THE JOSEPHY–NEWTON METHOD FOR SEMISMOOTH GENERALIZED EQUATIONS AND SEMISMOOTH SQP FOR OPTIMIZATION

A. F. Izmailov[†], A. S. Kurennoy[‡], and M. V. Solodov[§]

January 17, 2012

ABSTRACT

While generalized equations with differentiable single-valued base mappings and the associated Josephy–Newton method have been studied extensively, the setting with semismooth base mapping had not been previously considered (apart from the two special cases of usual nonlinear equations and of Karush-Kuhn-Tucker optimality systems). We introduce for the general semismooth case appropriate notions of solution regularity and prove local convergence of the corresponding Josephy–Newton method. As an application, we immediately recover the known primal-dual local convergence properties of semismooth SQP, but also obtain some new results that complete the analysis of the SQP primal rate of convergence, including its quasi-Newton variant.

Key words: generalized equation, B-differential, generalized Jacobian, BD-regularity, CD-regularity, strong regularity, semismoothness, Josephy–Newton method, SQP.
AMS subject classifications. 90C30, 90C55, 65K05.

---

[*] Research of the first two authors is supported by the Russian Foundation for Basic Research Grant 10-01-00251. The third author is supported in part by CNPq Grant 300513/2008-9, by PRONEX–Optimization, and by FAPERJ.
[†] Moscow State University, MSU, Uchebniy Korpus 2, VMK Faculty, OR Department, Leninskiye Gory, 119991 Moscow, Russia.
Email: `izmaf@ccas.ru`
[‡] Moscow State University, MSU, Uchebniy Korpus 2, VMK Faculty, OR Department, Leninskiye Gory, 119991 Moscow, Russia.
Email: `alex-kurennoy@yandex.ru`
[§] IMPA – Instituto de Matemática Pura e Aplicada, Estrada Dona Castorina 110, Jardim Botânico, Rio de Janeiro, RJ 22460-320, Brazil.
Email: `solodov@impa.br`

1 Introduction

We consider the generalized equation (GE)

$$\Phi(u) + N(u) \ni 0, \tag{1.1}$$

where $\Phi : \mathbb{R}^\nu \to \mathbb{R}^\nu$ is a (single-valued) *base mapping*, and $N(\cdot)$ is a *field multifunction* from $\mathbb{R}^\nu$ to the subsets of $\mathbb{R}^\nu$ (i.e., $N(u) \subset \mathbb{R}^\nu$ for each $u \in \mathbb{R}^\nu$). As is well known, this is a rather general framework [9]. For example, the case of usual nonlinear equations corresponds to $N(\cdot) = \{0\}$. More generally, when $N(\cdot) = N_Q(\cdot)$ is the normal map associated to a closed convex set $Q \subset \mathbb{R}^\nu$ then GE (1.1) is a variational inequality (VI)

$$u \in Q, \quad \langle \Phi(u),\, v - u\rangle \geq 0 \quad \forall\, v \in Q. \tag{1.2}$$

This in particular includes the Karush–Kuhn–Tucker (KKT) optimality conditions via the following well-known construction. Consider the problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & h(x) = 0,\ g(x) \leq 0, \end{array} \tag{1.3}$$

where the objective function $f : \mathbb{R}^n \to \mathbb{R}$ and the constraints mappings $h : \mathbb{R}^n \to \mathbb{R}^l$ and $g : \mathbb{R}^n \to \mathbb{R}^m$ are differentiable. Stationary points of problem (1.3) and the associated Lagrange multipliers are characterized by the KKT optimality system

$$\frac{\partial L}{\partial x}(x,\, \lambda,\, \mu) = 0, \quad h(x) = 0, \quad \mu \geq 0, \quad g(x) \leq 0, \quad \langle \mu,\, g(x)\rangle = 0, \tag{1.4}$$

where $L : \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m \to \mathbb{R}$ is the Lagrangian of problem (1.3):

$$L(x,\, \lambda,\, \mu) = f(x) + \langle \lambda,\, h(x)\rangle + \langle \mu,\, g(x)\rangle.$$

Then the KKT system (1.4) is a particular instance of GE (1.1) with the mapping $\Phi : \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m \to \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$ given by

$$\Phi(u) = \left(\frac{\partial L}{\partial x}(x,\, \lambda,\, \mu),\, -h(x),\, -g(x)\right), \quad u = (x,\, \lambda,\, \mu), \tag{1.5}$$

and with

$$N(\cdot) = N_Q(\cdot), \quad Q = \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m_+. \tag{1.6}$$

In this paper, we are interested in Newtonian methods for solving (1.1). An iteration of the Josephy–Newton method [16, 17, 3, 15] solves the following (partially) linearized GE:

$$\Phi(u^k) + J_k(u - u^k) + N(u) \ni 0, \tag{1.7}$$

where $u^k \in \mathbb{R}^\nu$ is the current approximation to a solution of (1.1) and $J_k \in \mathbb{R}^{\nu\times\nu}$. If Φ is differentiable then $J_k = \Phi'(u^k)$ is the basic choice. When in (1.1) we have $N(\cdot) = \{0\}$, then (1.7) is just the classical Newton iteration for nonlinear equations. If GE (1.1) is given by (1.5) and (1.6), i.e., it corresponds to KKT optimality conditions (1.4), then (1.7) is an iteration of the fundamental SQP algorithm [2] for optimization.

When the mapping Φ is not differentiable, a specific choice of J_k in the set of some generalized derivatives should be employed in (1.7) instead of $\Phi'(u^k)$. It appears that such methods have been previously studied only in the two special cases: that of the usual nonlinear equations (when $N(\cdot) = \{0\}$) [19, 20, 29, 25] and of the semismooth SQP (when (1.1) corresponds to KKT conditions) [27, 11]. The goal of this work is to develop the general semismooth Josephy–Newton framework. On the one hand, it extends the Josephy–Newton method [16, 17, 3, 15] to the case of a possibly nonsmooth base mapping Φ. On the other hand, it extends the semismooth Newton method for nonsmooth equations [19, 20, 29, 25] to the case of a GE. We shall also consider an application of this framework to optimization. As a by-product, we immediately recover the primal-dual local convergence result of [11] for semismooth SQP. We point out that this result follows here from a more general yet much shorter analysis. In addition, we obtain new and rather complete characterization of primal superlinear rate of convergence of semismooth SQP and its quasi-Newton variants.

In this work, we consider that Φ is only semismooth; differentiability of Φ is not assumed. One of the motivations to analyze this case comes from optimality systems for optimization problems with the objective function and constraints differentiable with locally Lipschitz-continuous first derivatives, but not necessarily twice differentiable. Problems with such smoothness properties arise in stochastic programming and optimal control (the so-called extended linear-quadratic problems [31, 32, 27]), in semi-infinite programming and in primal decomposition procedures (see [18, 26] and references therein). Once but not twice differentiable functions arise also when reformulating complementarity constraints as in [13] or in the lifting approach [33, 12]. Other possible sources are subproblems in penalty or augmented Lagrangian methods with lower-level constraints treated directly and upper-level inequality constraints treated via quadratic penalization or via augmented Lagrangian, which gives rise to certain terms that are not twice differentiable in general; see, e.g., [1].

The rest of the paper is organized as follows. In Section 2 we introduce the notion of strong regularity for GE with nondifferentiable base mapping, clarify its role and, in particular, what it means in the case of optimization problems. Section 3 constitutes convergence analysis of the semismooth Josephy–Newton method, including its perturbed and quasi-Newton variants. The application to the semismooth SQP for optimization is given in Section 4. We finish with some concluding remarks in Section 5. The appendix contains three technical lemmas concerned with partial derivatives and partial generalized Jacobians that are used in the paper.

Some final words about our notation are in order. The B-differential of $\Phi : \mathbb{R}^\nu \to \mathbb{R}^q$ at $u \in \mathbb{R}^\nu$ is the set

$$\partial_B \Phi(u) = \{J \in \mathbb{R}^{q\times\nu} \mid \exists\, \{u^k\} \subset \mathcal{S}_\Phi \text{ such that } \{u^k\} \to u,\ \{\Phi'(u^k)\} \to J\},$$

where $\mathcal{S}_\Phi$ is the set of points at which Φ is differentiable (this set is dense under our assumptions). Then the Clarke generalized Jacobian of Φ at u is given by

$$\partial \Phi(u) = \operatorname{conv} \partial_B \Phi(u),$$

where conv S stands for the convex hull of the set S. For a mapping $\Phi\colon \mathbb{R}^\nu \times \mathbb{R}^p \to \mathbb{R}^q$, the partial Clarke generalized Jacobian of Φ at $(u,\, v) \in \mathbb{R}^\nu \times \mathbb{R}^p$ with respect to u is the Clarke generalized Jacobian of the mapping $\Phi(\cdot,\, v)$, which we denote by $\partial_u \Phi(u,\, v)$.

The mapping $\Phi\colon \mathbb{R}^\nu \to \mathbb{R}^q$ is said to be *semismooth* [9, Section 7.4] at $u \in \mathbb{R}^\nu$ if it is locally Lipschitz-continuous around u, directionally differentiable at u in every direction, and satisfies the condition

$$\sup_{\Lambda \in \partial\Phi(u+v)} \|\Phi(u+v) - \Phi(u) - \Lambda v\| = o(\|v\|).$$

If the stronger condition

$$\sup_{\Lambda \in \partial\Phi(u+v)} \|\Phi(u+v) - \Phi(u) - \Lambda v\| = O(\|v\|^2)$$

holds, then Φ is said to be *strongly semismooth* at u.

We denote $B(u, \delta) = \{v \in \mathbb{R}^\nu \mid \|v - u\| \le \delta\}$, $u \in \mathbb{R}^\nu$, $\delta > 0$. The Euclidean projection of $u \in \mathbb{R}^\nu$ onto a closed convex set $S \subset \mathbb{R}^\nu$ is denoted by $\pi_S(u)$. Two properties that will be useful in the sequel are the following: if $C \subset \mathbb{R}^\nu$ is a closed convex cone then

$$\pi_C(u - \pi_C(u)) = 0 \quad \forall\, u \in \mathbb{R}^\nu, \tag{1.8}$$

and

$$\{u \in \mathbb{R}^\nu \mid \pi_C(u) = 0\} = C^\circ, \tag{1.9}$$

where $C^\circ = \{u \in \mathbb{R}^\nu \mid \langle u,\, v\rangle \le 0\ \forall\, v \in C\}$ is the negative dual cone to C.

2 Strong regularity

When Φ is differentiable, closely related to convergence of the Josephy–Newton scheme (1.7) is the notion of strong regularity, introduced in [30] (although, it should be mentioned that in the differentiable case convergence can be established under weaker assumptions [3]). Specifically, a solution $\bar{u}$ of GE (1.1) is referred to as *strongly regular* if for each $r \in \mathbb{R}^\nu$ close enough to 0 the perturbed (partially) linearized GE

$$\Phi(\bar{u}) + \Phi'(\bar{u})(u - \bar{u}) + N(u) \ni r$$

has near $\bar{u}$ the unique solution $u(r)$ and the mapping $u(\cdot)$ is locally Lipschitz-continuous at 0. Clearly, $\bar{u}$ is a strongly regular solution of GE (1.1) if and only if it is a strongly regular solution of its linearization

$$\Phi(\bar{u}) + \Phi'(\bar{u})(u - \bar{u}) + N(u) \ni 0.$$

Characterizations of strong regularity for generalized equations by means of generalized differentiation were derived in [21] (see also [22]).

We next introduce an appropriate generalization of the notion of regularity for the case when Φ is not differentiable.

Definition 2.1 A solution $\bar{u} \in \mathbb{R}^\nu$ of GE (1.1) is referred to as *strongly regular with respect to a set* $\Delta \subset \mathbb{R}^{\nu\times\nu}$ if for each $J \in \Delta$ the solution $\bar{u}$ of the GE

$$\Phi(\bar{u}) + J(u - \bar{u}) + N(u) \ni 0 \tag{2.1}$$

is strongly regular. (I.e., for each $J \in \Delta$ and for each $r \in \mathbb{R}^\nu$ close enough to 0, the perturbed partial linearization of (2.1)

$$\Phi(\bar{u}) + J(u - \bar{u}) + N(u) \ni r$$

has near $\bar{u}$ the unique solution $u_J(r)$ and the mapping $u_J(\cdot)$ is locally Lipschitz-continuous at 0.)

If $\Delta = \partial_B\Phi(\bar{u})$ ($\Delta = \partial\Phi(\bar{u})$) then $\bar{u}$ is referred to as a *BD-regular* (*CD-regular*) solution of GE (1.1).

Evidently, Definition 2.1 extends the following widely used notions: strong regularity [30] for the case of a smooth base mapping Φ and $\Delta = \{\Phi'(\bar{u})\}$, BD-regularity [24] and CD-regularity [28] for usual equations corresponding to $N(\cdot) = \{0\}$, $\Delta = \partial_B\Phi(\bar{u})$ and $\Delta = \partial\Phi(\bar{u})$, respectively.

The following result regarding the stability of strong regularity subject to small Lipschitzian perturbations follows, e.g., from [7, Theorem 1.4].

Proposition 2.1 *For given* $\Phi : \mathbb{R}^\nu \to \mathbb{R}^\nu$, $J \in \mathbb{R}^{\nu\times\nu}$ *and a multifunction* N *from* $\mathbb{R}^\nu$ *to the subsets of* $\mathbb{R}^\nu$, *let* $\bar{u}$ *be a strongly regular solution of GE* (2.1).

