SIAM J. OPTIM.
Vol. 31, No. 1, pp. 489–517




# ON A SEMISMOOTH* NEWTON METHOD FOR SOLVING GENERALIZED EQUATIONS*

HELMUT GFRERER† AND JIŘÍ V. OUTRATA‡

**Abstract.** In the paper, a Newton-type method for the solution of generalized equations (GEs) is derived, where the linearization concerns both the single-valued and the multivalued part of the considered GE. The method is based on the new notion of semismoothness*, which, together with a suitable regularity condition, ensures the local superlinear convergence. An implementable version of the new method is derived for a class of GEs, frequently arising in optimization and equilibrium models.

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

**AMS subject classifications.** 65K10, 65K15, 90C33

**DOI.** 10.1137/19M1257408

## 1. Introduction.

Starting in the seventies, we observed a considerable number of works devoted to the solution of generalized and nonsmooth equations via a Newton-type method (cf., e.g., the papers [18] and [22], the monographs [21] and [17], and the references therein). Concerning generalized equations (GEs), first results can be found in the papers of Josephy [19], [20]. The idea consists of the linearization of the single-valued part of the GE so that in the Newton step one solves typically an affine variational inequality or a linear complementarity problem. Other Newton-like schemes for the solution of GEs or inclusions with general multifunctions can be found, e.g., in [1], [3], or [22].

Concerning the solution of nonsmooth equations, various ideas have been developed starting with a pioneering paper by Kummer [24]. Let us mention at least the papers [25], [31], [33], and [16]. In [24] one finds a general approximation procedure for the Newton step, which is then specialized to continuous selection functions and locally Lipschitzian mappings. In these specializations and also in the above cited papers one makes use of the Clarke generalized Jacobians and directional/graphical derivatives. The approach via Clarke generalized Jacobians is closely related to the notion of semismoothness introduced by Mifflin [27] for real-valued functions. Later, this notion was extended to vector-valued mappings [32] and gave rise to a family of semismooth Newton-type methods which are based on the conceptual scheme from [24] and tailored to various types of nonsmooth equations.

In connection with the radius of metric subregularity analyzed in a recent paper [4] the authors made use of a special condition relating, for a given multifunction, the variables and the directions arising in the respective *directional limiting coderivative*

*Received by the editors April 19, 2019; accepted for publication (in revised form) October 28, 2020; published electronically February 2, 2021.
https://doi.org/10.1137/19M1257408
**Funding:** The research of the first author was supported by the Austrian Science Fund (FWF) under grant P29190-N32. The research of the second author was supported by the Grant Agency of the Czech Republic, project 17-04301S, and the Australian Research Council, project DP160100854.
†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, Ballarat, Vic 3350, Australia (outrata@utia.cas.cz).

[9], [10]. This condition defines a property of sets and mappings which amounts, when specified to Lipschitzian functions, to a slightly less restrictive version of the semismoothness property from [32]. That is why we call it semismoothness*. In fact, in the case of Lipschitzian vector-valued mappings this property already appeared (under a different name) in the literature. For example, it is equivalent to assumption (H2) in [16].

The semismoothness* enables us to construct a new Newton-type method for GEs which is substantially different from the methods mentioned above. In contrast to [19], [20], namely, also the multivalued part of the GE is approximated and our approximation is different from those which are used in [1] and [3]. Indeed, for instance, in [1] and [3] the authors work with appropriate selections of the graphical derivative of the considered multifunction, whereas our approximation is constructed on the basis of a finite number of points from the graph of the limiting (Mordukhovich) coderivative. This has the advantage that, in concrete situations, the rather rich calculus of limiting coderivatives can be employed and the Newton step reduces to the solution of a linear system. The used approximations for the set-valued mapping have further to be sufficiently accurate. In [1] this is, for instance, ensured by the concept of strict lower differentiability whereas we use the semismoothness* property.

Finally, in comparison with the general Newton-like scheme in [22], we provide here a precise description of the iteration process.

The outline of the paper is as follows. In the preliminary section 2 one finds the necessary background from variational analysis together with some useful auxiliary results. In section 3 we introduce the semismooth$^*$ sets and mappings, characterize them in terms of standard (regular and limiting) coderivatives, and investigate thoroughly their relationship to semismooth sets from [15] and the semismooth vector-valued mappings introduced in [32]. Moreover, in this section also some basic classes of semismooth$^*$ sets and mappings are presented. The main results are collected in sections 4 and 5. In particular, section 4 contains the basic conceptual version of the new method suggested for the numerical solution of the general inclusion

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

where $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$. In this version the "linearization" in the Newton step is performed on the basis of the limiting coderivative of $F$. In many situations of practical importance, however, $F$ is not semismooth$^*$ at the solution. Nevertheless, on the basis of a modified regular coderivative it is often possible to construct a modification of the limiting coderivative, with respect to which $F$ is semismooth$^*$ in a generalized sense. This enables us to suggest a generalized version of the new method which exhibits essentially the same convergence properties as the basic one.

Both the basic as well as the generalized version include the so-called approximation step in which one computes an approximative projection of the outcome from the Newton step onto the graph of $F$. This is a big difference with respect to most Newton-type methods in the literature, except, e.g., [1] and the modification of the (inexact) Josephy–Newton method made in [8], which are related to our approximation step.

The algorithms presented in section 4 are rather general and can be considered as a template for an actual implementation. Hence, in section 5 we apply the generalized variant to a frequently arising GE, where $F$ amounts to the sum of a smooth mapping and the normal-cone mapping related to a constraint system. A suitable modification of the regular coderivative is found and it is shown that F is semismooth$^*$ with respect to the respective modification of the limiting coderivative. Finally, we derive implementable procedures both for the approximation as well as for the Newton step. As a result one thus obtains a locally superlinearly convergent Newton-type method for a class of GEs without assuming the metric regularity of F. As shown by a simple example, the method of Josephy may not be always applicable to this class of problems because the linearized problems need not have a solution.

Our notation is standard. Given a linear space $\mathcal{L}$, $\mathcal{L}^\perp$ denotes its orthogonal complement and, for a closed cone $K$ with vertex at the origin, $K^\circ$ signifies its (negative) polar. $\mathcal{S}_{\mathbb{R}^n}$ stands for the unit sphere in $\mathbb{R}^n$ and $\mathcal{B}_\delta(x)$ denotes the closed ball around $x$ with radius $\delta$. Further, given a multifunction $F$, $\operatorname{gph} F := \{(x,y) \mid y \in F(x)\}$ stands for its graph. For an element $u \in \mathbb{R}^n$, $\|u\|$ denotes its Euclidean norm and $[u]$ is the linear space generated by $u$. In a product space we use the norm $\|(u,v)\| := \sqrt{\|u\|^2 + \|v\|^2}$. Given a matrix $A$, we employ the operator norm $\|A\|$ with respect to the Euclidean norm and the Frobenius norm $\|A\|_F$. $Id_s$ is the identity matrix in $\mathbb{R}^s$. Sometimes we write only $Id$. Given a set $\Omega \subset \mathbb{R}^s$, we define the distance of a point $x$ to $\Omega$ by $d_\Omega(x) := \operatorname{dist}(x,\Omega) := \inf\{\|y-x\| \mid y \in \Omega\}$.

## 2. Preliminaries.

Throughout the whole paper, we will make extensive use of 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 *$d$.*

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 $d \notin T_A(\bar{x})$.

In what follows, we will also employ the so-called critical cone. In the setting of Definition 2.1 with a given normal $d^* \in \widehat{N}_A(\bar{x})$, the cone

$$\mathcal{K}_A(\bar{x},d^*) := T_A(\bar{x}) \cap [d^*]^\perp$$

is called the *critical cone* to $A$ at $\bar{x}$ with respect to $d^*$.

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 $DF(\bar{x},\bar{y}) : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$, defined by*

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

*is called the* graphical derivative *of $F$ at $(\bar{x},\bar{y})$.*

(ii) *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 $(\bar{x},\bar{y})$.*

(iii) *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* $(\bar{x},\bar{y})$.

(iv) *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), (ii), and (iii) from Definition 2.2 we refer the interested reader to the monographs [35] and [28]. The directional limiting normal cone and coderivative were introduced by the first author in [9] and various properties of these objects can be found also in [13] 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 $v \notin DF(\bar{x},\bar{y})(u)$.

If $F$ is single-valued, $\bar{y} = F(\bar{x})$ and we write simply $DF(\bar{x})$, $\widehat{D}^*F(\bar{x})$, and $D^*F(\bar{x})$. If $F$ is Fréchet differentiable at $\bar{x}$, then

$$\widehat{D}^*F(\bar{x})(v^*) = \{\nabla F(\bar{x})^T v^*\}, \tag{2.1}$$

and if $F$ is even strictly differentiable at $\bar{x}$, then $D^*F(\bar{x})(v^*) = \{\nabla F(\bar{x})^T v^*\}$.

If a single-valued mapping $F$ is Lipschitzian near $\bar{x}$, denote by $\Omega_F$ the set

$$\Omega_F := \{x \in \mathbb{R}^n \mid F \text{ is differentiable at } x\}.$$

The set

$$\overline{\nabla} F(\bar{x}) := \{A \in \mathbb{R}^{m\times n} \mid \exists (u_k) \xrightarrow{\Omega_F} \bar{x} \text{ such that } \nabla F(u_k) \to A\}$$

is called the *B-subdifferential* of $F$ at $\bar{x}$. The *Clarke generalized Jacobian* of $F$ at $\bar{x}$ amounts then to $\operatorname{conv} \overline{\nabla} F(\bar{x})$. One can prove (see, e.g., [35, Theorem 9.62]) that

$$\operatorname{conv} D^*F(\bar{x})(v^*) = \{A^T v^* \mid A \in \operatorname{conv} \overline{\nabla} F(\bar{x})\}. \tag{2.2}$$

By the definition of $\overline{\nabla} F(\bar{x})$ and (2.1) we readily obtain

$$\{A^T v^* \mid A \in \overline{\nabla} F(\bar{x})\} \subseteq D^*F(\bar{x})(v^*).$$

The following iteration scheme, which goes back to Kummer [24], is an attempt for solving the nonlinear system $F(x) = 0$, where $F : \mathbb{R}^n \to \mathbb{R}^n$ is assumed to be locally Lipschitzian.

Algorithm 1 (Newton-type method for nonsmooth systems).

1. *Choose a starting point* $x^{(0)}$*; set the iteration counter* $k := 0$.
2. *Choose* $A^{(k)} \in \operatorname{conv} \overline{\nabla} F(x^{(k)})$ *and compute the new iterate* $x^{(k+1)} = x^{(k)} - {A^{(k)}}^{-1} F(x^{(k)})$.
3. *Set* $k := k+1$ *and go to* 2.

In order to ensure the locally superlinear convergence of this algorithm to a zero $\bar{x}$ one has to impose some assumptions. First, all the matrices ${A^{(k)}}^{-1}$ should be uniformly bounded, which can be ensured by the assumption that all matrices $A \in \operatorname{conv} \overline{\nabla} F(\bar{x})$ are nonsingular. Second, we need an estimate of the form

$$0 = F(\bar{x}) = F(x^{(k)}) + A^{(k)}(\bar{x} - x^{(k)}) + o(\|\bar{x} - x^{(k)}\|).$$

A popular tool for how the validity of this estimate could be ensured is the notion of semismoothness [27], [32].

Definition 2.3. *Let $U \subseteq \mathbb{R}^n$ be nonempty and open. A function $F : U \to \mathbb{R}^m$ is* semismooth *at $\bar{x} \in U$ if it is Lipschitz near $\bar{x}$ and if*

$$\lim_{\substack{A \in \operatorname{conv} \overline{\nabla} F(\bar{x}+tu') \\ u' \to u,\ t \downarrow 0}} Au'$$

*exists for all $u \in \mathbb{R}^n$. If $F$ is semismooth at all $\bar{x} \in U$, we call $F$* semismooth on $U$.

Given a closed convex cone $K \subset \mathbb{R}^n$ with vertex at the origin, then

$$\operatorname{lin} K := K \cap (-K)$$

denotes the *lineality* space of $K$, i.e., the largest linear space contained in $K$. Denoting by $\operatorname{span} K$ the linear space spanned by $K$, it holds that

$$\operatorname{span} K = K + (-K),\ (\operatorname{lin} K)^{\perp} = \operatorname{span} K^{\circ},\ (\operatorname{span} K)^{\perp} = \operatorname{lin} K^{\circ}.$$

A subset $C'$ of a convex set $C \subset \mathbb{R}^n$ is called a *face* of $C$ if it is convex and if for each line segment $[x, y] \subseteq C$ with $(x, y) \cap C' \neq \emptyset$ one has $x, y \in C'$. The faces of a polyhedral convex cone $K$ are exactly the sets of the form

$$\mathcal{F} = K \cap [v^*]^{\perp} \text{ for some } v^* \in K^{\circ}.$$

Lemma 2.4. *Let $D \subset \mathbb{R}^s$ be a convex polyhedral set. For every pair $(d, \lambda) \in \operatorname{gph} N_D$ there holds*

$$\operatorname{lin} T_D(d) = \operatorname{lin} \mathcal{K}_D(d, \lambda) \subseteq \mathcal{K}_D(d, \lambda) \subseteq T_D(d), \tag{2.3}$$

$$N_D(d) \subseteq \mathcal{K}_D(d, \lambda)^{\circ} \subseteq (\operatorname{lin} T_D(d))^{\perp} = \operatorname{span} N_D(d). \tag{2.4}$$

*Furthermore, for every $(\bar{d}, \bar{\lambda}) \in \operatorname{gph} N_D$ there is a neighborhood $U$ of $(\bar{d}, \bar{\lambda})$ such that for every $(d, \lambda) \in \operatorname{gph} N_D \cap U$ there is a face $\mathcal{F}$ of the critical cone $\mathcal{K}_D(\bar{x}, \bar{\lambda})$ such that $\operatorname{lin} T_D(d) = \operatorname{span} \mathcal{F}$ and consequently $\operatorname{span} N_D(d) = (\operatorname{span} \mathcal{F})^{\perp}$.*

*Proof.* For every $w \in \operatorname{lin} T_D(d)$ we have $\pm w \in T_D(d)$ and therefore $\pm\langle \lambda, w \rangle \leq 0$ because of $\lambda \in N_D(d)$. This yields $\langle \lambda, w \rangle = 0$ and consequently

$$\operatorname{lin} T_D(d) \subseteq \mathcal{K}_D(d, \lambda) \subseteq T_D(d)$$

and, by dualizing, (2.4) follows. Since we also have $\mathcal{K}_D(d, \lambda) \subseteq T_D(d)$, we obtain $\operatorname{lin} \mathcal{K}_D(d, \lambda) \subseteq \operatorname{lin} T_D(d) \subseteq \operatorname{lin} \mathcal{K}_D(d, \lambda)$ implying (2.3).

By [6, Lemma 4H.2] there is a neighborhood $U$ of $(\bar{d}, \bar{\lambda})$ such that for every $(d, \lambda) \in \operatorname{gph} N_D \cap U$ there are two faces $\mathcal{F}_2 \subseteq \mathcal{F}_1$ of $\mathcal{K}_D(\bar{d}, \bar{\lambda})$ such that $\mathcal{K}_D(d, \lambda) = \mathcal{F}_1 - \mathcal{F}_2$. We claim that $\operatorname{lin}(\mathcal{F}_1 - \mathcal{F}_2) = \mathcal{F}_2 - \mathcal{F}_2$. The inclusion $\operatorname{lin}(\mathcal{F}_1 - \mathcal{F}_2) \supseteq \mathcal{F}_2 - \mathcal{F}_2$ trivially holds since $\mathcal{F}_2 - \mathcal{F}_2 = \operatorname{span} \mathcal{F}_2$ is a subspace. Now consider $w \in \operatorname{lin}(\mathcal{F}_1 - \mathcal{F}_2) =$

$(\mathcal{F}_1 - \mathcal{F}_2) \cap (\mathcal{F}_2 - \mathcal{F}_1)$. Then there are $u_1, u_2 \in \mathcal{F}_1$ and $v_1, v_2 \in \mathcal{F}_2$ such that $w = u_1 - v_1 = v_2 - u_2$, implying $\frac{1}{2}u_1 + \frac{1}{2}u_2 = \frac{1}{2}(v_1 + v_2)$, i.e., the point $\frac{1}{2}(v_1 + v_2) \in \mathcal{F}_2$ is the midpoint of the line segment connecting $u_1, u_2 \in \mathcal{F}_1 \subseteq \mathcal{K}_D(g(\bar{x}), \bar{\lambda})$. Since $\mathcal{F}_2$ is a face of $\mathcal{K}_D(g(\bar{x}), \bar{\lambda})$, $u_1, u_2 \in \mathcal{F}_2$ follows and thus $w \in \mathcal{F}_2 - \mathcal{F}_2$. Thus our claim holds true and from (2.3) we obtain $\operatorname{lin} T_D(d) = \operatorname{lin} \mathcal{K}_D(d, \lambda) = \mathcal{F}_2 - \mathcal{F}_2 = \operatorname{span} \mathcal{F}_2$. This completes the proof of the lemma. ∎

## 3. On semismooth* sets and mappings.

Definition 3.1.

