

# Two Typical Implementable Semismooth\* Newton Methods for Generalized Equations are G-Semismooth Newton Methods

Liang Chen,<sup>a</sup> Defeng Sun,<sup>b,\*</sup> Wangyongquan Zhang<sup>b</sup>

<sup>a</sup>School of Mathematics, Hunan University, Changsha, China; <sup>b</sup> Department of Applied Mathematics, The Hong Kong Polytechnic University, Hung Hom, Hong Kong

\* Corresponding author

**Contact:** chl@hnu.edu.cn, <https://orcid.org/0000-0003-0398-4669> (LC); defeng.sun@polyu.edu.hk, <https://orcid.org/0000-0003-0481-272X> (DS); wangyongquan.zhang@connect.polyu.hk, <https://orcid.org/0000-0003-3085-9396> (WZ)

Authors are encouraged to submit new papers to INFORMS journals by means of a style file template, which includes the journal title. However, use of a template does not certify that the paper has been accepted for publication in the named journal. INFORMS journal templates are for the exclusive purpose of submitting to an INFORMS journal and are not intended to be a true representation of the article's final published form. Use of this template to distribute papers in print or online or to submit papers to another non-INFORM publication is prohibited.

**Abstract.** Semismooth\* Newton methods have been proposed in recent years targeting multi-valued inclusion problems and have been successfully implemented to deal with several concrete generalized equations. In this paper, we show that two typical implementations of them that are available are exactly the applications of G-semismooth Newton methods for solving nonsmooth equations localized from these generalized equations. This new understanding expands the breadth of G-semismooth Newton methods in theory, results in a few interesting problems regarding the two categories of nonsmooth Newton methods, and, more importantly, provides informative observations in facilitating the design and implementation of practical Newton-type algorithms for solving generalized equations.

**Funding:** This research was supported by the National Key R&D Program of China [No. 2021YFA1001300], the National Natural Science Foundation of China [No. 12271150], the Natural Science Foundation of Hunan Province [No. 2023JJ10001], the Science and Technology Innovation Program of Hunan Province [No. 2022RC1190], the Hong Kong RGC Senior Research Fellow Scheme [No. SRFS2223-5S02], and the GRF Grants [Nos. 15307822 and 15307523].

**Key words:** Semismooth\* Newton method, Semismooth Newton method, Generalized equation, Nonsmooth analysis, Variational analysis

*MSC2000 subject classification:* 49M15, 49J52, 49J53, 65K99

**History:** Received: Month DD, YYYY; Accepted: Month DD, YYYY; Published Online: Month DD, YYYY

**1. Introduction** Starting from the seminal work of Kummer [41], Newton-type methods for solving nonsmooth equations have evolved for several decades. The literature on this topic is

abundant; one may refer to [28, 29, 32, 42, 51, 54, 59] and references therein. Nowadays, nonsmooth Newton methods have been heavily incorporated into efficient numerical optimization software for large-scale optimization problems [44, 63, 64]. For generalized equations (GEs), nonsmooth Newton methods have also been extensively studied. In the pioneering works [33, 34] of this field, Josephy considered the GEs in the form of

$$0 \in H(x) + \Theta(x),$$

where  $H : \mathcal{X} \rightarrow \mathcal{Y}$  is a single-valued function,  $\Theta : \mathcal{X} \rightrightarrows \mathcal{Y}$  is a multifunction, and  $\mathcal{X}$  and  $\mathcal{Y}$  are finite-dimensional real Hilbert spaces, each endowed with an inner product and its induced norm. Studies in this direction include [1, 2, 5, 7, 9–15, 17–19, 30–32, 50, 61], to name only a few. In these algorithms, the single-valued part  $H$  is linearized while the multi-valued part  $\Theta$  is kept. So, the subproblems are linearized generalized equations. Alternatively, the constructions of Newton-type methods for nonsmooth GEs by approximating the multi-valued part  $\Theta$  have been investigated in [4, 8, 17, 28, 40]. Significant progress has also been made in Newton-type methods with subproblems being coderivative (or graphical derivative) inclusions [3, 35–38, 48, 49].

Recently, a semismooth\* Newton method was proposed in [23] for solving GEs in the form of

$$0 \in \Phi(x), \tag{1}$$

where  $\Phi : \mathcal{X} \rightrightarrows \mathcal{Y}$  is a set-valued mapping with closed graph, and was extended in [24] using subspace containing derivatives (SCD). A particular property of these semismooth\* Newton methods is that the subproblems involved are linear systems of equations, which is different from the methods mentioned above for solving (1). Moreover, since an inequality involving the target solution should be properly fulfilled in the “approximation step” at each iteration of these algorithms, specific implementations of these algorithms should be elaborately designed to make them practical. These implementations are rarely available in the literature. However, the two executable representatives of them, even seemingly sophisticated, are of practical value.

Specifically, on the one hand, the semismooth\* Newton method in [23] for (1) was implemented in [23, Section 5] for solving the GE

$$0 \in F(x) + \nabla G(x)N_D(G(x)), \tag{2}$$

where  $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$  is continuously differentiable,  $G : \mathbb{R}^n \rightarrow \mathbb{R}^s$  is twice continuously differentiable,  $D \subseteq \mathbb{R}^s$  is a convex polyhedral set,  $\nabla G(x)$  denotes the adjoint of the Jacobian operator  $G'(x)$ , and

$N_D(\cdot)$  denotes the normal cone mapping used in convex analysis. As mentioned in [23], the GE (2) arises frequently in optimization and equilibrium models, and it is equivalent to the GE given by  $0 \in F(x) + N_{G^{-1}(D)}(x)$  under certain constraint qualifications [22, Proposition 2.1]. On the other hand, in [25], the SCD semismooth\* Newton method was applied to the GE

$$0 \in F(x) + \partial q(x), \quad (3)$$

where  $\partial q$  is the subdifferential mapping of a given closed proper convex function  $q : \mathbb{R}^n \rightarrow (-\infty, \infty]$ . Such an implementation was further extended in [21] to a more general class of GEs.

The GEs (2) and (3) are of significant importance for taking a closer look at these semismooth\* Newton methods. Since the corresponding subproblems for computing the Newton directions are linear systems of equations, which are akin to the subproblems of semismooth Newton methods, it is natural to ask whether these implementations admit a connection to the existing Newton-type methods for solving nonsmooth equations. Note that, when  $\Phi$  in (1) is single-valued and locally Lipschitz continuous around a point  $\bar{x}$ , it is easy to see that the semismooth\* property of  $\Phi$  at  $\bar{x}$  is exactly G-semismoothness (semismoothness in the sense of Gowda [26]) of  $\Phi$  at  $\bar{x}$ , for example, by [23, Proposition 3.7]. Moreover, when solving a locally Lipschitz continuous equation, the relationship between the semismooth\* Newton method and the G-semismooth Newton method of Kummer [41] (c.f. Section 2.2 for details) has been discussed in [23, Sect. 4]. For solving GEs beyond nonsmooth equations, the relationship between the two types of nonsmooth Newton methods is unknown.

In this paper, by reformulating the GEs (2) and (3) as nonsmooth equations, which are proved to be locally Lipschitz continuous, we show that the corresponding practical implementations of semismooth\* Newton methods are exactly the applications of G-semismooth Newton methods. Specifically, we show that the algorithm implemented in [23, Section 5] for solving (2) is an application of a G-semismooth Newton method for solving an implicitly defined equation. Furthermore, for the algorithm in [25] for solving (3), we take the proximal residual mapping as the Lipschitz continuous localization of (3), and show that the implemented SCD semismooth\* Newton method is also an application of a G-semismooth Newton method. Additionally, we show that the conditions for ensuring the convergence of these semismooth\* Newton methods are sufficient for the corresponding applications of the G-semismooth Newton methods. Therefore, one can conclude that these implementable semismooth\* Newton methods are G-semismooth Newton methods. This leads to a concrete foundation for comprehending semismooth\* Newton methods and is beneficial

for developing practical nonsmooth Newton methods for solving GEs, especially considering globalization. Here, we emphasize that we focus on the local convergence properties. For globalizing the G-semismooth Newton method, one may refer to the (inexact) smoothing Newton methods studied in [16, 53] (note that though semismoothness was used in the cited two references, a quick examination reveals that G-semismoothness is sufficient for convergence and rate of convergence analysis). In addition, we only consider GEs or nonsmooth equations instead of  $C^{1,1}$  optimization problems, for which traditional globalized G-semismooth Newton methods generally require the directional differentiability of the gradient mapping. This requirement can be removed by involving the Lipschitz constant of the gradient mapping and a modulus for local stability in the line search as in the recent coderivative-based nonsmooth Newton methods for  $C^{1,1}$  optimization problems or composite optimization problems with  $C^{1,1}$  envelopes [36, 38, 48]. Most recently, the issue of globalizing the semismooth\* Newton method for nonconvex composite optimization problems has also been discussed [20].

The remaining parts of this paper are organized as follows. In Section 2, we collect some basic results in variational analysis and briefly introduce the G-semismooth Newton method.

In Section 3, the implementable semismooth\* Newton methods in [23, Section 5] and [25] are introduced, together with some intermediate results which are necessary for further discussions. In Section 4 and Section 5, we show that these executable implementations of semismooth\* Newton methods are applications of G-semismooth Newton methods for solving nonsmooth equations involving locally Lipschitz continuous functions. This constitutes the main contribution of this work. We conclude our paper in Section 6.

**2. Preliminaries** This section presents the definitions and necessary tools from variational analysis [47, 48, 57]. It also provides preliminary results and reviews the G-semismooth Newton method.

**2.1. Basic variational analysis** Let  $\mathcal{X}$  and  $\mathcal{Y}$  be two finite-dimensional real Hilbert spaces each equipped with an inner product  $\langle \cdot, \cdot \rangle$  and its induced norm  $\| \cdot \|$ . For any  $x \in \mathcal{X}$  and  $\delta > 0$ ,  $\mathbb{B}_\delta(x)$  denotes the closed ball centered at  $x$  with radius  $\delta$ , and  $\mathbb{B}_\mathcal{X}$  and  $\mathbb{B}_\mathcal{Y}$  are the unit balls in  $\mathcal{X}$  and  $\mathcal{Y}$ , respectively. Moreover,  $[x]$  denotes the subspace spanned by the given vector  $x \in \mathcal{X}$ . The notation  $(\cdot; \cdot)$  means two vectors or linear operators are stacked symbolically in column order. For a subspace  $\mathcal{X}_0$  of  $\mathcal{X}$ , we use  $\mathcal{X}_0^\perp$  to denote its orthogonal complement in  $\mathcal{X}$ . We use  $\mathbb{L}(\mathcal{X}, \mathcal{Y})$  to represent the space of all linear operators from  $\mathcal{X}$  to  $\mathcal{Y}$ , and write  $\mathbb{L}(\mathcal{X}) \equiv \mathbb{L}(\mathcal{X}, \mathcal{X})$  for convenience.

For an arbitrary linear operator  $V$ , we use  $\text{rge}V$  to denote its range space,  $\ker V$  to denote its null space. If  $V$  is a matrix, we use  $V^\top$  to denote its transpose.

For a nonempty set  $C \subseteq \mathcal{X}$ , we use  $\text{ri}C$  and  $\text{int}C$  to denote the relative interior and interior of  $C$ , respectively. The lineality space of  $C$ , denoted by  $\text{lin}C$ , is the largest linear subspace contained in  $C$ . Meanwhile, we use  $\text{span}C$  to denote the smallest linear subspace that contains  $C$ . When  $C$  is locally closed at  $\bar{x} \in C$ , the contingent (Bouligand) cone  $T_C(\bar{x})$ , the regular (Fréchet) normal cone  $\widehat{N}_C(\bar{x})$  and the limiting (Mordukhovich) normal cone  $N_C(\bar{x})$  to  $C$  at  $\bar{x}$  are defined, respectively, by

$$T_C(\bar{x}) := \limsup_{t \searrow 0} \frac{C - \bar{x}}{t}, \quad \widehat{N}_C(\bar{x}) := (T_C(\bar{x}))^\circ, \quad \text{and} \quad N_C(\bar{x}) := \limsup_{x \rightarrow \bar{x}, x \in C} \widehat{N}_C(x).$$

Furthermore,  $\mathcal{K}_C(\bar{x}, d) := T_C(\bar{x}) \cap [d]^\perp$  is the critical cone to  $C$  at  $\bar{x} \in C$  with respect to  $d \in \widehat{N}_C(\bar{x})$ .

If  $K \subseteq \mathcal{X}$  is a closed convex cone, we use  $K^\circ$  to denote its polar, i.e.,  $K^\circ := \{x \in \mathcal{X} \mid \langle x, x' \rangle \leq 0 \forall x' \in K\}$ . In this case, one has  $\text{lin}K = K \cap (-K)$  and  $\text{span}K = K + (-K)$ . Moreover, it holds  $(\text{lin}K)^\perp = \text{span}K^\circ$  and  $(\text{span}K)^\perp = \text{lin}K^\circ$ .

For a set-valued mapping  $\Phi : \mathcal{X} \rightrightarrows \mathcal{Y}$ , we use  $\text{gph}\Phi$  to denote its graph in  $\mathcal{X} \times \mathcal{Y}$ . The mapping  $\Phi$  is called outer semicontinuous at  $x$  if for any  $\varepsilon > 0$  there exists  $\delta > 0$  such that  $\Phi(x') \subseteq \Phi(x) + \varepsilon \mathbb{B}_{\mathcal{Y}}$  holds for all  $x' \in x + \delta \mathbb{B}_{\mathcal{X}}$ . When  $\text{gph}\Phi$  is (locally) closed, the regular (Fréchet) coderivative and the limiting (Mordukhovich) coderivative of  $\Phi$  at  $(\bar{x}, \bar{y})$  are the multifunctions  $\widehat{D}^*\Phi(\bar{x}, \bar{y}) : \mathcal{Y} \rightrightarrows \mathcal{X}$  and  $D^*\Phi(\bar{x}, \bar{y}) : \mathcal{Y} \rightrightarrows \mathcal{X}$ , respectively, such that

$$\begin{cases} \widehat{D}^*\Phi(\bar{x}, \bar{y})(v^*) := \{u^* \in \mathcal{X} \mid (u^*; -v^*) \in \widehat{N}_{\text{gph}\Phi}(\bar{x}, \bar{y})\} & \forall v^* \in \mathcal{Y}, \\ D^*\Phi(\bar{x}, \bar{y})(v^*) := \{u^* \in \mathcal{X} \mid (u^*; -v^*) \in N_{\text{gph}\Phi}(\bar{x}, \bar{y})\} & \forall v^* \in \mathcal{Y}. \end{cases}$$

If  $\Phi$  is single-valued, one can write the two coderivatives as  $\widehat{D}^*\Phi(\bar{x})$  and  $D^*\Phi(\bar{x})$  for simplicity. If  $\Phi$  is Fréchet differentiable at  $\bar{x}$ , by [47, Theorem 1.38] one has  $\widehat{D}^*\Phi(\bar{x})(v^*) = \{\nabla\Phi(\bar{x})v^*\}$ , where  $\nabla\Phi(\bar{x})$  is the adjoint of the Fréchet derivative  $\Phi'(\bar{x})$ . If  $\Phi$  is strictly differentiable at  $\bar{x}$ , one also has  $D^*\Phi(\bar{x})(v^*) = \{\nabla\Phi(\bar{x})v^*\}$ . If  $\Phi$  is Lipschitz continuous in an open neighborhood  $\Omega$  of  $\bar{x}$ , from Rademacher's theorem [55] we know  $\Phi$  is almost everywhere Fréchet differentiable in  $\Omega$ . In this case, the Bouligand subdifferential of  $\Phi(\cdot)$  at  $\bar{x}$  is defined by

$$\partial_B\Phi(\bar{x}) := \left\{ \lim_{k \rightarrow \infty} \Phi'(x^{(k)}) \mid \Phi \text{ is differentiable at } x^{(k)}, x^{(k)} \rightarrow \bar{x} \right\} \quad (4)$$

and Clarke's generalized Jacobian of  $\Phi$  at  $\bar{x}$  is defined by  $\partial\Phi(\bar{x}) := \text{conv}\partial_B\Phi(\bar{x})$ , i.e., by taking the convex hull of the Bouligand subdifferential.

In [23], a generalization of the coderivatives was introduced. Specifically, for  $\Phi : \mathcal{X} \rightrightarrows \mathcal{Y}$  with closed graph, one can let  $\widehat{\mathcal{D}}^*\Phi : \text{gph}\Phi \rightarrow (\mathcal{Y} \rightrightarrows \mathcal{X})$  be a mapping such that for every pair  $(x; y)$  in  $\text{gph}\Phi$ , the set  $\text{gph}\widehat{\mathcal{D}}^*\Phi(x, y)$  is a cone. One can define the associated limiting mapping  $\mathcal{D}^*\Phi : \text{gph}\Phi \rightarrow (\mathcal{Y} \rightrightarrows \mathcal{X})$  by

$$\text{gph}\mathcal{D}^*\Phi(x, y) := \limsup_{(x'; y') \xrightarrow{\text{gph}\Phi} (x; y)} \text{gph}\widehat{\mathcal{D}}^*\Phi(x', y').$$

Here,  $\widehat{\mathcal{D}}^*$  and  $\mathcal{D}^*$  serve as the generalizations of the regular and limiting coderivatives  $\widehat{D}^*$  and  $D^*$ , respectively. In [23], the notion of semismoothness\* was originally proposed for sets and is equivalent to the semismoothness of sets in [27, Definition 2.3]. The following definition of generalized semismoothness\* comes from its application and generalization to  $\text{gph}\Phi$ .

