0	0	0

Table 1: Cournot-Nash equilibrium strategies x_i, the corresponding objective values and costs of change $q_i(x_i)$ in case of convex and concave cost function c_1.

In the example depicted in Figure 1 one obtains in this way that

$$G_{ii} = \begin{cases} 1 & \text{if } x_i \in \mathscr{A}_i := \{b_i\} \cup \{a_i\} \cup \{d_i\} \\ 0 & \text{if } x_i \in (b_i, a_i) \cup (a_i, d_i), \end{cases} \tag{6.50}$$

which has been used in the computations discussed below.

Next we will present the numerical results for both the monotone and non-monotone case discussed above. In the former one we have used the data from [18, Section 5.1] with $t = 1$ and started the iteration process at the initial iterate $x^{(0)} = (75, 75, \ldots, 75)$. The results are displayed in the upper part of Table 1. Concerning the non-monotone case, we have replaced the original convex production cost function c_1 by a concave one, given by

$$c_1(x_1) := -(1/50)x_1^2 + 15x_1 \tag{6.51}$$

and started from the same vector $x^{(0)}$. The results are displayed in the lower part of Table 1. In both cases we have set $\gamma = 1$ in the approximation step and, as the stopping criterion, we have used the condition $||u|| \le \varepsilon = 10^{-10}$, where u is the output of the approximation step.

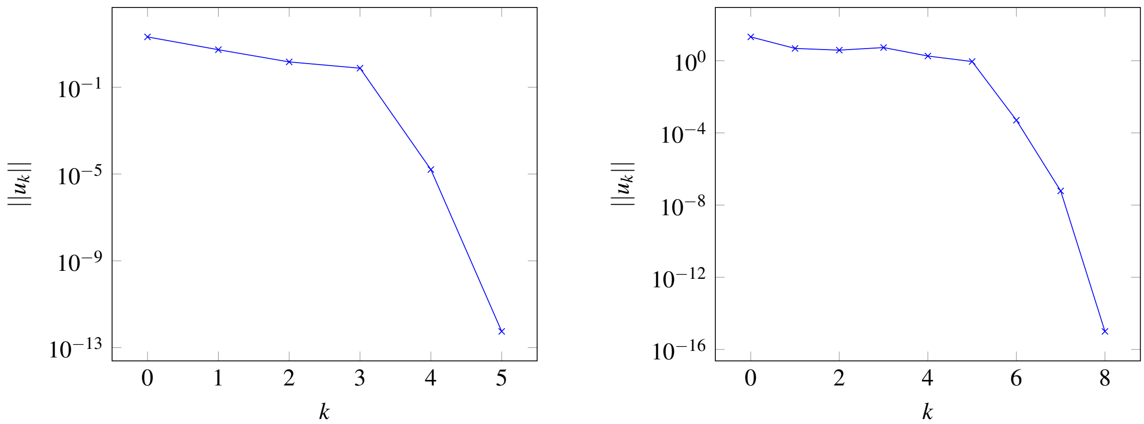

Figure 2: Convergence of $||u^{(k)}||$ in case of convex (left) or concave (right) cost functions c_1.

Figure 2 illustrates very well the superlinear convergence of Algorithm 1 whenever one reaches the respective neighborhood of the solution. Note that in our implementation we moved the stopping criterion behind the approximation step in order to dispose with the actual value of u. Further observe that the lowest eigenvalue of the symmetrized Jacobian of f at the solution is only 0.033474 in the non-monotone case whereas it amounts to 0.207905 in the case of convex c_1.

6.2 Randomly constructed test problems

Given the problem dimension n and a parameter $\beta > 0$, we construct an instance of the problem (1.2) as follows.

1. We randomly compute an $n \times n$ matrix C with elements uniformly distributed in $[-1, 1]$ and set

$$f(x) = \nabla h(x) + (C - C^T)x \text{ with } h(x) = (x^T A x)^2,\ A = \frac{\beta}{n} C C^T.$$

   Note that f is a maximal monotone operator but the skew-symmetric part $(C - C^T)x$ dominates $f(x)$ for small values of β.

2. The set-valued part is of the form $\partial q(x) = \prod_{i=1}^{n} \partial q_i(x)$, where each gph ∂q_i is a polygonal line in $\mathbb{R}^2$ connecting the points (5.35) as described in the previous subsection. The integer m_i is randomly chosen in $[1, 10]$, ξ_1 is randomly chosen from $[-\frac{m_i}{2}, \frac{m_i}{2}]$, η_1 is randomly chosen from $[-\frac{3\beta m_i}{2}, 0]$ and

$$\Delta\xi_j := \xi_{j+1} - \xi_j \begin{cases} \in [0,1] & \text{if } j \text{ even} \\ = 0 & \text{if } j \text{ odd} \end{cases}, \quad \Delta\eta_j := \eta_{j+1} - \eta_j \in [0, \beta], \quad j = 1, \ldots, 2m_i - 1.$$

The mapping ∂q is a maximal strongly monotone mapping (with probability 1) and so is the mapping H as well. Thus, the GE (1.2) has always a unique solution $\bar{x}$, but the problem characteristic will change with β and n.