Then for any fixed neighborhood W *of* $\bar{u}$ *and any sufficiently small* $\ell \geq 0$, *there exist* $\bar{\ell} > 0$ *and neighborhoods* U *of* $\bar{u}$ *and* V *of* 0 *such that for any mapping* $R : \mathbb{R}^\nu \to \mathbb{R}^\nu$ *which is Lipschitz-continuous on* W *with Lipschitz constant* ℓ, *and for any* $r \in R(\bar{u}) + V$, *the GE*

$$R(u) + \Phi(\bar{u}) + J(u - \bar{u}) + N(u) \ni r$$

has in U *the unique solution* $u(r)$, *and the mapping* $u(\cdot)$ *is Lipschitz-continuous on* $R(\bar{u}) + V$ *with Lipschitz constant* $\bar{\ell}$.

We next use Proposition 2.1 to prove solvability of perturbed linearized GEs for all points close enough to a strongly regular solution and all matrices J close enough to the associated set Δ.

Proposition 2.2 *Let* $\Phi : \mathbb{R}^\nu \to \mathbb{R}^\nu$ *be continuous at* $\bar{u} \in \mathbb{R}^\nu$. *For a given multifunction* N *from* $\mathbb{R}^\nu$ *to the subsets of* $\mathbb{R}^\nu$, *let* $\bar{u}$ *be a solution of GE* (1.1), *strongly regular with respect to a compact set* $\Delta \subset \mathbb{R}^{\nu\times\nu}$.

Then there exist $\varepsilon > 0$, $\bar{\ell} > 0$ *and neighborhoods* $\tilde{U}$ *and* U *of* $\bar{u}$ *and* V *of* 0 *such that for any* $\tilde{u} \in \tilde{U}$, *any* $J \in \mathbb{R}^{\nu\times\nu}$ *satisfying*

$$\operatorname{dist}(J, \Delta) < \varepsilon, \tag{2.2}$$

and any $r \in V$, *the GE*

$$\Phi(\tilde{u}) + J(u - \tilde{u}) + N(u) \ni r \tag{2.3}$$

has in U *the unique solution* $u(r)$, *and the mapping* $u(\cdot)$ *is Lipschitz-continuous on* V *with Lipschitz constant* $\bar{\ell}$.

Proof. Fix any $\bar{J} \in \Delta$. For each $\tilde{u} \in \mathbb{R}^\nu$ and $J \in \mathbb{R}^{\nu\times\nu}$ define the mapping $R : \mathbb{R}^\nu \to \mathbb{R}^\nu$,

$$R(u) = \Phi(\tilde{u}) - \Phi(\bar{u}) - J\tilde{u} + \bar{J}\bar{u} + (J - \bar{J})u. \tag{2.4}$$

For any pre-fixed $\ell > 0$, the mapping R is Lipschitz-continuous on $\mathbb{R}^\nu$ with Lipschitz constant ℓ provided J is close enough to $\bar{J}$. Note also that $R(\bar{u}) = \Phi(\tilde{u}) - \Phi(\bar{u}) - J(\tilde{u} - \bar{u})$ tends to 0 as $\tilde{u} \to \bar{u}$. Therefore, by Proposition 2.1 applied with $W = \mathbb{R}^\nu$, there exist $\varepsilon > 0$, $\bar{\ell} > 0$ and neighborhoods $\tilde{U}$ and U of $\bar{u}$ and V of 0 such that for any $\tilde{u} \in \tilde{U}$ and $J \in \mathbb{R}^{\nu\times\nu}$ such that $\|J - \bar{J}\| < \varepsilon$, and for any $r \in V$, the GE

$$R(u) + \Phi(\bar{u}) + \bar{J}(u - \bar{u}) + N(u) \ni r \tag{2.5}$$

has in U the unique solution $u(r)$, and the mapping $u(\cdot)$ is Lipschitz-continuous on V with Lipschitz constant $\bar{\ell}$. Substituting (2.4) into (2.5), observe that the latter coincides with (2.3).

Considering for each $\bar{J} \in \Delta$ the open ball in $\mathbb{R}^{\nu\times\nu}$ centered at $\bar{J}$ and of radius ε defined above, we obtain the open cover of the compact set Δ which has a finite subcover. We now re-define $\varepsilon > 0$ in such a way that any $J \in \mathbb{R}^{\nu\times\nu}$ satisfying (2.2) belongs to the specified finite subcover. Furthermore, we take the maximum value $\bar{\ell} > 0$ of the corresponding constants and the intersections $\tilde{U}$, U and V of the corresponding neighborhoods defined above over the centers $\bar{J}$ of the balls constituting this subcover. By further shrinking V (if necessary) in order to ensure that for any $\tilde{u} \in \tilde{U}$, any $J \in \mathbb{R}^{\nu\times\nu}$ satisfying (2.2), and any $r \in V$, the solution $u(r)$ of (2.3) corresponding to an appropriate element of the subcover belongs to U, we get all the ingredients for the stated assertion. ∎

Consider now the optimization problem (1.3), where the objective function $f : \mathbb{R}^n \to \mathbb{R}$ and the constraints mappings $h : \mathbb{R}^n \to \mathbb{R}^l$ and $g : \mathbb{R}^n \to \mathbb{R}^m$ are differentiable, with their derivatives being locally Lipschitz-continuous. As already noted, stationary points of problem (1.3) and the associated Lagrange multipliers are characterized by the KKT optimality system (1.4), which corresponds to the GE (1.1) with the base mapping Φ given by (1.5) and the field multifunction N given by (1.6).

For a feasible point $\bar{x}$ of problem (1.3), let

$$A(\bar{x}) = \{i = 1, \ldots, m \mid g_i(\bar{x}) = 0\}$$

stand for the set of indices of inequality constraints active at $\bar{x}$. Furthermore, for a Lagrange multiplier $\bar{\mu}$ associated with $\bar{x}$, set

$$A_+(\bar{x}, \bar{\mu}) = \{i \in A(\bar{x}) \mid \bar{\mu}_i > 0\}, \quad A_0(\bar{x}, \bar{\mu}) = \{i \in A(\bar{x}) \mid \bar{\mu}_i = 0\}.$$

Recall that the linear independence constraint qualification (LICQ) at $\bar{x}$ consists of saying that the gradients $h'_j(\bar{x})$, $j = 1, \ldots, l$, $g'_i(\bar{x})$, $i \in A(\bar{x})$, are linearly independent.

For optimization problems with twice differentiable data, characterization of strong regularity was derived in [30] (sufficiency) and in [4] (necessity). These facts imply the following result which, in turn, gives the characterization of strong regularity in the case of once differentiable data, in the sense of Definition 2.1.

Proposition 2.3 *Let* $f : \mathbb{R}^n \to \mathbb{R}$, $h : \mathbb{R}^n \to \mathbb{R}^l$ *and* $g : \mathbb{R}^n \to \mathbb{R}^m$ *be differentiable at* $\bar{x} \in \mathbb{R}^n$. *Let* $\bar{x}$ *be a stationary point of problem* (1.3), *and let* $(\bar{\lambda}, \bar{\mu}) \in \mathbb{R}^l \times \mathbb{R}^m$ *be an associated Lagrange multiplier. Let* $H \in \mathbb{R}^{n\times n}$ *be an arbitrary symmetric matrix and let*

$$J = \begin{pmatrix} H & (h'(\bar{x}))^{\mathrm{T}} & (g'(\bar{x}))^{\mathrm{T}} \\ -h'(\bar{x}) & 0 & 0 \\ -g'(\bar{x}) & 0 & 0 \end{pmatrix}. \tag{2.6}$$

If $\bar{x}$ *and* $(\bar{\lambda}, \bar{\mu})$ *satisfy LICQ and the condition*

$$\langle H\xi, \xi\rangle > 0 \quad \forall\, \xi \in C_+(\bar{x}, \bar{\mu}) \setminus \{0\}, \tag{2.7}$$

where

$$C_+(\bar{x}, \bar{\mu}) = \{\xi \in \mathbb{R}^n \mid h'(\bar{x})\xi = 0,\ g'_{A_+(\bar{x}, \bar{\mu})}(\bar{x})\xi = 0\},$$

then $\bar{u} = (\bar{x}, \bar{\lambda}, \bar{\mu})$ *is a strongly regular solution of GE* (2.1) *with* $\Phi(\cdot)$ *and* $N(\cdot)$ *defined according to* (1.5) *and* (1.6), *respectively.*

Moreover, LICQ is necessary for strong regularity of $\bar{u}$, *while the condition* (2.7) *is necessary for strong regularity of* $\bar{u}$ *if* $\bar{x}$ *is a local solution of the quadratic programming problem*

$$\begin{array}{ll} \text{minimize} & \langle f'(\bar{x}),\, x-\bar{x}\rangle + \frac{1}{2}\langle H(x-\bar{x}),\, x-\bar{x}\rangle \\ \text{subject to} & h'(\bar{x})(x-\bar{x}) = 0,\ g'_{A(\bar{x})}(\bar{x})(x-\bar{x}) \le 0. \end{array} \tag{2.8}$$

Proof. Problem (2.8) is locally (near $\bar{x}$) equivalent to the problem

$$\begin{array}{ll} \text{minimize} & \langle f'(\bar{x}),\, x-\bar{x}\rangle + \frac{1}{2}\langle H(x-\bar{x}),\, x-\bar{x}\rangle \\ \text{subject to} & h(\bar{x}) + h'(\bar{x})(x-\bar{x}) = 0,\ g(\bar{x}) + g'(\bar{x})(x-\bar{x}) \le 0. \end{array} \tag{2.9}$$

It can be easily seen that the KKT system for problem (2.9) can be stated as the GE (2.1) with $\Phi(\cdot)$ and $N(\cdot)$ defined according to (1.5) and (1.6), respectively, and with J defined in (2.6). Moreover, stationarity of $\bar{x}$ in problem (1.3) with an associated Lagrange multiplier $(\bar{\lambda}, \bar{\mu})$ is equivalent to stationarity of $\bar{x}$ in problem (2.9) with the same Lagrange multiplier $(\bar{\lambda}, \bar{\mu})$; the sets of active at $\bar{x}$ inequality constraints of the two problems are the same; LICQ for the two problems at $\bar{x}$ means the same; and finally, condition (2.7) coincides with the so-called strong second-order optimality condition for problem (2.9). The needed assertions now follow applying the results of [30] and [4] to problem (2.9) (this can be done, since (2.9) is a quadratic program and thus satisfies the smoothness assumptions in [30, 4]). ■

Remark 2.1 It can be seen that for any $u = (x, \lambda, \mu) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$ it holds that

$$\partial\Phi(u) = \left\{ \begin{pmatrix} H & (h'(x))^{\mathrm{T}} & (g'(x))^{\mathrm{T}} \\ -h'(x) & 0 & 0 \\ -g'(x) & 0 & 0 \end{pmatrix} \,\middle|\, H \in \partial_x \frac{\partial L}{\partial x}(x, \lambda, \mu) \right\}. \tag{2.10}$$

Indeed, the inclusion of the left-hand side into the right-hand side follows from Lemma A.2 in the Appendix, while the converse inclusion follows by the fact that a mapping of two variables, which is differentiable with respect to one variable and affine with respect to the other, is necessarily differentiable with respect to the aggregated variable.

By equality (2.10), Proposition 2.3 immediately implies the following: If $\bar{x}$ and $(\bar{\lambda}, \bar{\mu})$ satisfy LICQ and the strong second-order sufficient optimality condition (SSOSC)

$$\forall\, H \in \partial_x \frac{\partial L}{\partial x}(\bar{x}, \bar{\lambda}, \bar{\mu}) \quad \langle H\xi, \xi\rangle > 0 \quad \forall\, \xi \in C_+(\bar{x}, \bar{\mu}) \setminus \{0\}, \tag{2.11}$$

then $\bar{u} = (\bar{x}, \bar{\lambda}, \bar{\mu})$ is a CD-regular solution of GE (2.1) with $\Phi(\cdot)$ and $N(\cdot)$ defined according to (1.5) and (1.6), respectively. In particular, in the case of twice differentiable data, Proposition 2.3 applied with $H = \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda}, \bar{\mu})$ recovers the characterization of strong regularity obtained in [30] and [4].

In the sequel, along with SSOSC (2.11) we shall employ the weaker second-order sufficient optimality condition (SOSC)

$$\forall\, H \in \partial_x \frac{\partial L}{\partial x}(\bar{x}, \bar{\lambda}, \bar{\mu}) \quad \langle H\xi, \xi\rangle > 0 \quad \forall\, \xi \in C(\bar{x}) \setminus \{0\}, \tag{2.12}$$

where

$$C(\bar{x}) = \{\xi \in \mathbb{R}^n \mid h'(\bar{x})\xi = 0,\ g'_{A(\bar{x})}(\bar{x})\xi \leq 0,\ \langle f'(\bar{x}), \xi\rangle \leq 0\}$$

is the critical cone of problem (1.3) at $\bar{x}$. Recall that the critical cone has the equivalent representation

$$C(\bar{x}) = \{\xi \in \mathbb{R}^n \mid h'(\bar{x})\xi = 0,\ g'_{A_+(\bar{x}, \bar{\mu})}(\bar{x})\xi = 0,\ g'_{A_0(\bar{x}, \bar{\mu})}(\bar{x})\xi \leq 0\} \tag{2.13}$$

for any Lagrange multiplier $(\bar{\lambda}, \bar{\mu})$ associated with a stationary point $\bar{x}$. Condition (2.12) is indeed sufficient for local optimality of a stationary point $\bar{x}$, as established in [18].