1. *A set* $A \subseteq \mathbb{R}^s$ *is called* semismooth* *at a point* $\bar{x} \in A$ *if for all* $u \in \mathbb{R}^s$ *it holds that*

$$\langle x^*, u \rangle = 0 \ \forall x^* \in N_A(\bar{x}; u). \tag{3.1}$$

2. *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* $\operatorname{gph} F$ *is semismooth* at* $(\bar{x}, \bar{y})$*, i.e., for all* $(u, v) \in \mathbb{R}^n \times \mathbb{R}^m$ *we have*

$$\langle u^*, u \rangle = \langle v^*, v \rangle \ \forall (v^*, u^*) \in \operatorname{gph} D^* F((\bar{x}, \bar{y}); (u, v)). \tag{3.2}$$

In the above definition the semismooth* sets and mappings have been defined via directional limiting normal cones and coderivatives. In some situations, however, it is convenient to make use of equivalent characterizations in terms of standard (regular and limiting) normal cones and coderivatives, respectively.

Proposition 3.2. *Let* $A \subset \mathbb{R}^s$ *and* $\bar{x} \in A$ *be given. Then the following three statements are equivalent:*

(i) *A is semismooth* at* $\bar{x}$.

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

$$|\langle x^*, x - \bar{x} \rangle| \leq \epsilon \|x - \bar{x}\| \|x^*\| \ \forall x \in \mathcal{B}_\delta(\bar{x}) \ \forall x^* \in \widehat{N}_A(x). \tag{3.3}$$

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

$$|\langle x^*, x - \bar{x} \rangle| \leq \epsilon \|x - \bar{x}\| \|x^*\| \ \forall x \in \mathcal{B}_\delta(\bar{x}) \ \forall x^* \in N_A(x). \tag{3.4}$$

*Proof.* Assuming that $A$ is not semismooth* at $\bar{x}$, there is $u \neq 0$, $0 \neq u^* \in N_A(\bar{x}; u)$ such that $\epsilon' := |\langle u^*, u \rangle| > 0$. By the definition of directional limiting normals there are sequences $t_k \downarrow 0$, $u_k \to u$, $u_k^* \to u^*$ such that $u_k^* \in \widehat{N}_A(\bar{x} + t_k u_k)$. Then for all $k$ sufficiently large we have $|\langle u_k^*, u_k \rangle| > \epsilon'/2$, implying

$$|\langle u_k^*, (\bar{x} + t_k u_k) - \bar{x} \rangle| > \frac{\epsilon'}{2} t_k = \frac{\epsilon'}{2\|u_k^*\| \|u_k\|} \|(\bar{x} + t_k u_k) - \bar{x}\| \|u_k^*\|.$$

Hence statement (ii) does not hold for $\epsilon = \epsilon' / (4\|u^*\| \|u\|)$ and the implication (ii)⇒(i) is shown.

In order to prove the reverse implication we assume that (ii) does not hold, i.e., there is some $\epsilon > 0$ together with sequences $x_k \to \bar{x}$ and $x_k^*$ such that $x_k^* \in \widehat{N}_A(x_k)$ and

$$|\langle x_k^*, x_k - \bar{x} \rangle| > \epsilon \|x_k - \bar{x}\| \|x_k^*\|$$

holds for all $k$. It follows that $x_k - \bar{x} \neq 0$ and $x_k^* \neq 0$ for all $k$ and, by possibly passing to a subsequence, we can assume that the sequences $(x_k - \bar{x})/\|x_k - \bar{x}\|$ and $x_k^*/\|x_k^*\|$

converge to some $u$ and $u^*$, respectively. Then $u^* \in N_A(\bar{x}; u)$ and

$$|\langle u^*, u\rangle| = \lim_{k\to\infty} \frac{|\langle x_k^*, x_k - \bar{x}\rangle|}{\|x_k - \bar{x}\|\|x_k^*\|} > \epsilon,$$

showing that $A$ is not semismooth* at $\bar{x}$. This proves the implication (i)⇒(ii).

Finally, the equivalence between (ii) and (iii) is an immediate consequence of the definition of limiting normals. □

By simply using Definition 3.1 (part 2) we obtain from Proposition 3.2 the following corollary.

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* $\epsilon > 0$ *there is some* $\delta > 0$ *such that*

$$(3.5)\quad |\langle x^*, x - \bar{x}\rangle - \langle y^*, y - \bar{y}\rangle| \le \epsilon\|(x, y) - (\bar{x}, \bar{y})\|\|(x^*, y^*)\|\ \forall (x, y) \in \mathcal{B}_\delta(\bar{x}, \bar{y})\ \forall (y^*, x^*) \in \operatorname{gph} \widehat{D}^* F(x, y).$$

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

$$(3.6)\quad |\langle x^*, x - \bar{x}\rangle - \langle y^*, y - \bar{y}\rangle| \le \epsilon\|(x, y) - (\bar{x}, \bar{y})\|\|(x^*, y^*)\|\ \forall (x, y) \in \mathcal{B}_\delta(\bar{x}, \bar{y})\ \forall (y^*, x^*) \in \operatorname{gph} D^* F(x, y).$$

On the basis of Definition 3.1, Proposition 3.2, and Corollary 3.3 we may now specify some fundamental classes of semismooth* sets and mappings.

PROPOSITION 3.4. *Let* $A \subset \mathbb{R}^s$ *be a closed convex set. Then* $A$ *is semismooth** *at each* $\bar{x} \in A$.

*Proof.* Since $N_A(\bar{x}; u) = \{x^* \in N_A(\bar{x}) | \langle x^*, u\rangle = 0\}$ by virtue of [10, Lemma 2.1], the statement follows immediately from the definition. □

PROPOSITION 3.5. *Assume that we are given closed sets* $A_i \subset \mathbb{R}^s$, $i = 1, \dots p$, *and* $\bar{x} \in A := \bigcup_{i=1}^p A_i$. *If the sets* $A_i$, $i \in \bar{I} := \{j \mid \bar{x} \in A_j\}$, *are semismooth** *at* $\bar{x}$, *then so is the set* $A$.

*Proof.* Fix any $\epsilon > 0$ and choose according to Proposition 3.2 $\delta_i > 0$, $i \in \bar{I}$, such that for every $i \in \bar{I}$, every $x \in \mathcal{B}_{\delta_i}(\bar{x})$, and every $x^* \in \widehat{N}_{A_i}(x)$ there holds

$$|\langle x^*, x - \bar{x}\rangle| \le \epsilon\|x^*\|\|x - \bar{x}\|.$$

Since the sets $A_i$, $i = 1, \dots, p$, are assumed to be closed, there is some $0 < \delta \le \min\{\delta_i \mid i \in \bar{I}\}$ such that

$$I(x) := \{j \mid x \in A_j\} \subset \bar{I}\ \forall x \in \mathcal{B}_\delta(\bar{x}).$$

Using the identity $\widehat{N}_A(x) = \bigcap_{i\in I(x)} \widehat{N}_{A_i}(x)$ valid for every $x \in A$ it follows that (3.3) holds. Thus the assertion follows from Proposition 3.2. □

Thus, in particular, the union of finitely many closed convex sets is semismooth* at every point. We obtain the following:

1. A closed convex multifunction $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is semismooth* at every point $(\bar{x}, \bar{y}) \in \operatorname{gph} F$.

2. A polyhedral multifunction[1] $F:\mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is semismooth$^*$ at every point $(\bar{x},\bar{y}) \in \operatorname{gph} F$. In particular, for every convex polyhedral set $D \subset \mathbb{R}^s$ the normal cone mapping $N_D$ is semismooth$^*$ at every point of its graph.

Since the semismoothness$^*$ of mappings is defined via the graph, it follows from Corollary 3.3 that $F:\mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is semismooth$^*$ at $(\bar{x},\bar{y}) \in \operatorname{gph} F$ if and only if $F^{-1}:\mathbb{R}^m \rightrightarrows \mathbb{R}^n$ is semismooth$^*$ at $(\bar{y},\bar{x})$. Indeed, the relation (3.6) can be rewritten

$$|\langle y^*, y-\bar{y}\rangle - \langle x^*, x-\bar{x}\rangle| \le \varepsilon \|(y,x)-(\bar{y},\bar{x})\| \; \|(y^*,x^*)\| \forall (y,x) \in \mathcal{B}_\delta(\bar{y},\bar{x})$$
$$\forall (-x^*,-y^*) \in \operatorname{gph} D^*F^{-1}(y,x),$$

which is, in turn, equivalent to the semismoothness$^*$ of $F^{-1}$ at $(\bar{y},\bar{x})$.

In some cases of practical importance one has

$$F(x) = f(x) + Q(x),$$

where $f:\mathbb{R}^n \to \mathbb{R}^n$ is continuously differentiable and $Q:\mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is a closed-graph multifunction.

PROPOSITION 3.6. *Let $\bar{y} \in F(\bar{x})$ and $Q$ be semismooth$^*$ at $(\bar{x},\bar{y}-f(\bar{x}))$. Then $F$ is semismooth$^*$ at $(\bar{x},\bar{y})$.*

*Proof.* Let $(u,v)$ be an arbitrary pair of directions and $u^* \in D^*F((\bar{x},\bar{y});(u,v))(v^*)$. By virtue of [13, formula (2.4)] it holds that

$$D^*F((\bar{x},\bar{y});(u,v))(v^*) = \nabla f(\bar{x})^T v^* + D^*Q((\bar{x},\bar{y}-f(\bar{x}));(u,v-\nabla f(\bar{x})u))(v^*).$$

Thus, $\langle u^*,u\rangle = \langle \nabla f(\bar{x})^T v^* + z^*, u\rangle$ with some $z^* \in D^*Q((\bar{x},\bar{z});(u,w))(v^*)$, where $\bar{z} = \bar{y} - f(x)$ and $w = v - \nabla f(\bar{x})u$. It follows that

$$\langle u^*,u\rangle = \langle v^*, \nabla f(\bar{x})u\rangle + \langle z^*,u\rangle = \langle v^*, \nabla f(\bar{x})u\rangle + \langle v^*,w\rangle$$

due to the assumed semismoothness$^*$ of $Q$ at $(\bar{x},\bar{y}-f(\bar{x}))$. We conclude that $\langle u^*,u\rangle = \langle v^*,v\rangle$ and the proof is complete. ∎

From this statement and the previous development we easily deduce that the solution map $S: y \mapsto x$, related to the canonically perturbed GE

$$y \in f(x) + N_\Gamma(x),$$

is semismooth$^*$ at any $(\bar{y},\bar{x}) \in \operatorname{gph} S$ provided $\Gamma$ is convex polyhedral. Results of this sort in terms of the standard semismoothness property can be found, e.g., in [30, Theorems 6.20 and 6.21].

Let us now figure out the relationship of semismoothness$^*$ and the classical semismoothness in case of single-valued mappings (Definition 2.3). To this purpose note that for a continuous single-valued mapping $F:\mathbb{R}^n \to \mathbb{R}^m$ condition (3.6) is equivalent (with a possibly different $\delta$) to the requirement

$$(3.7)\; |\langle x^*, x-\bar{x}\rangle - \langle y^*, F(x)-F(\bar{x})\rangle|$$
$$\le \epsilon \|(x,F(x)) - (\bar{x},F(\bar{x}))\| \|(x^*,y^*)\| \;\forall x \in \mathcal{B}_\delta(\bar{x}) \;\forall (y^*,x^*) \in \operatorname{gph} D^*F(x).$$

PROPOSITION 3.7. *Assume that $F:\mathbb{R}^n \to \mathbb{R}^m$ is a single-valued mapping which is Lipschitzian near $\bar{x}$. Then the following two statements are equivalent:*

[1] A mapping whose graph is the union of finitely many convex polyhedral sets.

(i) *$F$ is semismooth$^*$ at $\bar{x}$.*
(ii) *For every $\epsilon > 0$ there is some $\delta > 0$ such that*

$$\|F(x) - F(\bar{x}) - C(x - \bar{x})\| \le \epsilon\|x - \bar{x}\| \ \forall x \in \mathcal{B}_\delta(\bar{x}) \ \forall C \in \operatorname{conv} \overline{\nabla} F(x). \tag{3.8}$$

*Proof.* Let $L$ denote the modulus of Lipschitz continuity of $F$ in some neighborhood of $\bar{x}$. In order to show the implication (i)⇒(ii), fix any $\epsilon' > 0$ and choose $\delta > 0$ such that (3.7) holds with $\epsilon = \epsilon'/(1 + L^2)$. Consider $x \in \mathcal{B}_\delta(\bar{x})$, $C \in \operatorname{conv} \overline{\nabla} F(x)$ and choose $y^* \in \mathcal{S}_{\mathbb{R}^m}$ with

$$\|F(x) - F(\bar{x}) - C(x - \bar{x})\| = \langle y^*, F(x) - F(\bar{x}) - C(x - \bar{x})\rangle.$$

By (2.2) there holds $C^T y^* \in \operatorname{conv} D^*F(x)(y^*)$ and therefore, by the Carathéodory theorem, there are elements $x_i^* \in D^*F(x)(y^*)$ and scalars $\alpha_i \ge 0$, $i = 1, \ldots, N$, with $\sum_{i=1}^N \alpha_i = 1$ and $C^T y^* = \sum_{i=1}^N \alpha_i x_i^*$. It follows from (3.7) that

$$\begin{aligned}
&\|F(x) - F(\bar{x}) - C(x - \bar{x})\| \\
&\quad = \langle y^*, F(x) - F(\bar{x}) - C(x - \bar{x})\rangle = \langle y^*, F(x) - F(\bar{x})\rangle - \langle C^T y^*, x - \bar{x}\rangle \\
&\quad = \sum_{i=1}^N \alpha_i \big(\langle y^*, F(x) - F(\bar{x})\rangle - \langle x_i^*, x - \bar{x}\rangle\big) \\
&\quad \le \sum_{i=1}^N \alpha_i \epsilon \|(x, F(x)) - (\bar{x}, F(\bar{x}))\| \|(x_i^*, y^*)\| \le \epsilon(1 + L^2)\|x - \bar{x}\| \\
&\quad = \epsilon' \|x - \bar{x}\|,
\end{aligned}$$

where we have taken into account $\|x_i^*\| \le L\|y^*\| = L$ and $\|F(x) - F(\bar{x})\| \le L\|x - \bar{x}\|$. This inequality justifies (3.8) and the implication (i)⇒(ii) is verified.

