A hybrid semismooth quasi-Newton method for structured nonsmooth operator equations in Banach spaces

Florian Mannel[*] Armin Rund[*]

Preprint, April 20, 2021

Abstract

We present an algorithm for the solution of structured nonsmooth operator equations in Banach spaces. Specifically, we seek roots of mappings that involve the composition of a smooth outer and a semismooth inner map. To exploit this structure we propose a hybrid approach in which the semismooth part is linearized in the same way as in semismooth Newton methods while the smooth part is handled by a Broyden–like method. The resulting algorithm is a semismooth Newton-type method that does not require the evaluation of the derivative of the smooth part.

We prove local q-linear and q-superlinear convergence results for the hybrid algorithm. In particular, this is the first work that establishes superlinear convergence of a semismooth quasi-Newton method in an infinite-dimensional setting. The convergence results also extend known finite-dimensional ones in that the structure of the equation and the algorithm under consideration are more general than those available in the literature. In addition, it is shown that q-linear convergence of the iterates and compactness of the initial operator discrepancy of the smooth part implies q-superlinear convergence without the assumption that the initial operator discrepancy is small in norm, which is a new type of result for semismooth quasi-Newton methods. The convergence theory is developed under mild assumptions, which yields extensions of available results for semismooth quasi-Newton methods as well as for Broyden–like methods.

The benefit of the method in practical applications is addressed in a complementary paper. There, we show on problems from optimal control that the assumptions for q-superlinear convergence are satisfied and that the hybrid approach leads to highly competitive numerical schemes that have substantially lower runtimes than state-of-the-art semismooth Newton methods.

Key words. Semismooth Newton methods, semismooth Newton-type methods, Broyden–like method, quasi-Newton methods, superlinear convergence, nonsmooth operator equations

AMS subject classifications. 47J25, 47N10, 49J27, 49J52, 49M15, 49M27, 65J15, 90C30, 90C48, 90C53, 90C56

[*]University of Graz, Heinrichstr. 36, 8010 Graz, Austria (florian.mannel/armin.rund@uni-graz.at).

1 Introduction

In this paper we combine a quasi-Newton and a semismooth Newton method to form a superlinearly convergent algorithm for the solution of structured nonsmooth operator equations in Banach spaces. Throughout this work we consider equations of the form

$$H(q) := F(G(q)) + \hat{G}(q) = 0, \tag{P}$$

where $G : Q \to U$ and $\hat{G} : Q \to V$ are semismooth, $F : U \to V$ is smooth, Q and V are Banach spaces, and U is a Hilbert space; more details are provided in Section 3. The structure of (P) is related to Robinson's normal maps, cf. [Rob92], and there is a vast amount of practically relevant problems that lead to equations of this form, including generalized variational inequalities as well as problems from nonsmooth optimization and optimal control. Examples can be found in [HS97, MR20], for instance.

Under a mild additional assumption the mapping $H : Q \to V$ is semismooth, which can be used to establish local q-superlinear convergence of semismooth Newton methods applied to (P). These methods require the evaluation of F' in each iteration since the step s^k in iteration k solves

$$\left[F'(G(q^k))M_k + \hat{M}_k\right]s^k = -H(q^k), \tag{1}$$

where $M_k \in \partial G(q^k)$ and $\hat{M}_k \in \partial\hat{G}(q^k)$ for appropriate generalized derivatives ∂G and $\partial\hat{G}$. In the semismooth Newton-type method that we develop in this paper the operator $F'(G(q^k))$ in (1) is replaced by a quasi-Newton approximation B_k, while M_k and $\hat{M}_k$ are left unchanged. This eliminates the need to evaluate F', but still yields local q-superlinear convergence as we will show. Accordingly, the runtime of the hybrid method can be expected to be lower than that of semismooth Newton methods whenever the evaluation of F' is expensive. Such is the case, for instance, in optimal control, where the hybrid approach is several times faster than state-of-the-art semismooth Newton methods. The application to optimal control including numerous numerical results is presented in the complementary work [MR20]. The present paper, however, is devoted to the convergence analysis of the method. It is the first work that proves superlinear convergence of a semismooth quasi-Newton method in an infinite-dimensional setting, and this is done under the same assumptions on G and $\hat{G}$, respectively, F that semismooth Newton methods, respectively, quasi-Newton methods require to achieve this rate of convergence. That is, the combination of the two methods does not introduce additional assumptions. In fact, the assumptions that we impose on F are quite mild, which yields extensions of well-known convergence results for semismooth quasi-Newton methods and for Broyden–like methods.

We also obtain a type of result that seems to be new for semismooth quasi-Newton methods: We show that if $B_0 - F'(G(\bar{q}))$ is compact and (q^k) converges q-linearly to $\bar{q}$ with $H(\bar{q}) = 0$, then (q^k) converges q-superlinearly even if the initial point q^0 and the initial operator B_0 are not close in norm to $\bar{q}$, respectively, to $F'(G(\bar{q}))$, provided the linear operators that are generated have uniformly bounded inverses or satisfy a Dennis–Moré-type condition. It is also noteworthy that (P) is quite general; several available semismooth quasi-Newton methods are analyzed for special cases of (P), but we have not seen a discussion of (P) in this context.

We emphasize that the choice to apply the quasi-Newton method only to the smooth part F, but not to the entire mapping H, is deliberate. In fact,

- standard quasi-Newton methods applied to semismooth equations cannot provide fast local convergence, in general;
- modified quasi-Newton methods for semismooth equations that are superlinearly convergent in infinite-dimensional spaces require strong assumptions and do not result in widely applicable numerical algorithms.

The first point is underlined by simple counterexamples which show that classical quasi-Newton methods—e.g., Broyden's method—do generally not converge superlinearly on semismooth equations, not even under favorable additional assumptions that ensure the superlinear convergence of semismooth Newton methods, cf. [Gri87, Introduction], [IS14, Example 2.40] and [AN18, Example 1]. For this reason several authors have developed modified quasi-Newton updates when dealing with nonsmooth problems. However, all but one of these methods are designed for finite-dimensional spaces and cannot be readily extended to infinite-dimensional spaces while retaining fast local convergence. The single exception is presented in [AN18], where a sound theoretical investigation of Newton-type methods for generalized equations with semismooth base mapping in Banach spaces is undertaken. Still, the proposed algorithms are not directly implementable except for particular problems, cf. [AN18, Remark 4]. In contrast, the hybrid method that we develop in this paper converges superlinearly and applies to the plethora of problems that amount to solving an equation of the form (P).