3 Semismooth Josephy–Newton method

In this section, along with the *semismooth Josephy–Newton method* given by (1.7) with some $J_k \in \partial\Phi(u^k)$, we shall also consider its perturbed generalization. Specifically, given the current iterate $u^k \in \mathbb{R}^\nu$, the next iterate u^{k+1} satisfies the GE

$$\omega^k + \Phi(u^k) + J_k(u - u^k) + N(u) \ni 0 \tag{3.1}$$

with some $J_k \in \partial\Phi(u^k)$, where $\omega^k \in \mathbb{R}^\nu$ is a perturbation term. The perturbation may be induced, for example, by inexact solution of the subproblem $\Phi(u^k) + J_k(u - u^k) + N(u) \ni 0$. Another possibility is the quasi-Newton variant that solves $\Phi(u^k) + J(u - u^k) + N(u) \ni 0$ with some $J \notin \partial\Phi(u^k)$. This corresponds to the perturbation term $\omega^k = (J - J_k)(u^{k+1} - u^k)$.

We start with the following *a posteriori* result concerned with superlinear rate of convergence, assuming convergence itself. Among other things, this line of analysis would turn convenient later, as it gives all the necessary convergence rate estimates once convergence is established.

Proposition 3.1 *Let $\Phi : \mathbb{R}^\nu \to \mathbb{R}^\nu$ be semismooth at $\bar{u} \in \mathbb{R}^\nu$. Let $\bar{u}$ be a solution of GE (1.1), strongly regular with respect to some closed set $\bar{\Delta} \subset \partial\Phi(\bar{u})$. Let a sequence $\{u^k\} \subset \mathbb{R}^\nu$ be convergent to $\bar{u}$, and assume that u^{k+1} satisfies (3.1) for each $k = 0, 1, \ldots,$ with some $J_k \in \partial\Phi(u^k)$ and $\omega^k \in \mathbb{R}^\nu$ such that*

$$\mathrm{dist}(J_k, \bar{\Delta}) \to 0 \text{ as } k \to \infty \tag{3.2}$$

and

$$\omega^k = o(\|u^{k+1} - u^k\| + \|u^k - \bar{u}\|). \tag{3.3}$$

Then the rate of convergence of $\{u^k\}$ is superlinear. Moreover, the rate of convergence is quadratic provided Φ is strongly semismooth at $\bar{u}$ and

$$\omega^k = O(\|u^{k+1} - u^k\|^2 + \|u^k - \bar{u}\|^2). \tag{3.4}$$

Proof. Define $\varepsilon > 0$, $\bar{\ell} > 0$, $\tilde{U}$, U and V according to Proposition 2.2 with $\Delta = \bar{\Delta}$. Then for any $u^k \in \tilde{U}$, any $J_k \in \partial\Phi(u^k)$ satisfying $\mathrm{dist}(J_k, \bar{\Delta}) < \varepsilon$, and any $r \in V$, the GE

$$\Phi(u^k) + J_k(u - u^k) + N(u) \ni r \tag{3.5}$$

has in U the unique solution $u(r)$ which is Lipschitz-continuous on V with Lipschitz constant $\bar{\ell}$. For each k, set

$$r^k = \Phi(u^k) - \Phi(\bar{u}) - J_k(u^k - \bar{u}). \tag{3.6}$$

Note that by the semismoothness of Φ at $\bar{u}$, it holds that

$$r^k = o(\|u^k - \bar{u}\|). \tag{3.7}$$

Note also that by (3.6),

$$0 \in \Phi(\bar{u}) + N(\bar{u}) = \Phi(u^k) + J_k(\bar{u} - u^k) + N(\bar{u}) - r^k. \tag{3.8}$$

By convergence of $\{u^k\}$ to $\bar{u}$, and by (3.2), (3.3) and (3.7), we conclude that for all k large enough it holds that $u^k,\, u^{k+1} \in \tilde{U} \cap U$, $\mathrm{dist}(J_k, \bar{\Delta}) < \varepsilon$, $-\omega^k \in V$ and $r^k \in V$. Hence, according to Proposition 2.2, u^{k+1} is the unique solution in U of GE (3.5) with $r = -\omega^k$, i.e., $u^{k+1} = u(-\omega^k)$, while by (3.8), $\bar{u}$ is the unique solution in U of GE (3.5) with $r = r^k$, i.e., $\bar{u} = u(r^k)$. Therefore,

$$\|u^{k+1} - \bar{u}\| = \|u(-\omega^k) - u(r^k)\| \leq \bar{\ell}\|\omega^k + r^k\| = o(\|u^{k+1} - u^k\| + \|u^k - \bar{u}\|), \tag{3.9}$$

where the last estimate is by (3.3) and (3.7).

The proof that (3.9) implies the superlinear rate is standard; see, e.g., [15, Proposition 2.1]. In addition, if Φ is strongly semismooth at $\bar{u}$ then for r^k defined in (3.6) it holds that

$$r^k = O(\|u^k - \bar{u}\|^2).$$

Combining this with (3.4), and with the inequality in (3.9), the quadratic rate of convergence follows. ∎

An immediate application of Proposition 3.1 is a Dennis–Moré-type result for the *semismooth quasi-Josephy–Newton method.* Let $\{J_k\} \subset \mathbb{R}^{\nu\times\nu}$ be a sequence of matrices. For the current iterate $u^k \in \mathbb{R}^\nu$, let the next iterate u^{k+1} be computed as a solution of GE (1.7), and assume that $\{J_k\}$ satisfies the Dennis–Moré-type condition:

$$\min_{J\in\partial\Phi(u^k)} \|(J_k - J)(u^{k+1} - u^k)\| = o(\|u^{k+1} - u^k\|). \tag{3.10}$$

In the following *a posteriori* result, we allow for a possibility of somewhat more special choices of matrices J_k.

Theorem 3.1 *Let $\Phi : \mathbb{R}^\nu \to \mathbb{R}^\nu$ be semismooth at $\bar{u} \in \mathbb{R}^\nu$. Let $\bar{u}$ be a solution of GE* (1.1)*, strongly regular with respect to some closed set $\bar{\Delta} \subset \partial\Phi(\bar{u})$. Let $\{J_k\} \subset \mathbb{R}^{\nu\times\nu}$ be a sequence of matrices, and let a sequence $\{u^k\} \subset \mathbb{R}^\nu$ be convergent to $\bar{u}$ and such that for all k large enough u^{k+1} satisfies* (1.7) *and there exist $\tilde{J}_k \in \partial\Phi(u^k)$ satisfying*

$$\operatorname{dist}(\tilde{J}_k, \bar{\Delta}) \to 0 \text{ as } k \to \infty \tag{3.11}$$

and

$$(J_k - \tilde{J}_k)(u^{k+1} - u^k) = o(\|u^{k+1} - u^k\|). \tag{3.12}$$

Then the rate of convergence of $\{u^k\}$ is superlinear.

Proof. For each k set

$$\omega^k = (J_k - \tilde{J}_k)(u^{k+1} - u^k).$$

Then (3.12) implies (3.3). Employing (3.11), the result now follows immediately from Proposition 3.1. ■

The assumption that for each k large enough there exist $\tilde{J}_k \in \partial\Phi(u^k)$ satisfying (3.12) is equivalent to (3.10). If $\bar{u}$ is a CD-regular solution of GE (1.1) then Theorem 3.1 is applicable with $\bar{\Delta} = \partial\Phi(\bar{u})$, and in this case (3.11) is automatic for any choice of $\tilde{J}_k \in \partial\Phi(u^k)$ according to upper semicontinuity of Clarke's generalized Jacobian. Another appealing possibility is to apply Theorem 3.1 with $\bar{\Delta} = \partial_B\Phi(\bar{u})$, assuming BD-regularity of the solution $\bar{u}$.

We proceed with *a priori* local analysis, i.e., sufficient conditions for convergence.

Theorem 3.2 *Let $\Phi : \mathbb{R}^\nu \to \mathbb{R}^\nu$ be semismooth at $\bar{u} \in \mathbb{R}^\nu$. Let $\bar{u}$ be a solution of GE* (1.1)*, strongly regular with respect to some closed set $\bar{\Delta} \subset \partial\Phi(\bar{u})$. Let Δ be a multifunction from $\mathbb{R}^\nu$ to the subsets of $\mathbb{R}^{\nu\times\nu}$, such that*

$$\Delta(u) \subset \partial\Phi(u) \quad \forall\, u \in \mathbb{R}^\nu \tag{3.13}$$

and for any $\varepsilon > 0$ there exists a neighborhood O of $\bar{u}$ such that

$$\operatorname{dist}(J, \bar{\Delta}) < \varepsilon \quad \forall\, J \in \Delta(u),\ \forall\, u \in O. \tag{3.14}$$

Then there exists $\delta > 0$ such that for any starting point $u^0 \in \mathbb{R}^\nu$ close enough to $\bar{u}$, for each $k = 0, 1, \ldots$ and any choice of $J_k \in \Delta(u^k)$, there exists the unique solution u^{k+1} of GE (1.7) satisfying

$$\|u^{k+1} - u^k\| \leq \delta; \tag{3.15}$$

the sequence $\{u^k\}$ generated this way converges to $\bar{u}$, and the rate of convergence is superlinear. Moreover, the rate of convergence is quadratic provided Φ is strongly semismooth at $\bar{u}$.

Proof. Define $\varepsilon > 0$, $\bar{\ell} > 0$, $\tilde{U}$, U and V according to Proposition 2.2 with $\Delta = \bar{\Delta}$. Moreover, let $\tilde{U}$ be such that (3.14) holds with $O = \tilde{U}$ and with the specified ε.

Then, according to Proposition 2.2, for any $u^k \in \tilde{U}$, any $J_k \in \Delta(u^k)$ and any $r \in V$, the GE

$$\Phi(u^k) + J_k(u - u^k) + N(u) \ni r \tag{3.16}$$

has in U the unique solution $u(r)$ which is Lipschitz-continuous on V with Lipschitz constant $\bar{\ell}$. In particular, GE (1.7) has in U the unique solution $u^{k+1} = u(0)$.

Defining r^k according to (3.6) and employing (3.13), by the semismoothness of Φ at $\bar{u}$ we conclude that (3.7) holds, and

$$0 \in \Phi(\bar{u}) + N(\bar{u}) = \Phi(u^k) + J_k(\bar{u} - u^k) + N(\bar{u}) - r^k.$$

Shrinking $\tilde{U}$ if necessary, by (3.7) we conclude that $r^k \in V$ provided $u^k \in \tilde{U}$, and hence, $\bar{u}$ is the unique solution of GE (3.16) with $r = r^k$, i.e., $\bar{u} = u(r^k)$. Therefore,

$$\|u^{k+1} - \bar{u}\| \leq \|u(r^k) - u(0)\| \leq \bar{\ell}\|r^k\| = o(\|u^k - \bar{u}\|), \tag{3.17}$$

where the last estimate is by (3.7).

From (3.17) we derive the following: for any $q \in (0, 1)$, there exists $\delta > 0$ such that $B(\bar{u}, \delta/2) \subset \tilde{U}$, $B(\bar{u}, 3\delta/2) \subset U$, and for any $u^k \in B(\bar{u}, \delta/2)$ it holds that

$$\|u^{k+1} - \bar{u}\| \leq q\|u^k - \bar{u}\|, \tag{3.18}$$

implying that $u^{k+1} \in B(\bar{u}, \delta/2)$. Then

$$\|u^{k+1} - u^k\| \leq \|u^{k+1} - \bar{u}\| + \|u^k - \bar{u}\| < \frac{\delta}{2} + \frac{\delta}{2} = \delta,$$

and hence, u^{k+1} is a solution of GE (1.7) satisfying (3.15). Moreover, for any point $u^{k+1} \in \mathbb{R}^n$ satisfying (3.15), it holds that

$$\|u^{k+1} - \bar{u}\| \leq \|u^{k+1} - u^k\| + \|u^k - \bar{u}\| \leq \delta + \frac{\delta}{2} = \frac{3\delta}{2},$$

and hence, $u^{k+1} \in U$, implying that u^{k+1} is a solution of GE (1.7) if and only if it coincides with $u^{k+1} = u(0)$. Thus, the latter is the unique solution of GE (1.7) satisfying (3.15).

Therefore, the inclusion $u^0 \in B(\bar{u}, \delta/2)$ implies that the entire sequence $\{u^k\}$ is uniquely defined (for any fixed rule of selecting $J_k \in \Delta(u^k)$ for all k) and is contained in $B(\bar{u}, \delta/2)$. Then (3.18) shows convergence of this sequence to $\bar{u}$. The convergence rate estimates now follow from Proposition 3.1. ∎

Remark 3.1 Theorem 3.2 generalizes local convergence results of the semismooth Newton method for usual equations [29, 25]. Indeed, the two basic options for $\Delta(\cdot)$ are $\partial_B\Phi(\cdot)$ and $\partial\Phi(\cdot)$. For the first option Theorem 3.2 is applicable with $\bar{\Delta} = \partial_B\Phi(\bar{u})$ assuming BD-regularity of $\bar{u}$, while for the second option it is applicable with $\bar{\Delta} = \partial\Phi(\bar{u})$ assuming CD-regularity of $\bar{u}$. In addition, other choices of $\Delta(\cdot)$ (e.g., related to the specific problem structure) are possible.

Remark 3.2 For generalized equations with smooth bases a subtler local convergence result was established in [3], where the assumptions are the semistability and hemistability of the solution. The combination of these two properties is generally weaker than strong regularity (in the smooth case). Note, however, that unlike in Theorem 3.2, local uniqueness of the subproblems' solutions does not hold under these assumptions.