Now let us show the reverse implication. Let $\epsilon > 0$ and choose $\delta > 0$ such that (3.8) holds. Consider $x \in \mathcal{B}_\delta(\bar{x})$ and $(y^*, x^*) \in \operatorname{gph} D^*F(x)$. Then by (2.2) there is some $C \in \operatorname{conv} \overline{\nabla} F(x)$ such that $x^* \in C^T y^*$ and we obtain

$$\begin{aligned}
&|\langle x^*, x - \bar{x}\rangle - \langle y^*, F(x) - F(\bar{x})\rangle| \\
&\quad = \langle y^*, C(x - \bar{x}) - (F(x) - F(\bar{x}))\rangle \le \|y^*\| \|F(x) - F(\bar{x}) - C(x - \bar{x})\| \\
&\quad \le \|y^*\| \epsilon \|x - \bar{x}\| \le \epsilon \|(x, F(x)) - (\bar{x}, F(\bar{x}))\| \|(x^*, y^*)\|.
\end{aligned}$$

Thus the implication (ii)⇒(i) is established and the proposition is shown. ∎

Condition (ii) of Proposition 3.7 can be equivalently written in the form that, for any $C \in \operatorname{conv} \overline{\nabla} F(\bar{x} + d)$,

$$\|F(\bar{x} + d) - F(\bar{x}) - Cd\| = o(\|d\|) \text{ as } d \to 0. \tag{3.9}$$

In the terminology of [22, Definition 2] this condition states that the mapping $x \rightrightarrows \operatorname{conv} \overline{\nabla} F(x)$ is a *Newton map* of $F$ at $\bar{x}$. This is one of the conditions used in [21, Lemma 10.1] for guaranteeing superlinear convergence of a generalized Newton method.

If the directional derivative $F'(\bar{x}; \cdot)$ exists (which is the same as the requirement that the graphical derivative $DF(\bar{x})(\cdot)$ is single-valued), then we have (cf. [36]) that

$$F(\bar{x} + d) - F(\bar{x}) - F'(\bar{x}; d) = o(\|d\|) \text{ as } d \to 0.$$

This relation, together with (3.9) and [32, Theorem 2.3], now leads directly to the following result.

COROLLARY 3.8. *Assume that $F : \mathbb{R}^n \to \mathbb{R}^m$ is a single-valued mapping which is Lipschitzian near $\bar{x}$. Then the following two statements are equivalent:*

(i) *$F$ is semismooth at $\bar{x}$ (Definition 2.3).*

(ii) *$F$ is semismooth$^*$ at $\bar{x}$ and $F'(\bar{x};\cdot)$ exists.*

In [23] one finds a Lipschitzian univariate function illustrating the difference between the semismoothness and the semismoothness*.

In Definition 3.1 we have started with semismoothness$^*$ of sets and extended this property to mappings via their graphs. For the reverse direction we may use the distance function.

PROPOSITION 3.9. *Let $A \subset \mathbb{R}^s$ be closed, $\bar{x} \in A$. Then $A$ is semismooth$^*$ at $\bar{x}$ if and only if the distance function $d_A$ is semismooth$^*$ at $\bar{x}$.*

*Proof.* The distance function $d_A(\cdot)$ is Lipschitzian with constant 1 and

$$\partial d_A(x) = \begin{cases} N_A(x) \cap \mathcal{B}_1(0) & \text{if } x \in A, \\ \frac{x - \Pi_A(x)}{d_A(x)} & \text{otherwise,} \end{cases} \tag{3.10}$$

where $\Pi_A(x) := \{z \in A \mid \|z - x\| = d_A(x)\}$ denotes the projection on $A$ (see, e.g., [28, Theorem 1.33]). Here, $\partial d_A(x)$ denotes the *limiting (Mordukhovich) subdifferential* of the distance function $d_A$ at $x$ (see, e.g., [28, Definition 1.18]). Further, by [35, Theorem 9.61] we have

$$\operatorname{conv} \overline{\nabla} d_A(x) = \operatorname{conv} \partial d_A(x)$$

for all $x$.

We first show the implication "$d_A$ is semismooth$^*$ at $\bar{x} \;\Rightarrow\; A$ is semismooth$^*$ at $\bar{x}$." For every $x \in A$ and every $0 \neq x^* \in N_A(x)$ we have $x^*/\|x^*\| \in \partial d_A(x) \subseteq \operatorname{conv} \overline{\nabla} d_A(x)$. Thus, if $d_A$ is semismooth$^*$ at $\bar{x}$, then it follows from Proposition 3.7 that for every $\epsilon > 0$ there is some $\delta > 0$ such that for every $x \in \mathcal{B}_\delta(\bar{x}) \cap A$ we have

$$|d_A(x) - d_A(\bar{x}) - \left\langle \frac{x^*}{\|x^*\|}, x - \bar{x} \right\rangle| = |\left\langle \frac{x^*}{\|x^*\|}, x - \bar{x} \right\rangle| \le \epsilon \|x - \bar{x}\| \;\forall 0 \neq x^* \in N_A(x).$$

By taking into account that (3.4) trivially holds for $x^* = 0$ and that $N_A(x) = \emptyset$ for $x \in \mathcal{B}_\delta(\bar{x}) \setminus A$, by virtue of Proposition 3.2 the set $A$ is semismooth$^*$ at $\bar{x}$.

In order to show the reverse implication, assume that $A$ is semismooth$^*$ at $\bar{x}$. Fix any $\epsilon > 0$ and choose $\delta > 0$ such that (3.4) holds. We claim that for every $x \in \mathcal{B}_{\delta/2}(\bar{x})$, and every $x^* \in \operatorname{conv} \overline{\nabla} d_A(x)$ there holds

$$|d_A(x) - d_A(\bar{x}) - \langle x^*, x - \bar{x} \rangle| \le 2\epsilon \|x - \bar{x}\|. \tag{3.11}$$

Consider $x \in \mathcal{B}_{\delta/2}(\bar{x})$. We first show the inequality (3.11) for $x^* \in \partial d_A(x)$. Indeed, if $x \in A$, then (3.4) implies

$$\begin{aligned} |\langle x^*, x - \bar{x} \rangle| &= |d_A(x) - d_A(\bar{x}) - \langle x^*, x - \bar{x} \rangle| \\ &\le \epsilon \|x^*\| \|x - \bar{x}\| \le \epsilon \|x - \bar{x}\| \;\forall x^* \in N_A(x) \cap \mathcal{B} = \partial d_A(x). \end{aligned}$$

Otherwise, if $x \notin A$, for every $x^* \in \partial d_A(x)$ there is some $x' \in \Pi_A(x)$ satisfying $x^* = (x - x')/d_A(x)$. The vector $x - x'$ is a so-called proximal normal to $A$ at $x'$ and therefore $x - x' \in \widehat{N}_A(x') \subset N_A(x')$ (see [35, Example 6.16]). From $\|x' - x\| \le \|\bar{x} - x\|$ we obtain $\|x' - \bar{x}\| \le 2\|x - \bar{x}\| \le \delta$ and we may conclude that

$$\begin{aligned} |\langle x - x', x' - \bar{x} \rangle| &= |\langle x - x', x' - x \rangle + \langle x - x', x - \bar{x} \rangle| \\ &= |-d_A(x)^2 + \langle x - x', x - \bar{x} \rangle| \le \epsilon \|x - x'\| \|x' - \bar{x}\| \\ &\le 2\epsilon d_A(x) \|x - \bar{x}\|. \end{aligned}$$

Dividing by $d_A(x)$ we infer

$$|d_A(x) - \left\langle \frac{x-x'}{d_A(x)}, x-\bar{x} \right\rangle| = |d_A(x) - d_A(\bar{x}) - \left\langle \frac{x-x'}{d_A(x)}, x-\bar{x} \right\rangle| \le 2\epsilon\|x-\bar{x}\|,$$

showing that (3.11) holds true in this case as well.

Now consider any $x^* \in \operatorname{conv} \overline{\nabla} d_A(x) = \operatorname{conv} \partial d_A(x)$. By the Carathéodory theorem there are finitely many elements $x_i^* \in \partial d_A(x)$ together with positive scalars $\alpha_i$, $i = 1, \ldots, N$, such that $\sum_{i=1}^N \alpha_i = 1$ and $x^* = \sum_{i=1}^N \alpha_i x_i^*$, implying

$$\begin{aligned}|d_A(x) - d_A(\bar{x}) - \langle x^*, x-\bar{x}\rangle| &= \left| \sum_{i=1}^N \alpha_i \big(d_A(x) - d_A(\bar{x}) - \langle x_i^*, x-\bar{x}\rangle\big)\right| \\ &\le \sum_{i=1}^N \alpha_i |d_A(x) - d_A(\bar{x}) - \langle x_i^*, x-\bar{x}\rangle| \\ &\le \sum_{i=1}^N \alpha_i 2\epsilon\|x-\bar{x}\| = 2\epsilon\|x-\bar{x}\|.\end{aligned}$$

Thus the claimed inequality (3.11) holds for all $x \in \mathcal{B}_{\delta/2}(\bar{x})$ and all $x^* \in \operatorname{conv} \overline{\nabla} d_A(x)$ and from Proposition 3.7 we conclude that $d_A$ is semismooth* at $\bar{x}$. □

*Remark* 3.10. Combining Proposition 3.2 with the formula (3.10) implies that a set $A$ is semismooth* at $\bar{x}$ if and only if for every $\epsilon > 0$ there is some $\delta > 0$ such that

$$\left|\left\langle x^*, \frac{x-\bar{x}}{\|x-\bar{x}\|}\right\rangle\right| \le \epsilon \ \forall x \in A \cap \mathcal{B}_\delta(\bar{x}),\ x \ne \bar{x},\ \forall x^* \in \partial d_A(x).$$

From this relation it follows that a set is semismooth* at $\bar{x}$ if and only if it is semismooth in the sense of [15, Definition 2.3].

## 4. A semismooth* Newton method.

Given a set-valued mapping $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ with closed graph, we want to solve the inclusion

$$0 \in F(x). \tag{4.1}$$

Given $(x, y) \in \operatorname{gph} F$, we denote by $\mathcal{A}F(x, y)$ the collection of all pairs of $n \times n$ matrices $(A, B)$, such that there are $n$ elements $(v_i^*, u_i^*) \in \operatorname{gph} D^*F(x, y)$, $i = 1, \ldots, n$, and the $i$th rows of $A$ and $B$ are ${u_i^*}^T$ and ${v_i^*}^T$, respectively. Further we denote

$$\mathcal{A}_{\mathrm{reg}}F(x, y) := \{(A, B) \in \mathcal{A}F(x, y) \mid A \text{ nonsingular}\}.$$

It turns out that the strong metric regularity of $F$ around $(x, y)$ is a sufficient condition for the nonemptiness of $\mathcal{A}_{\mathrm{reg}}F(x, y)$. Recall that a set-valued mapping $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is *strongly metrically regular* around $(x, y) \in \operatorname{gph} F$ (with modulus $\kappa$) if its inverse $F^{-1}$ has a Lipschitz continuous single-valued localization near $(y, x)$ (with Lipschitz constant $\kappa$) (cf. [28, Definition 5.12]).

THEOREM 4.1. *Assume that $F$ is strongly metrically regular around $(\hat{x}, \hat{y}) \in \operatorname{gph} F$ with modulus $\kappa > 0$. Then there is an $n \times n$ matrix $C$ with $\|C\| \le \kappa$ such that $(Id, C) \in \mathcal{A}_{\mathrm{reg}}F(\hat{x}, \hat{y})$.*

*Proof.* Note that $-y^* \in D^*F^{-1}(\hat{y}, \hat{x})(-x^*)$ if and only if $x^* \in D^*F(\hat{x}, \hat{y})(y^*)$ (cf. [35, equation 8(19)]). Let $s$ denote the single-valued localization of the inverse mapping $F^{-1}$ around $(\hat{y},\hat{x})$ which is Lipschitzian with modulus $\kappa$ near $\hat{y}$. Next take any element $C$ from the B-subdifferential $\overline{\nabla} s(\bar{y})$. Then $\|C\| \le \kappa$ and for any $u^*$ we have $-C^T u^* \in D^*F^{-1}(\hat{y},\hat{x})(-u^*)$ and consequently $u^* \in D^*F(\hat{x},\hat{y})(C^T u^*)$. Taking $u_i^*$ as the $i$th unit vector and $v_i^* = C^T u_i^*$, we obtain that $(Id, C) \in \mathcal{A}_{\rm reg}F(\hat{x},\hat{y})$. ◻

Corollary 4.2. *Let $(\hat{x},\hat{y}) \in \operatorname{gph} F$ and assume that there is $\kappa > 0$ and a sequence $(x_k,y_k) \in \operatorname{gph} F$ converging to $(\hat{x},\hat{y})$ such that for each $k$ the mapping $F$ is strongly metrically regular around $(x_k,y_k)$ with modulus $\kappa$. Then there is an $n\times n$ matrix $C$ with $\|C\| \le \kappa$ such that $(Id, C) \in \mathcal{A}_{\rm reg}F(\hat{x},\hat{y})$.*

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 $\epsilon > 0$ there is some $\delta > 0$ such that for every $(x,y) \in \operatorname{gph} F \cap \mathcal{B}_\delta(\bar{x},0)$ and every pair $(A,B) \in \mathcal{A}_{\rm reg}F(x,y)$ one has*

$$\|(x - A^{-1}By) - \bar{x}\| \le \epsilon \|A^{-1}\| \|(A \,\vdots\, B)\|_F \|(x,y) - (\bar{x},0)\|. \tag{4.2}$$

*Proof.* Let $\epsilon > 0$ be arbitrarily fixed, choose $\delta > 0$ such that (3.6) holds, and consider $(x,y) \in \operatorname{gph} F \cap \mathcal{B}_\delta(\bar{x},0)$ and $(A,B) \in \mathcal{A}_{\rm reg}F(x,y)$. By the definition of $\mathcal{A}F(x,y)$ we obtain that the $i$th component of the vector $A(x-\bar{x}) - By$ equals $\langle u_i^*, x-\bar{x}\rangle - \langle v_i^*, y-0\rangle$ and can be bounded by $\epsilon\|(x,y)-(\bar{x},0)\|\|(u_i^*,v_i^*)\|$ by (3.6). Since the Euclidean norm of the vector with components $\|(u_i^*,v_i^*)\|$ is exactly the Frobenius norm of the matrix $(A \,\vdots\, B)$, we obtain

$$\|A(x-\bar{x}) - By\| \le \epsilon \|(A \,\vdots\, B)\|_F \|(x,y) - (\bar{x},0)\|.$$

By taking into account that

$$\|(x - A^{-1}By) - \bar{x}\| = \|A^{-1}\big(A(x-\bar{x}) - By\big)\| \le \|A^{-1}\|\|A(x-\bar{x}) - By\|,$$

the estimate (4.2) follows. ◻

The Newton method for solving generalized equations is not uniquely defined in general. Given some iterate $x^{(k)}$, we cannot expect in general that $F(x^{(k)}) \ne \emptyset$ or that 0 is close to $F(x^{(k)})$, even if $x^{(k)}$ is close to a solution $\bar{x}$. Thus we first perform some step which yields $(\hat{x}^{(k)},\hat{y}^{(k)}) \in \operatorname{gph} F$ as an approximate projection of $(x^{(k)},0)$ on $\operatorname{gph} F$. Further we require that $\mathcal{A}_{\rm reg}F(\hat{x}^{(k)},\hat{y}^{(k)}) \ne \emptyset$ and compute the new iterate as $x^{(k+1)} = \hat{x}^{(k)} - A^{-1}B\hat{y}^{(k)}$ for some $(A,B) \in \mathcal{A}_{\rm reg}F(\hat{x}^{(k)},\hat{y}^{(k)})$.

Algorithm 2 (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. *Compute $(\hat{x}^{(k)},\hat{y}^{(k)}) \in \operatorname{gph} F$ close to $(x^{(k)},0)$ such that $\mathcal{A}_{\rm reg}F(\hat{x}^{(k)},\hat{y}^{(k)}) \ne \emptyset$.*
4. *Select $(A,B) \in \mathcal{A}_{\rm reg}F(\hat{x}^{(k)},\hat{y}^{(k)})$ 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.*

Of course, the heart of this algorithm is steps 3 and 4. We will call step 3 the *approximation step* and step 4 the *Newton step*.