For each $n \in \{150, 600, 2400\}$ and each $\beta \in \{1, 10^{-2}, 10^{-4}\}$ we constructed 5 test problems. In Table 2 we list the mean values for some characteristic values for different combinations of β and n.

n	$\beta = 1$		$\beta = 0.01$		$\beta = 10^{-4}$	
	$\lVert\nabla f(\bar{x})\rVert$ $\lVert\nabla f(\bar{x})^{-1}\rVert$	$\mu_f(\bar{x})$ $\mu_q(\bar{x})$	$\lVert\nabla f(\bar{x})\rVert$ $\lVert\nabla f(\bar{x})^{-1}\rVert$	$\mu_f(\bar{x})$ $\mu_q(\bar{x})$	$\lVert\nabla f(\bar{x})\rVert$ $\lVert\nabla f(\bar{x})^{-1}\rVert$	$\mu_f(\bar{x})$ $\mu_q(\bar{x})$
150	83.7 0.86	5.2×10^{-4} 5.8×10^{-2}	19.5 31.1	1.1×10^{-7} 5.9×10^{-4}	19.4 17.7	4.2×10^{-11} 5.2×10^{-6}
600	290 0.496	1.0×10^{-4} 7.5×10^{-3}	39.7 42.9	9.5×10^{-9} 8.3×10^{-5}	39.1 25.7	4.5×10^{-12} 4.4×10^{-7}
2400	1202 0.387	2.7×10^{-5} 2.9×10^{-3}	79.5 18.7	2.8×10^{-9} 8.8×10^{-5}	79.6 125	6.4×10^{-13} 5.3×10^{-7}

Table 2: Some characteristic values for the problems

Note that $\lVert\nabla f(\bar{x})\rVert$ acts as a local Lipschitz constant for f near the solution $\bar{x}$, whereas $\mu_f(\bar{x})$ and $\mu_q(\bar{x})$ given by (4.33),(4.34) are constants for local strong monotonicity for f and ∂q.

Theoretically the global convergent methods obey linear convergence properties. However, the available bounds for the convergence factors depend in some way on the ratio of the Lipschitz constant of f and the constants of strong monotonicity for f and q, c.f. [2],[22],[13]. We see from the table above that this ratio worsen for small values of β and large n and our numerical experiments confirm these estimates. In particular, for $\beta = 0.01$ and $\beta = 10^{-4}$ we could not observe linear convergence neither for the FB-splitting method nor the DR-splitting method and the hybrid projection method. These methods were not able to compute an accurate solution within a reasonable time.

In the following tables we display for different combinations of n and β the mean values for the number N of computed Newton directions, the number G of calls to the globally convergent method and the number F of evaluations of f as well as the mean CPU-time in seconds needed to reach a residual less than 10^{-8}. A time limit was set to $10^{-4}n^2$ seconds to perform this task. We tested the heuristic of Algorithm 2, the globally convergent hybrid algorithm 3 combined with any of the three globally convergent methods $\mathscr{T}_\gamma^{\rm FB}$, $\mathscr{T}_\gamma^{\rm DR}$ and $\mathscr{T}_\gamma^{\rm Pr}$ as described in Subsection 5.2, Algorithm 4 as well as all three globally convergent methods $\mathscr{T}_\gamma^{\rm FB}$, $\mathscr{T}_\gamma^{\rm DR}$ and $\mathscr{T}_\gamma^{\rm Pr}$ alone. We always choose $\nu = 0.1$ and $\gamma^{(k)} = \|\nabla f(x^{(k)})\|_1/\sqrt{n}$. In Algorithm 2 we set $\delta^{(k)} = 0.1/k$ and in Algorithm 3 we used $\delta^{(k)} \equiv 5 \times 10^{-4}$. Finally, the parameter ξ in Algorithm 4 was set to 0.9. For all test problems the origin was chosen as the starting point.