**DEFINITION 1** ([23, DEFINITION 4.8]). Let  $\Phi : \mathcal{X} \rightrightarrows \mathcal{Y}$  and  $(\bar{x}; \bar{y}) \in \text{gph}\Phi$ , which is nonempty and closed. Then,  $\Phi$  is called semismooth\* at  $(\bar{x}; \bar{y})$  with respect to  $\mathcal{D}^*\Phi$  if for every  $\epsilon > 0$  there is some  $\delta > 0$  such that

$$|\langle x^*, x - \bar{x} \rangle - \langle y^*, y - \bar{y} \rangle| \leq \epsilon \| (x; y) - (\bar{x}; \bar{y}) \| \| (x^*; y^*) \| \quad \forall (x; y) \in \mathbb{B}_\delta(\bar{x}; \bar{y}), \forall (y^*; x^*) \in \text{gph}\widehat{\mathcal{D}}^*\Phi(x, y).$$

Finally, we discuss proximal mappings and projections. For a maximal monotone mapping  $\mathcal{M} : \mathcal{X} \rightrightarrows \mathcal{X}$ , the corresponding proximal mapping is defined by  $\text{prox}_{\lambda\mathcal{M}} := (I + \lambda\mathcal{M})^{-1}$ ,  $\lambda > 0$ , which is single-valued and Lipschitz continuous with unit Lipschitz constant, where  $I$  represents the identity operator. Given a closed proper convex function  $q : \mathcal{X} \rightarrow (-\infty, +\infty]$  and a parameter  $\lambda > 0$ , its subdifferential mapping  $\partial q$  is always maximal monotone. Taking  $\mathcal{M} \equiv \partial q$ , it is easy to see that

$$\text{prox}_{\lambda\mathcal{M}}(x) = \mathcal{P}_{\lambda q}(x) := \arg \min_z \left\{ q(z) + \frac{1}{2\lambda} \|z - x\|^2 \right\} \quad \forall x \in \mathcal{X}. \quad (5)$$

Therefore, for a nonempty closed convex subset  $C \subseteq \mathcal{X}$  with  $\delta_C$  being its indicator function, the projection mapping (with respect to  $\|\cdot\|$ ) can be defined by  $\Pi_C(x) := \mathcal{P}_{\lambda\delta_C}(x)$  for any  $\lambda > 0$ . The following lemma on the Bouligand subdifferential of the projection mapping onto a convex polyhedral set is necessary for subsequent discussions.

**LEMMA 1.** Let  $Q \subset \mathbb{R}^l$  be a nonempty polyhedral convex set. Then for a given  $\mu \in \mathbb{R}^l$ ,

$$\partial_B \Pi_Q(\mu) = \{ \Pi_{\text{span}\mathcal{G}}(\cdot) \mid \mathcal{G} \text{ is a face of } K(\mu) := \mathcal{K}_Q(\Pi_Q(\mu), \mu - \Pi_Q(\mu)) \}.$$

*Proof* We know from [51, Lemma 5(i)] and the definition of  $K(\mu)$  that for any  $\Delta\mu \in \mathbb{R}^l$  with  $\|\Delta\mu\|$  being sufficiently small, it holds that

$$\Pi_Q(\mu + \Delta\mu) = \Pi_Q(\mu) + \Pi_{K(\mu)}(\Delta\mu). \quad (6)$$

Moreover, if  $\Pi_Q$  is differentiable at  $\mu + \Delta\mu$ , one has from [51, Lemma 5(ii)] that

$$\Pi'_Q(\mu + \Delta\mu) = \Pi_{\text{lin}K(\mu + \Delta\mu)}(\cdot) = \Pi_{\text{lin}T_Q(\Pi_Q(\mu + \Delta\mu))}(\cdot),$$

where the last equality comes from (2.3) of [23, Lemma 2.4]. Note that  $\lim_{\Delta\mu \rightarrow 0} \Pi_{K(\mu)}(\Delta\mu) = 0$ . Therefore, from [23, Lemma 2.4] we know that for every  $\Delta\mu$  with  $\|\Delta\mu\|$  sufficiently small one has  $\text{lin}K(\mu + \Delta\mu) = \text{span}\mathcal{G}$  with  $\mathcal{G}$  being a face of  $K(\mu)$ . Consequently, one has  $\Pi'_Q(\mu + \Delta\mu) = \Pi_{\text{span}\mathcal{G}}(\cdot)$ , so that

$$\partial_B \Pi_Q(\mu) \subseteq \{\Pi_{\text{span}\mathcal{G}}(\cdot) \mid \mathcal{G} \text{ is a face of } K(\mu)\}. \quad (7)$$

On the other hand, let  $\mathcal{G}$  be an arbitrary face of  $K(\mu)$ . It holds  $\mathcal{G} = K(\mu) \cap [\nu]^\perp$  for some  $\nu \in K(\mu)^\circ$ . Since  $\mathcal{G}$  is a closed convex set, one has from [58, Theorem 6.2] that  $\text{ri}\mathcal{G}$  is nonempty. Let  $\tilde{\mu} \in \text{ri}\mathcal{G} \subseteq K(\mu)$  be fixed. it holds that  $T_{\mathcal{G}}(\tilde{\mu}) = \text{span}\mathcal{G}$ . Moreover, from Moreau's decomposition theorem [58, Theorem 31.5] one can get  $\Pi_{K(\mu)}(\tilde{\mu} + \nu) = \tilde{\mu}$ . Then, by [51, Lemma 5.1(i)] we know that for all  $\Delta\mu \in \mathbb{R}^l$  with  $\|\Delta\mu\|$  sufficiently small, it holds that  $\Pi_{K(\mu)}(\tilde{\mu} + \nu + \Delta\mu) = \tilde{\mu} + \Pi_{T_{K(\mu)}(\tilde{\mu}) \cap [\nu]^\perp}(\Delta\mu) = \tilde{\mu} + \Pi_{\text{span}\mathcal{G}}(\Delta\mu)$ . Thus,  $\Pi_{K(\mu)}(\cdot)$  is differentiable at  $\tilde{\mu} + \nu$  with  $\Pi'_{K(\mu)}(\tilde{\mu} + \nu) = \Pi_{\text{span}\mathcal{G}}(\cdot)$ . Note that for any integer  $k > 0$ , one has  $\mathcal{G} = K(\mu) \cap [\nu/k]^\perp$ . Meanwhile, as  $\mathcal{G}$  is a closed convex cone, one has  $\tilde{\mu}/k \in \text{ri}\mathcal{G}$ . Therefore,  $\Pi_{K(\mu)}(\cdot)$  is differentiable at  $(\tilde{\mu} + \nu)/k$  for all  $k > 0$  with  $\Pi'_{K(\mu)}(\tilde{\mu}/k + \nu/k) = \Pi_{\text{span}\mathcal{G}}(\cdot)$ . Consequently, by (6) one has

$$\Pi'_Q(\mu + (\tilde{\mu}/k + \nu/k)) = \Pi'_{K(\mu)}(\tilde{\mu}/k + \nu/k) = \Pi_{\text{span}\mathcal{G}}(\cdot).$$

Taking limits in the above equality along with  $k \rightarrow \infty$ , one gets  $\Pi_{\text{span}\mathcal{G}}(\cdot) \in \partial_B \Pi_Q(\mu)$ . This, together with (7), completes the proof of the lemma.  $\square$

**2.2. A G-semismooth Newton method** The terminology ‘‘G-semismoothness’’ was coined in [52] for distinguishing the definition of semismoothness in [26] by Gowda from those in [46, 54] involving directional differentiability. Specifically, let  $\Omega \subseteq \mathcal{X}$  be an open set,  $H : \Omega \rightarrow \mathcal{Y}$  be a continuous function, and  $\mathcal{T} : \Omega \rightrightarrows \mathbb{L}(\mathcal{X}, \mathcal{Y})$  be a set-valued mapping. According to [26, Definition 2], we say  $H$  is called G-semismooth with respect to  $\mathcal{T}$  at  $x \in \Omega$  if for any  $h \rightarrow 0$  and  $V \in \mathcal{T}(x + h)$



it holds that  $H(x+h) - H(x) - Vh = o(\|h\|)$ . When  $H$  is locally Lipschitz continuous around  $x$ , it is simply called G-semismooth at  $x$  if  $\mathcal{T}$  is taken as  $\partial H$ , and this definition is invariant if  $\partial H$  is replaced by  $\partial_B H$  (cf. [26, Section 2.3]). The following G-semismooth Newton method, based on G-semismoothness, is a trivial inexact extension of [41].

---

**Algorithm 1** A G-semismooth Newton method for solving nonsmooth equations

---

**Input:**  $x^{(0)} \in \mathcal{X}$ ,  $H : \Omega \subseteq \mathcal{X} \rightarrow \mathcal{Y}$ ,  $\mathcal{T} : \mathcal{X} \rightrightarrows \mathbb{L}(\mathcal{X}, \mathcal{Y})$ , and  $\varrho \geq 0$ .

**Output:**  $\{x^{(k)}\}$ .

**For**  $k = 0, 1, \dots$ ,

1. if  $H(x^{(k)}) = 0$ , stop the algorithm;
  2. select  $V_k \in \mathbb{L}(\mathcal{X}, \mathcal{Y})$  such that  $\text{dist}(V_k, \mathcal{T}(x^{(k)})) \leq \varrho \|H(x^{(k)})\|$ , compute  $\Delta x^{(k)}$  via solving  $V_k \Delta x = -H(x^{(k)})$ , and obtain  $x^{(k+1)} := x^{(k)} + \Delta x^{(k)}$ .
- 

**THEOREM 1.** *Let  $\Omega \subseteq \mathcal{X}$  be an open set. Suppose that  $H : \Omega \rightarrow \mathcal{Y}$  is locally Lipschitz continuous (with modulus  $\vartheta > 0$ ) and G-semismooth with respect to  $\mathcal{T} : \Omega \rightrightarrows \mathbb{L}(\mathcal{X}, \mathcal{Y})$  at  $\bar{x}$  such that  $H(\bar{x}) = 0$ . Assume that  $\mathcal{T}(\cdot)$  is compact valued and outer semicontinuous at  $\bar{x}$ , and  $V^{-1}$  exists for all  $V \in \mathcal{T}(\bar{x})$ . Then there exists a neighborhood of  $\bar{x}$  such that, for any  $x^{(0)}$  in it, Algorithm 1 either terminates in finitely many steps or generates an infinite sequence  $\{x^{(k)}\}$  satisfying  $\|x^{(k+1)} - \bar{x}\| = o(\|x^{(k)} - \bar{x}\|)$  as  $k \rightarrow \infty$ .*

*Proof* Note that  $\|H(x)\| \leq \vartheta \|x - \bar{x}\|$  for all  $x$  sufficiently close to  $\bar{x}$ . Meanwhile, the G-semismoothness of  $H$  at  $\bar{x}$  implies that the multifunction  $\mathcal{T}(x) + \vartheta \varrho \|x - \bar{x}\| \mathbb{B}_{\mathbb{L}(\mathcal{X}, \mathcal{Y})}$  is a Newton map (c.f. [43] or [40, Definition 2] for the definition) for  $H$  at  $\bar{x}$ . As  $\mathcal{T}(\cdot)$  is compact valued and outer semicontinuous at  $\bar{x}$  and all  $V \in \mathcal{T}(\bar{x})$  are nonsingular, the Newton-regularity condition [40, Definition 3] holds at  $\bar{x}$ . So the convergence properties of Algorithm 1 follow from [40, Theorem 4] (or [39, Lemma 10.1]).  $\square$

**3. Implementable semismooth\* Newton methods for GEs** This section reviews the two typical semismooth\* Newton methods that are implementable to concrete GEs.

**3.1. A semismooth\* Newton method for the GE (2)** In [23, Section 5], the semismooth\* Newton method [23, Algorithm 3] for solving the GE (1) was implemented to (2) by introducing an auxiliary variable  $d \in \mathbb{R}^s$  and solving the equivalent problem

$$0 \in \mathcal{H}(x, d) := \begin{pmatrix} F(x) + \nabla G(x) N_D(d) \\ G(x) - d \end{pmatrix}. \quad (8)$$



Note that  $\bar{x}$  solves (2) if and only if  $(\bar{x}, \bar{d}) = (\bar{x}, G(\bar{x}))$  solves (8). For convenience, define the Lagrangian function

$$\mathcal{L}_\lambda(x) := F(x) + \nabla G(x)\lambda \quad \forall (x; \lambda) \in \mathbb{R}^n \times \mathbb{R}^s.$$

For a given point  $\hat{z} := ((\hat{x}, \hat{d}); (\hat{p}^*, G(\hat{x}) - \hat{d})) \in \text{gph}\mathcal{H}$ , one can choose  $\hat{\lambda} \in N_D(\hat{d})$  such that  $\hat{p}^* = \mathcal{L}_{\hat{\lambda}}(\hat{x})$ . Moreover, one can define for all  $(p; q^*) \in \mathbb{R}^n \times \mathbb{R}^s$  the mapping

$$\mathbb{T}(\hat{x}, \hat{d}, \hat{\lambda})(p, q^*) := \{(\nabla \mathcal{L}_{\hat{\lambda}}(\hat{x})p + \nabla G(\hat{x})q^*, d^*) \mid d^* + q^* \in \widehat{D}^* N_D(\hat{d}, \hat{\lambda})(G'(\hat{x})p)\}. \quad (9)$$

In this case, according to [23, eq. (5.5)], the regular coderivative of  $\mathcal{H}$  at  $\hat{z}$  satisfies

$$\widehat{D}^*\mathcal{H}(\hat{z})(p, q^*) \subseteq \mathbb{T}(\hat{x}, \hat{d}, \hat{\lambda})(p, q^*). \quad (10)$$

For implementing the semismooth\* Newton method [23, Algorithm 3] to (8), a mapping  $\widehat{D}^*\mathcal{H}$  that surrogates  $\widehat{D}^*\mathcal{H}$  has been specified based on (9). Furthermore, by defining  $\mathcal{D}^*\mathcal{H}$  as the outer limit of  $\widehat{D}^*\mathcal{H}$ , it is known from [23, Theorem 5.5] that the mapping  $\mathcal{H}$  in (8) is semismooth\* with respect to  $\mathcal{D}^*\mathcal{H}$  at every point  $((x, G(x)); (0, 0)) \in \text{gph}\mathcal{H}$ . Then, the “approximation step” in [23] was given as the following algorithm.

---

**Algorithm 2** An approximation step

---

**Input:**  $x \in \mathbb{R}^n$ .

**Output:**  $\hat{u} \in \mathbb{R}^n, \hat{x} \in \mathbb{R}^n, \hat{\lambda} \in \mathbb{R}^s, \hat{d} \in \mathbb{R}^s, \hat{p}^* \in \mathbb{R}^n$ .

1. compute

$$\hat{u} = \arg \min_{u \in \mathbb{R}^n} \left\{ \frac{1}{2} \|u\|^2 + \langle F(x), u \rangle \mid G(x) + G'(x)u \in D \right\} \quad (11)$$

together with a multiplier  $\hat{\lambda} \in N_D(G(x) + G'(x)\hat{u})$  satisfying  $\hat{u} + F(x) + \nabla G(x)\hat{\lambda} = \hat{u} + \mathcal{L}_{\hat{\lambda}}(x) = 0$ ;

2. set  $\hat{x} := x$ ,  $\hat{d} := G(x) + G'(x)\hat{u}$ ,  $\hat{p}^* := \mathcal{L}_{\hat{\lambda}}(\hat{x})$ , and  $\hat{y} := (\hat{p}^*, G(\hat{x}) - \hat{d})$ .

---

The semismooth\* Newton method in [23, Section 5] for solving (2) is given as Algorithm 3.

**Algorithm 3** An implementable semismooth\* Newton method for solving (2)**Input:**  $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$ ,  $G : \mathbb{R}^n \rightarrow \mathbb{R}^s$  and  $x^{(0)} \in \mathbb{R}^n$ .**Output:**  $\{x^{(k)}\}$ .**For**  $k = 0, 1, \dots$ ,

1. if  $x^{(k)}$  solves (2), stop the algorithm;
2. run Algorithm 2 with input  $x^{(k)}$  to compute  $\hat{\lambda}^{(k)}$ ,  $\hat{d}^{(k)}$  and  $\mathcal{L}_{\hat{\lambda}^{(k)}}(\hat{x}^{(k)})$ ;
3. set  $\hat{l}^{(k)} = \dim(\text{lin } T_D(\hat{d}^{(k)}))$  and compute an  $s \times (s - \hat{l}^{(k)})$  matrix  $\hat{W}^{(k)}$ , whose columns form a basis for  $\text{span } N_D(\hat{d}^{(k)})$  and then an  $n \times (n - (s - \hat{l}^{(k)}))$  matrix  $\hat{Z}^{(k)}$ , whose columns are an orthogonal basis for  $\ker(\hat{W}^{(k)\top} G'(x^{(k)}))$ ;
4. set  $x^{(k+1)} := x^{(k)} + \Delta x^{(k)}$  with the Newton direction  $\Delta x^{(k)}$  being a solution to the linear system

$$\begin{cases} \hat{Z}^{(k)\top} (\mathcal{L}'_{\hat{\lambda}^{(k)}}(x^{(k)}) \Delta x^{(k)} + \mathcal{L}_{\hat{\lambda}^{(k)}}(x^{(k)})) = 0, \\ \hat{W}^{(k)\top} (G(x^{(k)}) + G'(x^{(k)}) \Delta x^{(k)} - \hat{d}^{(k)}) = 0. \end{cases} \quad (12)$$

Recall that a point  $(x, d)$  is called nondegenerate with modulus  $\gamma > 0$  to the GE (2) if one has  $\|\nabla G(x)\lambda\| \geq \gamma\|\lambda\|$  for all  $\lambda \in \text{span } N_D(d)$ . It is called nondegenerate if the above condition holds with some  $\gamma > 0$ . The following assumption [23, Assumption 1] is essential for Algorithm 2.