Before we continue with the analysis of this algorithm let us consider the Newton step for the special case of a single-valued smooth mapping $F : \mathbb{R}^n \to \mathbb{R}^n$. We have $\hat{y}^{(k)} = F(\hat{x}^{(k)})$ and $D^*F(\hat{x}^{(k)})(v^*) = \nabla F(\hat{x}^{(k)})^T v^*$, yielding

$$\mathcal{A}F(\hat{x}^{(k)}, F(\hat{x}^{(k)})) = \{(B\nabla F(x^{(k)}), B) \mid B \text{ is } n\times n \text{ matrix}\}.$$

Thus the requirement $(A,B) \in \mathcal{A}_{\rm reg}F(\hat{x}^{(k)}, F(\hat{x}^{(k)}))$ means that $A = B\nabla F(\hat{x}^{(k)})$ is nonsingular, i.e., both $B$ and $\nabla F(x^{(k)})$ are nonsingular. Then the Newton step amounts to

$$x^{(k+1)} = \hat{x}^{(k)} - (B\nabla F(\hat{x}^{(k)}))^{-1} BF(\hat{x}^{(k)}) = \hat{x}^{(k)} - \nabla F(\hat{x}^{(k)})^{-1} F(\hat{x}^{(k)}).$$

We see that it coincides with the classical Newton step for smooth functions $F$. Note that the requirement that $B$ is nonsingular in order to have

$$(A,B) \in \mathcal{A}_{\rm reg}F(\hat{x}^{(k)}, F(\hat{x}^{(k)}))$$

is possibly not needed for general set-valued mappings $F$ (see (5.18) below).

Next let us consider the case of a single-valued Lipschitzian mapping $F : \mathbb{R}^n \to \mathbb{R}^n$. As before we have $\hat{y}^{(k)} = F(\hat{x}^{(k)})$ and for every $C \in \overline{\nabla} F(\hat{x}^{(k)})$ we have $C^T v^* \in D^*F(\hat{x}^{(k)})(v^*)$. Thus

$$\mathcal{A}F(\hat{x}^{(k)}, F(\hat{x}^{(k)})) \supseteq \bigcup_{C\in \overline{\nabla} F(\hat{x}^{(k)})} \{(BC, B) \mid B \text{ is an } n\times n \text{ matrix}\}. \tag{4.3}$$

Similarly as above we have that $(BC,B) \in \mathcal{A}_{\rm reg}F(\hat{x}^{(k)}, F(\hat{x}^{(k)}))$ if and only if both $B$ and $C$ are nonsingular and in this case the Newton step reads as $x^{(k+1)} = \hat{x}^{(k)} - C^{-1}F(\hat{x}^{(k)})$. Thus the classical semismooth Newton method of [32], restricted to the B-subdifferential $\overline{\nabla} F(\hat{x}^{(k)})$ instead of the generalized Jacobian conv $\overline{\nabla} F(\hat{x}^{(k)})$, fits into the framework of Algorithm 2. However, note that the inclusion (4.3) will be strict whenever $\overline{\nabla} F(\hat{x}^{(k)})$ is not a singleton: for every $u_i^*$, $i = 1,\dots,n$, forming the rows of the matrix $B$ we can take a different $C_i \in \overline{\nabla} F(\hat{x}^{(k)})$, $i = 1,\dots,n$, for generating the rows $C_i^T u_i^*$ of the matrix $A$. When using such a construction it is no longer mandatory to require $B$ nonsingular in order to have $(A,B) \in \mathcal{A}_{\rm reg}F(\hat{x}^{(k)}, F(\hat{x}^{(k)}))$ and thus Algorithm 2 offers a variety of other possibilities for how the Newton step can be performed.

Given two reals $L,\kappa > 0$ and a solution $\bar{x}$ of (4.1), we denote

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

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 $\mathcal{G}^{L,\kappa}_{F,\bar{x}}(x) \ne \emptyset$. Then there exists some $\delta > 0$ such that for every starting point $x^{(0)} \in \mathcal{B}_\delta(\bar{x})$ Algorithm 2 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 $(\hat{x}^{(k)}, \hat{y}^{(k)}, A, B) \in \mathcal{G}^{L,\kappa}_{F,\bar{x}}(x^{(k)})$.*

*Proof.* By Proposition 4.3, we can find some $\bar{\delta} > 0$ such that (4.2) holds with $\epsilon = \frac{1}{2L\kappa}$ for all $(x,y) \in \operatorname{gph} F \cap \mathcal{B}_{\bar{\delta}}(\bar{x},0)$ and all pairs $(A,B) \in \mathcal{A}_{\rm reg}F(x,y)$. Set $\delta := \bar{\delta}/L$ and consider an iterate $x^{(k)} \in \mathcal{B}_\delta(\bar{x}) \notin F^{-1}(0)$. Then

$$\|(\hat{x}^{(k)}, \hat{y}^{(k)}) - (\bar{x},0)\| \le L\|x^{(k)} - \bar{x}\| \le \bar{\delta}$$

and consequently

$$\|x^{(k+1)} - \bar{x}\| \le \frac{1}{2L\kappa}\|A^{-1}\|\|(A\,\vdots\,B)\|_F L\|x^{(k)} - \bar{x}\| \le \frac{1}{2}\|x^{(k)} - \bar{x}\|$$

by Proposition 4.3. It follows that for every starting point $x^{(0)} \in \mathcal{B}_\delta(\bar{x})$ Algorithm

2 either stops after finitely many iterations with a solution or produces a sequence $x^{(k)}$ converging to $\bar{x}$. The superlinear convergence of the sequence $x^{(k)}$ is now an easy consequence of Proposition 4.3. ◻

*Remark* 4.5. The bound $\|(\hat{x}-\bar{x},\hat{y})\| \le L\|x-\bar{x}\|$ is in particular fulfilled if $(\hat{x},\hat{y}) \in \operatorname{gph} F$ satisfies

$$\|(\hat{x}-x,\hat{y})\| \le \beta \operatorname{dist}((x,0),\operatorname{gph} F) \tag{4.4}$$

with some constant $\beta > 0$, because then we have

$$\|(\hat{x}-\bar{x},\hat{y})\| \le \|(\hat{x}-x,\hat{y})\| + \|x-\bar{x}\| \le \beta \operatorname{dist}((x,0),\operatorname{gph} F) + \|x-\bar{x}\| \le (\beta+1)\|x-\bar{x}\|.$$

According to Theorem 4.4, the outcome $(\hat{x}^{(k)},\hat{y}^{(k)}) \in \operatorname{gph} F$ from the approximation step has to fulfill the inequality

$$\|(\hat{x}^{(k)},\hat{y}^{(k)}) - (\bar{x},0)\| \le L\|x^{(k)}-\bar{x}\|. \tag{4.5}$$

This estimate, in view of (4.4), holds true in particular if

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

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

Note that the requirement (4.5) is stronger than the condition $\operatorname{dist}(\hat{x}^{(k)},F^{-1}(0)) \le L\operatorname{dist}(x^{(k)},F^{-1}(0))$ used in the modification of the Josephy–Newton method [8]. The reason is that we perform a linearization of the set-valued mapping $F$ around the point $(\hat{x}^{(k)},\hat{y}^{(k)}) \in \operatorname{gph} F$, whereas in the variant of the Josephy–Newton method from [8] the set-valued part of $F$ is not linearized.

*Remark* 4.6. Note that in the case of a single-valued mapping $F:\mathbb{R}^n \to \mathbb{R}^n$ an approximation step of the form $(\hat{x}^{(k)},\hat{y}^{(k)}) = (x^{(k)},F(x^{(k)}))$ requires $\|(x^{(k)}-\bar{x},F(x^{(k)}))\| \le L\|x^{(k)}-\bar{x}\|$, which is in general only fulfilled if $F$ is calm at $\bar{x}$, i.e., there is a positive real $L'$ such that $\|F(x)-F(\bar{x})\| \le L'\|x-\bar{x}\|$ for all $x$ sufficiently near $\bar{x}$.

THEOREM 4.7. *Assume that the mapping $F$ is both semismooth$^*$ at $(\bar{x},0)$ and strongly metrically regular around $(\bar{x},0)$. Then all assumptions of Theorem 4.4 are fulfilled.*

*Proof.* Let $s$ denote the single-valued Lipschitzian localization of $F^{-1}$ around $(0,\bar{x})$ and let $\kappa$ denote its Lipschitz constant. We claim that for every $\beta \ge 1$ the set $\mathcal{G}_{F,\bar{x}}^{1+\beta,\sqrt{n(1+\kappa^2)}}(x) \neq \emptyset$ for every $x$ sufficiently close to $\bar{x}$. Obviously there is a real $\rho > 0$ such that $s$ is a single-valued localization of $F^{-1}$ around $(\hat{y},\hat{x})$ for every $(\hat{x},\hat{y}) \in \operatorname{gph} F \cap \mathcal{B}_\rho(\bar{x},0)$ and, since $s$ is Lipschitzian with modulus $\kappa$, we obtain that $F$ is strongly metrically regular around $(\hat{x},\hat{y})$ with modulus $\kappa$. Consider now $x \in \mathcal{B}_{\rho'}(\bar{x})$, where $\rho' < \rho/(1+\beta)$ and $(\hat{x},\hat{y}) \in \operatorname{gph} F$ satisfying $\|(\hat{x}-x,\hat{y})\| \le \beta\operatorname{dist}((x,0),\operatorname{gph} F) \le \beta\|x-\bar{x}\|$. Then $\|\hat{x}-\bar{x},\hat{y}-0\| \le \beta\|x-\bar{x}\| + \|(x-\bar{x},0)\| = (1+\beta)\|x-\bar{x}\| < \rho$ and by Theorem 4.1 there is some matric $C$ with $\|C\| \le \kappa$ such that $(Id,C) \in \mathcal{A}_{\rm reg}F(\hat{x},\hat{y})$. Since $\|(Id \vdots C)\|_F^2 = n + \|C\|_F^2 \le n(1+\|C\|^2)$, we obtain $(\hat{x},\hat{y},Id,C) \in \mathcal{G}_{F,\bar{x}}^{1+\beta,\sqrt{n(1+\kappa^2)}}(x) \neq \emptyset$. ◻

To achieve superlinear convergence of the semismooth* Newton method, the conditions of Theorem 4.7 need not be fulfilled. We now introduce a generalization of the concept of semismoothness* which enables us to deal with mappings $F$ that are not semismooth* at $(\bar{x}, 0)$ with respect to the directional limiting coderivative in the sense of Definition 3.1. Our approach is motivated by the characterization of semismoothness* in Corollary 3.3. If $F$ fails to be semismooth* at $(\bar{x}, 0)$, then condition (3.5) does not hold for all $(x, y) \in \operatorname{gph} F \cap \mathcal{B}_\delta(\bar{x}, 0)$ and all elements $(y^*, x^*) \in \operatorname{gph} \widehat{D}^* F(x, y)$. However, we can possibly characterize the elements of the graph of the regular coderivative for which (3.5) holds true. If we use only those coderivatives in the algorithm, then we can again achieve superlinear convergence. Further, there is no reason to restrict ourselves to (regular) coderivatives; we can possibly use other objects which are easier to compute. Namely, the computation of coderivatives is in general a nontrivial task and in many cases we only have some inclusions at our disposal. However, we can use the elements from the right-hand side of this inclusion without hesitation as long as condition (3.5) is fulfilled.

In order to formalize these ideas we introduce the mapping $\widehat{\mathcal{D}}^* F : \operatorname{gph} F \to (\mathbb{R}^n \rightrightarrows \mathbb{R}^n)$ having the property that for every pair $(x, y) \in \operatorname{gph} F$ the set $\operatorname{gph} \widehat{\mathcal{D}}^* F(x, y)$ is a cone. Further we define the associated limiting mapping $\mathcal{D}^* F : \operatorname{gph} F \to (\mathbb{R}^n \rightrightarrows \mathbb{R}^n)$ via

$$\operatorname{gph} \mathcal{D}^* F(x, y) = \limsup_{(x', y') \overset{\operatorname{gph} F}{\longrightarrow} (x, y)} \operatorname{gph} \widehat{\mathcal{D}}^* F(x', y').$$

Definition 4.8. *The mapping $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is called semismooth* at $(\bar{x}, \bar{y}) \in \operatorname{gph} F$ with respect to $\mathcal{D}^* F$ if for every $\epsilon > 0$ there is some $\delta > 0$ such that*

$$(4.6)\quad |\langle x^*, x - \bar{x}\rangle - \langle y^*, y - \bar{y}\rangle| \\ \le \epsilon \|(x, y) - (\bar{x}, \bar{y})\| \|(x^*, y^*)\| \ \forall (x, y) \in \mathcal{B}_\delta(\bar{x}, \bar{y}) \ \forall (y^*, x^*) \in \operatorname{gph} \widehat{\mathcal{D}}^* F(x, y).$$

For an example of such a mapping $\widehat{\mathcal{D}}^* F$ we refer the reader to the next section.

Given $(x, y) \in \operatorname{gph} F$, we denote by $\mathcal{A}^{\mathcal{D}^*} F(x, y)$ the collection of all pairs of $n \times n$ matrices $(A, B)$, such that there are $n$ elements $(v_i^*, u_i^*) \in \operatorname{gph} \mathcal{D}^* F(x, y)$, $i = 1, \ldots, n$, and the $i$th row of $A$ and $B$ are ${u_i^*}^T$ and ${v_i^*}^T$, respectively. Further we denote

$$\mathcal{A}^{\mathcal{D}^*}_{\text{reg}} F(x, y) := \{(A, B) \in \mathcal{A}^{\mathcal{D}^*} F(x, y) \mid A \text{ nonsingular}\}.$$

Now we can generalize the previous results by replacing $\mathcal{A}_{\text{reg}} F$ by $\mathcal{A}^{\mathcal{D}^*}_{\text{reg}} F$.

Algorithm 3 (generalized semismooth* Newton-like 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. *Compute $(\hat{x}^{(k)}, \hat{y}^{(k)}) \in \operatorname{gph} F$ close to $(x^{(k)}, 0)$ such that $\mathcal{A}^{\mathcal{D}^*}_{\text{reg}} F(\hat{x}^{(k)}, \hat{y}^{(k)}) \neq \emptyset$.*
4. *Select $(A, B) \in \mathcal{A}^{\mathcal{D}^*}_{\text{reg}} F(\hat{x}^{(k)}, \hat{y}^{(k)})$ 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.*

Given two reals $L, \kappa > 0$ and a solution $\bar{x}$ of (4.1), we denote

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

THEOREM 4.9. *Assume that $F$ is semismooth$^*$ at $(\bar{x}, 0) \in \operatorname{gph} F$ with respect to $\mathcal{D}^*F$ and assume that there are $L, \kappa > 0$ such that for every $x \notin F^{-1}(0)$ sufficiently close to $\bar{x}$ we have $\mathcal{G}^{L,\kappa}_{F,\bar{x},\mathcal{D}^*}(x) \neq \emptyset$. Then there exists some $\delta > 0$ such that for every starting point $x^{(0)} \in \mathcal{B}_\delta(\bar{x})$ Algorithm 3 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 $(\hat{x}^{(k)}, \hat{y}^{(k)}, A, B) \in \mathcal{G}^{L,\kappa}_{F,\bar{x},\mathcal{D}^*}(x^{(k)})$.*

The proof can be conducted along the same lines as the proof of Theorem 4.4.

**5. Solving generalized equations.** The algorithms presented in the previous section are rather general and can be considered as a roadmap for solving the general inclusion (4.1). For an actual implementation of the approximation step and the Newton step we need, however, some more information about the mapping $F$. We will now illustrate Algorithm 3 by means of a frequently arising class of problems. Consider the GE

$$0 \in f(x) + \nabla g(x)^T N_D\big(g(x)\big), \tag{5.1}$$

where $f : \mathbb{R}^n \to \mathbb{R}^n$ is continuously differentiable, $g : \mathbb{R}^n \to \mathbb{R}^s$ is twice continuously differentiable, and $D \subseteq \mathbb{R}^s$ is a convex polyhedral set. Denoting $\Gamma := \{x \in \mathbb{R}^n \mid g(x) \in D\}$, we conclude $\nabla g(x)^T N_D\big(g(x)\big) \subseteq \widehat{N}_\Gamma(x) \subseteq N_\Gamma(x)$ (cf. [35, Theorem 6.14]). If in addition some constraint qualification is fulfilled, we also have $N_\Gamma(x) = \widehat{N}_\Gamma(x) = \nabla g(x)^T N_D(g(x))$ and in this case (5.1) is equivalent to the GE

$$0 \in f(x) + \widehat{N}_\Gamma(x). \tag{5.2}$$

Unfortunately, in many situations we cannot apply Algorithm 2 directly to the GE (5.1) since this would require finding some $\hat{x} \in g^{-1}(D)$ close to a given $x$ such that $\operatorname{dist}(0, f(\hat{x}) + \nabla g(\hat{x})^T N_D\big(g(\hat{x})\big))$ is small. This subproblem seems to be of the same difficulty as the original problem.