In the case of VI (1.2), the iteration (1.7) of the semismooth Josephy–Newton method takes the form of the (linearized) VI

$$u \in Q, \quad \langle \Phi(u^k) + J_k(u - u^k),\, v - u\rangle \geq 0 \quad \forall\, v \in Q \tag{3.19}$$

with some $J_k \in \partial\Phi(u^k)$. In particular, for a nonlinear complementarity problem (NCP)

$$u \geq 0, \quad \Phi(u) \geq 0, \quad \langle u,\, \Phi(u)\rangle = 0, \tag{3.20}$$

corresponding to VI (1.2) with $Q = \mathbb{R}^n_+$, the iteration (3.19) of the semismooth Josephy–Newton method takes the form of the linear complementarity problem

$$u \geq 0, \quad \Phi(u^k) + J_k(u - u^k) \geq 0, \quad \langle u,\, \Phi(u^k) + J_k(u - u^k)\rangle = 0.$$

According to Definition 2.1, strong regularity of a solution $\bar{u}$ of NCP (3.20) with respect to a set $\bar{\Delta} \subset \mathbb{R}^{\nu\times\nu}$ means that for any $J \in \bar{\Delta}$ the point $\bar{u}$ is a strongly regular solution of the linear complementarity problem

$$u \geq 0, \quad \Phi(\bar{u}) + J(u - \bar{u}) \geq 0, \quad \langle u,\, \Phi(\bar{u}) + J(u - \bar{u})\rangle = 0. \tag{3.21}$$

The algebraic characterization of strong regularity for semismooth NCP (3.20) readily follows applying the results of [30] to (3.21).

4 Semismooth SQP

In this section, we consider the *sequential quadratic programming* (SQP) algorithm [2] for problem (1.3), which is a special case of the Josephy–Newton method (1.7).

Algorithm 4.1 Choose $(x^0,\, \lambda^0,\, \mu^0) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m_+$ and set $k = 0$.

1. If $(x^k,\, \lambda^k,\, \mu^k)$ satisfies the KKT system (1.4), stop.

2. Choose a symmetric matrix $H_k \in \mathbb{R}^{n\times n}$ and compute $x^{k+1} \in \mathbb{R}^n$ as a stationary point of problem

$$\begin{array}{ll} \text{minimize} & f(x^k) + \langle f'(x^k),\, x - x^k\rangle + \frac{1}{2}\langle H_k(x - x^k),\, x - x^k\rangle \\ \text{subject to} & h(x^k) + h'(x^k)(x - x^k) = 0,\ g(x^k) + g'(x^k)(x - x^k) \leq 0, \end{array} \tag{4.1}$$

and $(\lambda^{k+1},\, \mu^{k+1}) \in \mathbb{R}^l \times \mathbb{R}^m_+$ as an associated Lagrange multiplier.

3. Adjust k by 1 and go to step 1.

By the basic *semismooth sequential quadratic programming* (semismooth SQP) method we mean Algorithm 4.1 with

$$H_k \in \partial_x \frac{\partial L}{\partial x}(x^k,\, \lambda^k,\, \mu^k). \tag{4.2}$$

Methods of this kind were considered, e.g., in [27, 11].

The KKT system of problem (4.1) has the form

$$\begin{array}{c} f'(x^k) + H_k(x - x^k) + (h'(x^k))^{\mathrm{T}}\lambda + (g'(x^k))^{\mathrm{T}}\mu = 0, \\ h(x^k) + h'(x^k)(x - x^k) = 0, \\ \mu \geq 0, \quad g(x^k) + g'(x^k)(x - x^k) \leq 0, \quad \langle \mu,\, g(x^k) + g'(x^k)(x - x^k)\rangle = 0. \end{array} \tag{4.3}$$

It can be seen that the latter system is equivalent to the iteration GE (1.7) of the Josephy–Newton method with $u^k = (x^k,\, \lambda^k,\, \mu^k)$, $\Phi(\cdot)$ and $N(\cdot)$ defined according to (1.5) and (1.6), respectively, and with

$$J_k = \left(\begin{array}{ccc} H_k & (h'(x^k))^{\mathrm{T}} & (g'(x^k))^{\mathrm{T}} \\ -h'(x^k) & 0 & 0 \\ -g'(x^k) & 0 & 0 \end{array}\right). \tag{4.4}$$

Semismoothness of the derivatives of f, h and g at $\bar{x}$ implies semismoothness of Φ at $\bar{u} = (\bar{x},\, \bar{\lambda},\, \bar{\mu})$. Moreover, taking into account Proposition 2.3 and Remark 2.1, Theorem 3.2 is applicable with $\bar{\Delta} = \partial\Phi(\bar{x})$ and $\Delta(\cdot) = \partial\Phi(\cdot)$ provided $\bar{x}$ and $(\bar{\lambda},\, \bar{\mu})$ satisfy LICQ and SSOSC (2.11). Therefore, we immediately obtain the local convergence and rate of convergence result for the basic semismooth SQP algorithm.

Theorem 4.1 *Let $f : \mathbb{R}^n \to \mathbb{R}$, $h : \mathbb{R}^n \to \mathbb{R}^l$ and $g : \mathbb{R}^n \to \mathbb{R}^l$ be differentiable in a neighborhood of $\bar{x} \in \mathbb{R}^n$, with their derivatives being semismooth at $\bar{x}$. Let $\bar{x}$ be a local solution of problem* (1.3) *satisfying LICQ, and let SSOSC* (2.11) *hold for the associated Lagrange multiplier $(\bar{\lambda},\, \bar{\mu}) \in \mathbb{R}^l \times \mathbb{R}^m$.*

Then there exists $\delta > 0$ such that for any starting point $(x^0,\, \lambda^0,\, \mu^0) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$ close enough to $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$, for each $k = 0,\, 1,\, \ldots$ and any choice of H_k satisfying (4.2)*, there exists the unique stationary point x^{k+1} of problem* (4.1) *and the unique associated Lagrange multiplier $(\lambda^{k+1},\, \mu^{k+1})$ satisfying*

$$\|(x^{k+1} - x^k,\, \lambda^{k+1} - \lambda^k,\, \mu^{k+1} - \mu^k)\| \leq \delta; \tag{4.5}$$

the sequence $\{(x^k,\, \lambda^k,\, \mu^k)\}$ converges to $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$, and the rate of convergence is superlinear. Moreover, the rate of convergence is quadratic provided the derivatives of f, h and g are strongly semismooth at $\bar{x}$.

Theorem 4.1 essentially recovers the local superlinear convergence result in [11], which was obtained by direct (and rather involved) analysis. Here, this property is an immediate consequence of the general local convergence theory for the semismooth Josephy–Newton method, given by Theorem 3.2. A similar result was derived in [27], but under stronger assumptions including the strict complementarity condition.

For optimization problems with twice differentiable data, Theorem 4.1 can be sharpened. Specifically, it was demonstrated in [3] that LICQ can be replaced by the generally weaker strict Mangasarian–Fromovitz constraint qualification, while SSOSC can be replaced by the usual second-order sufficient optimality condition. However, unlike in Theorem 4.1, these assumptions cannot guarantee uniqueness of the iteration sequence satisfying the localization condition (4.5).

As is well known (e.g., [5, Exercise 14.8]), superlinear or quadratic Q-rate of convergence of the primal-dual sequence does not necessarily imply superlinear (or even linear) Q-rate for the primal part. At the same time, primal convergence is often of particular importance. To that end, we proceed with primal superlinear convergence analysis for Algorithm 4.1.

Having in mind some potentially useful choices of H_k different from the basic (4.2), as well as truncation of subproblems solution (e.g., [14]), we consider the following perturbed version of semismooth SQP. For a given primal-dual iterate $(x^k,\, \lambda^k,\, \mu^k) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$, the next iterate $(x^{k+1},\, \lambda^{k+1},\, \mu^{k+1})$ satisfies the relations

$$\begin{gathered}\frac{\partial L}{\partial x}(x^k, \lambda^k, \mu^k) + W_k(x^{k+1} - x^k) + (h'(x^k))^{\mathrm{T}}(\lambda^{k+1} - \lambda^k) + (g'(x^k))^{\mathrm{T}}(\mu^{k+1} - \mu^k) + \omega_1^k = 0,\\ h(x^k) + h'(x^k)(x^{k+1} - x^k) + \omega_2^k = 0,\\ \mu^{k+1} \geq 0,\ g(x^k) + g'(x^k)(x^{k+1} - x^k) + \omega_3^k \leq 0,\ \langle \mu^{k+1}, g(x^k) + g'(x^k)(x^{k+1} - x^k) + \omega_3^k \rangle = 0,\end{gathered} \tag{4.6}$$

with some $W_k \in \partial_x \frac{\partial L}{\partial x}(x^k,\, \lambda^k,\, \mu^k)$, where $\omega_1^k \in \mathbb{R}^n$, $\omega_2^k \in \mathbb{R}^l$, and $\omega_3^k \in \mathbb{R}^m$ are perturbation terms.

We first establish necessary conditions for primal superlinear convergence of the iterates given by (4.6). Proposition 4.1 also suggests the proper form of the Dennis–Moré-type condition for the semismooth case.

Proposition 4.1 *Let $f : \mathbb{R}^n \to \mathbb{R}$, $h : \mathbb{R}^n \to \mathbb{R}^l$ and $g : \mathbb{R}^n \to \mathbb{R}^l$ be differentiable in a neighborhood of $\bar{x} \in \mathbb{R}^n$, with their derivatives being semismooth at $\bar{x}$. Let $\bar{x}$ be a stationary point of problem* (1.3), *and let $(\bar{\lambda},\, \bar{\mu}) \in \mathbb{R}^l \times \mathbb{R}^m$ be an associated Lagrange multiplier. Let $\{(x^k,\, \lambda^k,\, \mu^k)\} \subset \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$ be convergent to $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$, and assume that for each k large enough the triple $(x^{k+1},\, \lambda^{k+1},\, \mu^{k+1})$ satisfies the system* (4.6) *with some $W_k \in \partial_x \frac{\partial L}{\partial x}(x^k, \lambda^k, \mu^k)$ and some $\omega_1^k \in \mathbb{R}^n$, $\omega_2^k \in \mathbb{R}^l$ and $\omega_3^k \in \mathbb{R}^m$.*

If the rate of convergence of $\{x^k\}$ is superlinear then

$$\frac{\partial L}{\partial x}(x^{k+1}, \lambda^k, \mu^k) - \frac{\partial L}{\partial x}(x^k, \lambda^k, \mu^k) - W_k(x^{k+1} - x^k) = o(\|x^k - \bar{x}\|), \tag{4.7}$$

$$\omega_2^k = o(\|x^k - \bar{x}\|), \tag{4.8}$$

$$(\omega_3^k)_{A_+(\bar{x},\bar{\mu})} = o(\|x^k - \bar{x}\|). \tag{4.9}$$

If in addition

$$\{(\omega_3^k)_{\{1,\ldots,m\}\setminus A(\bar{x})}\} \to 0 \text{ as } k \to \infty, \tag{4.10}$$

then

$$\pi_{C(\bar{x})}(-\omega_1^k) = o(\|x^k - \bar{x}\|). \tag{4.11}$$

Proof. Let $\{\tilde{W}_k\}$ be an arbitrary sequence of matrices such that $\tilde{W}_k \in \partial_x \frac{\partial L}{\partial x}(x^{k+1}, \lambda^k, \mu^k)$ for each k. Since $\{(x^k,\, \lambda^k,\, \mu^k)\}$ is convergent to $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$ (and hence, $\{(\lambda^k,\, \mu^k)\}$ is bounded), and the derivatives of f, h and g are locally Lipschitz-continuous at $\bar{x}$, one can easily see that there exist a neighborhood U of $\bar{x}$ and $\ell > 0$ such that for all k the mapping $\frac{\partial L}{\partial x}(\cdot,\, \lambda^k,\, \mu^k)$ is Lipschitz-continuous on U with constant ℓ, and both x^k and x^{k+1} belong to U for all k large enough. This implies that $\|W_k\| \leq \ell$ and $\|\tilde{W}_k\| \leq \ell$ for all such k (since the matrices in the generalized Jacobian are bounded by the Lipschitz constant of the mapping in question). In particular, $\{W_k\}$ and $\{\tilde{W}_k\}$ are bounded sequences. Then, employing Lemma A.3 in the Appendix (with $p = r = n$, $q = l + m$, $K(x) = ((h'(x))^{\mathrm{T}},\, (g'(x))^{\mathrm{T}})$, $b(x) = f'(x)$, $y^k = (\lambda^k,\, \mu^k)$, $\bar{y} = (\bar{\lambda},\, \bar{\mu})$) and taking into account the superlinear convergence of $\{x^k\}$ to $\bar{x}$, we obtain

$$\begin{aligned}
& \frac{\partial L}{\partial x}(x^{k+1}, \lambda^k, \mu^k) - \frac{\partial L}{\partial x}(x^k, \lambda^k, \mu^k) - W_k(x^{k+1} - x^k) \\
= \; & \left(\frac{\partial L}{\partial x}(x^{k+1}, \lambda^k, \mu^k) - \frac{\partial L}{\partial x}(\bar{x}, \lambda^k, \mu^k) - \tilde{W}_k(x^{k+1} - \bar{x}) \right) \\
& - \left(\frac{\partial L}{\partial x}(x^k, \lambda^k, \mu^k) - \frac{\partial L}{\partial x}(\bar{x}, \lambda^k, \mu^k) - W_k(x^k - \bar{x}) \right) - (W_k - \tilde{W}_k)(x^{k+1} - \bar{x}) \\
= \; & o(\|x^{k+1} - \bar{x}\|) + o(\|x^k - \bar{x}\|) + O(\|x^{k+1} - \bar{x}\|) \\
= \; & o(\|x^k - \bar{x}\|),
\end{aligned} \tag{4.12}$$

which gives (4.7).