**ASSUMPTION 1.**  $(\bar{x}, G(\bar{x}))$  is a nondegenerate solution to (8) with modulus  $\tilde{\gamma} > 0$ .

**REMARK 1.** The point  $(x, d)$  is called nondegenerate if and only if  $G'(x)\mathbb{R}^n + \text{lin } T_D(d) = \mathbb{R}^s$ . Moreover, from [23, Remark 5.3 and Lemma 5.4] it holds that for any  $\hat{z} := ((\hat{x}, \hat{d}); (\hat{p}^*, G(\hat{x}) - \hat{d})) \in \text{gph } \mathcal{H}$  with  $(\hat{x}, \hat{d})$  being nondegenerate, one has (10) holds as equality and there exists only one  $\lambda \in N_D(\hat{d})$ , denoted by  $\hat{\lambda}(\hat{x}, \hat{d}, \hat{p}^*)$ , such that  $\hat{p}^* = F(\hat{x}) + \nabla G(\hat{x})\lambda$ .

Then one has the following result (there is a typo in [23, eq. (5.13)], and we take the revised form).

**PROPOSITION 1 ([23, Proposition 5.7]).** Under Assumption 1, there exists a positive radius  $\omega$  and positive reals  $\beta, \beta_u$ , and  $\beta_\lambda$  such that for all  $x \in \mathbb{B}_\omega(\bar{x})$ , the quadratic program in (11) is well-defined and admits a unique solution  $\hat{u}$ , and the output of Algorithm 2 satisfies  $\|\hat{u}\| \leq \beta_u \|\hat{x} - \bar{x}\|$ ,

$$\|((\hat{x}, \hat{d}); \hat{y}) - ((\bar{x}, G(\bar{x})); (0; 0))\| \leq \beta \|\hat{x} - \bar{x}\|, \quad \text{and} \quad \|\hat{\lambda} - \bar{\lambda}\| \leq \beta_\lambda \|\hat{x} - \bar{x}\|,$$

where  $\bar{\lambda}$  is the unique multiplier for  $\bar{x}$ .

Further,  $(\hat{x}, \hat{d})$  is nondegenerate with modulus  $\tilde{\gamma}/2$  and  $N_D(\hat{d}) \subseteq N_D(G(\bar{x}))$ .

The following assumption [23, Assumption 2] provides a regularity condition that guarantees that the linear system (12) admits a unique solution (there is a typo in [23] and we use the corrected form here).

**ASSUMPTION 2.** *For any 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  $\{u \mid G'(\bar{x})u \in \text{span}\mathcal{F}\}$ , such that the matrix  $Z_{\mathcal{F}}^T \mathcal{L}'_{\bar{\lambda}}(\bar{x}) Z_{\mathcal{F}}$  is nonsingular.*

According to [23, Theorem 5.12], under Assumptions 1 and 2, there exists a neighborhood  $U$  of  $\bar{x}$  such that for every starting point  $x^{(0)} \in U$ , Algorithm 3 either stops after finitely many iterations at a solution of (2) or produces a sequence  $\{x^{(k)}\}$  converging superlinearly to  $\bar{x}$ .

**3.2. An SCD semismooth\* Newton method for the GE (3)** In [25], the SCD semismooth\* Newton method proposed in [24] for (1) (with  $\mathcal{X} \equiv \mathcal{Y}$ ) was implemented to the GE (3). As was observed in [24], when applying the semismooth\* Newton methods to (3), it is advantageous to work with linear subspaces  $L \subseteq \mathcal{X} \times \mathcal{X}$  having the same dimension  $n$  with  $\mathcal{X}$  and contained in the graph of the limiting coderivative at a certain point  $(x; y) \in \text{gph}\Phi$ , i.e.,  $L \subseteq \text{gph}D^*\Phi(x, y)$ . Specifically, denote by  $\mathbb{Z}_n$  the metric space of all  $n$ -dimensional subspaces of  $\mathcal{X} \times \mathcal{X}$  equipped with the metric  $d_{\mathbb{Z}_n}(L_1, L_2) := \|\Pi_{L_1} - \Pi_{L_2}\|$ , where  $\Pi_{L_i}$  is the projection operator on  $L_i$ ,  $i = 1, 2$ . Further, define  $L^* := \{(-v^*; u^*) \mid (u^*; v^*) \in L^\perp\}$  for any  $L \in \mathbb{Z}_n$ . According to [24, Definition 3.3],  $\Phi : \mathcal{X} \rightrightarrows \mathcal{X}$  with closed graph is called graphically smooth of dimension  $n$  at  $(x, y)$ , if  $(x; y) \in \text{gph}\Phi$  and  $T_{\text{gph}\Phi}(x; y) \in \mathbb{Z}_n$ . Denote by  $\mathbb{O}_\Phi$  the set of all points where  $\Phi$  is graphically smooth of dimension  $n$ . Then one can define for  $\Phi$  the following set-valued mappings (from  $\text{gph}\Phi$  to  $\mathbb{Z}_n$ ):

$$\widehat{\mathcal{S}}_\Phi^*(x, y) := \begin{cases} \{(T_{\text{gph}\Phi}(x; y))^*\} & \text{if } (x, y) \in \mathbb{O}_\Phi, \\ \emptyset & \text{else,} \end{cases} \quad \mathcal{S}_\Phi^*(x, y) := \limsup_{(u; v) \xrightarrow{\text{gph}\Phi} (x; y)} \widehat{\mathcal{S}}_\Phi^*(u, v).$$

The following definition of SCD property also comes from [24, Definition 3.3].

**DEFINITION 2.**  $\Phi$  is said to have the SCD property at  $(x, y)$  if  $(x; y) \in \text{gph}\Phi$  and  $\mathcal{S}_\Phi^*(x, y) \neq \emptyset$ . It has the SCD property around  $(x, y)$  if  $(x; y) \in \text{gph}\Phi$  and there is a neighborhood  $\mathcal{N}$  of  $(x; y)$  such that  $\Phi$  has the SCD property at every  $(x'; y') \in \text{gph}\Phi \cap \mathcal{N}$ . It is called an SCD mapping if  $\Phi$  has the SCD property at every point  $(x, y)$  such that  $(x; y) \in \text{gph}\Phi$ .

The following definition of SCD regularity was given in [24, Definition 4.1].

DEFINITION 3. Define  $\mathbb{Z}_n^{\text{reg}} := \{L \in \mathbb{Z}_n \mid (y^*; 0) \in L \Rightarrow y^* = 0\}$ . A mapping  $\Phi : \mathcal{X} \rightrightarrows \mathcal{X}$  is called SCD regular around  $(x, y)$  if  $(x; y) \in \text{gph}\Phi$ ,  $\Phi$  has the SCD property around  $(x, y)$  and  $\mathcal{S}_\Phi^*(x, y) \subseteq \mathbb{Z}_n^{\text{reg}}$ . Moreover, the modulus of SCD regularity of  $\Phi$  around  $(x, y)$  is defined by

$$\text{scd reg } \Phi(x, y) := \sup\{\|y^*\| \mid (y^*; x^*) \in L, L \in \mathcal{S}_\Phi^*(x, y), \|x^*\| \leq 1\}.$$

According to [24, Lemma 3.7], the SCD property was coined because for any subspace  $L \in \mathcal{S}_\Phi^*(x, y)$  one has  $L \subseteq \text{gph}D^*\Phi(x, y)$ . Moreover, based on this property, the semismoothness\* in Definition 1 can be extended to the following SCD semismoothness\*.

DEFINITION 4 ([24, DEFINITION 5.1]). Let  $\Phi : \mathcal{X} \rightrightarrows \mathcal{X}$  and  $(\bar{x}; \bar{y}) \in \text{gph}\Phi$ , which is nonempty and closed. Then,  $\Phi$  is called SCD semismooth\* at  $(\bar{x}, \bar{y})$  if  $\Phi$  has the SCD property around  $(\bar{x}, \bar{y})$  and for every  $\epsilon > 0$  there exists some  $\delta > 0$  such that

$$\begin{aligned} |\langle x^*, x - \bar{x} \rangle - \langle y^*, y - \bar{y} \rangle| &\leq \epsilon \|(x; y) - (\bar{x}; \bar{y})\| \|(x^*; y^*)\| \\ \forall (x; y) \in \mathbb{B}_\delta(\bar{x}; \bar{y}), \quad \forall (y^*; x^*) \in \bigcup \mathcal{S}_\Phi^*(x, y) &\subseteq \text{gph}D^*\Phi(x, y). \end{aligned}$$

Note that the GE (3) can be solved equivalently via finding  $(x; d) \in \mathbb{R}^n \times \mathbb{R}^n$  such that

$$0 \in \mathcal{J}(x, d) := \begin{pmatrix} F(x) + \partial q(d) \\ x - d \end{pmatrix}. \quad (13)$$

Given  $\gamma > 0$ , define the proximal residual mapping  $u_\gamma : \mathbb{R}^n \rightarrow \mathbb{R}^n$  by

$$u_\gamma(x) := \mathcal{P}_{\gamma^{-1}q}(x - \gamma^{-1}F(x)) - x \quad \forall x \in \mathbb{R}^n. \quad (14)$$

In [25, Sect. 5], the following implementable SCD semismooth\* Newton method was proposed for (3), which is an application of [24, Algorithm 1] to (13).

---

**Algorithm 4** An implementable SCD semismooth\* Newton method for solving (3)

---

**Input:**  $x^{(0)} \in \mathbb{R}^n$ ,  $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$  and  $q : \mathbb{R}^n \rightarrow \bar{\mathbb{R}}$ .

**Output:**  $\{x^{(k)}\}$ .

**For**  $k = 0, 1, \dots$ ,

1. if  $0 \in F(x^{(k)}) + \partial q(x^{(k)})$ , stop the algorithm;
2. select  $\gamma^{(k)} > 0$ , and compute  $u^{(k)} := u_{\gamma^{(k)}}(x^{(k)})$ ,  $\hat{d}^{(k)} := x^{(k)} + u^{(k)}$  and  $\hat{d}^{*(k)} := -\gamma^{(k)}u^{(k)} - F(x^{(k)})$ ;
3. select  $(Y^{*(k)}, X^{*(k)})$  with  $\text{rge}(Y^{*(k)}; X^{*(k)}) \in \mathcal{S}_{\partial q}^*(\hat{d}^{(k)}, \hat{d}^{*(k)})$ , and compute the Newton direction  $\Delta x^{(k)}$  from

$$(Y^{*(k)\top} F'(x^{(k)}) + X^{*(k)\top}) \Delta x^{(k)} = (\gamma^{(k)} Y^{*(k)\top} + X^{*(k)\top}) u^{(k)}, \quad (15)$$

and obtain the new iterate via  $x^{(k+1)} := x^{(k)} + \Delta x^{(k)}$ .

---

One has the following convergence theorem for Algorithm 4.

**THEOREM 2 ([25, Theorem 5.2]).** *Let  $\bar{x}$  be a solution of (3) and assume that  $\partial q$  is SCD semismooth\* at  $(\bar{x}, -F(\bar{x}))$ . In addition, suppose that  $F + \partial q$  is SCD regular around  $(\bar{x}, 0)$ . Then for every pair  $\underline{\gamma}, \bar{\gamma}$  with  $0 < \underline{\gamma} \leq \bar{\gamma}$  there exists a neighborhood  $U$  of  $\bar{x}$  such that for every starting point  $x^{(0)} \in U$  Algorithm 4 produces a sequence  $\{x^{(k)}\}$  converging superlinearly to  $\bar{x}$ , provided we choose in every iteration step  $\gamma^{(k)} \in [\underline{\gamma}, \bar{\gamma}]$ .*

**4. Proof of Algorithm 3 as a G-semismooth Newton method** In this section, we demonstrate that Algorithm 3 can be treated as an application of the G-semismooth Newton method (Algorithm 1).

**4.1. Lipschitz continuous localization** Let  $F$ ,  $G$ , and  $D$  be the functions and the polyhedral set in (2).

Suppose that Assumption 1 holds and  $\omega > 0$  is the radius given by Proposition 1. Recall from Proposition 1 that the output  $\hat{u}$  by running Algorithm 2 can be locally represented via the solution mapping

$$S(x) := \{u \mid 0 \in u + F(x) + \nabla G(x)N_D(G(x) + G'(x)u)\} \quad \forall x \in \mathbb{B}_\omega(\bar{x}), \quad (16)$$

which is well-defined and single-valued. Note that  $S(x) = 0$  if and only if  $x$  is a solution of the GE (2) on  $\mathbb{B}_\omega(\bar{x})$ . Moreover, the corresponding multiplier calculated from Algorithm 2 can also

be defined as a single-valued mapping  $\Lambda(\cdot)$  on  $\mathbb{B}_\omega(\bar{x})$ . It has been established in Proposition 1 that  $S(\cdot)$  is isolated calm at point  $\bar{x}$  in the sense that  $\|S(x)\| \leq \beta_u \|x - \bar{x}\| \quad \forall x \in \mathbb{B}_\omega(\bar{x})$ . In fact, a more robust result can be obtained, for which the following consequence of the reduction approach [6, Example 3.139] is necessary.

**LEMMA 2.** *Suppose that Assumption 1 holds and  $\omega > 0$  is the radius given by Proposition 1. Define for  $x \in \mathbb{B}_\omega(\bar{x})$  the parameterized nonsmooth equation*

$$\tilde{\mathcal{H}}(u, \lambda; x) := \begin{pmatrix} u + F(x) + \nabla G(x)\lambda \\ G(x) + G'(x)u - \Pi_D(G(x) + G'(x)u + \lambda) \end{pmatrix} = 0. \quad (17)$$

*Then, for any  $\tilde{x} \in \mathbb{B}_\omega(\bar{x})$ , the nondegeneracy condition holds at  $(\tilde{x}, \tilde{d}) := (\tilde{x}, G(\tilde{x}) + G'(\tilde{x})S(\tilde{x}))$  for all  $\tilde{x} \in \mathbb{B}_\omega(\bar{x})$ , where  $S(\cdot)$  is defined by (16), in the sense that*

$$G'(\tilde{x})\mathbb{R}^n + \text{lin}T_D(\tilde{d}) = \mathbb{R}^s \quad \forall \tilde{x} \in \mathbb{B}_\omega(\bar{x}). \quad (18)$$

*Moreover, let  $\tilde{l}$  be the dimension of  $\text{lin}T_D(\tilde{d})$  and define  $Q := \tilde{W}^\top(T_D(\tilde{d}))$ , where  $\tilde{W} \in \mathbb{R}^{s \times s - \tilde{l}}$  is any matrix whose columns are linearly independent such that  $\text{rge}\tilde{W} = \text{span}N_D(\tilde{d})$ . Then the nonsmooth equation (17) is locally equivalent to*

$$\tilde{\mathcal{H}}_{\tilde{W}}(u, \mu; x) := \begin{pmatrix} u + F(x) + \nabla G(x)\tilde{W}\mu \\ \tilde{W}^\top(G(x) + G'(x)u - \tilde{d}) - \Pi_Q(\tilde{W}^\top(G(x) + G'(x)u - \tilde{d}) + \mu) \end{pmatrix} = 0, \quad (19)$$

*in the sense that, when  $x$  is sufficiently close to  $\tilde{x}$ ,  $(u, \lambda)$  solves (17) if and only if  $(u, \mu)$  solves (19) and  $\lambda = \tilde{W}\mu$ .*

*Proof* Since Assumption 1 holds, according to the proof of [23, Proposition 5.7], for any  $\tilde{x} \in \mathbb{B}_\omega(\bar{x})$  one has that  $(\tilde{u}, \tilde{\lambda}) := (S(\tilde{x}), \Lambda(\tilde{x}))$  is the unique point such that (17) holds at  $(\tilde{u}, \tilde{\lambda}; \tilde{x})$ . Moreover, from Remark 1 and the proof of [23, Proposition 5.7] we know that the nondegeneracy condition (18) holds at  $(\tilde{x}, \tilde{d})$ . From the definitions of  $\tilde{d}$  and  $\tilde{W}$  we know that the mapping  $d \rightarrow \tilde{W}^\top(d - \tilde{d})$  meets the requirements in [6, Definition 3.135], so that  $D$  is  $C^\infty$ -cone reducible to  $Q$ , which is a pointed closed convex cone. Then, it comes from [60, Section 4] that the nonsmooth equation (17) is locally equivalent to (19).  $\square$

Based on the above lemma, the following result holds.

**PROPOSITION 2.** *Suppose that Assumption 1 holds and  $\omega > 0$  is the radius given by Proposition 1. Then, both the mapping  $S(\cdot)$  defined in (16) and the multiplier mapping  $\Lambda(\cdot)$  are Lipschitz continuous in  $\mathbb{B}_\omega(\bar{x})$ . Moreover,  $S(\cdot)$  is G-semismooth in  $\text{int}\mathbb{B}_\omega(\bar{x})$ .*

*Proof* Let  $\tilde{x} \in \mathbb{B}_\omega(\bar{x})$ , and define  $(\tilde{u}, \tilde{\lambda}) := (S(\tilde{x}), \Lambda(\tilde{x}))$  and  $\tilde{d} := G(\tilde{x}) + G'(\tilde{x})\tilde{u}$ . With  $\tilde{l}$ ,  $Q$  and  $\tilde{W}$  being defined the same as those in Lemma 2, it comes from this lemma that the parameterized nonsmooth equation (17) is locally equivalent to (19) around  $\tilde{x}$ , and (18) holds. In particular,  $\tilde{\mathcal{H}}_{\tilde{W}}(\tilde{u}, \tilde{\mu}; \tilde{x}) = 0$  with  $\tilde{\mu} \in \mathbb{R}^{s-\tilde{l}}$  being the unique vector such that  $\tilde{\lambda} = \tilde{W}\tilde{\mu}$ .