A widespread approach is to introduce multipliers and to consider, e.g., the problem

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} \in \tilde{F}(x, \lambda) := \begin{pmatrix} f(x) + \nabla g(x)^T \lambda \\ (g(x), \lambda) \end{pmatrix} - \{0\} \times \operatorname{gph} N_D. \tag{5.3}$$

We suggest here another equivalent reformulation

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} \in F(x, d) := \begin{pmatrix} f(x) + \nabla g(x)^T N_D(d) \\ g(x) - d \end{pmatrix} \tag{5.4}$$

which avoids the introduction of multipliers as problem variables. Obviously, $\bar{x}$ solves (5.1) if and only if $(\bar{x}, g(\bar{x}))$ solves (5.4).

In what follows we define for every $\lambda \in \mathbb{R}^s$ the *Lagrangian* $\mathcal{L}_\lambda : \mathbb{R}^n \to \mathbb{R}^n$ by

$$\mathcal{L}_\lambda(x) := f(x) + \nabla g(x)^T \lambda.$$

Next let us consider the regular coderivative of $F$ at some point $\hat{z} := ((\hat{x}, \hat{d}), (\hat{p}^*, g(\hat{x}) - \hat{d})) \in \operatorname{gph} F$ and choose any $\hat{\lambda} \in N_D(\hat{d})$ with $\hat{p}^* = \mathcal{L}_{\hat{\lambda}}(\hat{x})$. If $(x^*, d^*) \in$

$\widehat{D}^*F(\hat{z})(p,q^*)$, we have

$$\begin{aligned}0 &\geq \limsup_{((x,d),(p^*,g(x)-d))\overset{\text{gph}\,F}{\longrightarrow}\hat{z}} \frac{\langle x^*, x-\hat{x}\rangle + \langle d^*, d-\hat{d}\rangle - \langle p, p^*-\hat{p}^*\rangle - \langle q^*, g(x)-d-(g(\hat{x})-\hat{d})\rangle}{\|(x-\hat{x}, d-\hat{d}, p^*-\hat{p}^*, g(x)-d-(g(\hat{x})-\hat{d}))\|}\\ &\geq \limsup_{\substack{x\to\hat{x}\\ (d,\lambda)\overset{\text{gph}\,N_D}{\longrightarrow}(\hat{d},\hat{\lambda})}} \frac{\langle x^*, x-\hat{x}\rangle + \langle d^*, d-\hat{d}\rangle - \langle p, \mathcal{L}_\lambda(x)-\mathcal{L}_{\hat{\lambda}}(\hat{x})\rangle - \langle q^*, g(x)-d-(g(\hat{x})-\hat{d})\rangle}{\|(x-\hat{x}, d-\hat{d}, \mathcal{L}_\lambda(x)-\mathcal{L}_{\hat{\lambda}}(\hat{x}), g(x)-d-(g(\hat{x})-\hat{d}))\|}\\ &= \limsup_{\substack{x\to\hat{x}\\ (d,\lambda)\overset{\text{gph}\,N_D}{\longrightarrow}(\hat{d},\hat{\lambda})}} \frac{\langle x^* - \nabla\mathcal{L}_{\hat{\lambda}}(\hat{x})^T p - \nabla g(\hat{x})^T q^*, x-\hat{x}\rangle + \langle d^*+q^*, d-\hat{d}\rangle - \langle \nabla g(x)p, \lambda-\hat{\lambda}\rangle}{\|(x-\hat{x}, d-\hat{d}, \mathcal{L}_\lambda(x)-\mathcal{L}_{\hat{\lambda}}(\hat{x}), g(x)-d-(g(\hat{x})-\hat{d}))\|}.\end{aligned}$$

Fixing $(d,\lambda) = (\hat{d},\hat{\lambda})$, we obtain

$$0 \geq \limsup_{x\to\hat{x}} \frac{\langle x^* - \nabla\mathcal{L}_{\hat{\lambda}}(\hat{x})^T p - \nabla g(\hat{x})^T q^*, x-\hat{x}\rangle}{\|(x-\hat{x}, 0, \mathcal{L}_{\hat{\lambda}}(x)-\mathcal{L}_{\hat{\lambda}}(\hat{x}), g(x)-g(\hat{x}))\|}.$$

By our differentiability assumption, $\mathcal{L}_{\hat{\lambda}}$ and $g$ are Lipschitzian near $\hat{x}$ and therefore we have

$$x^* = \nabla\mathcal{L}_{\hat{\lambda}}(\hat{x})^T p + \nabla g(\hat{x})^T q^*.$$

Similarly, when fixing $x = \hat{x}$, we may conclude that

$$0 \geq \limsup_{(d,\lambda)\overset{\text{gph}\,N_D}{\longrightarrow}(\hat{d},\hat{\lambda})} \frac{\langle d^*+q^*, d-\hat{d}\rangle - \langle \nabla g(\hat{x})p, \lambda-\hat{\lambda}\rangle}{\|(0, d-\hat{d}, \nabla g(\hat{x})^T(\lambda-\hat{\lambda}), d-\hat{d})\|},$$

implying $d^*+q^* \in \widehat{D}^*N_D(\hat{d},\hat{\lambda})(\nabla g(\hat{x})p)$. Thus we have shown the inclusion

(5.5)
$$\begin{aligned}\widehat{D}^*F(\hat{z})(p,q^*) &\subseteq T(\hat{x},\hat{d},\hat{\lambda})(p,q^*)\\ &:= \left\{(\nabla\mathcal{L}_{\hat{\lambda}}(\hat{x})^T p + \nabla g(\hat{x})^T q^*, d^*) \mid d^*+q^* \in \widehat{D}^*N_D(\hat{d},\hat{\lambda})(\nabla g(\hat{x})p)\right\}.\end{aligned}$$

It is clear from the existing theory on coderivatives that this inclusion is strict in general. In order to proceed we introduce the following nondegeneracy condition.

Definition 5.1. *We say that* $(x,d) \in \mathbb{R}^n \times \mathbb{R}^s$ *is* nondegenerate with modulus $\gamma > 0$ *if*

$$\|\nabla g(x)^T\mu\| \geq \gamma\|\mu\| \;\forall \mu \in \operatorname{span} N_D(d). \tag{5.6}$$

*We simply say that* $(x,d)$ *is nondegenerate if* (5.6) *holds with some modulus* $\gamma > 0$.

*Remark* 5.2. The point $(\hat{x},\hat{d})$ is nondegenerate if and only if

$$\ker \nabla g(\hat{x})^T \cap \operatorname{span} N_D(\hat{d}) = \{0\},$$

which in turn is equivalent to $\nabla g(\hat{x})\mathbb{R}^n + \operatorname{lin} T_D(\hat{d}) = \mathbb{R}^s$. Thus, $(\hat{x},\hat{d})$ is nondegenerate if and only if $\hat{x}$ is a nondegenerate point in the sense of [34] (or [29, Assumption (A2)]) of the mapping $g(x) - (g(\hat{x}) - \hat{d})$ with respect to $D$. By [2, equation (4.172)], this is also related to the nondegenerate points in the sense of [2, Definition 4.70] without describing the $C^1$-reducibility of the set $D$.

*Remark* 5.3. It is not difficult to show that (5.5) holds with equality if $(\hat{x},\hat{d})$ is nondegenerate. However, this property is not important for the subsequent analysis.

LEMMA 5.4. *Consider* $((\hat{x},\hat{d}),(\hat{p}^*,g(\hat{x})-\hat{d}))\in \operatorname{gph} F$ *and assume that* $(\hat{x},\hat{d})$ *is nondegenerate. Then the system*

$$\hat{p}^* = f(\hat{x}) + \nabla g(\hat{x})^T\lambda (= \mathcal{L}_\lambda(\hat{x})),\ \ \lambda\in N_D(\hat{d}) \tag{5.7}$$

*has a unique solution denoted by* $\hat{\lambda}(\hat{x},\hat{d},\hat{p}^*)$.

*Proof.* By the definition of $F$, system (5.7) has at least one solution. Now let us assume that there are two distinct solutions $\lambda_1\neq\lambda_2$. Then $0\neq\lambda_1-\lambda_2\in \operatorname{span} N_D(\hat{d})$ and

$$\nabla g(\hat{x})^T(\lambda_1-\lambda_2) = f(\hat{x})+\nabla g(\hat{x})^T\lambda_1 - (f(\hat{x})+\nabla g(\hat{x})^T\lambda_2) = \hat{p}^*-\hat{p}^* = 0,$$

contradicting the nondegeneracy of $(\hat{x},\hat{d})$. Hence the solution to (5.7) is unique. ☐

We are now in the position to define the mapping $\widehat{\mathcal{D}}^*F$. Given some real $\hat{\gamma}>0$, we define

$$\begin{aligned}&\widehat{\mathcal{D}}^*F(\hat{z})(p,q^*)\\ &:= \begin{cases} T(\hat{x},\hat{d},\hat{\lambda}(\hat{x},\hat{d},\hat{p}^*))(p,q^*) & \text{if } (\hat{x},\hat{d}) \text{ is nondegenerate with modulus } \hat{\gamma},\\ \{(0,0)\} & \text{if } (\hat{x},\hat{d}) \text{ is not nondegenerate}\\ & \text{with modulus } \hat{\gamma} \text{ and } (p,q^*)=(0,0),\\ \emptyset & \text{otherwise}\end{cases}\end{aligned} \tag{5.8}$$

for every $\hat{z}:=(\hat{x},\hat{d},\hat{p}^*,g(\hat{x})-\hat{d})\in\operatorname{gph} F$ with $T$ given by (5.5). We neglect in the notation the dependence on $\hat{\gamma}$, which will be specified later.

THEOREM 5.5. *The mapping* $F$ *is semismooth*$^*$ *with respect to* $\mathcal{D}^*F$ *at every point* $(\bar{x},g(\bar{x}),0,0)$.

*Proof.* The proof is by contraposition. Assume on the contrary that there is a solution $(\bar{x},g(\bar{x}))$ to (5.4) together with $\epsilon>0$ and sequences

$$\begin{aligned}&((x_k,d_k),(p_k^*,g(x_k)-d_k))\stackrel{\operatorname{gph} F}{\longrightarrow}((\bar{x},g(\bar{x})),(0,0)),\\ &(x_k^*,d_k^*,p_k,q_k^*)\in\operatorname{gph}\widehat{\mathcal{D}}^*F((x_k,d_k),(p_k^*,g(x_k)-d_k))\end{aligned}$$

such that

$$\begin{aligned}&|\langle x_k^*,x_k-\bar{x}\rangle+\langle d_k^*,d_k-g(\bar{x})\rangle-\langle p_k,p_k^*\rangle-\langle q_k^*,g(x_k)-d_k\rangle|\\ &\quad>\epsilon\|(x_k-\bar{x},d_k-g(\bar{x}),p_k^*,g(x_k)-d_k)\|\|(x_k^*,d_k^*,p_k,q_k^*)\|\ \forall k.\end{aligned} \tag{5.9}$$

We may conclude that $(x_k^*,d_k^*,p_k,q_k^*)\neq(0,0,0,0)$ and consequently $(x_k,d_k)$ is nondegenerate with modulus $\hat{\gamma}$. It follows that the sequence $\lambda_k:=\hat{\lambda}(x_k,d_k,p_k^*)$ defined by Lemma 5.4 fulfills

$$\hat{\gamma}\|\lambda_k\|\le\|\nabla g(x_k)^T\lambda_k\| = \|p_k^*-f(x_k)\|$$

and hence it is bounded. By possibly passing to a subsequence we can assume that $\lambda_k$ converges to some $\bar{\lambda}$. It is easy to see that $\bar{\lambda}\in N_D(g(\bar{x}))$ and $\mathcal{L}_{\bar{\lambda}}(\bar{x})=0$ and by the definition of $\widehat{\mathcal{D}}^*F$ we obtain from (5.9)

$$\begin{aligned}
&|\langle x_k^*, x_k-\bar x\rangle + \langle d_k^*, d_k-g(\bar x)\rangle - \langle p_k, p_k^*\rangle - \langle q_k^*, g(x_k)-d_k\rangle|\\
&= |\langle \nabla\mathcal{L}_{\lambda_k}(x_k)^T p_k + \nabla g(x_k)^T q_k^*, x_k-\bar x\rangle + \langle d_k^*+q_k^*, d_k-g(\bar x)\rangle - \langle q_k^*, g(x_k)-g(\bar x)\rangle\\
&\qquad - \langle p_k, \mathcal{L}_{\lambda_k}(x_k)-\mathcal{L}_{\bar\lambda}(\bar x)\rangle|\\
&= |\langle p_k, \mathcal{L}_{\lambda_k}(\bar x)-\mathcal{L}_{\lambda_k}(x_k)+\nabla\mathcal{L}_{\lambda_k}(x_k)(x_k-\bar x)\\
&\qquad + (\nabla g(x_k)-\nabla g(\bar x))^T(\lambda_k-\bar\lambda) - \nabla g(x_k)^T(\lambda_k-\bar\lambda)\rangle\\
&\qquad + \langle q_k^*, g(\bar x)-g(x_k)+\nabla g(x_k)(x_k-\bar x)\rangle + \langle d_k^*+q_k^*, d_k-g(\bar x)\rangle|\\
&> \epsilon\|(x_k-\bar x, d_k-g(\bar x), \mathcal{L}_{\lambda_k}(x_k), g(x_k)-d_k)\|\|(x_k^*, d_k^*, p_k, q_k^*)\|\\
&\geq \epsilon\|(x_k-\bar x, d_k-g(\bar x), \mathcal{L}_{\lambda_k}(x_k))\|\|(d_k^*, p_k, q_k^*)\|.
\end{aligned}$$

For all $k$ sufficiently large we have

$$\begin{aligned}
&|\langle p_k, \mathcal{L}_{\lambda_k}(\bar x)-\mathcal{L}_{\lambda_k}(x_k)+\nabla\mathcal{L}_{\lambda_k}(x_k)(x_k-\bar x) + (\nabla g(x_k)-\nabla g(\bar x))^T(\lambda_k-\bar\lambda)\rangle\\
&\qquad + \langle q_k^*, g(\bar x)-g(x_k)+\nabla g(x_k)(x_k-\bar x)\rangle|\\
&\leq \frac{\epsilon}{2}\|x_k-\bar x\|\|(p_k,q_k^*)\| \leq \frac{\epsilon}{2}\|(x_k-\bar x, d_k-g(\bar x), \mathcal{L}_{\lambda_k}(x_k))\|\|(d_k^*, p_k, q_k^*)\|,
\end{aligned}$$

implying

(5.10)
$$|\langle d_k^*+q_k^*, d_k-g(\bar x)\rangle - \langle \nabla g(x_k)p_k, \lambda_k-\bar\lambda\rangle| > \frac{\epsilon}{2}\|(x_k-\bar x, d_k-g(\bar x), \mathcal{L}_{\lambda_k}(x_k))\|\|(d_k^*, p_k, q_k^*)\|.$$

Next observe that

$$\|\mathcal{L}_{\lambda_k}(x_k)\| = \|\mathcal{L}_{\lambda_k}(x_k)-\mathcal{L}_{\bar\lambda}(\bar x)\| = \|\nabla g(x_k)^T(\lambda_k-\bar\lambda)+\mathcal{L}_{\bar\lambda}(x_k)-\mathcal{L}_{\bar\lambda}(\bar x)\|$$

and let $L>0$ denote some real such that $\|\mathcal{L}_{\bar\lambda}(x_k)-\mathcal{L}_{\bar\lambda}(\bar x)\| \leq L\|x_k-\bar x\|\ \forall k$. If $\|x_k-\bar x\| < \|\nabla g(x_k)^T(\lambda_k-\bar\lambda)\|/(L+1)$, then we have

$$\|\mathcal{L}_{\lambda_k}(x_k)\| \geq \|\nabla g(x_k)^T(\lambda_k-\bar\lambda)\| - \|\mathcal{L}_{\bar\lambda}(x_k)-\mathcal{L}_{\bar\lambda}(\bar x)\| > \|\nabla g(x_k)^T(\lambda_k-\bar\lambda)\|/(L+1),$$

implying

$$\|(x_k-\bar x, \mathcal{L}_{\lambda_k}(x_k))\| \geq \|\nabla g(x_k)^T(\lambda_k-\bar\lambda)\|/(L+1).$$