Furthermore, from (4.6), employing superlinear convergence of $\{x^k\}$ to $\bar{x}$, boundedness of $\{W_k\}$, Lemma A.3 and local Lipschitz-continuity of the derivatives of h and g at $\bar{x}$, we obtain that

$$\begin{aligned}
-\omega_1^k \; = \; & \frac{\partial L}{\partial x}(x^k, \lambda^k, \mu^k) + W_k(x^{k+1} - x^k) + (h'(x^k))^{\mathrm{T}}(\lambda^{k+1} - \lambda^k) + (g'(x^k))^{\mathrm{T}}(\mu^{k+1} - \mu^k) \\
= \; & (h'(\bar{x}))^{\mathrm{T}}(\lambda^k - \bar{\lambda}) + (g'(\bar{x}))^{\mathrm{T}}(\mu^k - \bar{\mu}) + (h'(x^k))^{\mathrm{T}}(\lambda^{k+1} - \lambda^k) + (g'(x^k))^{\mathrm{T}}(\mu^{k+1} - \mu^k) \\
& + \left(\frac{\partial L}{\partial x}(x^k, \lambda^k, \mu^k) - \frac{\partial L}{\partial x}(\bar{x}, \lambda^k, \mu^k) - W_k(x^k - \bar{x}) \right) + W_k(x^{k+1} - \bar{x}) \\
= \; & ((h'(x^k))^{\mathrm{T}} - (h'(\bar{x}))^{\mathrm{T}})(\lambda^{k+1} - \lambda^k) + (h'(\bar{x}))^{\mathrm{T}}(\lambda^{k+1} - \bar{\lambda}) \\
& + ((g'(x^k))^{\mathrm{T}} - (g'(\bar{x}))^{\mathrm{T}})(\mu^{k+1} - \mu^k) + (g'(\bar{x}))^{\mathrm{T}}(\mu^{k+1} - \bar{\mu}) + o(\|x^k - \bar{x}\|) \\
= \; & (h'(\bar{x}))^{\mathrm{T}}(\lambda^{k+1} - \bar{\lambda}) + (g'(\bar{x}))^{\mathrm{T}}(\mu^{k+1} - \bar{\mu}) + o(\|x^k - \bar{x}\|),
\end{aligned} \tag{4.13}$$

and

$$\begin{aligned}
\omega_2^k \; = \; & -h(x^k) - h'(x^k)(x^{k+1} - x^k) \\
= \; & -h(x^k) + h(\bar{x}) + h'(\bar{x})(x^k - \bar{x}) + (h'(x^k) - h'(\bar{x}))(x^k - \bar{x}) - h'(x^k)(x^{k+1} - \bar{x}) \\
= \; & o(\|x^k - \bar{x}\|).
\end{aligned}$$

The latter relation givies (4.8).

Moreover, since $\bar{\mu}_{A_+(\bar{x},\bar{\mu})} > 0$, we have that $\mu^k_{A_+(\bar{x},\bar{\mu})} > 0$ for all k large enough, and it then follows from the last line in (4.6) that

$$\begin{aligned}(\omega_3^k)_{A_+(\bar{x},\bar{\mu})} &= -g_{A_+(\bar{x},\bar{\mu})}(x^k) - g'_{A_+(\bar{x},\bar{\mu})}(x^k)(x^{k+1} - x^k)\\ &= -g_{A_+(\bar{x},\bar{\mu})}(x^k) + g_{A_+}(\bar{x}) + g'_{A_+(\bar{x},\bar{\mu})}(\bar{x})(x^k - \bar{x})\\ &\quad +(g'_{A_+(\bar{x},\bar{\mu})}(x^k) - g'_{A_+(\bar{x},\bar{\mu})}(\bar{x}))(x^k - \bar{x}) - g'_{A_+(\bar{x},\bar{\mu})}(x^k)(x^{k+1} - \bar{x})\\ &= o(\|x^k - \bar{x}\|),\end{aligned}$$

which gives (4.9).

For each k set

$$\tilde{\omega}_1^k = (h'(\bar{x}))^{\mathrm{T}}(\lambda^{k+1} - \bar{\lambda}) + (g'(\bar{x}))^{\mathrm{T}}(\mu^{k+1} - \bar{\mu}). \tag{4.14}$$

Then from (4.13) it follows that

$$\omega_1^k + \tilde{\omega}_1^k = o(\|x^k - \bar{x}\|). \tag{4.15}$$

If (4.10) holds then, since $\{g_{\{1,\dots,m\}\setminus A(\bar{x})}(x^k)\} \to g_{\{1,\dots,m\}\setminus A(\bar{x})}(\bar{x}) < 0$, the last line in (4.6) implies that $\mu^k_{\{1,\dots,m\}\setminus A(\bar{x})} = 0$ for all k large enough. Taking this into account, we obtain from (2.13) and (4.14) that for all such k, for any $\xi \in C(\bar{x})$ it holds that

$$\langle \tilde{\omega}_1^k,\, \xi\rangle = \langle \lambda^{k+1} - \bar{\lambda},\, h'(\bar{x})\xi\rangle + \langle \mu^{k+1} - \bar{\mu},\, g'(\bar{x})\xi\rangle = \langle \mu^k_{A_0(\bar{x},\bar{\mu})},\, g'_{A_0(\bar{x},\bar{\mu})}(\bar{x})\xi\rangle \le 0,$$

where the inequality $\mu^{k+1} \ge 0$ was also employed. Therefore, $\tilde{\omega}_1^k \in (C(\bar{x}))^\circ$. Hence, according to (1.9), $\pi_{C(\bar{x})}(\tilde{\omega}_1^k) = 0$. Combining the latter with (4.15), and taking into account the fact that $\pi_{C(\bar{x})}(\cdot)$ is nonexpansive, we obtain the last needed estimate (4.11). ■

We now proceed with sufficient conditions for primal superlinear convergence. Following [10], in this analysis we only assume that the limiting stationary point $\bar{x}$ and the associated limiting multiplier $(\bar{\lambda},\, \bar{\mu})$ satisfy SOSC (2.12). Note that even in the twice differentiable case other results in the literature (e.g., [5, Theorem 15.7]; see also [3], [23, Theorem 18.5] for related statements) require LICQ in addition to SOSC. The analysis relies on the following primal error bound result that generalizes [10] with respect to its smoothness assumptions.

Proposition 4.2 *Let $f : \mathbb{R}^n \to \mathbb{R}$, $h : \mathbb{R}^n \to \mathbb{R}^l$ and $g : \mathbb{R}^n \to \mathbb{R}^l$ be differentiable in a neighborhood of $\bar{x} \in \mathbb{R}^n$, with their derivatives being semismooth at $\bar{x}$. Let $\bar{x}$ be a stationary point of problem* (1.3)*, let $(\bar{\lambda},\, \bar{\mu}) \in \mathbb{R}^l \times \mathbb{R}^m$ be an associated Lagrange multiplier, and assume that SOSC* (2.12) *holds.*

Then the estimate

$$\|x - \bar{x}\| = O\left(\left\|\begin{pmatrix}\pi_{C(\bar{x})}\left(\frac{\partial L}{\partial x}(x, \lambda, \mu)\right)\\ h(x)\\ \min\{\mu,\, -g(x)\}\end{pmatrix}\right\|\right) \tag{4.16}$$

holds for all $(x,\, \lambda,\, \mu) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$ close enough to $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$.

Proof. We argue by contradiction. Suppose that (4.16) does not hold. Then there exist a sequence $\{(x^k, \lambda^k, \mu^k)\} \subset \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$ tending to $(\bar{x}, \bar{\lambda}, \bar{\mu})$ and a sequence $t_k \to +\infty$ such that for all k

$$\|x^k - \bar{x}\| > t_k \left\| \begin{pmatrix} \pi_{C(\bar{x})}\left(\frac{\partial L}{\partial x}(x^k, \lambda^k, \mu^k)\right) \\ h(x^k) \\ \min\{\mu^k, -g(x^k)\} \end{pmatrix} \right\|.$$

This is further equivalent to

$$\pi_{C(\bar{x})}\left(\frac{\partial L}{\partial x}(x^k, \lambda^k, \mu^k)\right) = o(\|x^k - \bar{x}\|), \tag{4.17}$$

$$h(x^k) = o(\|x^k - \bar{x}\|), \tag{4.18}$$

$$\min\{\mu^k, -g(x^k)\} = o(\|x^k - \bar{x}\|). \tag{4.19}$$

From (4.18) it follows that

$$0 = h(\bar{x}) + h'(\bar{x})(x^k - \bar{x}) + o(\|x^k - \bar{x}\|) = h'(\bar{x})(x^k - \bar{x}) + o(\|x^k - \bar{x}\|). \tag{4.20}$$

Moreover, since $g_{A_+(\bar{x},\bar{\mu})}(\bar{x}) = 0 < \bar{\mu}_{A_+(\bar{x},\bar{\mu})}$, from (4.19) we obtain that for all k large enough

$$\begin{aligned} 0 &= \min\{\mu^k_{A_+(\bar{x},\bar{\mu})}, -g_{A_+(\bar{x},\bar{\mu})}(x^k)\} + o(\|x^k - \bar{x}\|) \\ &= -g_{A_+(\bar{x},\bar{\mu})}(x^k) + o(\|x^k - \bar{x}\|) \\ &= -g_{A_+(\bar{x},\bar{\mu})}(\bar{x}) - g'_{A_+(\bar{x},\bar{\mu})}(\bar{x})(x^k - \bar{x}) + o(\|x^k - \bar{x}\|) \\ &= -g'_{A_+(\bar{x},\bar{\mu})}(\bar{x})(x^k - \bar{x}) + o(\|x^k - \bar{x}\|), \end{aligned} \tag{4.21}$$

and similarly, since $g_{\{1,\ldots,m\}\setminus A(\bar{x})}(\bar{x}) < 0 = \bar{\mu}_{\{1,\ldots,m\}\setminus A(\bar{x})}$,

$$\begin{aligned} 0 &= \min\{\mu^k_{\{1,\ldots,m\}\setminus A(\bar{x})}, -g_{\{1,\ldots,m\}\setminus A(\bar{x})}(x^k)\} + o(\|x^k - \bar{x}\|) \\ &= \mu^k_{\{1,\ldots,m\}\setminus A(\bar{x})} + o(\|x^k - \bar{x}\|). \end{aligned} \tag{4.22}$$

Since the number of different partitions of the set $A_0(\bar{x}, \bar{\mu})$ is finite, passing onto a subsequence if necessary, we can assume that there exist index sets I_1 and I_2 such that $I_1 \cup I_2 = A_0(\bar{x}, \bar{\mu})$, $I_1 \cap I_2 = \emptyset$, and for each k it holds that

$$\mu^k_{I_1} \geq -g_{I_1}(x^k), \quad \mu^k_{I_2} < -g_{I_2}(x^k). \tag{4.23}$$

Then by (4.19) we have

$$\begin{aligned} 0 &= \min\{\mu^k_{I_1}, -g_{I_1}(x^k)\} + o(\|x^k - \bar{x}\|) \\ &= -g_{I_1}(x^k) + o(\|x^k - \bar{x}\|) \\ &= -g_{I_1}(\bar{x}) - g'_{I_1}(\bar{x})(x^k - \bar{x}) + o(\|x^k - \bar{x}\|) \\ &= -g'_{I_1}(\bar{x})(x^k - \bar{x}) + o(\|x^k - \bar{x}\|), \end{aligned} \tag{4.24}$$

$$0 = \min\{\mu^k_{I_2}, -g_{I_2}(x^k)\} + o(\|x^k - \bar{x}\|) = \mu^k_{I_2} + o(\|x^k - \bar{x}\|). \tag{4.25}$$

Finally, from (4.23) it also follows that

$$\begin{aligned} -\mu^k_{I_2} &> g_{I_2}(x^k) \\ &= g_{I_2}(\bar{x}) + g'_{I_2}(\bar{x})(x^k - \bar{x}) + o(\|x^k - \bar{x}\|) \\ &= g'_{I_2}(\bar{x})(x^k - \bar{x}) + o(\|x^k - \bar{x}\|), \end{aligned}$$

and hence, by (4.25),

$$g'_{I_2}(\bar{x})(x^k - \bar{x}) \leq o(\|x^k - \bar{x}\|). \tag{4.26}$$

Without loss of generality we can assume that $x^k \neq \bar{x}$ for all k, and $(x^k - \bar{x})/\|x^k - \bar{x}\|$ converges to some $\xi \in \mathbb{R}^n \setminus \{0\}$ ($\|\xi\| = 1$). Then by (2.13), (4.20), (4.21), (4.24) and (4.26) we conclude that $\xi \in C(\bar{x}) \setminus \{0\}$.

Furthermore, employing (1.8) and (4.17) we obtain

$$\begin{aligned} 0 &= \pi_{C(\bar{x})}\left(\frac{\partial L}{\partial x}(x^k,\, \lambda^k,\, \mu^k) - \pi_{C(\bar{x})}\left(\frac{\partial L}{\partial x}(x^k,\, \lambda^k,\, \mu^k)\right)\right) \\ &= \pi_{C(\bar{x})}\left(\frac{\partial L}{\partial x}(x^k,\, \lambda^k,\, \mu^k) + o(\|x^k - \bar{x}\|)\right), \end{aligned}$$

and hence, by (1.9) and by Lemma A.3,

$$\begin{aligned} (C(\bar{x}))^\circ &\ni \frac{\partial L}{\partial x}(x^k,\, \lambda^k,\, \mu^k) + o(\|x^k - \bar{x}\|) \\ &= \left(\frac{\partial L}{\partial x}(x^k,\, \lambda^k,\, \mu^k) - \frac{\partial L}{\partial x}(\bar{x},\, \lambda^k,\, \mu^k)\right) + \left(\frac{\partial L}{\partial x}(\bar{x},\, \lambda^k,\, \mu^k) - \frac{\partial L}{\partial x}(\bar{x},\, \bar{\lambda},\, \bar{\mu})\right) \\ &\quad + o(\|x^k - \bar{x}\|) \\ &= W_k(x^k - \bar{x}) + (h'(\bar{x}))^{\mathrm{T}}(\lambda^k - \bar{\lambda}) + (g'(\bar{x}))^{\mathrm{T}}(\mu^k - \bar{\mu}) + o(\|x^k - \bar{x}\|), \end{aligned} \tag{4.27}$$

where $\{W_k\}$ is any sequence of matrices such that $W_k \in \partial_x \frac{\partial L}{\partial x}(x^k,\, \lambda^k,\, \mu^k)$ for each k.