Moreover, it holds that

$$\langle \Delta u, ((\tilde{u} + F(\tilde{x}))'_u + (\nabla G(\tilde{x})\tilde{W}\tilde{\mu})'_u) \Delta u \rangle = \|\Delta u\|^2.$$

Therefore, by following the proof of [45, Proposition 2] and using [45, Corollary 2] we know that there exists a neighborhood  $\mathbb{O}(\tilde{x})$  of  $\tilde{x}$  such that  $S(\cdot)$  and  $\Lambda(\cdot)$  are Lipschitz continuous. Since  $\mathbb{B}_\omega(\bar{x})$  is a compact set, for all  $\tilde{x} \in \mathbb{B}_\omega(\bar{x})$ , one can always find a finite collection  $\chi$  of  $\tilde{x}$  such that the union of these open neighborhoods  $\cup_{\tilde{x} \in \chi} \mathbb{O}(\tilde{x})$  covers  $\mathbb{B}_\omega(\bar{x})$ . Therefore,  $S(\cdot)$  and  $\Lambda(\cdot)$  are Lipschitz continuous in  $\mathbb{B}_\omega(\bar{x})$ . Finally, as the projection onto  $Q$  is strongly G-semismooth, by [45, Corollary 2] we know that  $S(\cdot)$  is G-semismooth with respect to  $\partial_B S$  or  $\partial S$ . This completes the proof.  $\square$

According to Rademacher's theorem, the solution mapping  $S(\cdot)$  defined in (16) is almost everywhere differentiable in  $\mathbb{B}_\omega(\bar{x})$  due to Proposition 2. It is not hard to compute the Frechét derivative of  $S$  when it is differentiable, but it is hard to directly compute the corresponding Bouligand sub-differential (4) or Clarke's generalized Jacobian by taking limits, since  $S(\cdot)$  is implicitly defined. Therefore, to implement Algorithm 1 to solve the nonsmooth equation  $S(x) = 0$ , the corresponding mapping  $\mathcal{T}$  should be explicitly computed, which will be done in the next part.

**4.2. G-Semismoothness** The analysis in [45] for locally Lipschitz continuous homeomorphisms can be utilized to compute a set-valued mapping such that the solution mapping  $S(\cdot)$  defined in (16) is G-semismooth with respect to it. Specifically, we have the following key result, which gives the G-semismoothness of  $S$  around  $\bar{x}$ .

**PROPOSITION 3.** *Suppose that Assumption 1 holds and  $\omega > 0$  is the radius given by Proposition 1. Then the solution mapping  $S(\cdot)$  defined in (16) is G-semismooth at every  $\tilde{x} \in \text{int}\mathbb{B}_\omega(\bar{x})$  with respect to the set-valued mapping*

$$\mathcal{T}_S(x) := \left\{ \begin{array}{l} -\mathcal{L}'_\lambda(x) + \nabla G(x)W[W^\top G'(x)\nabla G(x)W]^{-1}W^\top(G'(x)\mathcal{L}'_\lambda(x) - [G(x) + G'(x)u]'_x) \\ | W \text{ has full column rank, } \text{rge}W = (\text{span}\mathcal{F})^\perp \text{ with } \mathcal{F} \text{ being a face of } \mathcal{K}_D(d, \lambda) \end{array} \right\}, \quad (20)$$

where  $u := S(x)$ ,  $d := G(x) + G'(x)S(x)$  and  $\lambda := \Lambda(x)$ . Moreover,  $\mathcal{T}_S(\cdot)$  is outer semicontinuous at every  $\tilde{x} \in \text{int}\mathbb{B}_\omega(\bar{x})$ .



*Proof* Let  $\tilde{x} \in \text{int}\mathbb{B}_\omega(\bar{x})$  and define  $\tilde{u} := S(\tilde{x})$ ,  $\tilde{d} := G(\tilde{x}) + G'(\tilde{x})\tilde{u}$ , and  $\tilde{\lambda} := \Lambda(\tilde{x})$ . Let  $\tilde{I}$ ,  $\tilde{W}$  and  $Q$  be the same as those defined in Lemma 2. We know from Lemma 2 that the parameterized nonsmooth equation (17) is locally equivalent to (19) around  $\tilde{x}$ , and (18) holds. Then by following the analysis of [45, Proposition 2] the mapping

$$\Psi(v, \varsigma, x) := \begin{pmatrix} v + F(x) + \nabla G(x)\tilde{W}\varsigma \\ \tilde{W}^\top [G(x) + G'(x)v - \tilde{d}] - \Pi_Q(\tilde{W}^\top [G(x) + G'(x)v - \tilde{d}] + \varsigma) \\ x \end{pmatrix}$$

is locally Lipschitz homeomorphism around  $(\tilde{u}, \tilde{\mu}, \tilde{x})$ , where  $\tilde{\mu}$  is the unique vector such that  $\tilde{\lambda} = \tilde{W}\tilde{\mu}$ . Note that

$$(\tilde{W}^\top [G(x) + G'(x)v - \tilde{d}] + \varsigma)'(v, \varsigma, x) = (\tilde{W}^\top G'(x), I, \tilde{W}^\top [G(x) + G'(x)v]_x'),$$

which is always surjective. Consequently, from [62, Lemma 2.1] we know that

$$\partial_B \Psi(v, \varsigma, x) = \left\{ \begin{pmatrix} I & \nabla G(x)\tilde{W} & \mathcal{L}'_{\tilde{W}\varsigma}(x) \\ (I - \Xi)\tilde{W}^\top G'(x) & -\Xi & (I - \Xi)\tilde{W}^\top [G(x) + G'(x)v]_x' \\ 0 & 0 & I \\ \mid \Xi \in \partial_B \Pi_Q(\tilde{W}^\top [G(x) + G'(x)v - \tilde{d}] + \varsigma) \end{pmatrix} \right\}. \quad (21)$$

Since the projection operator  $\Pi_Q$  is (strongly) semismooth, one has  $\Psi$  is also semismooth, and it follows from [45, Theorem 2] that  $\Psi^{-1}$  is semismooth at  $(0, 0, \tilde{x}) \in \mathbb{R}^n \times \mathbb{R}^{s-\tilde{l}} \times \mathbb{R}^n$ . Moreover, from [45, Lemma 2] we know that all the elements of  $\partial_B \Psi(v, \varsigma, x)$  are nonsingular whenever  $(v, \varsigma, x)$  is sufficiently close to  $(\tilde{u}, \tilde{\mu}, \tilde{x})$ . It can be observed from (21) that each element of  $\partial_B \Psi$  is nonsingular at  $(v, \varsigma, x)$  if and only if each matrix  $\Xi + (I - \Xi)\tilde{W}^\top G'(x)\nabla G(x)\tilde{W}$  is nonsingular for all  $\Xi \in \partial_B \Pi_Q(\tilde{W}^\top [G(x) + G'(x)v - \tilde{d}] + \varsigma)$ . Therefore, for all  $(v, \varsigma, x)$  sufficiently close to  $(\tilde{u}, \tilde{\mu}, \tilde{x})$ , it holds by elementary column transformations that

$$\begin{aligned} & \partial_B \Psi(v, \varsigma, x) \\ &= \left\{ \begin{pmatrix} I - \nabla G(x)\tilde{W}(\Gamma(\Xi))^{-1}(I - \Xi)\tilde{W}^\top G'(x) & \nabla G(x)\tilde{W}(\Gamma(\Xi))^{-1} & \mathcal{E}_S(v, \varsigma, \Xi) \\ (\Gamma(\Xi))^{-1}(I - \Xi)\tilde{W}^\top G'(x) & -(\Gamma(\Xi))^{-1} & \mathcal{E}_\Lambda(v, \varsigma, \Xi) \\ 0 & 0 & I \\ \mid \Gamma(\Xi) := \Xi + (I - \Xi)\tilde{W}^\top G'(x)\nabla G(x)\tilde{W} \text{ with } \Xi \in \partial_B \Pi_Q(\tilde{W}^\top [G(x) + G'(x)u - \tilde{d}] + \varsigma) \end{pmatrix}^{-1} \right\}, \end{aligned}$$

where

$$\begin{cases} \mathcal{E}_S(v, \varsigma, \Xi) := -\mathcal{L}'_{\tilde{W}\varsigma}(x) + \nabla G(x)\tilde{W}(\Gamma(\Xi))^{-1}(I - \Xi)\tilde{W}^\top(G'(x)\mathcal{L}'_{\tilde{W}\varsigma}(x) - [G(x) + G'(x)v]'_x), \\ \mathcal{E}_\Lambda(v, \varsigma, \Xi) := -(\Gamma(\Xi))^{-1}(I - \Xi)\tilde{W}^\top(G'(x)\mathcal{L}'_{\tilde{W}\varsigma}(x) - [G(x) + G'(x)v]'_x). \end{cases}$$

Note that for all  $x$  sufficiently close to  $\tilde{x}$ , it holds that  $(u; \mu; x) = \Psi^{-1}(0, 0, x)$  where  $\mu$  is the unique vector such that  $\lambda = \tilde{W}\mu$ . Since  $\Psi^{-1}$  is G-semismooth, the solution mapping  $S(\cdot)$  is G-semismooth at  $\tilde{x}$  with respect to

$$\begin{aligned} \tilde{\mathcal{E}}_S(x) := & \left\{ \begin{aligned} & -\mathcal{L}'_\lambda(x) + \nabla G(x)\tilde{W}(\Gamma(\Xi))^{-1}(I - \Xi)\tilde{W}^\top(G'(x)\mathcal{L}'_\lambda(x) - [G(x) + G'(x)u]'_x) \\ & \mid \Gamma(\Xi) := \Xi + (I - \Xi)\tilde{W}^\top G'(x)\nabla G(x)\tilde{W} \text{ with } \Xi \in \partial_B \Pi_Q(\tilde{W}^\top[G(x) + G'(x)u - \tilde{d}] + \mu) \end{aligned} \right\}. \end{aligned}$$

Moreover,  $\tilde{\mathcal{E}}_S(\cdot)$  is outer semicontinuous around  $\tilde{x}$ . According to Lemma 1, one has for all  $x$  sufficiently close to  $\tilde{x}$ ,

$$\begin{aligned} \tilde{\mathcal{E}}_S(x) &= \left\{ \begin{aligned} & -\mathcal{L}'_\lambda(x) + \nabla G(x)\tilde{W}(\tilde{\Gamma}_{\mathcal{G}})^{-1}\Pi_{(\text{span}\mathcal{G})^\perp}\tilde{W}^\top(G'(x)\mathcal{L}'_\lambda(x) - [G(x) + G'(x)u]'_x) \\ & \mid \tilde{\Gamma}_{\mathcal{G}} := \Pi_{\text{span}\mathcal{G}} + \Pi_{(\text{span}\mathcal{G})^\perp}\tilde{W}^\top G'(x)\nabla G(x)\tilde{W}, \mathcal{G} \text{ is a face of } \mathcal{K}_Q(\tilde{W}^\top[d - \tilde{d}], \mu) \end{aligned} \right\}. \end{aligned} \quad (22)$$

Let  $\mathcal{G}$  be a face of  $\mathcal{K}_Q(\tilde{W}^\top[d - \tilde{d}], \mu)$ . One can find two matrices  $U_1$  and  $U_2$ , whose columns form an orthonormal basis of  $\text{span}\mathcal{G}$  and  $(\text{span}\mathcal{G})^\perp$ , respectively. Note that  $\Pi_{\text{span}\mathcal{G}} = U_1 U_1^\top$  and  $\Pi_{(\text{span}\mathcal{G})^\perp} = U_2 U_2^\top$ . Consequently,

$$(\tilde{\Gamma}_{\mathcal{G}})^{-1}\Pi_{(\text{span}\mathcal{G})^\perp} = [U_1 U_1^\top + U_2 U_2^\top \tilde{W}^\top G'(x)\nabla G(x)\tilde{W}]^{-1} U_2 U_2^\top.$$

For any  $w$  and  $v$  such that  $v = [U_1 U_1^\top + U_2 U_2^\top \tilde{W}^\top G'(x)\nabla G(x)\tilde{W}]^{-1} U_2 U_2^\top w$ , it holds that

$$\begin{aligned} & U_1 U_1^\top v + U_2 U_2^\top \tilde{W}^\top G'(x)\nabla G(x)\tilde{W} v = U_2 U_2^\top w \\ \Rightarrow & U_1^\top v = 0 \text{ and } U_2^\top \tilde{W}^\top G'(x)\nabla G(x)\tilde{W} v = U_2^\top w \\ \Rightarrow & v = U_2 \xi \text{ for some } \xi, \text{ and } U_2^\top \tilde{W}^\top G'(x)\nabla G(x)\tilde{W} U_2 \xi = U_2^\top w. \end{aligned}$$

Therefore, one gets  $v = U_2 \xi = U_2 [U_2^\top \tilde{W}^\top G'(x)\nabla G(x)\tilde{W} U_2]^{-1} U_2^\top w$ , so that

$$\tilde{W}(\tilde{\Gamma}_{\mathcal{G}})^{-1}\Pi_{(\text{span}\mathcal{G})^\perp}\tilde{W}^\top = \tilde{W} U_2 [U_2^\top \tilde{W}^\top G'(x)\nabla G(x)\tilde{W} U_2]^{-1} U_2^\top \tilde{W}^\top. \quad (23)$$

Recall that  $\lambda = \tilde{W}\mu$ , so that  $[\lambda]^\perp = \{e \mid \langle \tilde{W}^\top e, \mu \rangle = 0\}$ , which implies  $\tilde{W}^\top([\lambda]^\perp) = [\mu]^\perp$ . One has from [57, Theorem 6.31] that

$$\mathcal{K}_D(d, \lambda) = \{e \mid e \in T_D(d), \langle e, \lambda \rangle = 0\} = \{e \mid \tilde{W}^\top e \in T_Q(\tilde{W}^\top[d - \tilde{d}]), \langle \tilde{W}^\top e, \mu \rangle = 0\}. \quad (24)$$

Therefore, it holds that

$$\tilde{W}^\top \mathcal{K}_D(d, \lambda) = \mathcal{K}_Q(\tilde{W}^\top[d - \tilde{d}], \mu). \quad (25)$$

Consequently, one has for any given  $v \in \mathbb{R}^{s-\tilde{l}}$ ,

$$\sup_{e \in \mathcal{K}_D(d, \lambda)} \{\langle \tilde{W}v, e \rangle\} = \sup_{e \in \mathcal{K}_D(d, \lambda)} \{\langle v, \tilde{W}^\top e \rangle \mid e \in \mathcal{K}_D(d, \lambda)\} = \sup_{v \in \mathbb{R}^{s-\tilde{l}}} \{\langle v, v \rangle \mid v \in \mathcal{K}_Q(\tilde{W}^\top[d - \tilde{d}], \mu)\}.$$

Therefore,  $v \in \mathcal{K}_Q(\tilde{W}^\top[d - \tilde{d}], \mu)^\circ$  if and only if  $\tilde{W}v \in \mathcal{K}_D(d, \lambda)^\circ$ . Recall that  $\mathcal{G}$  is a face of  $\mathcal{K}_Q(\tilde{W}^\top[d - \tilde{d}], \mu)$ , i.e., there exists a nonzero vector  $\tilde{v} \in \mathbb{R}^{s-\tilde{l}}$  in its polar such that  $\mathcal{G} = \mathcal{K}_Q(\tilde{W}^\top[d - \tilde{d}], \mu) \cap [\tilde{v}]^\perp$ . Note that

$$\tilde{\mathcal{F}} := \{p \mid \tilde{W}^\top p \in \mathcal{G}\} = \{p \mid \tilde{W}^\top p \in \mathcal{K}_Q(\tilde{W}^\top[d - \tilde{d}], \mu) \cap [\tilde{v}]^\perp\} = \mathcal{K}_D(d, \lambda) \cap [\tilde{W}\tilde{v}]^\perp.$$

Since  $\tilde{W}\tilde{v} \in \mathcal{K}_D(d, \lambda)^\circ$ , the set  $\tilde{\mathcal{F}}$  is a face of  $\mathcal{K}_D(d, \lambda)$ . Moreover, one has  $\text{span}\tilde{\mathcal{F}} = \{p \mid \tilde{W}^\top p \in \text{span}\mathcal{G}\}$ . Recall that  $\text{rge}U_1 = \text{span}\mathcal{G}$ , so that  $\text{span}\tilde{\mathcal{F}} = \{p \mid \tilde{W}^\top p \in \text{rge}U_1\} = \{p \mid U_2^\top \tilde{W}^\top p = 0\}$ . Therefore, if  $W$  is a matrix with full column rank such that  $\text{rge}W = (\text{span}\tilde{\mathcal{F}})^\perp = \text{rge}(\tilde{W}U_2)$ , there exists a nonsingular square matrix  $P$  such that  $W = \tilde{W}U_2P$ . In this case, one has

$$\begin{aligned} W[W^\top G'(x)\nabla G(x)W]^{-1}W^\top &= \tilde{W}U_2P[P^\top(\tilde{W}U_2)^\top G'(x)\nabla G(x)\tilde{W}U_2P]^{-1}P^\top(\tilde{W}U_2)^\top \\ &= \tilde{W}U_2P[(\tilde{W}U_2)^\top G'(x)\nabla G(x)\tilde{W}U_2P]^{-1}(P^\top)^{-1}P^\top(\tilde{W}U_2)^\top \\ &= \tilde{W}U_2[(\tilde{W}U_2)^\top G'(x)\nabla G(x)\tilde{W}U_2]^{-1}(\tilde{W}U_2)^\top. \end{aligned}$$