All tests were performed in MATLAB using a desktop equipped with an i7-7700 CPU, 3.6 GHZ and 32GB RAM.

	n=150 N/G/F CPU	n=600 N/G/F CPU	n=2400 N/G/F CPU
Alg.4	5.2/5.2/41.4 0.029	5.8/5.8/43.6 0.394	6/6/40.6 7.07
Alg. 2, Alg. 3	$7.2/-/8.2$ 0.012	$7.6/-/8.6$ 0.099	$7.6/-/8.6$ 2.15
$\mathscr{T}^{\rm FB}$	$-/117.8/118.8$ 0.164	$-/158/159$ 0.935	$-/171.4/172.4$ 7.35
$\mathscr{T}^{\rm DR}$	$-/89/351.2$ 0.157	$-/106.2/406.2$ 3.18	$-/114.2/423.8$ 49.8
$\mathscr{T}^{\rm Pr}$	$-/487.6/978.2$ 0.526	$-/1322/2648$ 7.84	$-/1445/2894$ 92.7

Table 3: Test results for $\beta = 1$

In case when $\beta = 1$ all tested methods found a solution with the prescribed tolerance within the given time limit. The heuristic Algorithm 2 as well as the hybrid Algorithm 3 executed only Newton steps with stepsize $\alpha^{(k)} = 1$, i.e., they behave like the pure semismooth* Newton method of Algorithm 1, and showed the best performance of all methods. The second best one was Algorithm 4 which was in turn faster than any of the three globally convergent methods.

The results for $\beta = 10^{-2}$ are listed in Table 4. Algorithm 4 was the fastest one, whereas the three globally convergent method did not find a solution within the time limit. Note that in case $n = 2400$ only (damped) Newton steps were performed and therefore no calls to the global convergent method were done. For $\beta = 10^{-4}$ Algorithm 4 was again the fastest method, cf. Table 5. Now, in addition to the three globally convergent methods, also the heuristic Algorithm 2 failed.

7 Conclusion

The theoretical background of the semismooth* Newton method has been established in [8]. The main aim of this paper is to implement this method to a class of VIs of the second kind and to examine its numerical properties via extensive numerical experiments. The performed tests show in a convincing way that the new Newton method represents an efficient numerical tool for a number

	n=150 N/G/F CPU	n=600 N/G/F CPU	n=2400 N/G/F CPU
Alg.4	45.8/45.8/333.4 0.181	87.8/87.8/625 4.34	105.8/105.8/771.4 85.7
Alg. 2	439.6/ − /1982 1.34	986.4/ − /4260 22.4	716.2/ − /2748 265
Alg. 3($\mathscr{T}^{\mathrm{FB}}$)	134.6/91.2/565 0.469	482.2/42.6/1660 9.82	624/ − /1829 209
Alg. 3($\mathscr{T}^{\mathrm{DR}}$)	155.8/9.2/596.2 0.393	451.2/9.6/1565 8.95	624/ − /1829 209
Alg. 3($\mathscr{T}^{\mathrm{Pr}}$)	165.8/9.6/593.6 0.417	459/8.4/1564 9.25	624/ − /1829 209

Table 4: Test results for $\beta = 10^{-2}$

	n=150 N/G/F CPU	n=600 N/G/F CPU	n=2400 N/G/F CPU
Alg.4	67.2/67.2/463.8 0.271	153.4/153.4/1053 6.45	315.8/315.8/2115 207
Alg. 3($\mathscr{T}^{\mathrm{DR}}$)	210.4/132/1028 0.636	458.8/321.2/2203 13.9	962.6/641.2/4506 482
Alg. 3($\mathscr{T}^{\mathrm{Pr}}$)	176.2/109/761.6 0.514	695.4/479.4/3088 15.6	1316/912.2/5913 509

Table 5: Test results for $\beta = 10^{-4}$

of complicated equilibrium problems. It can be used, e.g., for the computation of Nash equilibria in case of nonsmooth (and even nonconvex) objectives of the players. Further, in combination with some splitting algorithms, it exhibits remarkable (global and local) convergence properties when applied to rather complicated family of monotone variational inequalities of the second kind.

Acknowledgements. The research of the first author was supported by the Austrian Science Fund (FWF) under grant P29190-N32. The research of the third author was supported by the Grant Agency of the Czech Republic and the Austrian Science Fund, project GACR-FWF 19-29646L.

References

[1] J.-P. AUBIN, *Optima and Equilibria*, Springer, Berlin, 1998.

[2] G. H.-G. CHEN, R. T. ROCKAFELLAR, *Convergence rates in forward-backward splitting*, SIAM J. Optim. 7 (1997), pp. 421-444.

[3] F. FACCHINEI, J.-S. PANG, *Finite-Dimensional Variational Inequalities and Complementarity Problems*, vol. I+II, Springer, New York, 2003.

[4] S.D. FLÅM, *Games and cost of change*, Ann. Oper. Res. (2020). `https://doi.org/10.1007/s10479-020-03585-w`.

[5] D. GABAY, *Applications of the method of multipliers to variational inequalities*, in Augmented Lagrangian Methods: Applications to the Solution of Boundary Value Problems, M. Fortin and R. Glowinski, eds., North-Holland, Amsterdam, 1983.