By Lemma A.2 there exists a sequence $\{\bar{W}_k\}$ such that $\bar{W}_k \in \partial_x \frac{\partial L}{\partial x}(x^k,\, \bar{\lambda},\, \bar{\mu})$ for k large enough and $W_k - \bar{W}_k = O(\|(\lambda^k - \bar{\lambda},\, \mu^k - \bar{\mu})\|)$. Then, since the sequence $\{\bar{W}_k\}$ is bounded (by Lipschitz continuity of $\frac{\partial L}{\partial x}(\cdot,\, \bar{\lambda},\, \bar{\mu})$ on some neighborhood of $\bar{x}$), and since $\{(\lambda^k,\, \mu^k)\}$ converges to $(\bar{\lambda},\, \bar{\mu})$, passing to a subsequence if necessary, we may assume that both $\{\bar{W}_k\}$ and $\{W_k\}$ converge to some $\bar{W} \in \mathbb{R}^{n \times n}$. By upper semicontinuity of the generalized Jacobian it follows that $\bar{W} \in \partial_x \frac{\partial L}{\partial x}(\bar{x},\, \bar{\lambda},\, \bar{\mu})$. Taking into account (2.13), (4.22) and (4.25), the inclusion $\xi \in C(\bar{x})$ and the equalities $g'_{I_1}(\bar{x})\xi = 0$ (see (4.24)) and $\bar{\mu}_{\{1,\, \ldots,\, m\} \setminus A_+(\bar{x},\, \bar{\mu})} = 0$, from (4.27) we obtain

$$\begin{aligned} 0 &\geq \langle W_k(x^k - \bar{x}),\, \xi\rangle + \langle \lambda^k - \bar{\lambda},\, h'(\bar{x})\xi\rangle + \langle \mu^k - \bar{\mu},\, g'(\bar{x})\xi\rangle + o(\|x^k - \bar{x}\|) \\ &= \langle W_k(x^k - \bar{x}),\, \xi\rangle + \langle \mu^k_{I_2 \cup (\{1,\, \ldots,\, m\} \setminus A(\bar{x}))},\, g'_{I_2 \cup (\{1,\, \ldots,\, m\} \setminus A(\bar{x}))}(\bar{x})\xi\rangle + o(\|x^k - \bar{x}\|) \\ &= \langle W_k(x^k - \bar{x}),\, \xi\rangle + o(\|x^k - \bar{x}\|), \end{aligned}$$

Dividing the obtained relation by $\|x^k - \bar{x}\|$ and passing onto the limit, we conclude that

$$\langle \bar{W}\xi, \xi\rangle \leq 0,$$

which contradicts SOSC (2.12) because $\xi \in C(\bar{x}) \setminus \{0\}$. ■

We are now in position to give conditions that are sufficient for primal superlinear convergence of perturbed semismooth SQP.

Theorem 4.2 *Under the assumptions of Proposition 4.1, let SOSC* (2.12) *hold.*
If (4.10) *holds and*

$$\pi_{C(\bar{x})}\left(\frac{\partial L}{\partial x}(x^{k+1}, \lambda^k, \mu^k) - \frac{\partial L}{\partial x}(x^k, \lambda^k, \mu^k) - W_k(x^{k+1} - x^k) - \omega_1^k\right) = o(\|x^{k+1} - x^k\| + \|x^k - \bar{x}\|), \tag{4.28}$$

$$\omega_2^k = o(\|x^{k+1} - x^k\| + \|x^k - \bar{x}\|), \tag{4.29}$$

$$(\omega_3^k)_{A(\bar{x})} = o(\|x^{k+1} - x^k\| + \|x^k - \bar{x}\|), \tag{4.30}$$

then the rate of convergence of $\{x^k\}$ *is superlinear.*

Proof. Employing convergence of $\{(x^k,\, \lambda^k,\, \mu^k)\}$ to $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$ and local Lipschitz-continuity of the derivatives of h and g at $\bar{x}$, from the first and the second equalities in (4.6) we derive

$$\begin{aligned}
\frac{\partial L}{\partial x}(x^{k+1},\, \lambda^{k+1},\, \mu^{k+1}) &= \frac{\partial L}{\partial x}(x^{k+1},\, \lambda^k,\, \mu^k) \\
&\quad + (h'(x^{k+1}))^{\mathrm{T}}(\lambda^{k+1} - \lambda^k) + (g'(x^{k+1}))^{\mathrm{T}}(\mu^{k+1} - \mu^k) \\
&= \frac{\partial L}{\partial x}(x^{k+1},\, \lambda^k,\, \mu^k) - \frac{\partial L}{\partial x}(x^k,\, \lambda^k,\, \mu^k) \\
&\quad + \frac{\partial L}{\partial x}(x^k,\, \lambda^k,\, \mu^k) + (h'(x^k))^{\mathrm{T}}(\lambda^{k+1} - \lambda^k) + (g'(x^k))^{\mathrm{T}}(\mu^{k+1} - \mu^k) \\
&\quad + o(\|x^{k+1} - x^k\|) \\
&= \frac{\partial L}{\partial x}(x^{k+1},\, \lambda^k,\, \mu^k) - \frac{\partial L}{\partial x}(x^k,\, \lambda^k,\, \mu^k) - W_k(x^{k+1} - x^k) - \omega_1^k \\
&\quad + o(\|x^{k+1} - x^k\|)
\end{aligned} \tag{4.31}$$

and

$$h(x^{k+1}) = h(x^k) + h'(x^k)(x^{k+1} - x^k) + o(\|x^{k+1} - x^k\|) = -\omega_2^k + o(\|x^{k+1} - x^k\|). \tag{4.32}$$

Since $\{(\omega_3^k)_{\{1,\,\ldots,\,m\}\setminus A(\bar{x})}\} \to 0$ (by (4.10)) and $\{g_{\{1,\,\ldots,\,m\}\setminus A(\bar{x})}(x^k)\} \to g_{\{1,\,\ldots,\,m\}\setminus A(\bar{x})}(\bar{x}) < 0$, we then conclude that for all k large enough $\mu^{k+1}_{\{1,\,\ldots,\,m\}\setminus A(\bar{x})} = 0$. Hence,

$$\min\{\mu^{k+1}_{\{1,\,\ldots,\,m\}\setminus A(\bar{x})},\, -g_{\{1,\,\ldots,\,m\}\setminus A(\bar{x})}(x^{k+1})\} = 0. \tag{4.33}$$

Observe that the last line in (4.6) can be written in the form

$$\min\{\mu^{k+1},\, -g(x^k) - g'(x^k)(x^{k+1} - x^k) - \omega_3^k\} = 0.$$

For each $i \in A(\bar{x})$, employing this equality and the property

$$|\min\{a,\, b\} - \min\{a,\, c\}| \leq |b - c| \quad \forall\, a,\, b,\, c \in \mathbb{R},$$

we obtain the estimate

$$\begin{aligned}\|\min\{\mu^{k+1}_{A(\bar{x})},\, -g_{A(\bar{x})}(x^{k+1})\}\| &= \|\min\{\mu^{k+1}_{A(\bar{x})},\, -g_{A(\bar{x})}(x^k) - g'_{A(\bar{x})}(x^k)(x^{k+1} - x^k)\\ &\quad +o(\|x^{k+1} - x^k\|)\}\\ &\quad - \min\{\mu^{k+1}_{A(\bar{x})},\, -g_{A(\bar{x})}(x^k) - g'_{A(\bar{x})}(x^k)(x^{k+1} - x^k)\\ &\quad -(\omega_3^k)_{A(\bar{x})}\}\|\\ &\leq \|(\omega_3^k)_{A(\bar{x})}\| + o(\|x^{k+1} - x^k\|). \end{aligned} \tag{4.34}$$

Combining Proposition 4.2 and relations (4.28)–(4.32) and (4.33)–(4.34), we conclude that

$$\|x^{k+1} - \bar{x}\| = o(\|x^{k+1} - x^k\| + \|x^k - \bar{x}\|) = o(\|x^{k+1} - \bar{x}\| + \|x^k - \bar{x}\|),$$

i.e., there exists a sequence $\{t_k\}$ of nonnegative reals such that $t_k \to 0$ and

$$\|x^{k+1} - \bar{x}\| \leq t_k(\|x^{k+1} - \bar{x}\| + \|x^k - \bar{x}\|).$$

for all k large enough. This implies that

$$(1 - t_k)\|x^{k+1} - \bar{x}\| \leq t_k\|x^k - \bar{x}\|,$$

and hence, for all k large enough

$$\|x^{k+1} - \bar{x}\| \leq \frac{t_k}{1 - t_k}\|x^k - \bar{x}\|,$$

i.e.,

$$\|x^{k+1} - \bar{x}\| = o(\|x^k - \bar{x}\|),$$

which completes the proof. ∎

Remark 4.1 Condition (4.28) follows from (4.7) and (4.11). Therefore, according to Proposition 4.1, it is in fact also necessary for the primal superlinear convergence rate (assuming (4.10)).

Remark 4.2 In Theorem 4.2 SOSC (2.12) can be replaced by the following sequential second-order condition:

$$\liminf_{k\to\infty} \max_{W \in \partial_x \frac{\partial L}{\partial x}(x^k,\, \lambda^k,\, \mu^k)} \langle W\xi,\, \xi\rangle > 0 \quad \forall\, \xi \in C(\bar{x}) \setminus \{0\} \tag{4.35}$$

(employing Lemma A.2, one can easily see that (4.35) is implied by (2.12)). This would require the development of a sequential counterpart of the primal error bound established in Proposition 4.2. We ommit the details.

The analysis of primal superlinear convergence developed above for the general perturbed semismooth SQP framework (4.6) can be applied to some more specific algorithms. In particular, Algorithm 4.1 can be viewed as a special case of this framework with

$$\omega_1^k = (H_k - W_k)(x^{k+1} - x^k), \quad \omega_2^k = 0, \quad \omega_3^k = 0,$$

where $\{W_k\}$ is an arbitrary sequence of matrices such that $W_k \in \partial_x \frac{\partial L}{\partial x}(x^k,\, \lambda^k,\, \mu^k)$ for each k. From Proposition 4.1, Theorem 4.2, and Remarks 4.1 and 4.2, it follows that under (4.35) primal superlinear convergence of *quasi-Newton semismooth SQP* is characterized by the condition

$$\pi_{C(\bar{x})}\left(\frac{\partial L}{\partial x}(x^{k+1}, \lambda^k, \mu^k) - \frac{\partial L}{\partial x}(x^k, \lambda^k, \mu^k) - H_k(x^{k+1} - x^k)\right) = o(\|x^{k+1} - x^k\|). \tag{4.36}$$

This can be regarded as a natural generalization of the Dennis–Moré-type condition for smooth quasi-Newton SQP methods [23, Theorem 18.5] to the case of semismooth first derivatives.

Recalling the usual Dennis–Moré condition for the smooth case, one might think of replacing (4.36) by something like

$$\max_{W \in \partial_x \frac{\partial L}{\partial x}(x^k,\, \lambda^k,\, \mu^k)} \|\pi_{C(\bar{x})}((W - H_k)(x^{k+1} - x^k))\| = o(\|x^{k+1} - x^k\|). \tag{4.37}$$

This condition corresponds to the one used for similar purposes in [12], where it was shown that (4.37) is necessary and, under (4.35), sufficient for primal superlinear convergence of Algorithm 4.1 in the case when there are no inequality constraints. If f, h and g are twice continuously differentiable near $\bar{x}$, then by the Mean-Value Theorem one can easily see that conditions (4.36) and (4.37) are equivalent. In the semismooth case the relationship between these two conditions is not so clear. From Proposition 4.1 it easily follows that (4.37) is necessary for primal superlinear convergence. Therefore, it is implied by (4.36) provided that (4.35) holds. The converse implication might not be true, but according to the discussion below, it appears difficult to give an example of the lack of this implication. Namely, we shall show that under a certain reasonable additional assumption, (4.37) is sufficient for primal superlinear convergence and thus implies (4.36).

Specifically, assume that the set

$$A_k = \{i = 1,\, \ldots,\, m \mid g_i(x^k) + g_i'(x^k)(x^{k+1} - x^k) = 0\} \tag{4.38}$$

of indices of active inequality constraints of the semismooth SQP subproblems (4.1) stabilizes, i.e., it holds that $A_k = A$ for some fixed $A \subset \{1,\, \ldots,\, m\}$ and all k large enough. According to the last line in (4.3), by continuity, the inclusions

$$A_+(\bar{x},\, \bar{\mu}) \subset A_k \subset A(\bar{x}) \tag{4.39}$$

always hold for all k large enough. Therefore, the stabilization property is automatic with $A = A(\bar{x})$ when $\{(\lambda^k,\, \mu^k)\}$ converges to a multiplier $(\bar{\lambda},\, \bar{\mu})$ satisfying the strict complementarity condition, i.e., such that $\bar{\mu}_{A(\bar{x})} > 0$ (and hence, $A_+(\bar{x},\, \bar{\mu}) = A(\bar{x})$). In other cases, the stabilization property may not hold, but this still seems to be reasonable numerical behavior, which should be quite typical. Note also that if this stabilization property does not hold, one should hardly expect convergence of the dual sequence, in general.