Let us set our work in perspective with the literature. Although the main focus of this paper is on infinite-dimensional settings, we start with what is available for finite-dimensional problems. There are many contributions on modified and unmodified quasi-Newton methods for nonsmooth equations in finite dimensions, e.g. [IK92, CQ94, Qi97, Che90, LF00, SH97, HS97, PQS98, AB03, Che97, LF01, QJ97, WML11, CY92]. In particular, the idea to apply a quasi-Newton method to the smooth part of a structured nonsmooth equation appears in [CY92, WML11, QJ97, HS97, SH97]. Among these five papers, [HS97] is the closest to our work. However, [HS97] does not address the infinite-dimensional setting and, in finite dimensions, is less general than the approach that we present here. On the other hand, [HS97] offers a deeper treatment of the specific setting that is investigated there, which is normal maps with polyhedral sets. Concerning the use of quasi-Newton methods on nonsmooth equations in infinite-dimensional spaces, we are only aware of the two recent papers [AN18, MHMP13], the former of which we discussed above. In [MHMP13] the authors propose to use a quasi-Newton method on the smooth part of a structured nonsmooth equation, but the structure of the equation is different from the one we use and superlinear convergence is not established. For completeness we mention that the results on linear convergence in [Gri87] also allow a certain degree of nonsmoothness, cf. [Gri87, (1.12) and (1.13)].

Infinite-dimensional quasi-Newton methods and semismooth Newton methods are studied in [Sac86, KS91, Gri87, HK92, ABDL14], respectively, [Ulb11]. Since semismooth Newton–type methods can be viewed as inexact semismooth Newton methods and vice versa, we also mention the works [MQ95, CQ10, Ulb11, LG06] on inexact semismooth Newton methods.

This paper is organized as follows. In Section 2 we provide necessary concepts and auxiliary results. In Section 3 we introduce the hybrid method and discuss the assumptions that underlie its convergence analysis. Section 4 is devoted to proving local linear and superlinear convergence, and the final Section 5 contains a summary of this paper as well as an outlook on the complementary paper [MR20].

Notation We use $\mathbb{N}$ for the natural numbers and $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$. The open ball with radius δ around x is $\mathbb{B}_\delta(x)$, respectively, $\mathbb{B}^*_\delta(x) := \mathbb{B}_\delta(x) \setminus \{x\}$ if x is excluded. Given normed linear spaces X and Y, the space of continuous linear mappings from X to Y is indicated by $\mathcal{L}(X,Y)$. In a Hilbert space U we use $(v,w)_U$ for the scalar product and $(v,\cdot)_U$ for the linear functional $w \mapsto (v,w)_U$.

2 Preliminaries

We collect definitions and results that are used in this work. Throughout Section 2, X, Y, Z are normed linear spaces.

2.1 Semismoothness and a stronger version of strict differentiability

We use semismoothness in the sense of Ulbrich [Ulb11] in this work.

Definition 2.1 (Cf. [Ulb11, Definition 3.1])**.** Let $\bar{x} \in X$ and let $g : X \to Y$ be continuous in an open neighborhood of $\bar{x} \in X$. Moreover, let $\partial g : X \rightrightarrows \mathcal{L}(X,Y)$ satisfy $\partial g(x) \neq \emptyset$ for all $x \in X$. We call g *semismooth at $\bar{x}$ with respect to ∂g* iff

$$\sup_{M \in \partial g(\bar{x}+h)} \|g(\bar{x}+h) - g(\bar{x}) - Mh\|_Y = o(\|h\|_X) \qquad \text{for } \|h\|_X \to 0.$$

The set-valued mapping $\partial g : X \rightrightarrows \mathcal{L}(X,Y)$ is called a *generalized derivative of g*.

Remark 2.2. Definition 2.1 includes as special cases Newton differentiability, cf. [IK08, Definition 8.10], semismoothness based on slant derivatives, cf. [CNQ00], as well as most finite-dimensional notions of semismoothness. At the expense of additional technicalities it is possible to extend the results of this paper to the more general notion of semismoothness introduced in [Kru18, Definition 3].

We will use the following differentiability concept that is inspired by [Qi97].

Definition 2.3. Let $\bar{x} \in X$ and $\eta > 0$. We say that $f : X \to Y$ is *η-strictly differentiable at $\bar{x}$* iff there exist $f'(\bar{x}) \in \mathcal{L}(X,Y)$ and $C_f, \delta_f > 0$ such that

$$\|f(y) - f(x) - f'(\bar{x})(y-x)\|_Y \le C_f \|y-x\|_X \max\{\|y-\bar{x}\|_X, \|x-\bar{x}\|_X\}^\eta$$

is satisfied for all $x, y \in \mathbb{B}_{\delta_f}(\bar{x})$.

We state properties of η-strictly differentiable functions. For the concept of strict differentiability we refer to [Mor05, Definition 1.13] and [DR14, Section 1.4].

Lemma 2.4.

1) *If f satisfies Definition 2.3, then it is semismooth at $\bar{x}$ wrt. $\partial f(x) := \{f'(\bar{x})\}$, $x \in X$, strictly differentiable at $\bar{x}$, and Fréchet differentiable at $\bar{x}$. In particular, $f'(\bar{x})$ is unique.*

2) *If $f : X \to Y$ is Hölder continuously Fréchet differentiable in a neighborhood of $\bar{x}$, then it is η-strictly differentiable at $\bar{x}$ with η equal to the Hölder exponent.*

Proof. All claims follow from the respective definitions. □

2.2 Local calmness and local metric subregularity

We introduce local calmness and local metric subregularity, cf. [DR14, Section 1.3 and Section 3.8], and provide sufficient conditions for these properties in case of semismooth mappings.

Definition 2.5. Let $D \subset X$, $g : D \to Y$, and $\bar{x} \in D$.

1) g is *calm at* $\bar{x}$ iff there exists $L_g > 0$ such that for all $x \in D$

$$\|g(x) - g(\bar{x})\|_Y \leq L_g \|x - \bar{x}\|_X.$$

2) g is *metrically subregular at* $\bar{x}$ iff there exists $\kappa_g > 0$ such that for all $x \in D$

$$\|x - \bar{x}\|_X \leq \kappa_g \|g(x) - g(\bar{x})\|_Y.$$

3) g is *locally calm (locally metrically subregular) at* $\bar{x}$ iff there exists $\delta_g > 0$ such that g restricted to $\mathbb{B}_{\delta_g}(\bar{x})$ is calm (metrically subregular) at $\bar{x}$.

We provide sufficient conditions for local calmness and local metric subregularity in the presence of semismoothness based on the following concepts.

Definition 2.6. Let $g : X \to Y$ be semismooth at $\bar{x} \in X$.

1) We say ∂g has a *uniformly bounded selection near* $\bar{x}$ iff there are $C_M, \delta_M > 0$ such that for every $x \in \mathbb{B}_{\delta_M}(\bar{x})$ there is $M \in \partial g(x)$ with $\|M\|_{\mathcal{L}(X,Y)} \leq C_M$.
2) If the inequality in 1) holds for all $x \in \mathbb{B}^*_{\delta_M}(\bar{x})$ ($x \in \mathbb{B}_{\delta_M}(\bar{x})$) and all $M \in \partial g(x)$, then ∂g is called *uniformly* (uniformly) bounded near* $\bar{x}$.
3) We say ∂g has a *uniformly* (uniformly) invertible selection near* $\bar{x}$ iff there are $C_{M^{-1}}, \delta_{M^{-1}} > 0$ such that for every $x \in \mathbb{B}^*_{\delta_{M^{-1}}}(\bar{x})$ ($x \in \mathbb{B}_{\delta_{M^{-1}}}(\bar{x})$) there is an invertible $M \in \partial g(x)$ with $\|M^{-1}\|_{\mathcal{L}(Y,X)} \leq C_{M^{-1}}$.
4) If the inequality in 3) holds for all $x \in \mathbb{B}^*_{\delta_{M^{-1}}}(\bar{x})$ ($x \in \mathbb{B}_{\delta_{M^{-1}}}(\bar{x})$) and all $M \in \partial g(x)$, then ∂g is called *uniformly* (uniformly) invertible near* $\bar{x}$.