Obviously this inequality holds as well when $\|x_k-\bar x\| \geq \|\nabla g(x_k)^T(\lambda_k-\bar\lambda)\|/(L+1)$. Further, by Lemma 2.4 for every $k$ sufficiently large there is a face $\mathcal{F}_k$ of $\mathcal{K}_D(g(\bar x),\bar\lambda)$ with $\operatorname{span} N_D(d_k) = (\operatorname{span}\mathcal{F}_k)^\perp$ and, since a convex polyhedral set has only finitely many faces, by possibly passing to a subsequence, we may assume that $\mathcal{F}_k = \mathcal{F}$ $\forall k$. Then $\bar\lambda = \lim_{k\to\infty}\lambda_k \in (\operatorname{span}\mathcal{F})^\perp$ and consequently $\lambda_k-\bar\lambda \in (\operatorname{span}\mathcal{F})^\perp = \operatorname{span} N_D(d_k)$. This yields $\|\nabla g(x_k)^T(\lambda_k-\bar\lambda)\| \geq \hat\gamma\|\lambda_k-\bar\lambda\|$ by the nondegeneracy of $(x_k,d_k)$ and we obtain the inequality

$$\|(x_k-\bar x, d_k-g(\bar x), \mathcal{L}_{\lambda_k}(x_k))\| \geq \min\{\frac{\hat\gamma}{L+1}, 1\}\|(d_k-g(\bar x), \lambda_k-\bar\lambda)\|.$$

Now let us choose some upper bound $C\geq 1$ for the bounded sequence $\|\nabla g(x_k)\|$ in order to obtain

$$\|(d_k^*, p_k, q_k^*)\| \geq \frac{\|(d_k^*+q_k^*, p_k)\|}{\sqrt2} \geq \frac{\|(d_k^*+q_k^*, \nabla g(x_k)p_k)\|}{C\sqrt2}.$$

Thus we derive from (5.10)

$$|\langle d_k^* + q_k^*, d_k - g(\bar{x})\rangle - \langle \nabla g(x_k)p_k, \lambda_k - \bar{\lambda}\rangle| \\ > \frac{\epsilon}{2\sqrt{2}C}\min\{\frac{\hat{\gamma}}{L+1}, 1\}\|d_k - g(\bar{x}), \lambda_k - \bar{\lambda}\|\|(d_k^* + q_k^*, \nabla g(x_k)p_k)\|,$$

showing, together with $d_k^* + q_k^* \in \widehat{D}^* N_D(d_k, \lambda_k)(\nabla g(x_k)p_k)$, that the mapping $N_D$ is not semismooth$^*$ at $(g(\bar{x}), \bar{\lambda})$. This contradicts our result from section 3.1 and the theorem is proven. ∎

Note that the mapping $F$ will in general not be semismooth$^*$ in the sense of Definition 3.1 at a solution $(\bar{x}, g(\bar{x}))$ to (5.4), provided $(\bar{x}, g(\bar{x}))$ is not nondegenerate.

It is quite surprising that no constraint qualification is required in Theorem 5.5. In fact, there is a constraint qualification hidden in our assumption because usually we are interested in solutions of (5.2) and here we assume that a solution to (5.1) is given. Based on Theorem 5.5, in a forthcoming paper we will present a locally superlinearly convergent Newton-type algorithm which does not require, apart from the solvability of (5.1), any other constraint qualification. In this paper we want just to demonstrate the basic principles of how the approximation step and the Newton step can be performed. Therefore, for the ease of presentation, in the remainder of this section we will impose the following.

*Assumption* 1. $(\bar{x}, g(\bar{x}))$ is a nondegenerate solution to (5.4) with modulus $\bar{\gamma}$.

In the following lemma we summarize two easy consequences of Assumption 1. Recall that a mapping $G : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is *metrically regular* around $(\bar{x}, \bar{y}) \in \operatorname{gph} G$ if there are neighborhoods $U$ of $\bar{x}$ and $V$ of $\bar{y}$ along with a positive real $\kappa$ such that

$$\operatorname{dist}(x, G^{-1}(y)) \le \kappa \operatorname{dist}(y, G(x)) \;\forall (x, y) \in U \times V.$$

LEMMA 5.6. *Assume that Assumption* 1 *is fulfilled. Then there is a neighborhood* $W$ *of* $(\bar{x}, g(\bar{x}))$ *such that all points* $(x, d) \in W$ *are nondegenerate with modulus* $\bar{\gamma}/2$. *Further, the mapping* $x \rightrightarrows g(x) - D$ *is metrically regular around* $(\bar{x}, 0)$ *and the mapping* $u \rightrightarrows g(\bar{x}) + \nabla g(\bar{x})u - D$ *is metrically regular around* $(0, 0)$.

*Proof.* We show the first assertion by contraposition. Assume on the contrary that there are sequences $(x_k, d_k) \to (\bar{x}, g(\bar{x}))$ and $(\mu_k)$ such that $\mu_k \in \operatorname{span} N_D(d_k) \cap \mathcal{S}_{\mathbb{R}^s}$ and $\|\nabla g(x_k)^T \mu_k\| < \bar{\gamma}/2$ for all $k$. By possibly passing to a subsequence we can assume that $\mu_k$ converges to some $\bar{\mu} \in \mathcal{S}_{\mathbb{R}^s}$ satisfying $\|\nabla g(\bar{x})^T \bar{\mu}\| \le \bar{\gamma}/2$. Since $D$ is polyhedral we have $N_D(d_k) \subset N_D(g(\bar{x}))$ for all $k$ sufficiently large implying $\bar{\mu} \in \operatorname{span} N_D(g(\bar{x}))$, which contradicts our assumption on the modulus of nondegeneracy at $(\bar{x}, g(\bar{x}))$. In order to show the metric regularity property of the two mappings just note that Assumption 1 implies

$$\nabla g(\bar{x})^T \mu = 0, \mu \in N_D(g(\bar{x})) \subset \operatorname{span} N_D(g(\bar{x})) \;\Rightarrow \mu = 0.$$

Now the assertion follows from [35, Example 9.44]. ∎

We now want to specialize the approximation step and the Newton step for the GE (5.4). In this case the approximation step can be performed as follows.

ALGORITHM 4 (approximation step). *Input:* $x \in \mathbb{R}^n$.

1. *Compute a solution* $\hat{u}$ *of the strictly convex quadratic program*

$$QP(x) \qquad \min_{u \in \mathbb{R}^n} \frac{1}{2}\|u\|^2 + f(x)^T u \\ \text{subject to } g(x) + \nabla g(x)u \in D$$

*together with an associated multiplier* $\hat\lambda \in N_D(g(x)+\nabla g(x)\hat u)$ *satisfying*

$$\hat u + f(x) + \nabla g(x)^T\hat\lambda = \hat u + \mathcal{L}_{\hat\lambda}(x) = 0. \tag{5.11}$$

2. *Set* $\hat x := x$, $\hat d := g(x)+\nabla g(x)\hat u$, $\hat p^* := \mathcal{L}_{\hat\lambda}(\hat x)$, $\hat y := (\hat p^*, g(\hat x)-\hat d)$.

Obviously we have $\big((\hat x,\hat d),\hat y\big) \in \operatorname{gph} F$. Program $QP(x)$ amounts exactly to problem $\mathcal{P}(x)$ in [8, p. 708] in the case when (5.1) describes first-order optimality conditions for a nonlinear program, i.e., when $f = \nabla\phi$ for some objective $\phi$ and $D = \mathbb{R}^s_-$. In fact, both $QP(x)$ and $\mathcal{P}(x)$ serve the same purpose, namely, the computation of an appropriate modification of the current iterate. The modified Wilson method, investigated in [8], is, however, completely different from our method, e.g., in [8] a system in variables $(x,\lambda)$ related to (5.3) is solved and we consider the inclusion (5.4) with variables $(x,d)$.

In the following proposition we state some properties of the output of Algorithm 4 when the input $x$ is sufficiently close to $\bar x$. We denote by $\bar\lambda := \hat\lambda(\bar x, g(\bar x), 0)$ the unique multiplier associated with the nondegenerate solution $(\bar x, g(\bar x))$ of (5.4) (cf. Lemma 5.4).

Proposition 5.7. *Under Assumption* 1 *there is a positive radius* $\rho$ *and positive reals* $\beta, \beta_u$, *and* $\beta_\lambda$ *such that for all* $x \in \mathcal{B}_\rho(\bar x)$ *the problem* $QP(x)$ *has a unique solution and the output* $\hat x, \hat d, \hat\lambda, \hat y$, *and* $\hat u$ *of Algorithm* 4 *fulfills*

$$\|\hat u\| \le \beta_u \|x-\bar x\|, \tag{5.12}$$

$$\|\big((\hat x,\hat d),\hat y\big) - \big((\bar x, g(\bar x)),(0,0)\big)\| \le \beta\|\hat u\|, \tag{5.13}$$

$$\|\hat\lambda - \bar\lambda\| \le \beta_\lambda \|x-\bar x\|. \tag{5.14}$$

*Further,* $(\hat x,\hat d)$ *is nondegenerate with modulus* $\bar\gamma/2$ *and* $N_D(\hat d) \subseteq N_D(g(\bar x))$.

*Proof.* Let $\tilde\Gamma(x) := \{u \mid \tilde g(x,u) := g(x)+\nabla g(x)u \in D\}$ denote the feasible region of the problem $QP(x)$. By Lemma 5.6 the mapping $u \rightrightarrows \tilde g(\bar x,u) - D$ is metrically regular around $(0,0)$. Considering $x$ as a parameter and $u$ as the decision variable, by [11, Corollary 3.7] the system $\tilde g(x,u) \in D$ has the so-called Robinson stability property at $(\bar x, 0)$, implying $\tilde\Gamma(x) \ne \emptyset$ for all $x$ belonging to some neighborhood $U'$ of $\bar x$. Thus the feasible region of the quadratic program $QP(x)$ is not empty and, since the objective is strictly convex, for every $x \in U'$ the existence of a unique solution $\hat u$ follows. Obviously, $\hat u = 0$ is the unique solution of $QP(\bar x)$. The convexity of the quadratic program $QP(x)$ ensures that $\hat u$ is a solution if and only if the first-order optimality condition

$$0 \in \tilde f(x,\hat u) + \nabla_u \tilde g(x,\hat u)^T N_D(\tilde g(x,\hat u)) \tag{5.15}$$

with $\tilde f(x,u) := u + f(x)$ is fulfilled. Defining for every $\lambda \in \mathbb{R}^s$ the linear mapping $\tilde{\mathcal{F}}_\lambda : \mathbb{R}^n \to \mathbb{R}^n$ by $\tilde{\mathcal{F}}_\lambda v := \nabla_u \tilde f(\bar x, 0)v + \nabla_u^2\langle\lambda^T\tilde g(\cdot)\rangle(\bar x,0)v = v$, we obviously have $\langle\tilde{\mathcal{F}}_\lambda v, v\rangle = \|v\|^2 > 0\ \forall v \ne 0$ and therefore all assumptions of [12, Theorem 6.2] for the isolated calmness property of the solution map to the parameterized variational system (5.15) are fulfilled. Thus there is a positive radius $\rho'$ and some constant $\beta_u > 0$ such that $\mathcal{B}_{\rho'} \subset U'$ and for every $x \in B_{\rho'}(\bar x)$ the solution $\hat u$ to $QP(x)$ fulfills the inequality $\|\hat u\| \le \beta_u\|x-\bar x\|$. Setting $L := \sup\{\|\nabla g(x)\| \mid x \in \mathcal{B}_{\rho'}(\bar x)\}$, we obtain

$$\|\hat d - g(\bar x)\| \le \|g(x)-g(\bar x)\| + \|\nabla g(x)\hat u\| \le L(\|x-\bar x\| + \|\hat u\|) \le L(1+\beta_u)\|x-\bar x\|$$

and

$$\|\hat{y}\| = \|(\hat{p}^*, g(\hat{x}) - \hat{d})\| = \| - (\hat{u}, \nabla g(x)\hat{u})\| \le \sqrt{1+L^2}\|\hat{u}\| \le \sqrt{1+L^2}\beta_u\|x-\bar{x}\|,$$

implying that (5.13) holds with $\beta^2 = 1 + L^2(1+\beta_u)^2 + (1+L^2)\beta_u^2$. Next we choose $0 < \rho \le \rho'$ such that $\mathcal{B}_\rho(\bar{x}) \times \mathcal{B}_{\beta\rho}(g(\bar{x}))$ is contained in the neighborhood $W$ given by Lemma 5.6. Then $(\hat{x}, \hat{d})$ is nondegenerate with modulus $\bar{\gamma}/2$ and we obtain

$$\frac{\bar{\gamma}}{2}\|\hat{\lambda}\| \le \|\nabla g(x)^T\hat{\lambda}\| = \| - \hat{u} - f(x)\|,$$

$D$ is polyhedral, there is some neighborhood $O$ of $g(\bar{x})$ such that $N_D(d) \subseteq N_D(g(\bar{x}))$ for all $d \in D \cap O$, and we may assume that $\rho$ is chosen small enough so that $\mathcal{B}_{\beta\rho}(g(\bar{x})) \subset O$. Then $\hat{\lambda} - \bar{\lambda} \in \operatorname{span} N_D(g(\bar{x}))$ and we obtain

$$\begin{aligned}
&\bar{\gamma}\|\hat{\lambda} - \bar{\lambda}\| \\
&\quad \le \|\nabla g(\bar{x})^T(\hat{\lambda} - \bar{\lambda})\| \le \|\nabla g(x)^T\hat{\lambda} - \nabla g(\bar{x})^T\bar{\lambda}\| + \|\nabla g(x)^T\hat{\lambda} - \nabla g(\bar{x})^T\hat{\lambda}\| \\
&\quad = \| - f(x) - \hat{u} + f(\bar{x})\| + \|\nabla g(x) - \nabla g(\bar{x})\|\|\hat{\lambda}\| \le (L_f + \beta_u + L_{\nabla g}\|\hat{\lambda}\|)\|x - \bar{x}\|,
\end{aligned}$$

where $L_f$ and $L_{\nabla g}$ denote the Lipschitz moduli of $f$ and $\nabla g$ in $\mathcal{B}_\rho(\bar{x})$, respectively. This implies (5.14). ☐

Having performed the approximation step, we now turn to the Newton step. We start with the following auxiliary lemma.