The following result extends the sufficiency part of [12, Theorem 2.2] to the case when inequality constraints can be present.

Theorem 4.3 *Let $f : \mathbb{R}^n \to \mathbb{R}$, $h : \mathbb{R}^n \to \mathbb{R}^l$ and $g : \mathbb{R}^n \to \mathbb{R}^l$ be differentiable in a neighborhood of $\bar{x} \in \mathbb{R}^n$, with their derivatives being semismooth at $\bar{x}$. Let $\bar{x}$ be a stationary point of problem* (1.3)*, and let $(\bar{\lambda}, \bar{\mu}) \in \mathbb{R}^l \times \mathbb{R}^m$ be an associated Lagrange multiplier. Let a sequence $\{(x^k, \lambda^k, \mu^k)\} \subset \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$ generated by Algorithm 4.1 be convergent to $(\bar{x}, \bar{\lambda}, \bar{\mu})$. Assume that* (4.35) *and* (4.37) *hold and that there exists an index set $A \subset \{1, \ldots, m\}$ such that $A_k = A$ for all k large enough, where the index sets A_k are defined according to* (4.38).

Then the rate of convergence of $\{x^k\}$ is superlinear.

Proof. Define the set

$$\tilde{C}(\bar{x}) = \{\xi \in \mathbb{R}^n \mid h'(\bar{x})\xi = 0,\ g'_A(\bar{x})\xi = 0,\ g'_{A(\bar{x})\setminus A}(\bar{x})\xi \leq 0\}.$$

By Hoffman's error bound for linear systems (e.g., [9, Lemma 3.2.3]) we have that

$$\begin{aligned}\operatorname{dist}(x^{k+1} - \bar{x}, \tilde{C}(\bar{x})) &= O(\|h'(\bar{x})(x^{k+1} - \bar{x})\| + \|g'_A(\bar{x})(x^{k+1} - \bar{x})\| \\ &\quad + \|\max\{0,\ g'_{A(\bar{x})\setminus A}(\bar{x})(x^{k+1} - \bar{x})\}\|).\end{aligned} \tag{4.40}$$

From the second line in (4.3), and from local Lipschitz-continuity of the derivative of h at $\bar{x}$, we obtain

$$\begin{aligned}h'(\bar{x})(x^{k+1} - \bar{x}) &= h'(\bar{x})(x^{k+1} - x^k) + h'(\bar{x})(x^k - \bar{x}) - h(x^k) - h'(x^k)(x^{k+1} - x^k) \\ &= -(h'(x^k) - h'(\bar{x}))(x^{k+1} - x^k) - (h(x^k) - h(\bar{x}) - h'(\bar{x})(x^k - \bar{x})) \\ &= o(\|x^k - \bar{x}\|).\end{aligned} \tag{4.41}$$

For any sufficiently large k it holds that $g_A(x^k) + g'_A(x^k)(x^{k+1} - x^k) = 0$, and similarly to (4.41) it follows that

$$g'_A(\bar{x})(x^{k+1} - \bar{x}) = o(\|x^k - \bar{x}\|). \tag{4.42}$$

Finally, if $i \in A(\bar{x}) \setminus A$ and $\langle g'_i(\bar{x}),\ x^{k+1} - \bar{x}\rangle > 0$, taking into account the last line of (4.3) and local Lipschitz-continuity of the derivative of g at $\bar{x}$, we obtain

$$\begin{aligned}\max\{0,\ \langle g'_i(\bar{x}),\ x^{k+1} - \bar{x}\rangle\} &= \langle g'_i(\bar{x}),\ x^{k+1} - \bar{x}\rangle \\ &= \langle g'_i(\bar{x}),\ x^{k+1} - x^k\rangle + \langle g'_i(\bar{x}),\ x^k - \bar{x}\rangle \\ &\leq \langle g'_i(\bar{x}),\ x^{k+1} - x^k\rangle + \langle g'_i(\bar{x}),\ x^k - \bar{x}\rangle \\ &\quad - g_i(x^k) - \langle g'_i(x^k),\ x^{k+1} - x^k\rangle \\ &= -\langle g'_i(x^k) - g'_i(\bar{x}),\ x^{k+1} - x^k\rangle \\ &\quad - (g_i(x^k) - g_i(\bar{x}) - \langle g'_i(\bar{x}),\ x^k - \bar{x}\rangle) \\ &= o(\|x^k - \bar{x}\|).\end{aligned} \tag{4.43}$$

Relations (4.40)–(4.43) imply that

$$\operatorname{dist}(x^{k+1}-\bar{x},\, \tilde{C}(\bar{x})) = o(\|x^k-\bar{x}\|).$$

The latter means that for all k there exists $\xi^k \in \tilde{C}(\bar{x})$ such that

$$x^{k+1}-\bar{x} = \xi^k + o(\|x^k-\bar{x}\|). \tag{4.44}$$

From the first line of (4.3) and from semismoothness of the derivatives of f, h and g at $\bar{x}$, employing Lemma A.3 and convergence of $\{(\lambda^k,\, \mu^k)\}$ to $(\bar{\lambda},\, \bar{\mu})$ we derive that for any choice of matrices $W_k \in \partial_x \frac{\partial L}{\partial x}(x^k,\, \lambda^k,\, \mu^k)$ it holds that

$$\begin{aligned} -H_k(x^{k+1}-x^k) &= \frac{\partial L}{\partial x}(x^k,\, \lambda^k,\, \mu^k) + (h'(x^k))^T(\lambda^{k+1}-\lambda^k) + (g'(x^k))^T(\mu^{k+1}-\mu^k) \\ &= \frac{\partial L}{\partial x}(x^k,\, \lambda^k,\, \mu^k) - \frac{\partial L}{\partial x}(\bar{x},\, \lambda^k,\, \mu^k) - W_k(x^k-\bar{x}) \\ &\quad + \frac{\partial L}{\partial x}(\bar{x},\, \lambda^k,\, \mu^k) - \frac{\partial L}{\partial x}(\bar{x}, \bar{\lambda}, \bar{\mu}) + (h'(x^k))^T(\lambda^{k+1}-\lambda^k) \\ &\quad + (g'(x^k))^T(\mu^{k+1}-\mu^k) + W_k(x^k-\bar{x}) \\ &= W_k(x^k-\bar{x}) + (h'(\bar{x}))^T(\lambda^k-\bar{\lambda}) + (g'(\bar{x}))^T(\mu^k-\bar{\mu}) \\ &\quad + (h'(\bar{x}))^T(\lambda^{k+1}-\lambda^k) + (g'(\bar{x}))^T(\mu^{k+1}-\mu^k) + o(\|x^k-\bar{x}\|) \\ &= W_k(x^k-\bar{x}) + (h'(\bar{x}))^T(\lambda^{k+1}-\bar{\lambda}) + (g'(\bar{x}))^T(\mu^{k+1}-\bar{\mu}) \\ &\quad + o(\|x^k-\bar{x}\|). \end{aligned}$$

Therefore,

$$\begin{aligned} W_k(x^{k+1}-\bar{x}) &= (W_k-H_k)(x^{k+1}-x^k) - (h'(\bar{x}))^T(\lambda^{k+1}-\bar{\lambda}) \\ &\quad - (g'(\bar{x}))^T(\mu^{k+1}-\bar{\mu}) + o(\|x^k-\bar{x}\|). \end{aligned} \tag{4.45}$$

From the definition of set A it follows that $g_{\{1,\, \ldots,\, m\}\setminus A}(x^k) + g'_{\{1,\, \ldots,\, m\}\setminus A}(x^k)(x^{k+1}-x^k) < 0$ for all k large enough. Then, by the last line in (4.3),

$$\mu^{k+1}_{\{1,\, \ldots,\, m\}\setminus A} = 0 \tag{4.46}$$

for all such k. Moreover, according to (4.39) it holds that $\tilde{C}(\bar{x}) \subset C(\bar{x})$, and therefore, (4.37) remains true with $C(\bar{x})$ substituted for $\tilde{C}(\bar{x})$. Then, employing (4.45), (4.46), and the fact that $\langle x,\, \xi\rangle \leq \langle \pi_{\tilde{C}(\bar{x})}(x),\, \xi\rangle$ for all $x \in \mathbb{R}^n$ and all $\xi \in \tilde{C}(\bar{x})$ (see (1.8) and (1.9)), we further obtain

$$\begin{aligned} \langle W_k\xi^k,\, \xi^k\rangle &= \langle W_k(x^{k+1}-\bar{x}),\, \xi^k\rangle + o(\|x^k-\bar{x}\|\|\xi^k\|) \\ &= \langle (W_k-H_k)(x^{k+1}-x^k),\, \xi^k\rangle \\ &\quad - \langle (h'(\bar{x}))^T(\lambda^{k+1}-\bar{\lambda}) + (g'(\bar{x}))^T(\mu^{k+1}-\bar{\mu}),\, \xi^k\rangle + o(\|x^k-\bar{x}\|\|\xi^k\|) \\ &\leq \langle \pi_{\tilde{C}(\bar{x})}((W_k-H_k)(x^{k+1}-x^k)),\, \xi^k\rangle + o(\|x^k-\bar{x}\|\|\xi^k\|) \\ &= o((\|x^{k+1}-x^k\| + \|x^k-\bar{x}\|)\|\xi^k\|). \end{aligned} \tag{4.47}$$

From (4.35) and the inclusion $\tilde{C}(\bar{x}) \subset C(\bar{x})$ it further follows that there exist $\gamma > 0$ and a sequence $\{W_k\}$ of matrices such that $W_k \in \partial_x \frac{\partial L}{\partial x}(x^k, \lambda^k, \mu^k)$ and for all k large enough

$$\langle W_k \xi^k, \xi^k \rangle \geq \gamma \|\xi^k\|^2.$$

Then (4.47) implies

$$\xi^k = o(\|x^{k+1} - x^k\| + \|x^k - \bar{x}\|),$$

and hence, given (4.44),

$$\|x^{k+1} - \bar{x}\| = o(\|x^{k+1} - x^k\| + \|x^k - \bar{x}\|).$$

Repeating the argument completing the proof of Theorem 4.2, we obtain the superlinear convergence rate of $\{x^k\}$. ∎

5 Concluding Remarks

We have introduced the notion of solution regularity and developed local convergence theory for the Josephy–Newton method for generalized equations with semismooth base mappings. The special case of semismooth SQP for optimization was also considered, easily recovering its primal-dual convergence result and obtaining a new characterization of primal superlinear convergence rate.

6 Appendix

Lemma A.1 *Let $K\colon \mathbb{R}^p \to \mathbb{R}^{r \times q}$ be locally Lipschitz-continuous at $\bar{x} \in \mathbb{R}^p$ with Lipschitz constant $\ell_K > 0$ and $b\colon \mathbb{R}^p \to \mathbb{R}^r$ be an arbitrary map. Define the map $\Psi\colon \mathbb{R}^p \times \mathbb{R}^q \mapsto \mathbb{R}^r$,*

$$\Psi(x, y) = K(x)y + b(x). \tag{A.1}$$

If Ψ is differentiable with respect to x at $(\bar{x}, y^1) \in \mathbb{R}^p \times \mathbb{R}^q$ and $(\bar{x}, y^2) \in \mathbb{R}^p \times \mathbb{R}^q$ with some $y^1, y^2 \in \mathbb{R}^q$ then

$$\left\| \frac{\partial \Psi}{\partial x}(\bar{x}, y^1) - \frac{\partial \Psi}{\partial x}(\bar{x}, y^2) \right\| \leq \ell_K \|y^1 - y^2\|. \tag{A.2}$$

Proof. Differentiability of Ψ with respect to x at $(\bar{x}, y^1)$ and $(\bar{x}, y^2)$ means that for any $\xi \in \mathbb{R}^p$

$$\begin{aligned} & K(\bar{x}+\xi)y^j + b(\bar{x}+\xi) - K(\bar{x})y^j - b(\bar{x}) - \frac{\partial \Psi}{\partial x}(\bar{x}, y^j)\xi \\ = \; & \Psi(\bar{x}+\xi, y^j) - \Psi(\bar{x}+\xi, y^j) - \frac{\partial \Psi}{\partial x}(\bar{x}, y^j)\xi \\ = \; & o(\|\xi\|), \quad j = 1, 2. \end{aligned}$$

This implies the relation

$$(K(\bar{x}+\xi) - K(\bar{x}))(y^1 - y^2) - \left(\frac{\partial \Psi}{\partial x}(\bar{x}, y^1) - \frac{\partial \Psi}{\partial x}(\bar{x}, y^2) \right) \xi = o(\|\xi\|). \tag{A.3}$$

Fix an arbitrary $\xi \in \mathbb{R}^p$, $\|\xi\| = 1$. By (A.3), employing the fact that K is locally Lipschitz-continuous at $\bar{x}$ with Lipschitz constant ℓ_K, we have for all $t > 0$

$$
\begin{aligned}
t\left\|\left(\frac{\partial\Psi}{\partial x}(\bar{x},\, y^1) - \frac{\partial\Psi}{\partial x}(\bar{x},\, y^2)\right)\xi\right\| &\leq \|K(\bar{x}+t\xi) - K(\bar{x})\|\|y^1 - y^2\| + o(t) \\
&\leq \ell_K t\|y^1 - y^2\| + o(t).
\end{aligned}
$$