Lemma 2.7. *Let $g : X \to Y$ be semismooth at $\bar{x} \in X$. Then:*

1) *g is locally calm at $\bar{x}$ if ∂g admits a uniformly bounded selection near $\bar{x}$.*
2) *g is locally metrically subregular at $\bar{x}$ if ∂g admits a uniformly* invertible selection near $\bar{x}$.*

Proof. The proofs are quite simple, so we prove only 2). By the selection property there are $C_{M^{-1}}, \delta_{M^{-1}} > 0$ such that for every $x \in \mathbb{B}^*_{\delta_{M^{-1}}}(\bar{x})$ there is at least one $M_x \in \partial g(x)$ that is invertible with $\|M_x^{-1}\|_{\mathcal{L}(Y,X)} \leq C_{M^{-1}}$. The semismoothness yields $\hat{\delta} \in (0, \delta_{M^{-1}}]$ such that

$$\begin{aligned}\frac{\|x-\bar{x}\|_X}{C_{M^{-1}}} \leq \|M_x(x-\bar{x})\|_Y &\leq \|g(x)-g(\bar{x})-M_x(x-\bar{x})\|_Y + \|g(x)-g(\bar{x})\|_Y \\ &\leq \frac{\|x-\bar{x}\|_X}{2C_{M^{-1}}} + \|g(x)-g(\bar{x})\|_Y\end{aligned}$$

is satisfied for all $x \in \mathbb{B}_{\hat{\delta}}(\bar{x})$, which implies the claim. □

Corollary 2.8. *Let $f : X \to Y$ be η-strictly differentiable at $\bar{x} \in X$. Then:*

1) f is locally calm at $\bar{x}$.

2) f is locally metrically subregular at $\bar{x}$ if $f'(\bar{x})$ is invertible.

Proof. Apply Lemma 2.7 to f and $\partial f(x) := \{f'(\bar{x})\}$, $x \in X$. □

2.3 A chain rule for semismooth mappings

The following chain rule for semismooth mappings generalizes [Ulb11, Proposition 3.8].

Lemma 2.9. *Let $g : X \to Y$ be semismooth at $\bar{x} \in X$ and locally calm at $\bar{x}$. Let $f : Y \to Z$ be semismooth at $\bar{y} := g(\bar{x})$ with uniformly bounded generalized derivative near $\bar{y}$. Then $h : X \to Z$, $h := f(g(x))$ is semismooth at $\bar{x}$ with respect to*

$$\partial h : X \rightrightarrows \mathcal{L}(X,Z), \qquad \partial h(x) := \left\{M_f \circ M_g : \ M_f \in \partial f(g(x)),\ M_g \in \partial g(x)\right\}.$$

In addition, h is locally calm at $\bar{x}$.

Proof. The semismoothness can be proven as in [Ulb11, Proposition 3.8]. Since f is locally calm at $\bar{y}$ by Lemma 2.7, it follows readily that $h = f \circ g$ is locally calm at $\bar{x}$. □

3 The hybrid semismooth quasi-Newton method

3.1 Problem setting and algorithm

In the remainder of this work we consider the following setting. Given Banach spaces Q, V and a Hilbert space U as well as mappings $G : Q \to U$, $F : U \to V$ and $\hat{G} : Q \to V$, we want to find $\bar{q} \in Q$ such that

$$H(\bar{q}) = 0, \tag{P}$$

where $H : Q \to V$, $H(q) := F(G(q)) + \hat{G}(q)$. F shall be smooth and $G, \hat{G}$ shall be semismooth at $\bar{q}$ with respect to generalized derivatives ∂G and $\partial \hat{G}$. For concrete problem classes that are covered by (P) we refer to [HS97, MR20].

The hybrid semismooth quasi-Newton method is obtained by replacing the operator $F'(u^k)$ in semismooth Newton methods by a Broyden–like approximation.

Algorithm SQN: Hybrid semismooth quasi-Newton method

Input: $q^0 \in Q$, $B_0 \in \mathcal{L}(U,V)$.

```
1  Let u^0 := G(q^0).
2  for k = 0, 1, 2, ... do
3      if H(q^k) = 0 then let q* := q^k; STOP.
4      Choose M_k ∈ ∂G(q^k) and M̂_k ∈ ∂Ĝ(q^k).
5      Let M̃_k := B_k M_k + M̂_k.
6      Solve M̃_k s^k = −H(q^k) for s^k.
7      Let q^{k+1} := q^k + s^k and u^{k+1} := G(q^{k+1}).
8      Let s_u^k := u^{k+1} − u^k and y^k := F(u^{k+1}) − F(u^k).
9      Choose σ_k ∈ [0, 2].
10     if s_u^k ≠ 0 then let B_{k+1} := B_k + σ_k (y^k − B_k s_u^k) (s_u^k, ·)_U / ‖s_u^k‖_U^2;
11     else let B_{k+1} := B_k.
12 end
```

Output: q^*.

Since B_k is used instead of $F'(u^k)$ in Line 5, the evaluation of F' is not required. Practical aspects of Algorithm SQN including globalization and application in large-scale optimal control, where the evaluation of F' is expensive, are treated in [MR20].

The idea to generalize the Broyden update through the parameter σ_k is well-known, cf. [MT76], [Sac85, Section 6], [HK92], and [LF00, Algorithm 1]; the use of σ_k ensures that B_{k+1} will be invertible if B_k is invertible. For semismooth quasi-Newton methods, however, this generalized update has not been considered before. Under appropriate assumptions we will show q-linear convergence for arbitrary choices of $(\sigma_k) \subset [0,2]$ and q-superlinear convergence if $\liminf_{k\to\infty} \sigma_k > 0$ and $\limsup_{k\to\infty} \sigma_k < 2$ hold.

3.2 Main assumptions and consequences

The convergence analysis rests on two sets of assumptions. The first one reads as follows.

Assumption 3.1. Suppose that

- Q, V are Banach spaces and U is a Hilbert space;
- mappings $G : Q \to U$, $F : U \to V$ and $\hat{G} : Q \to V$ are given;
- there is $\bar{q} \in Q$ with $H(\bar{q}) = 0$, where $H := F \circ G + \hat{G}$;
- G and $\hat{G}$ are semismooth at $\bar{q}$ wrt. generalized derivatives ∂G and $\partial\hat{G}$;
- ∂G is uniformly* bounded near $\bar{q}$ with constants $C_M, \delta_M > 0$;
- F is η-strictly differentiable at $\bar{u} := G(\bar{q})$ with constants $C_F, \delta_F > 0$;

- the generalized derivative $\partial H : Q \rightrightarrows \mathcal{L}(Q, V)$ is given by

$$\partial H(q) := \left\{ F'(\bar{u}) \circ M + \hat{M} : \quad M \in \partial G(q),\, \hat{M} \in \partial \hat{G}(q) \right\}. \tag{2}$$

Remark 3.2. Since ∂H involves the unknown point $\bar{u}$, we stress that ∂H is not used in Algorithm SQN, but for the convergence analysis only. Using $\bar{u}$ instead of a full neighborhood reduces the requirements for F and ∂H.