[6] H. GFRERER, *On directional metric regularity, subregularity and optimality conditions for non-smooth mathematical programs*, Set-Valued Var. Anal., 21 (2013), 151–176.

[7] H. GFRERER, J.V. OUTRATA, *On Lipschitzian properties of implicit multifunctions*, SIAM J. Optim., 26 (2016), pp. 2160–2189.

[8] H. GFRERER, J. V. OUTRATA, *On a semismooth* Newton method for solving generalized equations*, submitted, arXiv:1904.09167.

[9] R. GLOWINSKI, *Numerical Methods for Nonlinear Variational Problems,* Springer, New York, 1984.

[10] J. HASLINGER, M. MIETTINEN, P. D. PANAGIOTOPOULOS (eds.), *Finite Element Method for Hemivariational Inequalities. Theory, Methods and Applications*, Kluwer, Dordrecht, 1999.

[11] K. ITO, K. KUNISCH, *On a semi-smooth Newton method and its globalization*, Math. Program., 118 (2009), pp. 347–370.

[12] C. KANZOW, A. SCHWARTZ, *Spieltheorie*, Springer Nature, Cham, 2018.

[13] P. L. LIONS, B. MERCER, *Splitting algorithms for the sum of two monotone operators*, SIAM J. Numer. Anal. 16(1979), pp. 964–979.

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

[15] B. S. MORDUKHOVICH, *Variational Analysis and Applications*, Springer, Cham (2018)

[16] F. H. MURPHY, H. D. SHERALI, A. L. SOYSTER, *A mathematical programming qpproach for determining oligopolistic market equilibrium*, Math. Program., 24 (1982), pp. 2–106.

[17] J. V. OUTRATA, M. KOČVARA, J. ZOWE, *Nonsmooth Approach to Optimization Problems with Equilibrium Constraints*, Kluwer Academic Publishers, Dordrecht, 1998.

[18] J. V. OUTRATA, J. VALDMAN, *On computation of optimal strategies in oligopolistic markets respecting the cost of change*, accepted in Math. Meth. Oper. Res.

[19] L. QI, J. SUN, *A nonsmooth version of Newton's method*, Math. Program., 58 (1993), pp. 353–367.

[20] S. M. ROBINSON, *A point-of-attraction result for Newton's method with point-based approximations*, Optimization, 60 (2011), pp. 89-99.

[21] R. T. ROCKAFELLAR, R. J.-B. WETS , *Variational Analysis*, Springer, Berlin, 1998.

[22] M. V. SOLODOV, B. F. SVAITER, *A hybrid projection–proximal point algorithm*, J. Conv. Anal. 6(1999), pp. 59–70.

[23] X. XIAO, Y. LI, Z. WEN, L. ZHANG, *A Regularized Semi-Smooth Newton Method with Projection Steps for Composite Convex Programs*, J. Sci. Comput., 76 (2018), pp. 364–389.

Latest Reports in this series

2009 - 2018

[..]

2019

2019-01	Helmut Gfrerer and Jiří V. Outrata *On a Semismooth* Newton Method for Solving Generalized Equations*	April 2019
2019-02	Matúš Benko, Michal Červinka and Tim Hoheisel *New Verifiable Sufficient Conditions for Metric Subregularity of Constraint Systems with Applications to Disjunctive Programs*	June 2019
2019-03	Matúš Benko *On Inner Calmness*, Generalized Calculus, and Derivatives of the Normal-cone Map*	October 2019
2019-04	Rainer Schneckenleitner and Stefan Takacs *Condition number bounds for IETI-DP methods that are explicit in h and p*	December 2019
2019-05	Clemens Hofreither, Ludwig Mitter and Hendrik Speleers *Local multigrid solvers for adaptive Isogeometric Analysis in hierarchical spline spaces*	December 2019

2020

2020-01	Ioannis Toulopoulos *Viscoplastic Models and Finite Element Schemes for the Hot Rolling Metal Process*	February 2020
2020-02	Rainer Schneckenleitner and Stefan Takacs *Convergence Theory for IETI-DP Solvers for Discontinuous Galerkin Isogeometric Analysis That Is Explicit in h and p*	May 2020
2020-03	Svetoslav Nakov and Ioannis Toulopoulos *Convergence Estimates of Finite Elements for a Class of Quasilinear Elliptic Problems*	May 2020
2020-04	Helmut Gfrerer, Jiří V. Outrata and Jan Valdman *On the Application of the Semismooth* Newton Method to Variational Inequalities of the Second Kind*	July 2020

From 1998 to 2008 reports were published by SFB013. Please see
`http://www.sfb013.uni-linz.ac.at/index.php?id=reports`
From 2004 on reports were also published by RICAM. Please see
`http://www.ricam.oeaw.ac.at/publications/`
For a complete list of NuMa reports see
`http://www.numa.uni-linz.ac.at/Publications/List/`