Dividing both sides by t and passing onto the limit, we obtain

$$
\left\|\left(\frac{\partial\Psi}{\partial x}(\bar{x},\, y^1) - \frac{\partial\Psi}{\partial x}(\bar{x},\, y^2)\right)\xi\right\| \leq \ell_K\|y^1 - y^2\|.
$$

Since ξ is arbitrary, the required estimate (A.2) follows. ■

Lemma A.2 *Let $K\colon \mathbb{R}^p \to \mathbb{R}^{r\times q}$ and $b\colon \mathbb{R}^p \to \mathbb{R}^r$ be locally Lipschitz-continuous at $\bar{x} \in \mathbb{R}^p$, and define the map $\Psi\colon \mathbb{R}^p\times\mathbb{R}^q \mapsto \mathbb{R}^r$ according to* (A.1). *Let the sequences $\{(x^k,\, y_1^k)\} \subset \mathbb{R}^p \times \mathbb{R}^q$ and $\{(x^k,\, y_2^k)\} \subset \mathbb{R}^p \times \mathbb{R}^q$ be both convergent to $(\bar{x},\, \bar{y})$ with some $\bar{y} \in \mathbb{R}^q$.*

Then for any sequence of matrices $\{W_k^1\} \subset \mathbb{R}^{r\times p}$ such that $W_k^1 \in \partial_x\Psi(x^k,\, y_1^k)$ for all k, there exists a sequence of matrices $\{W_k^2\} \subset \mathbb{R}^{r\times p}$ such that $W_k^2 \in \partial_x\Psi(x^k,\, y_2^k)$ for all k large enough, and

$$
\|W_k^1 - W_k^2\| = O(\|y_1^k - y_2^k\|).
$$

Proof. Let U be a neighborhood of $\bar{x}$, such that K and b are Lipschitz-continuous on U. Then for all k the mapping $\Psi(\cdot,\, y_2^k)$ is evidently Lipschitz-continuous on U. Therefore, by Rademacher's theorem, it is differentiable everywhere on $U \setminus \Gamma$, where the Lebesgue measure of the set $\Gamma \subset U$ is zero. Let $\mathcal{D}_k \subset U$ stand for the set of points of differentiability of $\Psi(\cdot,\, y_1^k)$. Since Clarke's generalized Jacobian is "blind" to sets of Lebesgue measure zero [8], for any k large enough (so that $x^k \in U$) and for any matrix $W_k^1 \in \partial_x\Psi(x^k,\, y_1^k)$ there exist a positive integer s_k, matrices $W_{k,\,i}^1 \in \mathbb{R}^{r\times q}$ and reals $\alpha_{k,\,i} \geq 0$, $i = 1,\, \ldots,\, s_k$, such that $\sum_{i=1}^{s_k} \alpha_{k,\,i} = 1$, $W_k^1 = \sum_{i=1}^{s_k} \alpha_{k,\,i} W_{k,\,i}^1$, and for each $i = 1,\, \ldots,\, s_k$, there exists a sequence $\{x_j^{k,\,i}\} \subset \mathcal{D}_k \setminus \Gamma$ convergent to x^k and such that $\{\frac{\partial\Psi}{\partial x}(x_j^{k,\,i},\, y_1^k)\} \to W_{k,\,i}^1$ as $j \to \infty$.

Furthermore, by Lemma A.1, for any k and j large enough it holds that

$$
\left\|\frac{\partial\Psi}{\partial x}(x_j^{k,\,i},\, y_1^k) - \frac{\partial\Psi}{\partial x}(x_j^{k,\,i},\, y_2^k)\right\| \leq \ell_K\|y_1^k - y_2^k\| \quad \forall\, i = 1,\, \ldots,\, s_k, \tag{A.4}
$$

where ℓ_K is the Lipschitz constant for K on U. For all k large enough, since $\Psi(\cdot,\, y_2^k)$ is locally Lipschitz-continuous at x^k, for all $i = 1,\, \ldots,\, s_k$ the sequence $\{\frac{\partial\Psi}{\partial x}(x_j^{k,\,i},\, y_2^k)\}$ is bounded and therefore, passing to a subsequence if necessary, we can assume that each of these sequences converges to some $W_{k,\,i}^2$ as $j \to \infty$. Then by passing onto the limit in (A.4) we derive the estimate

$$
\|W_{k,\,i}^1 - W_{k,\,i}^2\| \leq \ell_K\|y_1^k - y_2^k\| \quad \forall\, i = 1,\, \ldots,\, s_k. \tag{A.5}
$$

Moreover, by the definition of B-differential we obtain that $W^2_{k,\,i} \in (\partial_x)_B \Psi(x^k,\, y^k_2)$. Hence, by the definition of the generalized Jacobian, the convex combination $W^2_k = \sum_{i=1}^{s_k} \alpha^i_k W^2_{k,\,i}$ belongs to $\partial_x \Psi(x^k,\, y^k_2)$, and employing (A.5) we derive the estimate

$$\|W^1_k - W^2_k\| = \left\| \sum_{i=1}^{s_k} \alpha_{k,\,i} W^1_{k,\,i} - \sum_{i=1}^{s_k} \alpha_{k,\,i} W^2_{k,\,i} \right\| \leq \sum_{i=1}^{s_k} \alpha_{k,\,i} \|W^1_{k,\,i} - W^2_{k,\,i}\| \leq \ell_K \|y^k_1 - y^k_2\|.$$

■

Lemma A.3 *Let* $K\colon \mathbb{R}^p \to \mathbb{R}^{r\times q}$ *and* $b\colon \mathbb{R}^p \to \mathbb{R}^r$ *be semismooth at* $\bar{x} \in \mathbb{R}^p$*, and define the map* $\Psi\colon \mathbb{R}^p \times \mathbb{R}^q \mapsto \mathbb{R}^r$ *according to* (A.1). *Let a sequence* $\{(x^k,\, y^k)\} \subset \mathbb{R}^p \times \mathbb{R}^q$ *be convergent to* $(\bar{x},\, \bar{y})$ *with some* $\bar{y} \in \mathbb{R}^q$.

Then for any sequence of matrices $\{W_k\} \subset \mathbb{R}^{r\times p}$ *such that* $W_k \in \partial_x \Psi(x^k,\, y^k)$ *for all* k*, it holds that*

$$\Psi(x^k,\, y^k) - \Psi(\bar{x},\, y^k) - W_k(x^k - \bar{x}) = o(\|x^k - \bar{x}\|).$$

Proof. Applying Lemma A.2 with $y^k_1 = y^k$ and $y^k_2 = \bar{y}$ for all k, we conclude that there exists a sequence of matrices $\{\bar{W}_k\} \subset \mathbb{R}^{r\times p}$ such that $\bar{W}_k \in \partial_x \Psi(x^k,\, \bar{y})$ for all sufficiently large k and $W_k - \bar{W}_k = O(\|y^k - \bar{y}\|)$. Employing (A.1) and semismoothness of K and b at $\bar{x}$, we then derive the estimate

$$\begin{aligned}
\|\Psi(x^k,\, y^k) - \Psi(\bar{x},\, y^k) - W_k(x^k - \bar{x})\| &\leq \|(\Psi(x^k,\, y^k) - \Psi(x^k,\, \bar{y})) - (\Psi(\bar{x},\, y^k) - \Psi(\bar{x}, \bar{y}))\| \\
&\quad + \|(W_k - \bar{W}_k)(x^k - \bar{x})\| \\
&\quad + \|\Psi(x^k,\, \bar{y}) - \Psi(\bar{x},\, \bar{y}) - \bar{W}_k(x^k - \bar{x})\| \\
&= \|(K(x^k) - K(\bar{x}))(y^k - \bar{y})\| \\
&\quad + O(\|x^k - \bar{x}\|\|y^k - \bar{y}\|) + o(\|x^k - \bar{x}\|) \\
&= O(\|x^k - \bar{x}\|\|y^k - \bar{y}\|) + o(\|x^k - \bar{x}\|) \\
&= o(\|x^k - \bar{x}\|).
\end{aligned}$$

■

References

[1] R. Andreani, E.G. Birgin, J.M. Martínez, and M.L. Schuverdt. On Augmented Lagrangian methods with general lower-level constraints. SIAM J. Optim. 18 (2007), 1286–1309.

[2] B.T. Boggs and J.W. Tolle. Sequential quadratic programming. Acta Numerica 4 (1996), 1–51.

[3] J.F. Bonnans. Local analysis of Newton-type methods for variational inequalities and nonlinear programming. Appl. Math. Optim. 29 (1994), 161–186.

[4] J.F. Bonnans and A. Sulem. Pseudopower expansion of solutions of generalized equations and constrained optimization. Math. Program. 70 (1995), 123–148.

[5] J.F. Bonnans, J.Ch. Gilbert, C. Lemaréchal, and C. Sagastizábal. Numerical Optimization: Theoretical and Practical Aspects. Springer–Verlag, Berlin, Germany, 2006. Second Edition.

[6] F.H. Clarke. Optimization and Nonsmooth Analysis. John Wiley & Sons, New York, 1983.

[7] A.L. Dontchev and R.T. Rockafellar. Newton's method for generalized equations: a sequential implicit function theorem. Math. Program. 123 (2010), 139–159.

[8] M. Fabian, D. Preiss. On the Clarke's generalized Jacobian. Proceedings of the 14th Winter School on Abstract Analysis. Circolo Matematico di Palermo, Palermo, 1987. Rendiconti del Circolo Matematico di Palermo, Ser. II, Number 14, pp. 305–307.

[9] F. Facchinei and J.-S. Pang. Finite-Dimensional Variational Inequalities and Complementarity Problems. Springer-Verlag, New York, 2003.

[10] D. Fernández, A.F. Izmailov, and M.V. Solodov. Sharp primal superlinear convergence results for some Newtonian methods for constrained optimization. SIAM J. Optim. 20 (2010), 3312–3334.

[11] J. Han and D. Sun. Superlinear convergence of approximate Newton methods for LC^1 optimization problems without strict complementarity. *Recent Advances in Nonsmooth Optimization*, D.-Z. Du, L. Qi, and R.S. Womersley, eds., V. 58, pp. 353–367. World Scientific Publishing Co, Singapur, 1993.

[12] A.F. Izmailov, A.L. Pogosyan, and M.V. Solodov. Semismooth Newton method for the lifted reformulation of mathematical programs with complementarity constraints. Comput. Optim. Appl. 51 (2012), 199–221.

[13] A.F. Izmailov and M.V. Solodov. The theory of 2-regularity for mappings with Lipschitzian derivatives and its applications to optimality conditions. Math. Oper. Res. 27 (2002), 614–635.

[14] A.F. Izmailov and M.V. Solodov. A truncated SQP method based on inexact interior-point solutions of subproblems. SIAM J. Optim. 20 (2010), P. 2584–2613.

[15] A.F. Izmailov and M.V. Solodov. Inexact Josephy–Newton framework for generalized equations and its applications to local analysis of Newtonian methods for constrained optimization. Comput. Optim. Appl. 46 (2010), P. 347–368.

[16] N.H. Josephy. Newton's method for generalized equations. Technical Summary Report no. 1965. Mathematics Research Center, University of Wisconsin, Madison, 1979.

[17] N.H. Josephy. Quasi-Newton methods for generalized equations. Technical Summary Report no. 1966. Mathematics Research Center, University of Wisconsin, Madison, 1979.

[18] D. Klatte and K. Tammer. On the second order sufficient conditions to perturbed $C^{1,1}$ optimization problems. Optimization 19 (1988), 169–180.

[19] B. Kummer. Newton's method for nondifferentiable functions. In J. Guddat, B. Bank, H. Hollatz, P. Kall, D. Klatte, B. Kummer, K. Lommalzsch, L. Tammer, M. Vlach, and K. Zimmerman, eds., Advances in Mathematical Optimization, V. 45, pp. 114–125. Akademie-Verlag, Berlin, 1988.

[20] B. Kummer. Newton's method based on generalized derivatives for nonsmooth functions. In W. Oettli and D. Pallaschke, eds., Advances in Optimization, pp. 171–194. Springer-Verlag, Berlin, 1992.

[21] B. Mordukhovich. Stability theory for parametric generalized equations and variational inequalities via nonsmooth analysis. Transactions of the American Mathematical Society 343 (1994), 609–657.

[22] B.S. Mordukhovich. Variational Analysis and Generalized Differentiation. Springer, Berlin, 2006.

[23] J. Nocedal and S.J. Wright. Numerical Optimization. Springer, New York, 2006. Second Edition.

[24] J.-S. Pang and L. Qi. Nonsmooth equations: motivation and algorithms. SIAM J. Optim. 3 (1993), 443–465.

[25] L. Qi. Convergence analysis of some algorithms for solving nonsmooth equations. Math. Oper. Res. 18 (1993), 227–244.

[26] L. Qi. LC^1 functions and LC^1 optimization problems. Technical Report AMR 91/21. School of Mathematics, The University of New South Wales, Sydney, 1991.

[27] L. Qi. Superlinearly convergent approximate Newton methods for LC^1 optimization problems. Math. Program. 64 (1994), 277–294.

[28] L. Qi and H. Jiang. Semismooth Karush–Kuhn–Tucker equations and convergence analysis of Newton and quasi-Newton methods for solving these equations. Math. Oper. Res. 22 (1997), 301–325.

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

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

[31] R.T. Rockafellar. Computational schemes for solving large-scale problems in extended linear-quadratic programming. Math. Program. 48 (1990), 447–474.

[32] R.T. Rockafellar and J.B. Wets. Generalized linear-quadratic problems of deterministic and stochastic optimal control in discrete time. SIAM J. Control Optim. 28 (1990), 810–922.

[33] O. Stein. Lifting mathematical programs with complementarity constraints. Math. Program. 2010. DOI 10.1007/s10107-010-0345-y.