We will prove two types of superlinear convergence results. In the first type we *assume* that all $\tilde{M}_k$ generated by Algorithm SQN are invertible and that the sequence $(\|\tilde{M}_k^{-1}\|_{\mathcal{L}(V,Q)})$ is bounded. In the second type we *prove* that these properties are satisfied under an additional assumption. The first type holds under Assumption 3.1, while the second requires the following stronger assumption.

Assumption 3.3. Assumption 3.1 holds and ∂H is uniformly* invertible near $\bar{q}$ with constants $C_{\bar{M}^{-1}}, \delta_{\bar{M}^{-1}} > 0$.

Remark 3.4. If $q \mapsto \partial H(q)$ is upper semicontinuous at $\bar{q}$, then ∂H is uniformly invertible near $\bar{q}$ if there exists $C_{\bar{M}^{-1}} > 0$ such that every $\bar{M} \in \partial H(\bar{q})$ is invertible with $\|\bar{M}^{-1}\|_{\mathcal{L}(V,Q)} \le C_{\bar{M}^{-1}}$. This follows from the Banach lemma.

It is fundamental for the convergence analysis that H is semismooth at $\bar{q}$.

Lemma 3.5. *Let Assumption 3.1 hold. Then H is semismooth at $\bar{q}$ with respect to ∂H defined in (2), and F, G, and $F \circ G$ are locally calm at $\bar{u}$ and $\bar{q}$, respectively.*

Proof. The uniform* boundedness of ∂G near $\bar{q}$ implies by Lemma 2.7 that G is locally calm at $\bar{q}$. Corollary 2.8 shows that F is locally calm at $\bar{u}$. To prove semismoothness of H, note that $\hat{G}$ is semismooth at $\bar{q}$ by assumption, hence it is sufficient to show that $F \circ G$ is semismooth at $\bar{q}$. By Lemma 2.4, F is semismooth at $\bar{u}$ wrt. $\partial F(u) := \{F'(\bar{u})\}$, $u \in U$. Since ∂F is trivially uniformly bounded near $\bar{u}$, the semismoothness of $F \circ G$ at $\bar{q}$ follows from Lemma 2.9. In addition, this lemma yields that $F \circ G$ is locally calm at $\bar{q}$. □

The previous lemma implies that we can assume without loss of generality that the constant δ_M in Assumption 3.1 is so small that G is locally calm at $\bar{q}$ in $\mathbb{B}_{\delta_M}(\bar{q})$.

Notation 3.6. If Assumption 3.1 holds, then we write L_F for the constant of local calmness of F at $\bar{u}$ and L_G for the constant of local calmness of G at $\bar{q}$ in $\mathbb{B}_{\delta_M}(\bar{q})$.

For later use let us also record the following properties of H.

Lemma 3.7. *If Assumption 3.3 holds, then H is locally metrically subregular at $\bar{q}$. If Assumption 3.1 holds and $\hat{G}$ is locally calm at $\bar{q}$, then H is locally calm at $\bar{q}$.*

Proof. Under Assumption 3.1, $F \circ G$ is locally calm at $\bar{q}$ by Lemma 3.5. Together with the calmness of $\hat{G}$ at $\bar{q}$ this implies the claimed calmness of H. Under Assumption 3.3, ∂H is uniformly* invertible near $\bar{q}$. Thus, Lemma 2.7 yields that H is locally metrically subregular at $\bar{q}$, where we used that H is semismooth at $\bar{q}$ by Lemma 3.5. □

4 Convergence analysis

This is the main section of this work, where we develop the local convergence theory for Algorithm SQN. We use the following notation.

Notation 4.1. Whenever Assumption 3.1 holds and Algorithm SQN has generated iterates q^{k+1}, q^k, u^{k+1}, u^k and an operator B_k, we define

$$s^k := q^{k+1} - q^k, \qquad s_u^k := u^{k+1} - u^k, \qquad \bar{s}^k := q^k - \bar{q}, \qquad \bar{s}_u^k := u^k - \bar{u},$$

as well as

$$\hat{s}_u^k := \frac{s_u^k}{\|s_u^k\|_U} \;\text{ if } s_u^k \neq 0 \qquad \text{and} \qquad \hat{s}_u^k := 0 \;\text{ if } s_u^k = 0,$$

and

$$E_k := B_k - F'(\bar{u}).$$

4.1 Linear convergence

In this section we prove local q-linear convergence of Algorithm SQN. The proof uses the following consequence of Banach's lemma.

Lemma 4.2. *Let Q and V be Banach spaces and $A, B \in \mathcal{L}(Q,V)$. If A is invertible with $\|A^{-1}\|_{\mathcal{L}(V,Q)} \leq C$ for some $C > 0$ and B satisfies $\|A-B\|_{\mathcal{L}(Q,V)} \leq (2C)^{-1}$, then B is invertible with $\|B^{-1}\|_{\mathcal{L}(V,Q)} \leq 2C$.*

Proof. This follows readily from [Cia13, Theorem 3.6-3]. □

A common tool to show linear convergence of Broyden–like methods is the *bounded deterioration principle*, cf., e.g., [BDM73, Theorem 3.2], [Kel95, Section 7.2.1], and the set-valued version [AN18, Theorem 4]. We will use the following variant of it.

Lemma 4.3. *Let U be a Hilbert space and V be Banach space. Let $F: U \to V$ be η-strictly differentiable at $\bar{u}$ with constants $C_F, \delta_F > 0$. Then for all $u, u_+ \in \mathbb{B}_{\delta_F}(\bar{u})$, all $\sigma \in [0,2]$, and all $B \in \mathcal{L}(U,V)$ the linear operator*

$$B_+ := \begin{cases} B & \text{if } s = 0, \\ B + \sigma\,(y - Bs)\,\frac{(s,\cdot)_U}{\|s\|_U^2} & \text{else,} \end{cases}$$

where $s := u_+ - u$ and $y := F(u_+) - F(u)$, satisfies

$$\|B_+ - F'(\bar{u})\|_{\mathcal{L}(U,V)} \leq \|B - F'(\bar{u})\|_{\mathcal{L}(U,V)} + \sigma C_F \max\{\|u_+ - \bar{u}\|_U\,, \|u - \bar{u}\|_U\}^{\eta}.$$

Proof. For all $u_+, u \in \mathbb{B}_{\delta_F}(\bar u)$ the η-strict differentiability of F provides

$$\|F(u_+) - F(u) - F'(\bar u)(u_+ - u)\|_V \le C_F \|u_+ - u\|_U \max\{\|u_+ - \bar u\|_U, \|u - \bar u\|_U\}^\eta. \tag{3}$$

Let $u, u_+ \in \mathbb{B}_{\delta_F}(\bar u)$, $\sigma \in [0,2]$, $B \in \mathcal{L}(U,V)$, and set $s := u_+ - u$. Since the claim is trivial for $s = 0$, we can assume $s \neq 0$. Defining $\hat s := \frac{s}{\|s\|_U}$ we readily confirm that

$$B_+ - F'(\bar u) = \left[B - F'(\bar u)\right]\left[I - \sigma \hat s\,(\hat s, \cdot)_U\right] + \sigma\left[F(u_+) - F(u) - F'(\bar u)s\right]\frac{(\hat s, \cdot)_U}{\|s\|_U}.$$

It is elementary to see that $\|I - \sigma \hat s\,(\hat s, \cdot)_U\|_{\mathcal{L}(U,U)} \le 1$. With (3) this yields

$$\|B_+ - F'(\bar u)\|_{\mathcal{L}(U,V)} \le \|B - F'(\bar u)\|_{\mathcal{L}(U,V)} + \sigma C_F \max\{\|u_+ - \bar u\|_U, \|u - \bar u\|_U\}^\eta. \qquad \square$$

As main result of this section we prove q-linear convergence of Algorithm SQN.