LEMMA 5.8. *Let $\hat{d} \in D$ and let $\hat{l} := \dim(\operatorname{lin} T_D(\hat{d}))$. Then for every $s \times (s - \hat{l})$ matrix $\hat{W}$, whose columns belong to $N_D(\hat{d})$ and form a basis for* $\operatorname{span} N_D(\hat{d})$*, and for every $\hat{x} \in \mathbb{R}^n$ there holds*

$$\{u \mid \nabla g(\hat{x})u \in \operatorname{lin} T_D(\hat{d})\} = \{u \mid \hat{W}^T\nabla g(\hat{x})u = 0\}. \tag{5.16}$$

*Moreover, if $(\hat{x}, \hat{d})$ is nondegenerate, then $\hat{W}^T\nabla g(\hat{x})$ has full row rank $s - \hat{l}$.*

*Proof.* Equation (5.16) is an immediate consequence of the relation

$$\begin{aligned}
&\nabla g(\hat{x})u \in \operatorname{lin} T_D(\hat{d}) = (\operatorname{span} N_D(\hat{d}))^\perp = (\operatorname{Range} \hat{W})^\perp \;\Leftrightarrow\; \forall w \in \mathbb{R}^{s-\hat{l}} \langle \hat{W}w, \nabla g(\hat{x})u\rangle = 0 \\
&\Leftrightarrow\; \hat{W}^T\nabla g(\hat{x})u = 0.
\end{aligned}$$

Now assume that $\hat{W}^T\nabla g(\hat{x})$ does not have full row rank and there is some $0 \ne \mu \in \mathbb{R}^{s-\hat{l}}$ with $\mu^T\hat{W}^T\nabla g(\hat{x}) = 0$. Then $0 \ne \hat{W}\mu \in \operatorname{span} N_D(\hat{d})$ and $\nabla g(\hat{x})^T(\hat{W}\mu) = 0$, contradicting the nondegeneracy of $(\hat{x}, \hat{d})$. ☐

Assume now that $(\hat{x}, \hat{d})$ is nondegenerate with modulus $\hat{\gamma}$. One can extract from [5, proof of Theorem 2] that

$$\widehat{N}_{\operatorname{gph} N_D}(\hat{d}, \hat{\lambda}) = \mathcal{K}_D(\hat{d}, \hat{\lambda})^\circ \times \mathcal{K}_D(\hat{d}, \hat{\lambda}).$$

Thus, by (5.8), for $(p, q^*) \in \mathbb{R}^n \times \mathbb{R}^s$ the set $\hat{\mathcal{D}}F\big((\hat{x}, \hat{d}), (\hat{p}^*, g(\hat{x}) - \hat{d})\big)(p, q^*)$ consists of the elements $(\nabla\mathcal{L}_{\hat{\lambda}}(\hat{x})^Tp + \nabla g(\hat{x})^Tq^*, d^*)$ such that

$$d^* + q^* \in \widehat{D}^*N_D(\hat{d}, \hat{\lambda})(\nabla g(\hat{x})p) = \begin{cases} \mathcal{K}_D(\hat{d}, \hat{\lambda})^\circ & \text{if } -\nabla g(\hat{x})p \in \mathcal{K}_D(\hat{d}, \hat{\lambda}), \\ \emptyset & \text{else.} \end{cases} \tag{5.17}$$

We have to compute suitable matrices $(A,B)\in\mathcal{A}^{\mathcal{D}^*}_{\rm reg}F\big((\hat x,\hat d),(\hat p^*,g(\hat x)-\hat d)\big)$. This is done by choosing suitable elements $(p_i,q_i^*,d_i^*)$, $i=1,\dots,n+s$, fulfilling (5.17), and setting $A_i, B_i$, the $i$th row of $A$ and $B$, respectively, to

$$A_i := \left(p_i^T\nabla\mathcal{L}_{\hat\lambda}(\hat x)+{q_i^*}^T\nabla g(\hat x)\ \vdots\ {d_i^*}^T\right),\quad B_i:=\left(p_i^T\ \vdots\ {q_i^*}^T\right),\ i=1,\dots,n+s.$$

Denoting $\hat l:=\dim(\operatorname{lin}T_D(\hat d))$, we have $\dim(\operatorname{span}N_D(\hat d))=s-\hat l$ and we can find an $s\times(s-\hat l)$ matrix $\hat W$, whose columns belong to $N_D(\hat d)$ and form a basis for $\operatorname{span}N_D(\hat d)$ (cf. Lemma 5.8). The $(s-\hat l)\times n$ matrix $\hat W^T\nabla g(x)$ has full row rank $s-\hat l$ and thus we can find vectors $p_i$, $i=1,\dots,n-(s-\hat l)$, constituting an orthonormal basis for $\ker \hat W^T\nabla g(x)$ and set $d_i^*=q_i^*=0$, $i=1,\dots,n-(s-\hat l)$. By (2.3) and (5.16) we have $-\nabla g(\hat x)p_i\in\mathcal{K}_D(\hat d,\hat\lambda)$ and $d_i^*+q_i^*=0\in\mathcal{K}_D(\hat d,\hat\lambda)^\circ$ trivially holds. The next elements $p_i$, $i=n-(s-\hat l)+1,\dots,n+s$, are all chosen as 0. Further we choose the $s-\hat l$ elements $q_i^*$, $i=n-(s-\hat l)+1,\dots,n$, as the columns of the matrix $\hat W$ and set $d_i^*=0$. Finally we set $q_i^*:=-d_i^*:=e_{i-n}$, $i=n+1,\dots,n+s$, where $e_j$ denotes the $j$th unit vector.

With this choice, the corresponding matrices $(A,B)\in\mathcal{A}^{\mathcal{D}^*}F\big((\hat x,\hat d),(\hat p^*,g(\hat x)-\hat d)\big)$ are given by

$$A=\begin{pmatrix}\hat Z^T\nabla\mathcal{L}_{\hat\lambda}(\hat x) & \vdots & 0\\ \cdots & & \cdots\\ \hat W^T\nabla g(\hat x) & \vdots & 0\\ \cdots & & \cdots\\ \nabla g(\hat x) & \vdots & -Id_s\end{pmatrix},\quad B=\begin{pmatrix}\hat Z^T & \vdots & 0\\ \cdots & & \cdots\\ 0 & \vdots & \hat W^T\\ \cdots & & \cdots\\ 0 & \vdots & Id_s\end{pmatrix},\tag{5.18}$$

where $\hat Z$ is the $n\times(n-(s-\hat l))$ matrix with columns $p_i$, $i=1,\dots,n-(s-\hat l)$. In particular, we have $\hat Z^T\hat Z=Id_{n-(s-\hat l)}$ and $\hat W^T\nabla g(x)\hat Z=0$. Note that the matrix $B$ in (5.18) is certainly not nonsingular.

Lemma 5.9. *Assume that the matrix $G:=\hat Z^T\nabla\mathcal{L}_{\hat\lambda}(\hat x)\hat Z$ is nonsingular. Then the matrix $A$ in* (5.18) *is nonsingular and*

$$A^{-1}=\begin{pmatrix}\hat ZG^{-1} & \vdots & (Id_n-\hat ZG^{-1}\hat Z^T\nabla\mathcal{L}_{\hat\lambda}(\hat x))C^\dagger & \vdots & 0\\ \cdots & & \cdots & & \cdots\\ \nabla g(\hat x)\hat ZG^{-1} & \vdots & \nabla g(\hat x)(Id_n-\hat ZG^{-1}\hat Z^T\nabla\mathcal{L}_{\hat\lambda}(\hat x))C^\dagger & \vdots & -Id_s\end{pmatrix},\tag{5.19}$$

*where the $n\times(s-\hat l)$ matrix $C^\dagger:=C^T(CC^T)^{-1}$ is the Moore–Penrose inverse of $C:=\hat W^T\nabla g(\hat x)$.*

*Proof.* The proof follows by the observation that the product of $A$ with the matrix on the right-hand side of (5.19) is the identity matrix. ∎

Since $D$ is polyhedral, there are only finitely many possibilities for $N_D(\hat d)$ and we assume that for identical normal cones we always use the same matrix $\hat W$.

Note that the matrix $\hat Z$ and consequently also the matrices $G$ and $G^{-1}$ are not uniquely given. Let $Z_1,Z_2$ be two $n\times(n-(s-\hat l))$ matrices whose columns form an orthogonal basis of $\ker C$ and $G_i:=Z_i^T\nabla\mathcal{L}_{\hat\lambda}(\hat x)Z_i$, $i=1,2$. Then $Z_2=Z_1V$, where the matrix $V := Z_1^T Z_2$ is orthogonal, and consequently

$$G_2 = V^T G_1 V,\ G_2^{-1} = V^T G_1^{-1} V,\ Z_2 G_2^{-1} = Z_1 G_1^{-1} V,\ Z_2 G_2^{-1} Z_2^T = Z_1 G_1^{-1} Z_1^T,$$
$$\|G_2\| = \|G_1\|,\ \|G_2\|_F = \|G_1\|_F,\ \|Z_2 G_2^{-1}\| = \|Z_1 G_1^{-1}\|,\ \|Z_2 G_2^{-1}\|_F = \|Z_1 G_1^{-1}\|_F.$$

It follows that the property of invertibility of $G$ (and consequently the invertibility of $A$), the matrix $\hat{Z} G^{-1} \hat{Z}^T$, and the quantity $\|A^{-1}\|_F \|(A \vdots B)\|_F$ are independent of the particular choice of $\hat{Z}$. In order to ensure that $A^{-1}$ exists and is bounded, a suitable second-order condition has to be imposed. By Lemma 5.9 this is ensured by the following assumption.

*Assumption* 2. For every face $\mathcal{F}$ of the critical cone $\mathcal{K}_D(g(\bar{x}), \bar{\lambda})$ there is a matrix $Z_{\mathcal{F}}$, whose columns form an orthogonal basis of $\{u \mid \nabla g(\bar{x}) u \in \operatorname{span} \mathcal{F}\}$, such that the matrix $Z_{\mathcal{F}}^T \mathcal{L}_{\bar{\lambda}}(\bar{x}) Z_{\mathcal{F}}$ is nonsingular.

In fact, if $Z_{\mathcal{F}}^T \mathcal{L}_{\bar{\lambda}}(\bar{x}) Z_{\mathcal{F}}$ is nonsingular, then $Z^T \mathcal{L}_{\bar{\lambda}}(\bar{x}) Z$ is nonsingular for every matrix $Z$ representing the subspace $\{u \mid \nabla g(\bar{x}) u \in \operatorname{span} \mathcal{F}\}$.

*Remark* 5.10. In the case when $D = \mathbb{R}^s_-$, let $\bar{I} := \{i \in \{1, \ldots, s\} \mid g_i(\bar{x}) = 0\}$ denote the index set of active inequality constraints and let $\bar{I}^+ := \{i \in \bar{I} \mid \bar{\lambda}_i > 0\}$ denote the index set of positive multipliers. Then the faces of $\mathcal{K}_{\mathbb{R}^s_-}(g(\bar{x}), \bar{\lambda})$ are exactly the sets

$$\{d \in \mathbb{R}^s \mid d_i = 0,\ i \in J, d_i \leq 0, i \in \bar{I} \setminus J\},\ \bar{I}^+ \subseteq J \subseteq \bar{I}.$$

Thus Assumption 2 says that for every index set $\bar{I}^+ \subseteq J \subseteq \bar{I}$ and every matrix $Z_J$, whose columns form an orthogonal basis of the subspace $\{u \mid \nabla g_i(\bar{x}) u = 0,\ i \in J\}$, the matrix $Z_J^T \mathcal{L}_{\bar{\lambda}}(\bar{x}) Z_J$ is nonsingular. However, we do not require that all the matrices $Z_J^T \mathcal{L}_{\bar{\lambda}}(\bar{x}) Z_J$ have the same determinantal sign and therefore Assumption 2 is weaker than the so-called *strong coherent orientation condition* (SCOC) of [7] (cf. the discussion on SCOC in [26]).

PROPOSITION 5.11. *Assume that at the solution $\bar{x}$ of* (5.1) *both Assumptions* 1 *and* 2 *are fulfilled and let $\widehat{\mathcal{D}}^* F$ be given by* (5.8) *with $\hat{\gamma} = \bar{\gamma}/2$. Then there are constants $\tilde{L}, \kappa > 0$ such that for every $x$ sufficiently close to $\bar{x}$ not solving* (5.1) *and for every $d \in \mathbb{R}^s_-$ the quadruple $((\hat{x}, \hat{d}), (\hat{p}^*, g(\hat{x}) - \hat{d}), A, B)$ belongs to $\mathcal{G}^{\tilde{L},\kappa}_{F,(\bar{x}, g(\bar{x})), \mathcal{D}^*}(x, d)$, where $\hat{x}, \hat{d}, \hat{p}^*$ are the result of Algorithms* 4 *and $A, B$ are given by* (5.18). *In particular, $\mathcal{G}^{\tilde{L},\kappa}_{F,(\bar{x}, g(\bar{x})), \mathcal{D}^*}(x, d) \neq \emptyset$.*

*Proof.* Let $\rho, \beta, \beta_u, \beta_\lambda$, and $U$ be as in Proposition 5.7 and Lemma 5.8, respectively, and set $\tilde{L} := \beta \beta_u$. By possibly reducing $\rho$ we may assume that $\mathcal{B}_\rho(\bar{x}) \times \mathcal{B}_{\beta_\lambda \rho}(\bar{\lambda}) \subset U$. Then, for every $x \in \mathcal{B}_\rho(\bar{x})$ and every $d \in \mathbb{R}^s_-$ we have

$$\|(\hat{x}, \hat{d}, \hat{p}^*, g(\hat{x}) - \hat{d}) - (\bar{x}, g(\bar{x}), 0, 0)\| \leq \beta \beta_u \|x - \bar{x}\| \leq \tilde{L} \|(x, d) - (\bar{x}, g(\bar{x}))\| \tag{5.20}$$

by Proposition 5.7 and it remains to show that $\|A^{-1}\|_F \|(A \vdots B)\|_F$ is uniformly bounded for $x$ close to $\bar{x}$. We consider the following possibility for computing a matrix $\hat{Z}$, whose columns are an orthonormal basis for a given $m \times n$ matrix $C$. Let $Q$ be an $n \times n$ orthogonal matrix such that $CQ = (L \vdots 0)$, where $L$ is an $m \times m$ lower triangular matrix. If $\operatorname{rank} C = m$, then $\hat{Z}$ can be taken as the last $n - m$ columns of $Q$ (cf. [14, section 5.1.3]). This can be practically done by so-called Householder transformations (see, e.g., [14, section 2.2.5.3]). When performing the Householder transformations, the signs of the diagonal elements of $L$ are usually chosen in such a way that cancellation errors are avoided. However, when modifying the Householder transformations in order to obtain nonnegative diagonal elements $L_{ii}$, it can be easily seen that the algorithm produces $Q$ and $L$ depending in a continuously differentiable way on $C$, provided $C$ has full row rank. Since the quantity $\|A^{-1}\|_F\|(A \vdots B)\|_F$ does not depend on the particular choice of $\hat{Z}$, we can assume that $\hat{Z}$ is computed in such a way. Now assume that the statement of the proposition does not hold true. In view of (5.20) there must be a sequence $x_k$ converging to $\bar{x}$ such that Algorithm 4 produces with input $x_k$ the quantities $\hat{x}_k$, $\hat{\lambda}_k$, $\hat{p}_k^*$, and $\hat{d}_k$, resulting by (5.18) in matrices $\hat{W}_k, \hat{Z}_k, A_k, B_k$, where either $A_k$ is singular or $\|{A_k}^{-1}\|_F\|(A_k \vdots B_k)\|_F \to \infty$ as $k \to \infty$. Since there are only finitely many possibilities for $\hat{W}_k$ and there are only finitely many faces of $\mathcal{K}_D(g(\bar{x}), \bar{\lambda})$, we can assume that $\hat{W}_k = \hat{W}$ and $\operatorname{lin} T_D(\hat{d}) = \operatorname{span} \mathcal{F}\ \forall k$ for some face $\mathcal{F}$ of $\mathcal{K}_D(g(\bar{x}), \bar{\lambda})$ by Lemma 2.4. In view of (5.16) we have $\{u \mid \nabla g(\bar{x})u \in \operatorname{span} \mathcal{F}\} = \ker(\hat{W}^T\nabla g(\bar{x}))$ and we can assume that the matrix $Z_{\mathcal{F}}$ is computed as above via an orthogonal factorization of the matrix $\hat{W}^T\nabla g(\bar{x})$. It follows that $\hat{Z}_k$ converges to $Z_{\mathcal{F}}$ and thus $\hat{Z}_k^T\mathcal{L}_{\hat{\lambda}_k}(\hat{x}_k)\hat{Z}_k$ converges to the nonsingular matrix $Z_{\mathcal{F}}^T\mathcal{L}_{\bar{\lambda}}(\bar{x})Z_{\mathcal{F}}$. Thus for all $k$ sufficiently large the matrices $\hat{Z}_k^T\mathcal{L}_{\hat{\lambda}_k}(\hat{x}_k)\hat{Z}_k$ are nonsingular and their inverses are uniformly bounded. Since the matrices $\hat{W}^T\nabla g(\hat{x}_k)$ converge to the matrix $\hat{W}^T\nabla g(\bar{x})$ having full row rank, its Moore–Penrose inverses converge to the one of $\hat{W}^T\nabla g(\bar{x})$. From Lemma 5.9 we may conclude that the matrices $A_k$ are nonsingular and $\|{A_k}^{-1}\|_F\|(A_k \vdots B_k)\|_F$ remains bounded. Thus the statement of the proposition must hold true. ☐

We are now in position to explicitly write down the Newton step. By Algorithm 3 the new iterate amounts to $(\hat{x}, \hat{d}) + (s_x, s_d)$ with

$$\begin{pmatrix} s_x \\ s_d \end{pmatrix} = -A^{-1}B \begin{pmatrix} \hat{p}^* \\ g(\hat{x}) - \hat{d} \end{pmatrix},$$

i.e., $(s_x, s_d)$ solves the linear system

$$\begin{aligned} &\hat{Z}^T\big(\nabla\mathcal{L}_{\hat{\lambda}}(\hat{x})s_x + \mathcal{L}_{\hat{\lambda}}(\hat{x})\big) = 0, \\ &\hat{W}^T(g(\hat{x}) + \nabla g(\hat{x})s_x - \hat{d}) = 0, \\ &g(\hat{x}) + \nabla g(\hat{x})s_x - (\hat{d} + s_d) = 0. \end{aligned}$$

Note that by the definition of $\hat{W}$ the second equation can be equivalently written as

$$g(\hat{x}) + \nabla g(\hat{x})s_x - \hat{d} \in \ker \hat{W}^T = (\operatorname{Range} W)^\perp = (\operatorname{span} N_D(\hat{d}))^\perp = \operatorname{lin} T_D(\hat{d}).$$

It appears that we need not compute the auxiliary variables $d$, $s_d$ and the columns of $\hat{W}$ need not necessarily belong to $N_D(\hat{d})$.