Such an equality, together with (20), (22), and (23), implies that

$$\tilde{\mathcal{E}}_S(x) \subseteq \mathcal{T}_S(x). \quad (26)$$

Next, we show that the inclusion in (26) is an equality. Let  $\mathcal{F}$  be an arbitrary face of  $\mathcal{K}_D(d, \lambda)$ , i.e., there exists  $\check{\lambda} \in \mathcal{K}_D(d, \lambda)^\circ = N_D(d) + \text{span}[\lambda] \subseteq \text{span}N_D(d) \subseteq \text{span}N_D(\tilde{d})$  such that  $\mathcal{F} = \mathcal{K}_D(d, \lambda) \cap [\check{\lambda}]^\perp$ . Since  $\check{\lambda} \in \text{rge}\tilde{W}$ , one has  $\check{\lambda} = \Pi_{\text{rge}\tilde{W}}(\check{\lambda}) = \tilde{W}[\tilde{W}^\top \tilde{W}]^{-1}\tilde{W}^\top \check{\lambda}$ . Then by (24) one can get

$$\mathcal{K}_D(d, \lambda) \cap [\check{\lambda}]^\perp = \{e \mid \tilde{W}^\top e \in T_Q(\tilde{W}^\top(d - \tilde{d})), \langle \tilde{W}^\top e, \mu \rangle = 0, \langle \tilde{W}^\top e, [\tilde{W}^\top \tilde{W}]^{-1}\tilde{W}^\top \check{\lambda} \rangle = 0\}.$$

Therefore, it holds that  $\tilde{W}^\top \mathcal{F} = \mathcal{K}_Q(\tilde{W}^\top [d - \tilde{d}], \mu) \cap [(\tilde{W}^\top \tilde{W})^{-1} \tilde{W}^\top \check{\lambda}]^\perp$ . Then, from (25) one has  $[\tilde{W}^\top \tilde{W}]^{-1} \tilde{W}^\top \check{\lambda} \in (\tilde{W}^\top \mathcal{K}_D(d, \lambda))^\circ = (\mathcal{K}_Q(\tilde{W}^\top [d - \tilde{d}], \mu))^\circ$ . Therefore, it holds that

$$\langle [\tilde{W}^\top \tilde{W}]^{-1} \tilde{W}^\top \check{\lambda}, \tilde{W}^\top \hat{d} \rangle = \langle \tilde{W} [\tilde{W}^\top \tilde{W}]^{-1} \tilde{W}^\top \check{\lambda}, \hat{d} \rangle = \langle \check{\lambda}, \hat{d} \rangle \leq 0 \quad \forall \hat{d} \in \mathcal{K}_D(d, \lambda).$$

Consequently,  $\tilde{W}^\top \mathcal{F}$  is exactly a face of  $\mathcal{K}_Q(\tilde{W}^\top [d - \tilde{d}], \mu)$ . Thus, the inclusion in (26) holds as an equality, i.e.,  $\mathcal{T}_S(x) = \tilde{\mathcal{E}}_S(x)$  for all  $x$  sufficiently close to  $\tilde{x}$ . Therefore,  $\mathcal{T}_S(\cdot)$  is also outer semicontinuous at any  $\tilde{x} \in \text{int}\mathbb{B}_\omega(\tilde{x})$ . This completes the proof.  $\square$

**4.3. Regularity conditions** The following result is crucial for using Algorithm 1 to solve  $S(x) = 0$ .

**PROPOSITION 4.** *Under Assumption 1, it holds that Assumption 2 is equivalent to the regularity condition that every element of  $\mathcal{T}_S(\bar{x})$  defined in (20) is nonsingular.*

*Proof* Since Assumption 1 holds, one has  $\bar{u} = S(\bar{x})$ . Then by Proposition 3 it holds that

$$\mathcal{T}_S(\bar{x}) = \left\{ \begin{array}{l} -\mathcal{L}'_{\bar{\lambda}}(\bar{x}) + \nabla G(\bar{x})W[W^\top G'(\bar{x})\nabla G(\bar{x})W]^{-1}W^\top(G'(\bar{x})\mathcal{L}'_{\bar{\lambda}}(\bar{x}) - G'(\bar{x})) \\ \equiv -(I - \Pi_{\text{rge}(\nabla G(\bar{x})W)})\mathcal{L}'_{\bar{\lambda}}(\bar{x}) - \Pi_{\text{rge}(\nabla G(\bar{x})W)} \\ | W \text{ has full column rank, } \text{rge}W = (\text{span}\mathcal{F})^\perp \text{ with } \mathcal{F} \text{ being a face of } \mathcal{K}_D(\bar{d}, \bar{\lambda}) \end{array} \right\},$$

where  $\bar{d} := G(\bar{x}) + G'(\bar{x})S(\bar{x})$  and  $\bar{\lambda} := \Lambda(\bar{x})$ . Let  $\mathcal{F}$  be an arbitrary face of  $\mathcal{K}_D(G(\bar{x}), \bar{\lambda})$  with  $W$  having full column rank such that  $\text{rge}W = (\text{span}\mathcal{F})^\perp$ . Let  $Z$  be an arbitrary matrix (with full column rank) such that  $\text{rge}Z = \{u \mid G'(\bar{x})u \in \text{span}\mathcal{F}\} = \ker(W^\top G'(\bar{x}))$ . Note that  $\text{rge}(\nabla G(\bar{x})W) = \ker Z^\top$ . Therefore, it holds that

$$\begin{aligned} & -(I - \Pi_{\text{rge}(\nabla G(\bar{x})W)})\mathcal{L}'_{\bar{\lambda}}(\bar{x})v - \Pi_{\text{rge}(\nabla G(\bar{x})W)}v \neq 0 \quad \forall v \neq 0 \\ \Leftrightarrow & (I - \Pi_{\text{rge}(\nabla G(\bar{x})W)})\mathcal{L}'_{\bar{\lambda}}(\bar{x})v \neq 0 \quad \forall v \neq 0 \text{ such that } \Pi_{\text{rge}(\nabla G(\bar{x})W)}v = 0 \\ \Leftrightarrow & Z(Z^\top Z)^{-1}Z^\top \mathcal{L}'_{\bar{\lambda}}(\bar{x})v \neq 0 \quad \forall 0 \neq v \in \text{rge}Z \\ \Leftrightarrow & Z^\top \mathcal{L}'_{\bar{\lambda}}(\bar{x})v \neq 0 \quad \forall 0 \neq v \in \text{rge}Z. \end{aligned}$$

Thus, Assumption 2 is equivalent to the condition that every element of  $\mathcal{T}_S(\bar{x})$  is nonsingular. This completes the proof.  $\square$

**4.4. Equivalence to a G-Semismooth Newton method** We are ready to show that the semismooth\* Newton method in algorithm 3 is exactly a special case of the G-semismooth Newton method in Algorithm 1. Suppose that Assumptions 1 and 2 hold. Let  $\omega$  be the parameter defined

in Proposition 1. For any  $\tilde{x} \in \mathbb{B}_\omega(\bar{x})$  with  $\tilde{d} = G(\tilde{x}) + G'(\tilde{x})S(\tilde{x})$  and  $\tilde{u} = S(\tilde{x})$ , one has  $\tilde{\lambda} \in N_D(\tilde{d})$ . Moreover, it is easy to see that  $\tilde{\mathcal{F}} := \text{lin}T_D(\tilde{d}) = \text{lin}\mathcal{K}_D(\tilde{d}, \tilde{\lambda})$  is a face of  $\mathcal{K}_D(\tilde{d}, \tilde{\lambda})$ . In fact, from the proof of Proposition 3 one can see that  $\tilde{\mathcal{F}}$  is exactly the face such that  $\tilde{W}^\top \tilde{\mathcal{F}} = \mathcal{G} \equiv \{0\}$ , where  $\tilde{W}$  is a matrix with full column rank such that  $\text{rge}\tilde{W} = \text{span}N_D(\tilde{d})$ , while  $\mathcal{G}$  is a face of  $\mathcal{K}_Q(0, \tilde{\mu}) = \tilde{W}^\top \mathcal{K}_D(\tilde{d}, \tilde{\lambda})$ . Consequently, the columns of  $\tilde{W}$  form a basis of  $(\text{span}\tilde{\mathcal{F}})^\perp$ , so that by (20) one can get

$$\begin{aligned} V_{\tilde{x}} &:= -\mathcal{L}'_{\tilde{\lambda}}(\tilde{x}) + \nabla G(\tilde{x})\tilde{W}[\tilde{W}^\top G'(\tilde{x})\nabla G(\tilde{x})\tilde{W}]^{-1}\tilde{W}^\top (G'(\tilde{x})\mathcal{L}'_{\tilde{\lambda}}(\tilde{x}) - [G(\tilde{x}) + G'(\tilde{x})\tilde{u}]'_x) \\ &= -\Pi_{\mathcal{Z}}\mathcal{L}'_{\tilde{\lambda}}(\tilde{x}) - \Pi_{\mathcal{W}} - \nabla G(\tilde{x})\tilde{W}[\tilde{W}^\top G'(\tilde{x})\nabla G(\tilde{x})\tilde{W}]^{-1}\tilde{W}^\top (G'(\tilde{x})\tilde{u})'_x \in \mathcal{T}_S(\tilde{x}), \end{aligned} \quad (27)$$

where  $\mathcal{W} := \text{rge}(\nabla G(\tilde{x})\tilde{W})$  and  $\mathcal{Z} := \mathcal{W}^\perp$ . Based on the results established in the previous two subsections, we can apply Algorithm 1 to solve problem (2). The resulting implementation is given as follows.

---

**Algorithm 5** A G-Semismooth Newton method for solving (2)

---

**Input:**  $F: \mathbb{R}^n \rightarrow \mathbb{R}^n$ ,  $G: \mathbb{R}^n \rightarrow \mathbb{R}^s$ ,  $D \subset \mathbb{R}^s$ ,  $x^{(0)} \in \mathbb{R}^n$ , and  $\varrho \geq 0$ .

**Output:**  $\{x^{(k)}\}$ .

**For**  $k = 0, 1, \dots$ ,

1. if  $x^{(k)}$  solves (2), stop the algorithm;
  2. run the approximation step in Algorithm 2 with input  $x^{(k)}$  to compute  $\hat{u}^{(k)}$ ,  $\hat{\lambda}^{(k)}$ ,  $\hat{d}^{(k)}$  and  $\mathcal{L}_{\hat{\lambda}^{(k)}}(x^{(k)})$ ;
  3. compute  $V^{(k)}$  such that  $\text{dist}(V^{(k)}, \mathcal{T}_S(x^{(k)})) \leq \varrho \|\hat{u}^{(k)}\|$  with  $\mathcal{T}_S$  being given in (20);
  4. compute the Newton direction  $\Delta x^{(k)}$  satisfying  $V^{(k)}\Delta x^{(k)} + \hat{u}^{(k)} = 0$ , and set  $x^{(k+1)} := x^{(k)} + \Delta x^{(k)}$ .
- 

The following result shows that Algorithm 3 is a special case of Algorithm 5.

**THEOREM 3.** Suppose that Assumptions 1 and 2 hold. Then, Algorithm 3 is an instance of Algorithm 1 (in the form of Algorithm 5) in the sense that the local superlinear convergence of Algorithm 3 can be obtained from Theorem 1.

*Proof* Let  $\{x^{(k)}\}$  be the sequence generated by Algorithm 3. For a fixed  $\bar{k} \geq 1$  such that  $x^{(\bar{k})}$  is well-defined, we assume that  $x^{(\bar{k}-1)}$  is also the output of the final step at the iteration indexed by  $(\bar{k} - 1)$  of Algorithm 5. Then, the first two steps at the iteration indexed by  $\bar{k}$  of Algorithm 5 are

the same as the first two steps at the iteration indexed by  $\bar{k}$  of Algorithm 3. During the third step of Algorithm 5 at iteration  $\bar{k}$ , one can take

$$\mathcal{W}^{(\bar{k})} := \text{rge}(\nabla G(x^{(\bar{k})})\hat{W}^{(\bar{k})}), \quad \mathcal{Z}^{(\bar{k})} := (\mathcal{W}^{(\bar{k})})^\perp \quad \text{and} \quad V^{(\bar{k})} := -\Pi_{\mathcal{Z}^{(\bar{k})}} \mathcal{L}'_{\hat{\lambda}^{(\bar{k})}}(x^{(\bar{k})}) - \Pi_{\mathcal{W}^{(\bar{k})}},$$

where  $\hat{W}^{(\bar{k})}$  comes from the third step of Algorithm 3 at the iteration indexed by  $\bar{k}$ . Then, from (27) one has

$$\begin{aligned} & \text{dist}(V^{(\bar{k})}, \mathcal{T}_S(x^{(\bar{k})})) \\ & \leq \left\| \nabla G(x^{(\bar{k})})\hat{W}^{(\bar{k})} [(\hat{W}^{(\bar{k})})^\top G'(x^{(\bar{k})})\nabla G(x^{(\bar{k})})\hat{W}^{(\bar{k})}]^{-1} (\hat{W}^{(\bar{k})})^\top (G'(x^{(\bar{k})})\hat{u}^{(\bar{k})})'_x \right\|, \end{aligned}$$

where  $\hat{u}^{(\bar{k})} := S(x^{(\bar{k})})$ . Then, the corresponding  $\Delta x^{(\bar{k})}$  computed by the final step at the iteration indexed by  $\bar{k}$  of Algorithm 5 is the same as the one calculated by (12) at the iteration indexed by  $\bar{k}$ . Let  $\omega$  be the parameter specified in Proposition 1. Recall that for any  $\tilde{x} \in \mathbb{B}_\omega(\bar{x})$ ,  $V_{\tilde{x}}$  in (27) is independent of the specific choice of the corresponding  $\tilde{W}$  in (27). Without loss of generality, one can assume that  $\tilde{W}$  in (27) is uniformly bounded. One has  $[\tilde{W}^\top G'(\tilde{x})\nabla G(\tilde{x})\tilde{W}]^{-1}$  is well-defined and uniformly bounded since the nondegeneracy condition holds and  $G$  is continuously differentiable. Furthermore, since  $G'(\cdot)$  is also continuously differentiable, one can get  $\text{dist}(V^{(\bar{k})}, \mathcal{T}_S(x^{(\bar{k})})) \leq \varrho \|\hat{u}^{(\bar{k})}\|$  with

$$\varrho := \sup_{\tilde{x} \in \mathbb{B}_\omega(\bar{x})} \left\{ \left\| \nabla G(\tilde{x})\tilde{W} [\tilde{W}^\top G'(\tilde{x})\nabla G(\tilde{x})\tilde{W}]^{-1} \tilde{W}^\top G''(\tilde{x}) \right\| \right\} < \infty.$$

Consequently, the iteration sequence  $\{x^{(k)}\}$  generated by Algorithm 3 can be viewed as the one generated by Algorithm 5.

Hence, if  $x^{(0)}$  is sufficiently close to  $\bar{x}$ , one has from Proposition 3 and Proposition 4 that the local superlinear convergence of Algorithm 3 is guaranteed by Theorem 1. This completes the proof.

□

**5. Proof of Algorithm 4 as a G-semismooth Newton method** In this section, we show that Algorithm 4 is also an application of Algorithm 1. The methodology developed here also can be used to show that the implementable SCD semismooth\* Newton method in the more recent work [21] is also a G-semismooth Newton method.

We first provide the following two preliminary results.

**LEMMA 3.** Let  $q$  be the function in (3) and  $\mathcal{P}_{\gamma^{-1}q}$  be the proximal mapping defined by (5) with  $\gamma > 0$ . For any  $y, z \in \mathbb{R}^n$  such that  $(z; y) \in \text{gph} \mathcal{P}_{\gamma^{-1}q}$ , it holds that

$$\begin{aligned} \{\text{rge}(I; B) \mid B \in \partial_B \mathcal{P}_{\gamma^{-1}q}(z)\} &= \mathcal{S}_{\mathcal{P}_{\gamma^{-1}q}}^*(z, y) \\ &= \left\{ \{(-(e^* + \gamma e); -\gamma e) \mid (e; e^*) \in L\} \mid L \in \mathcal{S}_{\partial q}^*(y, \gamma(z - y)) \right\}, \end{aligned} \quad (28)$$

where the definitions of  $\mathcal{S}_{\mathcal{P}_{\gamma^{-1}q}}^*$  and  $\mathcal{S}_{\partial q}^*$  come from Definition 2.