Theorem 4.4. *Let Assumption 3.3 hold and let $\beta \in (0,1)$. Then:*

1) *There exist $\delta, \varepsilon > 0$ with the following property: For every initial $(q^0, B_0) \in Q \times \mathcal{L}(U,V)$ with $\|q^0 - \bar q\|_Q < \delta$ and $\|E_0\|_{\mathcal{L}(U,V)} < \varepsilon$, Algorithm SQN is well-defined and either terminates with output $q^* = \bar q$ or it generates an infinite sequence (q^k), and there hold*

$$\|q^{k+1} - \bar q\|_Q \le \beta \|q^k - \bar q\|_Q, \qquad \max_{0 \le j \le k+1} \|E_j\|_{\mathcal{L}(U,V)} \le \frac{\beta}{4 C_M C_{\bar M^{-1}}}, \tag{4}$$

and

$$\|\tilde M_k^{-1}\|_{\mathcal{L}(V,Q)} \le 2 C_{\bar M^{-1}}$$

whenever q^{k+1} is generated for some $k \in \mathbb{N}_0$.

2) *If additionally F is Gâteaux differentiable in a neighborhood of $\bar u$ and the Gâteaux derivative is continuous at $\bar u$, then the condition $\|E_0\|_{\mathcal{L}(U,V)} < \varepsilon$ in 1) can be replaced by $\|B_0 - F'(u^0)\|_{\mathcal{L}(U,V)} < \varepsilon$. In particular, this replacement is possible if F is Hölder continuously Fréchet differentiable near $\bar u$.*

Proof. **Proof of 1):** The first task is to provide suitable values for δ and ε. By shrinking β if necessary, we can assume without loss of generality that $\beta \le \frac{1}{2^{1/\eta}}$, where $\eta > 0$ is the constant from Assumption 3.1. Thus, we have $\hat\beta := \beta^\eta \le \frac{1}{2}$. Since H is semismooth at $\bar q$ by Lemma 3.5, there is $\delta_H > 0$ such that

$$\sup_{\bar M \in \partial H(q)} \left\|H(q) - H(\bar q) - \bar M(q - \bar q)\right\|_V \le \frac{\beta}{4 C_{\bar M^{-1}}} \|q - \bar q\|_Q \tag{5}$$

holds for all $q \in \mathbb{B}_{\delta_H}(\bar q)$. We recall from Notation 3.6 that G is locally calm at $\bar q$ in $\mathbb{B}_{\delta_M}(\bar q)$ with constant $L_G > 0$ and from Assumption 3.1 that F is η-strictly differentiable at $\bar u$ with constants $C_F, \delta_F > 0$. Also, by Lemma 3.7 we have that H is locally metrically subregular at $\bar q$, which implies that there is $\delta_R > 0$ such that $\bar q$ is the only root of H in $\mathbb{B}_{\delta_R}(\bar q)$. Defining $\check C := 8 C_M C_{\bar M^{-1}} > 0$ and $\mathring{C} := 2 C_F L_G^\eta > 0$ we will show that the claims hold for

$$\varepsilon := \min\left\{\delta_F, \frac{\beta}{\check C}\right\} \qquad \text{and} \qquad \delta := \min\left\{\delta_R, \delta_M, \delta_{\bar M^{-1}}, \delta_H, \left(\frac{\beta}{2\check C \mathring{C}}\right)^{\frac{1}{\eta}}, \frac{\varepsilon}{L_G}\right\}.$$

To this end, let q^0 with $\|q^0 - \bar{q}\|_Q < \delta$ and B_0 with $\|E_0\|_{\mathcal{L}(U,V)} = \|F'(\bar{u}) - B_0\|_{\mathcal{L}(U,V)} < \varepsilon$ be given. Fix $k \in \mathbb{N}_0$ and assume by means of induction that q^{j+1} and B_{j+1} have been generated for all $0 \le j \le k-1$ and that there hold

$$\|q^{j+1} - \bar{q}\|_Q \le \beta \|q^j - \bar{q}\|_Q \quad \text{and} \quad \|E_{j+1}\|_{\mathcal{L}(U,V)} \le \|E_0\|_{\mathcal{L}(U,V)} + \mathring{C}\left(2 - \hat{\beta}^j\right)\|q^0 - \bar{q}\|_Q^\eta$$

for these j. We consider Line 2 to Line 11 in Algorithm SQN. If the algorithm terminates in Line 3, we need only show that $q^* = \bar{q}$. Since $\bar{q}$ is the only root of H in $\mathbb{B}_\delta(\bar{q})$, it is obvious that $q^* = \bar{q}$ in this case. In the remainder of the induction we may assume $q^k \ne \bar{q}$. Together with the induction assumption this implies $q^k \in \mathbb{B}_\delta^*(\bar{q})$. In particular, we have $q^k \in \mathbb{B}_{\delta_M}^*(\bar{q})$, which yields $\|M_k\|_{\mathcal{L}(Q,U)} \le C_M$ by Assumption 3.1. Moreover, we have $\bar{M}_k := F'(\bar{u})M_k + \hat{M}_k \in \partial H(q^k)$ and

$$\|\bar{M}_k - \tilde{M}_k\|_{\mathcal{L}(Q,V)} = \|\left(F'(\bar{u}) - B_k\right) M_k\|_{\mathcal{L}(Q,V)} \le \|E_k\|_{\mathcal{L}(U,V)} \|M_k\|_{\mathcal{L}(Q,U)}.$$