Algorithm 5 (semismooth* Newton method for solving (5.1)).

1. *Choose a starting point* $x^{(0)}$. *Set* $k := 0$.
2. *If* $x^{(k)}$ *is a solution of* (5.1), *stop the algorithm.*
3. *Run Algorithm* 4 *with input* $x^{(k)}$ *in order to compute* $\hat{\lambda}^{(k)}$, $\hat{d}^{(k)}$, *and* $\hat{p}^{*(k)} = \mathcal{L}_{\hat{\lambda}^{(k)}}(x^{(k)})$.
4. *Set* $\hat{l}^{(k)} = \dim(\operatorname{lin} T_D(\hat{d}^{(k)}))$ *and compute an* $s \times (s - \hat{l}^{(k)})$ *matrix* $\hat{W}^{(k)}$, *whose columns form a basis for* $\operatorname{span} N_D(\hat{d}^{(k)})$ *and then an* $n \times (n - (s - \hat{l}^{(k)}))$ *matrix* $\hat{Z}^{(k)}$, *whose columns are an orthogonal basis for* $\ker({\hat{W}^{(k)}}{}^T\nabla g(x^{(k)}))$.

5. *Compute the Newton direction* $s_x^{(k)}$ *by solving the linear system*

$$\hat{Z}^{(k)^T}\big(\nabla\mathcal{L}_{\hat{\lambda}^{(k)}}(x^{(k)})s_x + \mathcal{L}_{\hat{\lambda}^{(k)}}(x^{(k)})\big) = 0,$$
$$\hat{W}^{(k)^T}\big(g(x^{(k)}) + \nabla g(x^{(k)})s_x - \hat{d}^{(k)}\big) = 0$$

*and set* $x^{(k+1)} := x^{(k)} + s_x^{(k)}$.

6. *Increase* $k := k+1$ *and go to step* 2.

THEOREM 5.12. *Assume that* $\bar{x}$ *solves* (5.1) *and both Assumptions* 1 *and* 2 *are fulfilled. Then there is a neighborhood* $U$ *of* $\bar{x}$ *such that for every starting point* $x^{(0)} \in U$ *Algorithm* 5 *either stops after finitely many iterations at a solution of* (5.1) *or produces a sequence* $x^{(k)}$ *converging superlinearly to* $\bar{x}$.

*Proof.* The proof follows from Theorem 4.9 and Proposition 5.11. ∎

We now want to compare Algorithm 5 with the usual Josephy–Newton method for solving (5.1) via (5.3). Given an iterate $(x^{(k)}, \lambda^{(k)})$, the new iterate $(x^{(k+1)}, \lambda^{(k+1)})$ is computed as a solution of the partially linearized system

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} \in \begin{pmatrix} \mathcal{L}_{\lambda^{(k)}}(x^{(k)}) + \nabla\mathcal{L}_{\lambda^{(k)}}(x^{(k)})(x^{(k+1)} - x^{(k)}) + \nabla g(x^{(k)})^T(\lambda^{(k+1)} - \lambda^{(k)}) \\ (g(x^{(k)}) + \nabla g(x^{(k)})(x^{(k+1)} - x^{(k)}), \lambda^{(k+1)}) \end{pmatrix} - \{0\} \times \operatorname{gph} N_D, \tag{5.21}$$

i.e.,

$$\begin{aligned} 0 &= f(x^{(k)}) + \nabla\mathcal{L}_{\lambda^{(k)}}(x^{(k)})(x^{(k+1)} - x^{(k)}) + \nabla g(x^{(k)})^T\lambda^{(k+1)}, \\ \lambda^{(k+1)} &\in N_D(g(x^{(k)}) + \nabla g(x^{(k)})(x^{(k+1)} - x^{(k)})). \end{aligned} \tag{5.22}$$

In order to guarantee that (5.21) and (5.22), respectively, are solvable for all $(x^{(k)}, \lambda^{(k)})$ close to $(\bar{x}, \bar{\lambda})$ one has to impose an additional condition on $\tilde{F}$ given by (5.3), e.g., the metric regularity of $\tilde{F}$ around $\big((\bar{x}, \bar{\lambda}), (0,0)\big)$.

On the contrary the approximation step as described in Algorithm 4 requires only the solution of a strictly convex quadratic programming problem and the Newton step is performed by solving a linear system. Note that Assumptions 1 and 2 do not imply that multifunctions $x \rightrightarrows f(x) + \nabla g(x)^T N_D(g(x))$ or $(x,d) \rightrightarrows F(x,d)$ are metrically regular around $(\bar{x}, 0)$ and $((\bar{x}, g(\bar{x})), (0,0))$, respectively.

*Example* 5.13. Consider the nonlinear complementarity problem

$$0 \in -x - x^2 + N_{\mathbb{R}_-}(x) \tag{5.23}$$

and its solution $\bar{x} = 0$. Since $g(x) = x$, we have $\nabla g(x) = 1$ showing that $(0,0)$ is nondegenerate with modulus 1. We obtain $\bar{\lambda} = 0$ and Assumption 2 follows easily from $\nabla L_{\bar{\lambda}}(\bar{x}) = -1$. Thus Theorem 5.12 applies and we obtain local superlinear convergence of Algorithm 5. Indeed, given $x^{(k)}$, the quadratic program $QP(x^{(k)})$ amounts to

$$\min_{u\in\mathbb{R}} -(x^{(k)} + x^{(k)^2})u + \frac{1}{2}u^2 \text{ subject to } x^{(k)} + u \le 0,$$

which has the solution $u = \min\{x^{(k)} + x^{(k)^2}, -x^{(k)}\}$, resulting in $\hat{d}^{(k)} = x^{(k)} +$

$\min\{x^{(k)}+{x^{(k)}}^2, -x^{(k)}\} = \min\{2x^{(k)}+{x^{(k)}}^2, 0\}$. If $2x^{(k)}+{x^{(k)}}^2 < 0$, then $T_{\mathbb{R}_-}(\hat{d}^{(k)}) = \mathbb{R}$, $\hat{\lambda}^{(k)} = 0$, $\hat{l}^{(k)} = 1$, $\hat{Z}^{(k)} = 1$, and the Newton direction $s_x$ is given by

$$\hat{Z}^{(k)T}\big(\nabla\mathcal{L}_{\hat{\lambda}^{(k)}}(x^{(k)})s_x + \mathcal{L}_{\hat{\lambda}^{(k)}}(x^{(k)})\big) = -(1+2x^{(k)})s_x - (x^{(k)}+{x^{(k)}}^2) = 0 \;\Rightarrow\; s_x = -\frac{x^{(k)}+{x^{(k)}}^2}{1+2x^{(k)}}.$$

This yields $x^{(k+1)} = {x^{(k)}}^2/(1+2x^{(k)})$. On the other hand, if $2x^{(k)}+{x^{(k)}}^2 \geq 0$, then $T_{\mathbb{R}_-}(\hat{d}^{(k)}) = \mathbb{R}_-$, $\hat{\lambda}^{(k)} = 2x^{(k)}+{x^{(k)}}^2$, $\hat{l}^{(k)} = 0$, $\hat{W}^{(k)} = 1$, and the Newton direction $s_x$ is given by

$$\hat{W}^{(k)T}(x^{(k)} + s_x) = 0 \;\Rightarrow\; s_x = -x^{(k)},$$

resulting in $x^{(k+1)} = 0$. Hence we obtain in fact locally quadratic convergence of the sequence produced by Algorithm 5.

Now we want to demonstrate that the Newton–Josephy method does not work for this simple example. At the $k$th iterate the problem (5.22) reads as

$$\begin{aligned} 0 &\in -x^{(k)} - {x^{(k)}}^2 + (-1-2x^{(k)})(x^{(k+1)} - x^{(k)}) + \lambda^{(k+1)} \\ &= -(1+2x^{(k)})x^{(k+1)} + {x^{(k)}}^2 + \lambda^{(k+1)}, 0 \leq \lambda^{(k+1)} \perp x^{(k+1)} \leq 0 \end{aligned}$$

and this auxiliary problem is not solvable for any $x^{(k)}$ with $0 < |x^{(k)}| \leq \frac{1}{2}$. The reason is that the mapping $\tilde{F}$ is not metrically regular at $(\bar{x}, \bar{\lambda})$.

**6. Conclusion.** The crucial notion used in developing the new Newton-type method is the semismooth* property, which pertains not only to single-valued mappings (like the standard semismoothness) but also to sets and multifunctions. The second substantial ingredient in this development consists of a novel linearization of the set-valued part of the considered GE, which is performed on the basis of the respective limiting coderivative. Finally, also very important is the modification of the semismoothness* in Definition 4.8, which enables us to proceed even if the considered multifunction is not semismooth* in the original sense of Definition 3.1.

The new method contains, apart from the Newton step, also the so-called approximation step, having two principal goals. First, it ensures that in the next linearization we dispose with a feasible point and, second, it enables us to avoid points (if they exist) where the imposed regularity assumption is violated. In this way one obtains the local superlinear convergence without imposing a restrictive regularity assumption at the solution point (like the strong BD-regularity in [32]).

The application in section 5 illuminates the fact that the implementation to a concrete class of GEs may be quite demanding. On the other hand, the application area of the new method seems to be very large. It includes, among other things, various complicated GEs corresponding to variational inequalities of the second kind, hemivariational inequalities, etc. Their solution via an appropriate variant of the new method will be the subject of further research.

**Acknowledgments.** The authors are indebted to both reviewers for their valuable comments and suggestions. Further, they would like to express their gratitude to D. Klatte, A.Y. Kruger, and B. Kummer for their great help in the revision of this paper.

REFERENCES


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

[2] J. F. Bonnans and A. Shapiro, *Perturbation Analysis of Optimization Problems*, Springer, New York, 2000.

[3] S. Dias and G. Smirnov, *On the Newton method for set-valued maps*, Nonlinear Anal., 75 (2012), pp. 1219–1230.

[4] A. L. Dontchev, H. Gfrerer, A. Y. Kruger, and J. V. Outrata, *The radius of metric regularity*, Set-Valued Var. Anal., 28 (2020), pp. 451–473.

[5] A. L. Dontchev and R. T. Rockafellar, *Characterizations of strong regularity for variational inequalities over polyhedral convex sets*, SIAM J. Optim., 6 (1996), pp. 1087–1105.

[6] A. L. Dontchev and R. T. Rockafellar, *Implicit Functions and Solution Mappings*, Springer, New York, 2014.

[7] F. Facchinei and J.-S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems*, Vol. II, Springer, New York, 2003.

[8] A. Fischer, *Modified Wilson's method for nonlinear programs with nonunique multipliers*, Math. Oper. Res., 24 (1999), pp. 699–727.

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

[10] H. Gfrerer, *Optimality conditions for disjunctive programs based on generalized differentiation with application to mathematical programs with equilibrium constraints*, SIAM J. Optim., 24 (2014), pp. 898–931.

[11] H. Gfrerer and B. S. Mordukhovich, *Robinson stability of parametric constraint systems via variational analysis*, SIAM J. Optim., 27 (2017), pp. 438–465.

[12] H. Gfrerer and B. S. Mordukhovich, *Second-order variational analysis of parametric constraint and variational systems*, SIAM J. Optim., 29 (2019), pp. 423–53.

[13] H. Gfrerer and J. V. Outrata, *On Lipschitzian properties of implicit multifunctions*, SIAM J. Optim., 26 (2016), pp. 2160–2189.

[14] P. E. Gill, W. Murray, and M. H. Wright, *Practical Optimization*, Academic Press, London, 1981.

[15] R. Henrion and J.V. Outrata, *A subdifferential condition for calmness of multifunctions*, J. Math. Anal. Appl., 258 (2001), pp. 110–130.

[16] T. Hoheisel, Ch. Kanzow, B. S. Mordukhovich, and H. Phan, *Generalized Newton's method based on graphical derivatives*, Nonlinear Anal., 75 (2012), pp. 1324–1340.

[17] A. F. Izmailov and M. V. Solodov, *Newton-Type Methods for Optimization and Variational Problems*, Springer, New York, 2014.

[18] A. F. Izmailov and M. V. Solodov, *Newton-type methods: A broader view*, J. Optim. Theory Appl., 164 (2015), pp. 577–620.

[19] N. H. Josephy, *Newton's Method for Generalized Equations and the PIES Energy Model*, Ph.D. dissertation, Department of Industrial Engineering, University of Wisconsin-Madison, 1979.

[20] N. H. Josephy, *Quasi-Newton Method for Generalized Equations*, Technical summary report, Mathematics Research Center, University of Wisconsin-Madison, 1979.

[21] D. Klatte and B. Kummer, *Nonsmooth Equations in Optimization. Regularity, Calculus, Methods and Applications*, Nonconvex Optim. Appl. 60, Kluwer Academic Publishers, Boston, 2002.

[22] D. Klatte and B. Kummer, *Approximation and generalized Newton methods*, Math Program., 168 (2018), pp. 673–716.

[23] D. Klatte and B. Kummer, *Correction to: Approximations and generalized Newton methods*, Math Program., 179 (2020), pp. 469–471.

[24] B. Kummer, *Newton's method for non-differentiable functions*, in Advances in Mathematical Optimization, J. Guddat et al., eds., Ser. Math. Res. 45, Akademie-Verlag, Berlin, 1988, pp. 114–125.

[25] B. Kummer, *Newton's Method Based on Generalized Derivatives for Nonsmooth Functions: Convergence Analysis*, in Advances in Optimization, W. Oettli and D. Pallaschke, eds., Springer, New York, 1992, pp. 171–194.

[26] S. Lu and S. M. Robinson, *Variational inequalities over perturbed polyhedral convex sets*, Math. Oper. Res., 33 (208), pp. 689–711.

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

[28] B. S. Mordukhovich, *Variational Analysis and Applications*, Springer, New York, 2018.

[29] B. S. Mordukhovich, J. V. Outrata, and H. Ramirez C., *Graphical derivatives and stability analysis for parameterized equilibria with conic constraints*, Set-Valued Var. Anal., 23 (2015), pp. 678–704.

[30] J.V. Outrata, M. Kočvara, and J. Zowe, *Nonsmooth Approach to Optimization Problems with Equilibrium Constraints*, Kluwer Academic Publishers, Boston, 1998.

[31] J.-S. Pang, *Newton's method for B-differentiable equations*, Math. Oper. Res., 15 (1990), pp. 311–341.

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

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

[34] S. M. Robinson, *Constraint non-degeneracy in variational analysis*, Math. Oper. Res., 28 (2003), pp. 201–232.

[35] R. T. Rockafellar and R. J-B. Wets, *Variational Analysis*, Springer, New York, 1998.

[36] A. Shapiro, *On concepts of directional differentiability*, J. Optim. Theory Appl., 66 (1990), pp. 447–480.