*Proof* Since  $q$  is a closed proper convex function, one has for any  $y, z \in \mathbb{R}^n$ ,

$$(z; y) \in \text{gph} \mathcal{P}_{\gamma^{-1}q} \Leftrightarrow \gamma(z - y) \in \partial q(y) \Leftrightarrow \phi(z, y) := (y; \gamma(z - y)) \in \text{gph} \partial q. \quad (29)$$

Meanwhile, one has  $\phi'(z, y) = \begin{pmatrix} 0 & I \\ \gamma I & -\gamma I \end{pmatrix}$ , which is nonsingular for all  $\gamma > 0$ . Note that

$$\begin{pmatrix} 0 & -I \\ I & 0 \end{pmatrix} \begin{pmatrix} 0 & \gamma I \\ I & -\gamma I \end{pmatrix} \begin{pmatrix} 0 & -I \\ I & 0 \end{pmatrix}^\top \begin{pmatrix} e \\ e^* \end{pmatrix} = \begin{pmatrix} -(e^* + \gamma e) \\ -\gamma e \end{pmatrix}.$$

Thus from [24, Lemma 3.11 and Proposition 3.14] and [57, Theorem 13.52] one gets (28).  $\square$

**LEMMA 4.** Let  $u_\gamma$  be the function defined in (14) with  $\gamma > 0$ . Suppose that  $\bar{x}$  is a solution to the GE (3). For any  $x \in \mathbb{R}^n$ , by setting  $u := u_\gamma(x)$ ,  $z := x - \gamma^{-1}F(x)$  and  $\bar{z} := \bar{x} - \gamma^{-1}F(\bar{x})$ , one has

$$\|u + x - \bar{x}\|^2 + \|F(\bar{x}) - F(x) - \gamma u\|^2 \leq \max\{1, \gamma^2\} \|z - \bar{z}\|^2.$$

*Proof* Note that  $u + x = \mathcal{P}_{\gamma^{-1}q}(z)$ . One has  $(u + x; -F(x) - \gamma u) = (u + x; \gamma(z - (u + x))) \in \text{gph} \partial q$  from (29). Moreover, one has  $\bar{x} = \mathcal{P}_{\gamma^{-1}q}(\bar{z})$ , so that

$$\begin{aligned} &\|u + x - \bar{x}\|^2 + \|F(\bar{x}) - F(x) - \gamma u\|^2 \\ &= \|\mathcal{P}_{\gamma^{-1}q}(z) - \mathcal{P}_{\gamma^{-1}q}(\bar{z})\|^2 + \|F(\bar{x}) - F(x) - \gamma(\mathcal{P}_{\gamma^{-1}q}(z) - x)\|^2 \\ &= \|\mathcal{P}_{\gamma^{-1}q}(z) - \mathcal{P}_{\gamma^{-1}q}(\bar{z})\|^2 + \|F(\bar{x}) - \gamma\bar{x} + \gamma x - F(x) - \gamma(\mathcal{P}_{\gamma^{-1}q}(z) - \bar{x})\|^2 \\ &= \|\mathcal{P}_{\gamma^{-1}q}(z) - \mathcal{P}_{\gamma^{-1}q}(\bar{z})\|^2 + \gamma^2 \|z - \bar{z} - (\mathcal{P}_{\gamma^{-1}q}(z) - \mathcal{P}_{\gamma^{-1}q}(\bar{z}))\|^2 \leq \max\{1, \gamma^2\} \|z - \bar{z}\|^2, \end{aligned}$$

where the last inequality comes from [56, Proposition 1(c)]. This completes the proof.  $\square$

**5.1. Lipschitz continuous localization and G-semismoothness** Recall that the GE in (3) is equivalent to the nonsmooth equation  $u_\gamma(x) = 0$  (for any  $\gamma > 0$ ) with  $u_\gamma$  in (14). Let  $U$  be a



neighborhood of  $\bar{x}$  such that  $F$  is Lipschitz continuous on it with modulus  $\ell > 0$ . According to [25, eq. (5.13)] one has

$$\|u_\gamma(x) - u_\gamma(x')\| \leq 2\|x - x'\| + \gamma^{-1}\|F(x) - F(x')\| \leq (2 + \frac{\ell}{\gamma})\|x - x'\| \quad \forall x, x' \in U.$$

Therefore, for any  $\gamma > 0$ , the function  $u_\gamma(\cdot)$  is Lipschitz continuous on  $U$ . It is not easy to calculate the Bouligand subdifferential of  $u_\gamma(\cdot)$  although locally it is almost everywhere differentiable. Instead, it is more reasonable to consider using the composite mapping  $\mathcal{T}_{u_\gamma} : \mathbb{R}^n \rightrightarrows \mathbb{R}^{n \times n}$  defined by

$$\mathcal{T}_{u_\gamma}(x) = \partial_B \mathcal{P}_{\gamma^{-1}q}(\cdot) \big|_{x - \gamma^{-1}F(x)} \cdot (I - \gamma^{-1}F'(x)) - I. \quad (30)$$

Note that for any  $\gamma > 0$  the mapping  $\mathcal{T}_{u_\gamma}(\cdot)$  defined in (30) is outer semicontinuous at  $\bar{x}$ .

Next, we show that  $u_\gamma$  ( $\gamma > 0$ ) defined in (14) is G-semismooth at a solution  $\bar{x}$  to the GE (3) with respect to  $\mathcal{T}_{u_\gamma}$  defined by (30). For convenience of comparison, we take identical values for all parameters here to those used in the conditions in [25].

**PROPOSITION 5.** *Let  $\bar{x}$  be a solution to the GE (3), and  $\mathbb{B}_r(\bar{x})$  be the ball such that  $F$  is Lipschitz continuous on it with modulus  $\ell \geq 0$ . For any  $\epsilon > 0$ , let  $\delta$  and  $\delta_q$  be two positive constants (depending on  $\epsilon$ ) such that  $\delta \leq \min\{\frac{\delta_q}{1+\ell}, r\}$ ,*

$$\begin{aligned} |\langle e^*, d - \bar{x} \rangle - \langle e, d^* + F(\bar{x}) \rangle| &\leq \frac{\epsilon}{2\sqrt{2}(\ell+1)} \|(e; e^*)\| \|(d - \bar{x}; d^* + F(\bar{x}))\| \\ \forall (d; d^*) \in \text{gph} \partial q \cap \mathbb{B}_{\delta_q}(\bar{x}; -F(\bar{x})), \quad \forall (e; e^*) \in L \in \mathcal{S}_{\partial q}^*(d, d^*), \end{aligned} \quad (31)$$

and

$$\|F(x) - F(\bar{x}) - F'(x)(x - \bar{x})\| \leq \frac{\epsilon}{2\sqrt{2}(\ell+1)} \|x - \bar{x}\| \quad \forall x \in \mathbb{B}_\delta(\bar{x}). \quad (32)$$

Then, for any  $\hat{x} \in \mathbb{R}^n$  satisfying  $\|\hat{x} - \bar{x}\| \leq \min\{\frac{\min\{\delta_q, r\}}{(1+\frac{\ell}{\gamma})\max\{1, \gamma\}}, \delta\}$  the following results hold:

(a) For any  $v \in \mathbb{R}^n$  and any  $B \in \partial_B \mathcal{P}_{\gamma^{-1}q}(\hat{x} - \gamma^{-1}F(\hat{x}))$ , one has

$$\begin{aligned} & \left| \langle \gamma v, u_\gamma(\hat{x}) \rangle - (\langle \gamma Bv, (I - \gamma^{-1}F'(\hat{x}))(\hat{x} - \bar{x}) \rangle - \langle \gamma v, \hat{x} - \bar{x} \rangle) \right| \\ & \leq \frac{\epsilon}{2\sqrt{2}(\ell+1)} (\|(Bv; \gamma(I - B)v)\| \max\{1, \gamma\} (1 + \frac{\ell}{\gamma}) + \|Bv\|) \|\hat{x} - \bar{x}\|. \end{aligned} \quad (33)$$

(b) For any  $B \in \partial_B \mathcal{P}_{\gamma^{-1}q}(\hat{x} - \gamma^{-1}F(\hat{x}))$  such that  $C := B(I - \gamma^{-1}F'(\hat{x})) - I \in \mathcal{T}_{u_\gamma}(\hat{x})$  is nonsingular, by taking  $M := \gamma C^\top$  one has

$$\|C^{-1}u_\gamma(\hat{x}) - (\hat{x} - \bar{x})\| \leq \frac{\epsilon}{2\sqrt{2}(\ell+1)} (\max\{1, \gamma\} (1 + \frac{\ell}{\gamma}) \|(BM^{-1}; \gamma(I - B)M^{-1})\|_F + \|BM^{-1}\|_F) \|\hat{x} - \bar{x}\|.$$

*Proof* (a) Note that whenever  $\hat{x}$  satisfies  $\|\hat{x} - \bar{x}\| \leq \frac{\min\{\delta_q, r\}}{(1+\frac{\ell}{\gamma})\max\{1, \gamma\}}$ , one can get from Lemma 4 that

$$\begin{aligned} \|(\hat{y} - \bar{x}; \gamma(\hat{z} - \hat{y}) + F(\bar{x}))\| &= \|(\hat{u} + \hat{x} - \bar{x}; F(\bar{x}) - F(\hat{x}) - \gamma\hat{u})\| \\ &\leq \max\{1, \gamma\}\|\hat{z} - \bar{z}\| \leq \max\{1, \gamma\}(1 + \frac{\ell}{\gamma})\|\hat{x} - \bar{x}\| \leq \min\{\delta_q, r\}, \end{aligned}$$

where  $\hat{u} := u_\gamma(\hat{x})$ ,  $\hat{z} := \hat{x} - \gamma^{-1}F(\hat{x})$ ,  $\hat{y} := \hat{u} + \hat{x} = \mathcal{P}_{\gamma^{-1}q}(\hat{z})$ , and  $\bar{z} := \bar{x} - \gamma^{-1}F(\bar{x})$ . Therefore, whenever  $\hat{x}$  satisfies  $\|\hat{x} - \bar{x}\| \leq \frac{\min\{\delta_q, r\}}{(1+\frac{\ell}{\gamma})\max\{1, \gamma\}}$ , one can take  $(d; d^*) = (\hat{y}; \gamma(\hat{z} - \hat{y})) \in \text{gph}\partial q \cap \mathbb{B}_{\delta_q}(\bar{x}, -F(\bar{x}))$  in (31) such that for all  $(e; e^*) \in L \in \mathcal{S}_{\partial q}^*(d, d^*)$ ,

$$\begin{aligned} &|\langle \gamma e, \hat{z} - \bar{z} \rangle - \langle e^* + \gamma e, \hat{y} - \bar{x} \rangle| \\ &= |\langle e^* + \gamma e, \hat{y} - \bar{x} \rangle - \langle \gamma e, \hat{z} - \bar{x} + \frac{1}{\gamma}F(\bar{x}) \rangle| = |\langle e^*, \hat{y} - \bar{x} \rangle - \langle e, \gamma(\hat{z} - \hat{y}) + F(\bar{x}) \rangle| \\ &\leq \frac{\epsilon}{2\sqrt{2}(\ell+1)}\|(e; e^*)\| \|(\hat{y} - \bar{x}; \gamma(\hat{z} - \hat{y}) + F(\bar{x}))\| \leq \frac{\epsilon}{2\sqrt{2}(\ell+1)}\|(e; e^*)\| \max\{1, \gamma\}(1 + \frac{\ell}{\gamma})\|\hat{x} - \bar{x}\|. \end{aligned} \quad (34)$$

Also, by using (28) of Lemma 3 one can get

$$\mathcal{S}_{\partial q}^*(\hat{y}, \gamma(\hat{z} - \hat{y})) = \{\text{rge}(-\gamma^{-1}B; -(I - B)) \mid B \in \partial_B \mathcal{P}_{\gamma^{-1}q}(\hat{z})\} = \{\text{rge}(B; \gamma(I - B)) \mid B \in \partial_B \mathcal{P}_{\gamma^{-1}q}(\hat{z})\}.$$

Thus,  $(e; e^*) \in L$  if and only if  $(e; e^*) = (Bv; \gamma(I - B)v)$  for some  $v \in \mathbb{R}^n$ , where  $B \in \partial_B \mathcal{P}_{\gamma^{-1}q}(\hat{z})$  is the symmetric positive semidefinite  $n \times n$  matrix such that  $L = \text{rge}(B; \gamma(I - B))$ . Moreover, in this case, one has from (28) that  $(e^* + \gamma e; \gamma e) = (\gamma(I - B)v + \gamma Bv; \gamma Bv) = \gamma(v; Bv) \in \mathcal{S}_{\mathcal{P}_{\gamma^{-1}q}}^*(z, y)$ .

Therefore, if  $\hat{x}$  satisfies  $\|\hat{x} - \bar{x}\| \leq \frac{\min\{\delta_q, r\}}{(1+\frac{\ell}{\gamma})\max\{1, \gamma\}}$ , one can obtain from (34) that for any  $v \in \mathbb{R}^n$  and any  $B \in \partial_B \mathcal{P}_{\gamma^{-1}q}(\hat{x} - \gamma^{-1}F(\hat{x}))$  it holds that

$$|\langle \gamma Bv, \hat{z} - \bar{z} \rangle - \langle \gamma v, \hat{y} - \bar{x} \rangle| \leq \frac{\epsilon}{2\sqrt{2}(\ell+1)}\|(Bv; \gamma(I - B)v)\| \max\{1, \gamma\}(1 + \frac{\ell}{\gamma})\|\hat{x} - \bar{x}\|. \quad (35)$$

Note that both (32) and (35) hold if  $\hat{x}$  satisfies  $\|\hat{x} - \bar{x}\| \leq \min\{\frac{\min\{\delta_q, r\}}{(1+\frac{\ell}{\gamma})\max\{1, \gamma\}}, \delta\} \leq \min\{\frac{\delta_q}{1+\ell}, r\}$ . In this case, one has

$$\begin{aligned} &|\langle \gamma v, u_\gamma(\hat{x}) \rangle - (\langle \gamma Bv, (I - \gamma^{-1}F'(\hat{x}))(\hat{x} - \bar{x}) \rangle - \langle \gamma v, \hat{x} - \bar{x} \rangle)| \\ &= |\langle \gamma v, \hat{u} \rangle + \langle \gamma v, \hat{x} - \bar{x} \rangle - \langle \gamma Bv, \hat{x} - \bar{x} - \gamma^{-1}(F(\hat{x}) - F(\bar{x})) - \gamma^{-1}F'(\hat{x})(\hat{x} - \bar{x}) + \gamma^{-1}(F(\hat{x}) - F(\bar{x})) \rangle| \\ &\leq |\langle \gamma v, \hat{y} - \bar{x} \rangle - \langle \gamma Bv, \hat{z} - \bar{z} \rangle| + |\langle Bv, F(\hat{x}) - F(\bar{x}) - F'(\hat{x})(\hat{x} - \bar{x}) \rangle| \\ &\leq \frac{\epsilon}{2\sqrt{2}(\ell+1)}\|(Bv; \gamma(I - B)v)\| \max\{1, \gamma\}(1 + \frac{\ell}{\gamma})\|\hat{x} - \bar{x}\| + \frac{\epsilon}{2\sqrt{2}(\ell+1)}\|Bv\|\|\hat{x} - \bar{x}\|, \end{aligned}$$

which completes the proof of (a).

(b) When  $C$  is not singular, we can take  $v_i$  in (33) as the  $i$ -th column of  $M^{-1} = (\gamma C^\top)^{-1}$  for  $i = 1, \dots, n$ , that is,  $\gamma v_i^\top$  is the  $i$ -th row of  $C^{-1}$ . Consequently,

$$\begin{aligned}
 & \|C^{-1}u_\gamma(\hat{x}) - (\hat{x} - \bar{x})\| \\
 &= \|C^{-1}u_\gamma(\hat{x}) - C^{-1}B(I - \gamma^{-1}F'(\hat{x}))(\hat{x} - \bar{x}) + C^{-1}(\hat{x} - \bar{x})\| \\
 &= \left( \sum_{i=1}^n \left| \langle \gamma v_i, u_\gamma(\hat{x}) \rangle - \langle \gamma B v_i, (I - \gamma^{-1}F'(\hat{x}))(\hat{x} - \bar{x}) \rangle - \langle \gamma v_i, \hat{x} - \bar{x} \rangle \right|^2 \right)^{\frac{1}{2}} \\
 &\leq \frac{\epsilon}{2\sqrt{2}(\ell+1)} \left( \sum_{i=1}^n \left( \max\{1, \gamma\} \left(1 + \frac{\ell}{\gamma}\right) \|B v_i; \gamma(I - B)v_i\| + \|B v_i\| \right)^2 \right)^{\frac{1}{2}} \|\hat{x} - \bar{x}\| \\
 &\leq \frac{\epsilon}{2\sqrt{2}(\ell+1)} \left( \max\{1, \gamma\} \left(1 + \frac{\ell}{\gamma}\right) \left( \sum_{i=1}^n \|B v_i; \gamma(I - B)v_i\|^2 \right)^{\frac{1}{2}} + \left( \sum_{i=1}^n \|B v_i\|^2 \right)^{\frac{1}{2}} \right) \|\hat{x} - \bar{x}\| \\
 &= \frac{\epsilon}{2\sqrt{2}(\ell+1)} \left( \max\{1, \gamma\} \left(1 + \frac{\ell}{\gamma}\right) \|BM^{-1}; \gamma(I - B)M^{-1}\|_F + \|BM^{-1}\|_F \right) \|\hat{x} - \bar{x}\|,
 \end{aligned}$$

where the last inequality comes from the triangle inequality and the final equality comes from the definition of the Frobenius norm. This completes the proof of (b).  $\square$

The G-semismoothness of  $u_\gamma$  with respect to  $\mathcal{T}_{u_\gamma}$  is given as follows.