The induction assumption yields $\|E_k\|_{\mathcal{L}(U,V)} < \varepsilon + 2\mathring{C}\delta^\eta \le \frac{2\beta}{\breve{C}}$, hence

$$\|\bar{M}_k - \tilde{M}_k\|_{\mathcal{L}(Q,V)} \le \frac{\beta}{4C_{\bar{M}^{-1}}} \le \frac{1}{2C_{\bar{M}^{-1}}}, \tag{6}$$

where we used the definition of $\breve{C}$ to derive the first inequality. As $q^k \in \mathbb{B}_\delta^*(\bar{q})$ holds, $\bar{M}_k$ is invertible with $\|\bar{M}_k^{-1}\|_{\mathcal{L}(V,Q)} \le C_{\bar{M}^{-1}}$ by Assumption 3.3. From Lemma 4.2 we thus deduce that $\tilde{M}_k$ is invertible with $\|\tilde{M}_k^{-1}\|_{\mathcal{L}(V,Q)} \le 2C_{\bar{M}^{-1}}$. This establishes the inequality below (4) and implies that q^{k+1} is well-defined. Next we prove that $\|q^{k+1} - \bar{q}\|_Q \le \beta \|q^k - \bar{q}\|_Q$. From $q^k \in \mathbb{B}_\delta^*(\bar{q})$ we deduce $q^k \in \mathbb{B}_{\delta_H}^*(\bar{q})$, which allows us to apply (5). Using $H(\bar{q}) = 0$ we compute

$$\begin{aligned}
&\|q^{k+1} - \bar{q}\|_Q \\
&\quad = \|s^k + q^k - \bar{q}\|_Q = \left\| -\tilde{M}_k^{-1} H(q^k) + q^k - \bar{q} \right\|_Q \\
&\quad = \left\| -\tilde{M}_k^{-1} \left[H(q^k) - H(\bar{q}) - \tilde{M}_k \left(q^k - \bar{q}\right) \right] \right\|_Q \\
&\quad = \left\| -\tilde{M}_k^{-1} \left[H(q^k) - H(\bar{q}) - \bar{M}_k \left(q^k - \bar{q}\right) + (\bar{M}_k - \tilde{M}_k) \left(q^k - \bar{q}\right) \right] \right\|_Q \\
&\quad \le \left\| \tilde{M}_k^{-1} \right\|_{\mathcal{L}(V,Q)} \left[\left\| H(q^k) - H(\bar{q}) - \bar{M}_k \left(q^k - \bar{q}\right) \right\|_V + \left\| (\bar{M}_k - \tilde{M}_k) \left(q^k - \bar{q}\right) \right\|_V \right] \\
&\quad \le \left\| \tilde{M}_k^{-1} \right\|_{\mathcal{L}(V,Q)} \left[\frac{\beta}{4C_{\bar{M}^{-1}}} + \left\| \bar{M}_k - \tilde{M}_k \right\|_{\mathcal{L}(Q,V)} \right] \left\| q^k - \bar{q} \right\|_Q .
\end{aligned}$$

We have already established $\|\tilde{M}_k^{-1}\|_{\mathcal{L}(V,Q)} \le 2C_{\bar{M}^{-1}}$ and $\|\bar{M}_k - \tilde{M}_k\|_{\mathcal{L}(Q,V)} \le \frac{\beta}{4C_{\bar{M}^{-1}}}$. Inserting these inequalities on the right-hand side yields $\|q^{k+1} - \bar{q}\|_Q \le \beta\|q^k - \bar{q}\|_Q$, as desired. In particular, we can use $\|q^j - \bar{q}\|_Q < \delta \le \delta_M$ for all $0 \le j \le k+1$ in the remainder of the induction and, consequently, the local calmness of G at $\bar{q}$ with constant L_G is available for the iterates q^0, q^1, …, q^{k+1}. To complete the induction we have to show

$$\|E_{k+1}\|_{\mathcal{L}(U,V)} \le \|E_0\|_{\mathcal{L}(U,V)} + \mathring{C}(2 - \hat{\beta}^k)\|q^0 - \bar{q}\|_Q^\eta. \tag{7}$$

Due to $\|E_{k+1}\|_{\mathcal{L}(U,V)} \le \varepsilon + \frac{\beta}{\breve{C}} \le 2\frac{\beta}{\breve{C}} \le \frac{\beta}{4C_M C_{\bar{M}^{-1}}}$ and $\|E_0\|_{\mathcal{L}(U,V)} < \varepsilon$ this also implies the second claim of (4). If $s_u^k = 0$, then $E_{k+1} = E_k$ by definition of the update B_{k+1}, hence (7) holds by induction assumption. Thus, we suppose $s_u^k \ne 0$ in the following. Since

$$\|u^{k+1} - \bar{u}\|_U = \|G(q^{k+1}) - G(\bar{q})\|_U \le L_G\|q^{k+1} - \bar{q}\|_Q < L_G\delta \le \varepsilon \le \delta_F$$

and since the same upper bound holds for u^k instead of u^{k+1}, we can apply the deterioration estimate from Lemma 4.3. This produces

$$\begin{aligned}\|E_{k+1}\|_{\mathcal{L}(U,V)} &\le \|E_k\|_{\mathcal{L}(U,V)} + \sigma_k C_F \max\{\|u^{k+1}-\bar{u}\|_U, \|u^k-\bar{u}\|_U\}^\eta \\ &\le \|E_k\|_{\mathcal{L}(U,V)} + \sigma_k C_F L_G^\eta \|q^k-\bar{q}\|_Q^\eta.\end{aligned}$$

As $\|q^k-\bar{q}\|_Q \le \beta^k \|q^0-\bar{q}\|_Q$ and $\sigma_k \le 2$, we find

$$\|E_{k+1}\|_{\mathcal{L}(U,V)} \le \|E_k\|_{\mathcal{L}(U,V)} + 2C_F L_G^\eta \hat{\beta}^k \|q^0-\bar{q}\|_Q^\eta = \|E_k\|_{\mathcal{L}(U,V)} + \mathring{C}\hat{\beta}^k \|q^0-\bar{q}\|_Q^\eta.$$

Recalling that $\hat{\beta} \le \frac{1}{2}$, we have $1-\hat{\beta} \ge \frac{1}{2} \ge \hat{\beta}$. Together with the induction assumption $\|E_k\|_{\mathcal{L}(U,V)} \le \|E_0\|_{\mathcal{L}(U,V)} + \mathring{C}(2-\hat{\beta}^{k-1})\|q^0-\bar{q}\|_Q^\eta$, this implies

$$\begin{aligned}\|E_{k+1}\|_{\mathcal{L}(U,V)} &\le \|E_0\|_{\mathcal{L}(U,V)} + \mathring{C}(2-\hat{\beta}^{k-1}+\hat{\beta}^k)\|q^0-\bar{q}\|_Q^\eta \\ &\le \|E_0\|_{\mathcal{L}(U,V)} + \mathring{C}(2-\hat{\beta}^k)\|q^0-\bar{q}\|_Q^\eta,\end{aligned}$$

thereby concluding the induction and the proof of part 1).

Proof of 2): It is enough to show that for given $\delta, \varepsilon > 0$ there are $\hat{\delta}, \hat{\varepsilon} > 0$ such that

$$\Big[\|q^0-\bar{q}\|_Q < \hat{\delta} \ \wedge\ \|B_0 - F'(u^0)\|_{\mathcal{L}(U,V)} < \hat{\varepsilon}\Big] \implies \Big[\|q^0-\bar{q}\|_Q < \delta \ \wedge\ \|E_0\|_{\mathcal{L}(U,V)} < \varepsilon\Big].$$

Due to the continuity of F' at $\bar{u}$ there is $\hat{\varepsilon} > 0$ such that $\|u^0-\bar{u}\|_U < \hat{\varepsilon}$ implies $\|F'(u^0) - F'(\bar{u})\|_{\mathcal{L}(U,V)} < \frac{\varepsilon}{2}$. By decreasing $\hat{\varepsilon}$ if necessary, we can assume that $\hat{\varepsilon} \le \frac{\varepsilon}{2}$. We recall from Notation 3.6 that G is calm at $\bar{q}$ in $\mathbb{B}_{\delta_M}(\bar{q})$ with constant $L_G > 0$. Defining $\hat{\delta} := \min\{\delta, \delta_M, \frac{\hat{\varepsilon}}{L_G}\}$ it is clear that $\|q^0-\bar{q}\|_Q < \hat{\delta}$ yields $\|q^0-\bar{q}\|_Q < \delta$. It remains to establish $\|E_0\|_{\mathcal{L}(U,V)} < \varepsilon$. From $\|q^0-\bar{q}\|_Q < \hat{\delta} \le \frac{\hat{\varepsilon}}{L_G}$ and the calmness of G at $\bar{q}$ it follows that $\|u^0-\bar{u}\|_U = \|G(q^0)-G(\bar{q})\|_Q < \hat{\varepsilon}$. Since $\|u^0-\bar{u}\|_U < \hat{\varepsilon}$ implies $\|F'(u^0)-F'(\bar{u})\|_{\mathcal{L}(U,V)} < \frac{\varepsilon}{2}$ and since $\|B_0-F'(u^0)\|_{\mathcal{L}(U,V)} < \hat{\varepsilon} \le \frac{\varepsilon}{2}$, we obtain $\|B_0 - F'(\bar{u})\|_{\mathcal{L}(U,V)} < \varepsilon$, as desired. □

Remark 4.5.

1) After small modifications in the previous proof it follows that Theorem 4.4 stays valid for a given $\beta \in (0,1)$ if the semismoothness of G and $\hat{G}$ at $\bar{q}$ contained in Assumption 3.1 are replaced by (5).

2) It is possible to include a step length $\alpha_k > 0$ in Algorithm SQN, i.e., to put $q^{k+1} := q^k + \alpha_k s^k$ in Line 7. In fact, it can be shown that there are $\alpha_+ > 1$ and $0 < \alpha_- < 1$ such that the local linear convergence of Theorem 4.4 is preserved if $\alpha_k \in [\alpha_-, \alpha_+]$ holds for all k. The proof requires additionally that $\hat{G}$ is locally calm at $\bar{q}$. However, our main focus is on superlinear convergence, which requires $\alpha_k \to 1$, so we work with unit step length for simplicity.

3) Theorem 4.4 generalizes several results in the literature, both for semismooth quasi-Newton methods, e.g. [HS97, linear convergence part of Theorem 4], and for Broyden–like methods, e.g. [Den71, Theorem 5 and Corollary 5]. For the case of Broyden–like methods note that with $Q = U$, $G = \mathrm{id}$ and $\hat{G} \equiv 0$, Algorithm SQN reduces to a Broyden–like method for F.

4) For $(\sigma_k) \equiv 0$, Theorem 4.4 shows local q-linear convergence of a semismooth simplified Newton method, respectively, of a simplified Newton method.

The next result analyzes the convergence behavior of (u^k), $(F(u^k))$, and $(H(q^k))$. It involves multi-step q-linear convergence, which is stronger than r-linear convergence.

Definition 4.6. Let $(x^k) \subset X$ and $j \in \mathbb{N}$. We say that (x^k) *converges j-step q-linearly to $\bar{x} \in X$ with convergence factor $\beta \in [0,1)$* iff there is $K \in \mathbb{N}_0$ such that

$$\|x^{k+j} - \bar{x}\|_X \leq \beta \|x^k - \bar{x}\|_X$$

is satisfied for all $k \geq K$. For $j = 1$ the term "1-step" is omitted.

Since Theorem 4.4 yields q-linear convergence of (q^k), the following result can be used to obtain rates of convergence for (u^k), $(F(u^k))$, and $(H(q^k))$.

Lemma 4.7. *Let Assumption 3.1 hold and suppose that Algorithm SQN generates a sequence of iterates (q^k) that converges r-linearly to $\bar{q}$. Then:*

1) (u^k) and $(F(u^k))$ converge r-linearly to $\bar{u}$, respectively, $F(\bar{u})$, and there hold $\|u^k - \bar{u}\|_U \leq L_G \|q^k - \bar{q}\|_Q$, $\|F(u^k) - F(\bar{u})\|_V \leq L_F \|u^k - \bar{u}\|_U$, and $\|F(u^k) - F(\bar{u})\|_V \leq L_F L_G \|q^k - \bar{q}\|_Q$ for all k sufficiently large.

2) $(H(q^k))$ converges to zero. If $\hat{G}$ is locally calm at $\bar{q}$, then $(H(q^k))$ converges r-linearly. If $\hat{G}$ is locally calm at $\bar{q}$, Assumption 3.3 holds, and (q^k) converges multi-step q-linearly to $\bar{q}$, then $(H(q^k))$ converges multi-step q-linearly.

Proof. **Proof of 1):** The asserted estimates follow from the local calmness of F, G and $F \circ G$ at $\bar{u}$, respectively, $\bar{q}$, established in Lemma 3.5. The asserted r-linear convergence is an obvious consequence of these estimates.
Proof of 2): The continuity of H implies convergence of $(H(q^k))$. Recalling from Lemma 3.7 that H is locally calm at $\bar{q}$ if this is true for $\hat{G}$, the r-linear convergence of $(H(q^k))$ follows as in 1). To show multi-step q-linear convergence of $(H(q^k))$ let $\beta \in [0,1)$ and $K \in \mathbb{N}$ be such that $\|q^{k+K} - \bar{q}\|_Q \leq \beta \|q^k - \bar{q}\|_Q$ for all k sufficiently large. Let L_H and κ_H be constants of local calmness, respectively, local metric subregularity of H at $\bar{q}$; the latter exists under Assumption 3.3, cf. Lemma 3.7. Let $J \in \mathbb{N}$ be such that $\beta^J L_H \kappa_H < 1$ and set $\hat{K} := JK$. Then we find for all k sufficiently large

$$\begin{aligned}\|H(q^{k+\hat{K}}) - H(\bar{q})\|_V \leq L_H \|q^{k+JK} - \bar{q}\|_Q &\leq L_H \beta^J \|q^k - \bar{q}\|_Q \\ &\leq \beta^J L_H \kappa_H \|H(q^k) - H(\bar{q})\|_V.\end{aligned}$$

□

Remark 4.8. Similar as in the previous proof it is possible to obtain (possibly multi-step) q-linear convergence of (u^k) from multi-step q-linear convergence of (q^k) if G is locally metrically subregular at $\bar{q}$. Yet, in many applications G is a projection or a proximal mapping, so local metrical subregularity of G at $\bar{q}$ and thus q-linear convergence of (u^k) cannot be expected. On the other hand, in these applications there often holds $L_G = 1$, so $\|u^k - \bar{u}\|_U \leq \|q^k - \bar{q}\|_Q$, cf. for instance [MR20, Theorem 2].

Boundedness of $(\|E_k\|)$ holds under weaker conditions than those of Theorem 4.4.