**COROLLARY 1.** *Let  $\bar{x}$  be a solution to the GE (3), and  $\mathbb{B}_r(\bar{x})$  be the ball such that  $F$  is Lipschitz continuous on it with modulus  $\ell \geq 0$ . Assume that  $\partial q$  is SCD semismooth\* at  $(\bar{x}, -F(\bar{x}))$ . For any  $\gamma > 0$  the mapping  $u_\gamma$  defined in (14) is G-semismooth with respect to  $\mathcal{T}_{u_\gamma}$  given in (30) at  $\bar{x}$ .*

*Proof* Let  $\epsilon > 0$  be arbitrarily given. Note that one can find two positive constants  $\delta$  and  $\delta_q$  with  $\delta \leq \min\{\frac{\delta_q}{1+\ell}, r\}$  such that (31) and (32) hold. Then by Proposition 5(a) we know that for any  $\hat{x} \in \mathbb{R}^n$  satisfying  $\|\hat{x} - \bar{x}\| \leq \min\{\frac{\min\{\delta_q, r\}}{(1+\frac{\ell}{\gamma})\max\{1, \gamma\}}, \delta\}$ , it holds for any  $v \in \mathbb{R}^n$  and any  $B \in \partial_B \mathcal{P}_{\gamma^{-1}q}(\hat{x} - \gamma^{-1}F(\hat{x}))$  that

$$\begin{aligned}
 & \left| \langle \gamma v, u_\gamma(\hat{x}) \rangle - \langle \gamma B v, (I - \gamma^{-1}F'(\hat{x}))(\hat{x} - \bar{x}) \rangle - \langle \gamma v, \hat{x} - \bar{x} \rangle \right| \\
 & \leq \frac{\epsilon}{2\sqrt{2}(\ell+1)} \left( \|B v; \gamma(I - B)v\| \max\{1, \gamma\} \left(1 + \frac{\ell}{\gamma}\right) + \|B v\| \right) \|\hat{x} - \bar{x}\|.
 \end{aligned}$$

Since  $B$  is a firmly nonexpansive mapping by [24, Proposition 3.22], one can get that  $\|(B v; \gamma(I - B)v)\| \leq \max\{1, \gamma\}\|v\|$ . Therefore, by taking  $v$  as the vector such that  $\|\gamma v\| = 1$  and

$$\langle \gamma v, u_\gamma(\hat{x}) - (B(I - \gamma^{-1}F'(\hat{x})) - I)(\hat{x} - \bar{x}) \rangle = \|u_\gamma(\hat{x}) - (B(I - \gamma^{-1}F'(\hat{x})) - I)(\hat{x} - \bar{x})\|,$$

one can get  $\|B v\| \leq \max\{1, \frac{1}{\gamma}\}$  and

$$\|u_\gamma(\hat{x}) - (B(I - \gamma^{-1}F'(\hat{x})) - I)(\hat{x} - \bar{x})\| \leq \frac{\epsilon}{2\sqrt{2}(\ell+1)} \left( \max\{1, \gamma^2\}\|v\| \left(1 + \frac{\ell}{\gamma}\right) + \|B v\| \right) \|\hat{x} - \bar{x}\|.$$

Consequently,  $u_\gamma$  is G-semismooth with respect to  $\mathcal{T}_{u_\gamma}$ , and this completes the proof.  $\square$

**5.2. Regularity conditions** When using  $\mathcal{T}_{u_\gamma}$  in (30) as a generalized Jacobian in a G-semismooth Newton method for solving the nonsmooth equation  $u_\gamma(x) = 0$ , conditions for ensuring  $\mathcal{T}_{u_\gamma}$  being nonsingular around a reference solution should be verified. On the other hand, Algorithm 4 is well-defined only if the coefficient matrix of the linear equation (15) is nonsingular. In fact, we have the following results on the equivalence between the two regularity conditions mentioned above.

**LEMMA 5.** *Let  $\gamma > 0$  and  $u_\gamma$  be the function defined in (14). For any  $x \in \mathbb{R}^n$ , every element of  $\mathcal{T}_{u_\gamma}(x)$  is nonsingular if and only if  $(Y^{*\top} F'(x) + X^{*\top})$  is nonsingular for all  $X^* \in \mathbb{R}^{n \times n}$  and  $Y^* \in \mathbb{R}^{n \times n}$  such that  $\text{rge}(Y^*; X^*) \in \mathcal{S}_{\partial q}^*(x + u_\gamma(x), -\gamma u_\gamma(x) - F(x))$ .*

*Proof* For convenience, denote  $u := u_\gamma(x)$ ,  $z := x - \gamma^{-1} F(x)$ , and  $y := u + x = \mathcal{P}_{\gamma^{-1}q}(z)$ . It is easy to see from (5) that  $\gamma(z - y) \in \partial q(y)$ , so that  $(x + u_\gamma(x); -\gamma u_\gamma(x) - F(x)) = (y; \gamma(z - y)) \in \text{gph} \partial q$ . By using (28) of Lemma 3 we have

$$\mathcal{S}_{\partial q}^*(x + u_\gamma(x), -\gamma u_\gamma(x) - F(x)) = \{\text{rge}(B; \gamma(I - B)) \mid B \in \partial_B \mathcal{P}_{\gamma^{-1}q}(z)\}. \quad (36)$$

By (30) one has  $\mathcal{T}_{u_\gamma}(x) = \{B(I - \gamma^{-1} F'(x)) - I \mid B \in \partial_B \mathcal{P}_{\gamma^{-1}q}(z)\}$ . Thus, it is sufficient to prove that for every  $B \in \partial_B \mathcal{P}_{\gamma^{-1}q}(z)$ , the matrix  $B(I - \gamma^{-1} F'(x)) - I$  is nonsingular if and only if  $Y^{*\top} F'(x) + X^{*\top}$  is nonsingular for all  $X^* \in \mathbb{R}^{n \times n}$  and  $Y^* \in \mathbb{R}^{n \times n}$  such that  $\text{rge}(Y^*; X^*) = \text{rge}(B; \gamma(I - B))$ .

Fix  $B \in \partial_B \mathcal{P}_{\gamma^{-1}q}(z)$ . On the one hand, suppose that  $B(I - \gamma^{-1} F'(x)) - I \in \mathcal{T}_{u_\gamma}(x)$  is singular. By taking  $Y^* = B$  and  $X^* = \gamma(I - B)$  one has  $Y^{*\top} F'(x) + X^{*\top} = BF'(x) + \gamma(I - B) = -\gamma(B(I - \gamma^{-1} F'(x)) - I)$ , which is also singular. On the other hand, if  $Y^{*\top} F'(x) + X^{*\top}$  is singular with  $\text{rge}(Y^*; X^*) = \text{rge}(B; \gamma(I - B))$ , i.e., there exists a nonzero vector  $v \in \mathbb{R}^n$  such that  $Y^{*\top} F'(x)v + X^{*\top} v = 0$ , one has for every  $w \in \mathbb{R}^n$ ,

$$w^\top Y^{*\top} F'(x)v + w^\top X^{*\top} v = (Y^* w)^\top F'(x)v + (X^* w)^\top v = ((Y^* w)^\top F'(x) + (X^* w)^\top) v = 0.$$

Since  $\text{rge}(Y^*; X^*) = \text{rge}(B; \gamma(I - B))$ , one has

$$w^\top (BF'(x) + \gamma(I - B)) v = ((Bw)^\top F'(x) + (\gamma(I - B)w)^\top) v = 0 \quad \forall w \in \mathbb{R}^n.$$

Therefore, we have  $(BF'(x) + \gamma(I - B)) v = 0$ , which implies that the matrix  $-\gamma(B(I - \gamma^{-1} F'(x)) - I)$  is nonsingular. This completes the proof.  $\square$

**COROLLARY 2.** *Let  $\bar{x}$  be a solution to the GE (3). Then,  $F + \partial q$  is SCD regular around  $(\bar{x}, 0)$  if and only if every element of  $\mathcal{T}_{u_\gamma}(\bar{x})$  defined in (30) is nonsingular for any  $\gamma > 0$ .*

*Proof* According to [25, Proposition 5.1(ii)] we know that  $F + \partial q$  is SCD regular around  $(\bar{x}, 0)$  if and only if  $Y^{*\top} F'(\bar{x}) + X^{*\top}$  is nonsingular for all  $X^* \in \mathbb{R}^{n \times n}$  and  $Y^* \in \mathbb{R}^{n \times n}$  such that  $\text{rge}(Y^*; X^*) \in \mathcal{S}_{\partial q}^*(\bar{x}, -F(\bar{x}))$ . Thus, the conclusion follows from Lemma 5. This completes the proof.  $\square$

The following result is also a consequence of Lemma 5.

**LEMMA 6.** *Let  $u_\gamma$  ( $\gamma > 0$ ) be the function defined in (14), and  $\mathcal{J}$  be the mapping defined in (13). For any  $x \in \mathbb{R}^n$  such that every element of  $\mathcal{T}_{u_\gamma}(x)$  is nonsingular, one has*

$$\mathcal{S}_{\mathcal{J}}^*((x, x + u_\gamma(x)), (F(x) + d^*, -u_\gamma(x))) = \left\{ \text{rge}(C_B; I) \mid B \in \partial_B \mathcal{P}_{\gamma^{-1}q}(x - \gamma^{-1}F(x)) \right\}, \quad (37)$$

$$\text{where } C_B := \begin{pmatrix} B\tilde{M}^{-1} & B\tilde{M}^{-1} \\ \gamma(I-B)\tilde{M}^{-1} & \gamma(I-B)\tilde{M}^{-1} - I \end{pmatrix} \text{ with } \tilde{M} := F'(x)^\top B + \gamma(I-B).$$

*Proof* According to [25, Proposition 5.1(1)] we know that, for any  $(x; d) \in \mathbb{R}^{2n}$  and  $d^* \in \partial q(d)$ ,

$$\mathcal{S}_{\mathcal{J}}^*((x, d), (F(x) + d^*, x - d)) = \left\{ \text{rge} \left( \begin{pmatrix} Y^* & 0 \\ 0 & -I \end{pmatrix}; \begin{pmatrix} F'(x)^\top Y^* & -I \\ X^* & I \end{pmatrix} \right) \mid \text{rge}(Y^*; X^*) \in \mathcal{S}_{\partial q}^*(d, d^*) \right\}.$$

From Lemma 5 we know that every element of  $\mathcal{T}_{u_\gamma}(x)$  is nonsingular if and only if  $(Y^{*\top} F'(x) + X^{*\top})$  is nonsingular for all  $X^*$  and  $Y^*$  such that  $\text{rge}(Y^*; X^*) \subseteq \mathcal{S}_{\partial q}^*(x + u_\gamma(x), -\gamma u_\gamma(x) - F(x))$ . Moreover, (36) holds in this case, so that with  $z := x - \gamma^{-1}F(x)$  one has

$$\begin{aligned} & \mathcal{S}_{\mathcal{J}}^*((x, x + u_\gamma(x)), (F(x) + d^*, -u_\gamma(x))) \\ &= \left\{ \text{rge} \left( \begin{pmatrix} Y^* & 0 \\ 0 & -I \end{pmatrix} \begin{pmatrix} F'(x)^\top Y^* & -I \\ X^* & I \end{pmatrix}^{-1}; I \right) \mid \text{rge}(Y^*; X^*) = \text{rge}(B; \gamma(I-B)), \right. \\ & \quad \left. B \in \partial_B \mathcal{P}_{\gamma^{-1}q}(z) \right\}. \end{aligned} \quad (38)$$

Fix  $B \in \partial_B \mathcal{P}_{\gamma^{-1}q}(z)$  and choose  $(Y^*; X^*)$  such that  $\text{rge}(Y^*; X^*) = \text{rge}(B; \gamma(I-B))$ . Since that  $(B; \gamma(I-B))$  has full column rank, there exists a nonsingular matrix  $P \in \mathbb{R}^{n \times n}$  such that  $Y^* = BP$  and  $X^* = \gamma(I-B)P$ . Moreover, since  $M := F'(x)^\top Y^* + X^* = (F'(x)^\top B + \gamma(I-B))P$  is nonsingular, one has  $M^{-1} = P^{-1}(F'(x)^\top B + \gamma(I-B))^{-1}$  and

$$\begin{aligned} & \begin{pmatrix} Y^* & 0 \\ 0 & -I \end{pmatrix} \begin{pmatrix} F'(x)^\top Y^* & -I \\ X^* & I \end{pmatrix}^{-1} = \begin{pmatrix} BP & 0 \\ 0 & -I \end{pmatrix} \begin{pmatrix} F'(x)^\top BP & -I \\ \gamma(I-B)P & I \end{pmatrix}^{-1} \\ &= \begin{pmatrix} BP & 0 \\ 0 & -I \end{pmatrix} \begin{pmatrix} M^{-1} & M^{-1} \\ -\gamma(I-B)PM^{-1} & I - \gamma(I-B)PM^{-1} \end{pmatrix} = \begin{pmatrix} BPM^{-1} & BPM^{-1} \\ \gamma(I-B)PM^{-1} & \gamma(I-B)PM^{-1} - I \end{pmatrix} \\ &= \begin{pmatrix} B(F'(x)^\top B + \gamma(I-B))^{-1} & B(F'(x)^\top B + \gamma(I-B))^{-1} \\ \gamma(I-B)(F'(x)^\top B + \gamma(I-B))^{-1} & \gamma(I-B)(F'(x)^\top B + \gamma(I-B))^{-1} - I \end{pmatrix}. \end{aligned}$$

Therefore, one can remove the dependence on the precise choice of  $(Y^*; X^*)$  and use only the information of  $B$  in (38). So we obtain (37). This completes the proof.  $\square$

**5.3. Equivalence to a G-semismooth Newton method** Based on the discussions in the previous subsections, a G-semismooth Newton method (Algorithm 1) for solving (3) via its equivalent form (14) can be given as follows.

---

**Algorithm 6** A G-semismooth Newton method for solving (3)

---

**Input:**  $x^{(0)} \in \mathbb{R}^n$ ,  $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$ , and  $q : \mathbb{R}^n \rightarrow \bar{\mathbb{R}}$ .

**Output:**  $\{x^{(k)}\}$ .

**For**  $k = 0, 1, \dots$ ,

1. if  $0 \in F(x^{(k)}) + \partial q(x^{(k)})$ , stop the algorithm;
  2. select  $\gamma^{(k)} > 0$ , and compute  $u^{(k)} := u_{\gamma^{(k)}}(x^{(k)})$ ;
  3. select  $V^{(k)} \in \mathcal{T}_{u_{\gamma^{(k)}}}(x^{(k)})$  via (30), then compute the Newton direction  $\Delta x^{(k)}$  from  $V^{(k)} \Delta x^{(k)} = -u^{(k)}$ , and obtain the new iterate via  $x^{(k+1)} := x^{(k)} + \Delta x^{(k)}$ .
- 

Note that Algorithm 6 is essentially a “uniform” version of Algorithm 1 for solving a family of problems, i.e.,  $\{u_\gamma(x) = 0, \gamma > 0\}$ , sharing the common solutions. In each iteration, one selects one instance of these problems and performs the G-semismooth Newton step via  $\mathcal{T}_{u_\gamma}$ . In the following, we show that Algorithm 6 is well-defined if Algorithm 4 is, and vice versa. Moreover, a sequence generated by one of them can be treated as the one generated by the other.

**LEMMA 7.** *Given  $x^{(0)} \in \mathbb{R}^n$  and  $\{\gamma^{(k)}\}$ . Suppose that both Algorithm 4 and Algorithm 6 generate the same point  $x^{(\bar{k})}$  after the iteration indexed by  $(\bar{k} - 1) \geq 0$ , and  $B^{(\bar{k})} \in \partial_B \mathcal{P}_{\gamma^{(\bar{k})} q}(\cdot) \big|_{x^{(\bar{k}) - \gamma^{(\bar{k})} F(x^{(\bar{k})})}$  is chosen such that*

$$\text{rge}(Y^{*(\bar{k})}; X^{*(\bar{k})}) := \text{rge}(B^{(\bar{k})}; \gamma^{(\bar{k})}(I - B^{(\bar{k})})) \quad \text{and} \quad V^{(\bar{k})} := B^{(\bar{k})}(I - \gamma^{(\bar{k})^{-1}} F'(x^{(\bar{k})})) - I. \quad (39)$$

*Then, one has  $\text{rge}(Y^{*(\bar{k})}; X^{*(\bar{k})}) \in \mathcal{S}_{\partial q}^*(\hat{d}^{(\bar{k})}, \hat{d}^{*(\bar{k})})$  and  $V^{(\bar{k})} \in \mathcal{T}_{u_{\gamma^{(\bar{k})}}}(x^{(\bar{k})})$ . Moreover, both the two algorithms generate the same  $x^{(\bar{k}+1)}$  if (39) is used in them.*

*Proof* Recall that, in iteration indexed by  $\bar{k}$  of Algorithm 4, the Newton direction  $\Delta x^{(\bar{k})}$  generated by (15), i.e. the linear system

$$(Y^{*(\bar{k})^\top} F'(x^{(\bar{k})}) + X^{*(\bar{k})^\top}) \Delta x^{(\bar{k})} = (\gamma^{(\bar{k})} Y^{*(\bar{k})^\top} + X^{*(\bar{k})^\top}) u^{(\bar{k})}, \quad (40)$$

where  $u^{(\bar{k})} := u_{\gamma^{(\bar{k})}}(x^{(\bar{k})})$ ,  $\hat{d}^{(\bar{k})} := x^{(\bar{k})} + u^{(\bar{k})}$  and  $\hat{d}^{*(\bar{k})} := -\gamma^{(\bar{k})}u^{(\bar{k})} - F(x^{(\bar{k})})$ . Thus, by denoting  $W$  as the nonsingular transition matrix such that  $Y^{*(\bar{k})} = B^{(\bar{k})}W$  and  $X^{*(\bar{k})} = \gamma^{(\bar{k})}(I - B^{(\bar{k})})W$ , (40) is equivalent to

$$(W^\top B^{(\bar{k})} F'(x^{(\bar{k})}) + \gamma^{(\bar{k})} W^\top (I - B^{(\bar{k})})) \Delta x^{(\bar{k})} = (\gamma^{(\bar{k})} W^\top B^{(\bar{k})} + \gamma^{(\bar{k})} W^\top (I - B^{(\bar{k})})) u^{(\bar{k})},$$

or equivalently,  $(B^{(\bar{k})}(I - \frac{1}{\gamma^{(\bar{k})}} F'(x^{(\bar{k})})) - I) \Delta x^{(\bar{k})} = -u^{(\bar{k})}$ , which is exactly the Newton system of the G-semismooth Newton method using  $\mathcal{T}_{u_{\gamma^{(\bar{k})}}}$  (Algorithm 6). This completes the proof.  $\square$

Next, we show that the local superlinear convergence of Algorithm 6 can be obtained under the assumptions made in Theorem 2 (i.e., [25, Theorem 5.2]).

**PROPOSITION 6.** *Let  $\bar{x}$  be a solution of (3). Assume that  $\partial q$  is SCD semismooth\* at  $(\bar{x}, -F(\bar{x}))$  and  $\partial q + F$  is SCD regular around  $(\bar{x}, 0)$ . Then for every pair  $\underline{\gamma}, \bar{\gamma}$  with  $0 < \underline{\gamma} \leq \bar{\gamma}$  there exists a neighborhood  $\mathcal{U}$  of  $\bar{x}$  such that for every starting point  $x^{(0)} \in \mathcal{U}$  Algorithm 6 produces a sequence  $\{x^{(k)}\}$  converging superlinearly to  $\bar{x}$ , provided we choose in every iteration step  $\gamma^{(k)} \in [\underline{\gamma}, \bar{\gamma}]$ .*

*Proof* From [25, Proposition 5.1] we know  $\partial q + F$  is SCD regular around  $(\bar{x}, 0)$  if and only if the mapping  $\mathcal{J}$  defined in (13) is SCD regular around  $((\bar{x}, \bar{x}), (0, 0))$ . Moreover, according to [24, Proposition 4.8], for every  $\kappa > \text{scd reg } \mathcal{J}((\bar{x}, \bar{x}), (0, 0))$ , one can find a positive radius  $\rho > 0$  such that for every  $\gamma \in [\underline{\gamma}, \bar{\gamma}]$  and every  $x \in \mathbb{R}^n$  such that  $(x, x + u_\gamma(x)) \in \mathbb{B}_\rho(\bar{x}, \bar{x})$ , the mapping  $\mathcal{J}$  is also SCD regular around  $((x, x + u_\gamma(x)), (-\gamma u_\gamma(x), -u_\gamma(x)))$  and  $\kappa > \text{scd reg } \mathcal{J}((x, x + u_\gamma(x)), (-\gamma u_\gamma(x), -u_\gamma(x)))$ . Thus, by Lemma 5 and [25, Proposition 5.1(1)], each element of  $\mathcal{T}_{u_\gamma}(x)$  is nonsingular. Moreover, by combining [24, eq. (34)] and Lemma 6, we obtain for any  $B \in \partial_B \mathcal{P}_{\gamma^{-1}q}(\cdot) |_{x-\gamma^{-1}F(x)}$ ,

$$\left\| \begin{pmatrix} B[F'(x)^\top B + \gamma(I - B)]^{-1} & B[F'(x)^\top B + \gamma(I - B)]^{-1} \\ \gamma(I - B)[F'(x)^\top B + \gamma(I - B)]^{-1} & \gamma(I - B)[F'(x)^\top B + \gamma(I - B)]^{-1} - I \end{pmatrix} \right\| \leq \kappa. \quad (41)$$

Let  $\mathbb{B}_r(\bar{x})$  be the ball such that  $F$  is Lipschitz continuous on it with modulus  $\ell \geq 0$ . Take  $0 < \xi < 1$  and

$$\epsilon := \frac{2\sqrt{2}(\ell+1)\xi}{\sqrt{n}\kappa \max_{\gamma \in [\underline{\gamma}, \bar{\gamma}]} \left( \max\{1, \gamma\} \left(1 + \frac{\ell}{\gamma}\right) + 1 \right)}. \quad (42)$$

Then, for any  $\hat{x}$  satisfying  $\|\hat{x} - \bar{x}\| \leq \min\left\{\frac{\min\{\delta_q, r\}}{(1+\frac{\ell}{\gamma})\max\{1, \gamma\}}, \delta\right\}$  and  $(\hat{x}, \hat{x} + u_\gamma(\hat{x})) \in \mathbb{B}_\rho(\bar{x}, \bar{x})$ , where  $\delta_q$  and  $\delta \leq \min\left\{\frac{\delta_q}{1+\ell}, r\right\}$  are the positive constants (depending on  $\epsilon$ ) such that (31) and (32) holds, one can



obtain from Proposition 5(b) the nonsingularity of each element  $\widehat{C} := \widehat{B}(I - \gamma^{-1}F'(\hat{x})) - I \in \mathcal{T}_{u_\gamma}(\hat{x})$  with  $\widehat{B} \in \partial_B \mathcal{P}_{\gamma^{-1}q}(\hat{x} - \gamma^{-1}F(\hat{x}))$ , and

$$\begin{aligned} & \|\widehat{C}^{-1}u_\gamma(\hat{x}) - (\hat{x} - \bar{x})\| \\ & \leq \frac{\epsilon}{2\sqrt{2}(\ell+1)} \left( \max\{1, \gamma\} \left(1 + \frac{\ell}{\gamma}\right) \|(\widehat{B}M^{-1}; \gamma(I - \widehat{B})M^{-1})\|_F + \|\widehat{B}M^{-1}\|_F \right) \|\hat{x} - \bar{x}\| \\ & \leq \frac{\sqrt{n}\kappa\epsilon}{2\sqrt{2}(\ell+1)} \max_{\gamma \in [\underline{\gamma}, \bar{\gamma}]} \left( \max\{1, \gamma\} \left(1 + \frac{\ell}{\gamma}\right) + 1 \right) \|\hat{x} - \bar{x}\| = \xi \|\hat{x} - \bar{x}\|, \end{aligned}$$

where  $M := \gamma \widehat{C}^\top \equiv -F'(\hat{x})^\top \widehat{B} - \gamma(I - \widehat{B})$ , and the last inequality comes from (41). Then by letting

$$\mathcal{U} := \left\{ \hat{x} \mid \|\hat{x} - \bar{x}\| \leq \min\left\{ \frac{\min\{\delta_q, r\}}{(1+\frac{\ell}{\gamma})\max\{1, \gamma\}}, \delta \right\}, (\hat{x}, \hat{x} + u_\gamma(\hat{x})) \in \mathbb{B}_\rho(\bar{x}, \bar{x}) \quad \forall \gamma \in [\underline{\gamma}, \bar{\gamma}] \right\}, \quad (43)$$

one can get the convergence of Algorithm 6 provided that  $x^{(0)} \in \mathcal{U}$ . The superlinear convergence rate comes from further shrinking the value of  $\epsilon$ , and this completes the proof.  $\square$

Finally, we have the following result, showing that Algorithm 4 is an instance of Algorithm 6.

**THEOREM 4.** *Under the assumptions of Theorem 2, Algorithm 4 is an instance of Algorithm 6, and the local superlinear convergence of Algorithm 4 (i.e., Theorem 2) can be guaranteed by Proposition 6.*

*Proof* The conclusion of the theorem follows immediately from Lemma 7 and Proposition 6.

$\square$

**6. Conclusions** This paper showed that the two typical implementable semismooth\* Newton methods are applications of G-semismooth Newton methods. This further enriches the comprehension of G-semismooth Newton methods and helps design practical Newton-type methods for GEs. Accordingly, a natural question is whether an implementable semismooth\* Newton method is achievable for solving a GE that cannot be reformulated to locally Lipschitz continuous equations. Moreover, the relationship between the generic semismooth\* Newton methods and G-semismooth Newton methods is still unclear, so another question is whether one can obtain generalizations of G-semismooth Newton methods, involving certain tractable “approximation steps”, that can solve a broader class of problems. We leave these questions for future research.

## Acknowledgments

The authors would like to thank the three anonymous referees for their careful reading of the manuscript and the constructive suggestions and comments which have helped us to improve the quality of this work.

## References

- [1] Aragón Artacho FJ, Dontchev AL, Gaydu M, Geoffroy MH, Veliov VM (2011) Metric regularity of Newton's iteration. *SIAM J. Control Optim.* 49(2):339–362.
- [2] Aragón Artacho FJ, Belyakov A, Dontchev AL, Lopez M (2014) Local convergence of quasi-Newton methods under metric regularity. *Comput. Optim. Appl.* 58(1):225–247.
- [3] Aragón Artacho FJ, Mordukhovich BS, Pérez Aros P (2024) Coderivative-based semi-Newton method in nonsmooth difference programming. *Math. Program.* 1–48.
- [4] Azé D, Chou CC (1995) On a Newton type iterative method for solving inclusions. *Math. Oper. Res.* 20(4):790–800.
- [5] Bonnans JF (1994) Local analysis of Newton-type methods for variational inequalities and nonlinear programming. *Appl. Math. Optim.* 29:161–186.
- [6] Bonnans JF, Shapiro A (2000) *Perturbation Analysis of Optimization Problems* (Springer, New York).
- [7] Cibulka R, Dontchev AL, Geoffroy MH (2011) Inexact Newton methods and Dennis–Moré theorems for nonsmooth generalized equations. *SIAM J. Control Optim.* 49(2):339–362.
- [8] Dias JM, Smirnov G (2012) On the Newton method for set-valued maps. *Nonlinear Anal.* 75:1219–1230.
- [9] Dontchev AL (1996) Local convergence of the Newton method for generalized equations. *Comptes Rendus de l'Académie des Sciences - Series I - Mathematics.* 322(4):327–331.
- [10] Dontchev AL (1996) Local analysis of a Newton-type method based on partial linearization. Renegar J, Shub M, and Smale S, eds. *The mathematics of numerical analysis* (AMS, Providence, RI) 295–306.
- [11] Dontchev AL, Rockafellar RT (2010) Newton's method for generalized equations: a sequential implicit function theorem. *Math. Program.* 123(1):139–159.
- [12] Ferreira OP (2015) A robust semi-local convergence analysis of Newton's method for cone inclusion problems in Banach spaces under affine invariant majorant condition. *J. Comput. Appl. Math.* 279:318–335.
- [13] Ferreira OP, Silva GN (2017) Kantorovich's theorem on Newton's method for solving strongly regular generalized equation. *SIAM J. Optim.* 27(2):910–926.
- [14] Ferreira OP, Silva GN (2018) Local convergence analysis of Newton's method for solving strongly regular generalized equations. *J. Math. Anal. Appl.* 458(1):481–496.
- [15] Fischer A (1999) Modified Wilson's method for nonlinear programs with nonunique multipliers. *Math. Program.* 24:699–727.
- [16] Gao Y, Sun DF (2009), Calibrating least squares semidefinite programming with equality and inequality constraints. *SIAM J. Matrix Anal. Appl.* 31:1432–1457.
- [17] Gaydu M, Geoffroy MH (2013) A Newton iteration for differentiable set-valued maps. *J. Math. Anal. Appl.* 399(1):213–224.
- [18] Geoffroy MH, Piétrus A (2004) Local convergence of some iterative methods for generalized equations. *J. Math. Anal. Appl.* 290(2):497–505.
- [19] Geoffroy MH, Piétrus A (2005) A general iterative procedure for solving nonsmooth generalized equations. *Comput. Optim. Appl.* 31(1):57–67.
- [20] Gfrerer H. (2025) On a globally convergent semismooth\* Newton method in nonsmooth nonconvex optimization. *Comput. Optim. Appl.* 91:67–124
- [21] Gfrerer H, Mandlmayr M, Outrata JV, Valdman J (2023) On the SCD semismooth\* Newton method for generalized equations with application to a class of static contact problems with Coulomb friction. *Comput. Optim. Appl.* 86:1159–1191.
- [22] Gfrerer H., Mordukhovich BS (2019) Second-order variational analysis of parametric constraint and variational systems. *SIAM J. Optim.* 29(1):423–453.
- [23] Gfrerer H, Outrata JV (2021) On a semismooth\* Newton method for solving generalized equations. *SIAM J. Optim.* 31(1):489–517.
- [24] Gfrerer H, Outrata JV (2022) On (local) analysis of multifunctions via subspaces contained in graphs of generalized derivatives. *J. Math. Anal. Appl.* 508(2):125895.
- [25] Gfrerer H, Outrata JV, Valdman J (2022) On the application of the SCD semismooth\* Newton method to variational inequalities of the second kind. *Set-Valued Var. Anal.* 30:1453–1484.
- [26] Gowda MS (2004) Inverse and implicit function theorems for H-differentiable and semismooth functions. *Optim. Methods Softw.* 19:443–461.
- [27] Henrion R, Outrata JV (2001) A subdifferential condition for calmness of multifunctions. *J. Math. Anal. Appl.* 258:110–130.
- [28] Hoheisel T, Kanzow C, Mordukhovich BS, Phan H (2012) Generalized Newton's method based on graphical derivatives. *Nonlinear Anal.* 75(3):1324–1340.
- [29] Hoheisel T, Kanzow C, Mordukhovich BS, Phan H (2013) Erratum to “Generalized Newton's method based on graphical derivatives” [Nonlinear Anal. TMA 75 (2012) 1324–1340]. *Nonlinear Anal.* 86:157–158.
- [30] Izmailov A, Solodov M (2014) *Newton-Type Methods for Optimization and Variational Problems* (Springer, New York).
- [31] Izmailov A, Solodov M (2010) Inexact Josephy–Newton framework for generalized equations and its applications to local analysis of Newtonian methods for constrained optimization. *Comput. Optim. Appl.* 46(2):347–368.
- [32] Izmailov A, Solodov M (2015) Newton-type methods: A broader view. *J. Optim. Theory Appl.* 164:577–620.
- [33] Josephy N (1979) *Newton's Method for Generalized Equations And The PIES Energy Model*. Ph.D. dissertation (University of Wisconsin-Madison).
- [34] Josephy N (1979) Quasi-Newton method for generalized equations. *Technical summary report* ( University of Wisconsin-Madison).
- [35] Khanh PD, Mordukhovich BS, Phat VT (2023) A generalized Newton method for subgradient systems. *Math. Oper. Res.* 48(4):1811–1845.
- [36] Khanh PD, Mordukhovich BS, Phat VT, Tran DB (2023) Generalized damped Newton algorithms in nonsmooth optimization via second-order subdifferentials. *J. Global Optim.* 86(1): 93–122.
- [37] Khanh PD, Mordukhovich BS, Phat VT (2024) Coderivative-based Newton methods in structured nonconvex and nonsmooth optimization. *ArXiv preprint*, <https://arxiv.org/abs/2403.04262>.
- [38] Khanh PD, Mordukhovich BS, Phat VT, Tran DB (2024) Globally convergent coderivative-based generalized Newton methods in nonsmooth optimization. *Math. Program.* 205(1):373–429.
- [39] Klatte D, Kummer B (2002) *Nonsmooth Equations in Optimization* (Kluwer Academic Publishers, New York).

- [40] Klatte D, Kummer B (2018) Approximation and generalized Newton methods. *Math. Program.* 168:673–716.
- [41] Kummer B (1988) Newton’s method for non-differentiable functions. Guddat J et al., eds. *Advances in mathematical optimization* (De Gruyter, Berlin), 114–125.
- [42] Kummer B (1992) Newton’s method based on generalized derivatives for nonsmooth functions: convergence analysis. Oettli W, Pallaschke D, eds. *Advances in Optimization* (Springer, Berlin, Heidelberg), 171–194.
- [43] Kummer B (2000) Generalized Newton and NCP-methods: Convergence, regularity, actions. *Discuss. Math. Differ. Incl. Control Optim.* 20(2):209–244.
- [44] Li X, Sun DF, Toh K-C (2018) A highly efficient semismooth Newton augmented Lagrangian method for solving Lasso problems. *SIAM J. Optim.* 28(1):433–458.
- [45] Meng F, Sun DF, Zhao G (2005) Semismoothness of solutions to generalized equations and the Moreau-Yosida regularization. *Math. Program.* 104:561–581.
- [46] Mifflin R (1977) Semismooth and semiconvex functions in constrained optimization. *SIAM J. Control Optim.* 15:957–972.
- [47] Mordukhovich BS (2006) *Variational Analysis and Generalized Differentiation I: Basic Theory* (Springer, Berlin).
- [48] Mordukhovich BS (2024) *Second-Order Variational Analysis in Optimization, Variational Stability, and Control: Theory, Algorithms, Applications* (Springer, Cham, Switzerland).
- [49] Mordukhovich BS, Sarabi ME (2021) Generalized Newton algorithms for tilt-stable minimizers in nonsmooth optimization. *SIAM J. Optim.* 31(2):1184–1214.
- [50] de Oliveira F, Ferreira O, Silva G (2019) Newton’s method with feasible inexact projections for solving constrained generalized equations. *Comput. Optim. Appl.* 72:159–177.
- [51] Pang JS (1990) Newton’s method for B-differentiable equations. *Math. Oper. Res.* 15(2):311–341.
- [52] Pang JS, Sun DF, Sun J (2003) Semismooth homeomorphisms and strong stability of semidefinite and Lorentz complementarity problems. *Math. Oper. Res.* 28(1):39–63.
- [53] Qi L, Sun DF, Zhou G (2000) A new look at smoothing Newton methods for nonlinear complementarity problems and box constrained variational inequalities. *Math. Program.* 87:1–35.
- [54] Qi L, Sun J (1993) A nonsmooth version of Newton’s method. *Math. Program.* 58:353–367.
- [55] Rademacher H (1919) Über partielle und totale differenzierbarkeit von Funktionen mehrerer Variablen und über die Transformation der Doppelintegrale. *Math. Ann.* 79:340–359.
- [56] Rockafellar RT (1976) Monotone operators and the proximal point algorithm. *SIAM J. Control Optim.* 14(5):877–898.
- [57] Rockafellar RT, Wets RJB (1998) *Variational Analysis* (Springer, Berlin).
- [58] Rockafellar RT (1970) *Convex Analysis* (Princeton University Press, Princeton).
- [59] Robinson S (1994) Newton’s method for a class of nonsmooth functions. *Set-Valued Anal.* 2:291–305.
- [60] Shapiro A (2003) Sensitivity analysis of generalized equations. *J. Math. Sci. (N. Y.)* 115:2554–2565.
- [61] Solodov MV, Svaiter BF (2002) A new proximal-based globalization strategy for the Joseph-Newton method for variational inequalities. *Optim. Methods Softw.* 17(5):965–983.
- [62] Sun DF (2006) The strong second-order sufficient condition and constraint nondegeneracy in nonlinear semidefinite programming and their implications. *Math. Oper. Res.* 31(4):761–776.
- [63] Yang L, Sun DF, Toh K-C (2015) SDPNAL+: a majorized semismooth Newton-CG augmented Lagrangian method for semidefinite programming with nonnegative constraints. *Math. Program. Comput.* 7(3):1–36.
- [64] Zhao X, Sun DF, Toh K-C (2010) A Newton-CG augmented Lagrangian method for semidefinite programming. *SIAM J. Optim.* 20(4):1737–1765.