Lemma 4.9. *Let Assumption 3.1 hold. If Algorithm SQN generates (u^k) that satisfies $\sum_k \|u^k - \bar{u}\|_U^\eta < \infty$, then $(\|E_k\|_{\mathcal{L}(U,V)})$ is bounded.*

Proof. Since (u^k) converges to $\bar{u}$, there is $K \in \mathbb{N}_0$ such that Lemma 4.3 can be applied for all $k \geq K$. This yields for all $k \geq K$

$$\|E_{k+1}\|_{\mathcal{L}(U,V)} \leq \|E_K\|_{\mathcal{L}(U,V)} + 4C_F \sum_{j=K}^{k} \|u^j - \bar{u}\|_U^\eta \leq \|E_K\|_{\mathcal{L}(U,V)} + 4C_F \sum_{j=K}^{\infty} \|u^j - \bar{u}\|_U^\eta.$$

The right-hand side is independent of k, so the claim follows. □

4.2 Superlinear convergence

In the first part of this section we prepare and in the second part we prove the local superlinear convergence of Algorithm SQN. Readers that are only interested in the results may directly proceed to Section 4.2.2.

4.2.1 Preliminaries

The q-superlinear convergence of Algorithm SQN will be established by use of Dennis–Moré-type conditions, e.g., [DM74, Ulb11, Don12, AN18]. We provide two such conditions that will yield two different types of results on superlinear convergence. The first condition is in the spirit of [Ulb11, Theorem 3.18(a)].

Lemma 4.10. *Let Assumption 3.3 hold and let (q^k) be generated by Algorithm SQN. If (q^k) converges to $\bar{q}$ and satisfies both*

$$\lim_{k\to\infty} \frac{\|E_k M_k \bar{s}^k\|_V}{\|\bar{s}^k\|_Q} = 0 \qquad \text{and} \qquad \limsup_{k\to\infty} \frac{\|E_k M_k \bar{s}^{k+1}\|_V}{\|\bar{s}^{k+1}\|_Q} < \frac{1}{C_{\bar{M}^{-1}}}, \tag{DMT1}$$

then (q^k) converges q-superlinearly to $\bar{q}$.

Proof. Since Algorithm SQN has not terminated finitely, there holds $q^k \neq \bar{q}$ for all $k \in \mathbb{N}_0$. That is, $\bar{s}^k \neq 0$ for all $k \in \mathbb{N}_0$, so (DMT1) is sensible. We deduce from it the existence of a null sequence $(\alpha_k) \subset [0,\infty)$ and constants $c < 1$ and $K \in \mathbb{N}_0$ such that

$$C_{\bar{M}^{-1}} \|E_k M_k \bar{s}^k\|_V \leq \alpha_k \|\bar{s}^k\|_Q \qquad \text{and} \qquad C_{\bar{M}^{-1}} \|E_k M_k \bar{s}^{k+1}\|_V \leq c\, \|\bar{s}^{k+1}\|_Q \tag{8}$$

are satisfied for all $k \geq K$. Moreover, since $H = F \circ G + \hat{G}$ is semismooth at $\bar{q}$ by Lemma 3.5 wrt. $\partial H(q) = \{F'(\bar{u})M + \hat{M} : M \in \partial G(q),\, \hat{M} \in \partial \hat{G}(q)\}$ and since (q^k) converges to $\bar{q}$, there is a null sequence, wlog. (α_k), such that for every $k \in \mathbb{N}_0$

$$C_{\bar{M}^{-1}} \|H(q^k) - H(\bar{q}) - \bar{M}_k \bar{s}^k\|_V \leq \alpha_k \|\bar{s}^k\|_Q \tag{9}$$

is satisfied for all $\bar{M}_k \in \partial H(q^k)$. In particular, this is valid for $\bar{M}_k := F'(\bar{u})M_k + \hat{M}_k$, where $M_k \in \partial G(q^k)$ and $\hat{M}_k \in \hat{\partial} G(q^k)$ are the operators generated by Algorithm SQN in Line 4. Using $\tilde{M}_k s^k = -H(q^k)$ we compute for all K sufficiently large, wlog. $k \geq K$,

$$\|\bar{s}^{k+1}\|_Q = \|s^k + q^k - \bar{q}\|_Q = \left\|\bar{M}_k^{-1}\left[(\bar{M}_k - \tilde{M}_k)s^k + \tilde{M}_k s^k + \bar{M}_k(q^k - \bar{q})\right]\right\|_Q$$
$$\leq C_{\bar{M}^{-1}}\left[\|(\bar{M}_k - \tilde{M}_k)s^k\|_V + \|-H(q^k) + \bar{M}_k\bar{s}^k\|_V\right].$$

Since for all $k \in \mathbb{N}_0$ there holds

$$(\bar{M}_k - \tilde{M}_k)s^k = -E_k M_k s^k = -E_k M_k(\bar{s}^{k+1} - \bar{s}^k),$$

we deduce for all $k \geq K$

$$\|\bar{s}^{k+1}\|_Q \leq C_{\bar{M}^{-1}}\left[\|E_k M_k \bar{s}^{k+1}\|_V + \|E_k M_k \bar{s}^k\|_V + \|H(q^k) - H(\bar{q}) - \bar{M}_k\bar{s}^k\|_V\right].$$

Here, we have also used $H(\bar{q}) = 0$. By means of (8) and (9) this implies

$$(1-c)\,\|\bar{s}^{k+1}\|_Q \leq \alpha_k\|\bar{s}^k\|_Q + \alpha_k\|\bar{s}^k\|_Q$$

for all $k \geq K$. Since $c < 1$ is independent of k, we obtain the assertion. □

We now replace the lim sup condition of (DMT1) by uniform invertibility of $(\tilde{M}_k)$. Since uniform invertibility of $(\tilde{M}_k)$ is *assumed*, we can work with Assumption 3.1.

Lemma 4.11. *Let Assumption 3.1 hold and let (q^k) be generated by Algorithm SQN. If $(\|\tilde{M}_k^{-1}\|_{\mathcal{L}(V,Q)})$ is bounded and if (q^k) converges to $\bar{q}$ and satisfies*

$$\lim_{k\to\infty} \frac{\left\|E_k M_k \bar{s}^k\right\|_V}{\|\bar{s}^k\|_Q} = 0, \tag{DMT2}$$

then (q^k) converges q-superlinearly to $\bar{q}$.

Proof. As Algorithm SQN has not terminated finitely, we have $\bar{s}^k \neq 0$ for all $k \in \mathbb{N}_0$, so (DMT2) is sensible. Let $C_{\tilde{M}^{-1}} > 0$ denote an upper bound of $(\|\tilde{M}_k^{-1}\|_{\mathcal{L}(V,Q)})$. Using $\tilde{M}_k s^k = -H(q^k)$ and $H(\bar{q}) = 0$ we compute for all k sufficiently large

$$\|\bar{s}^{k+1}\|_Q = \|q^{k+1} - \bar{q}\|_Q = \left\|\tilde{M}_k^{-1}\left[\tilde{M}_k\bar{s}^k - H(q^k)\right]\right\|_Q$$
$$\leq C_{\tilde{M}^{-1}}\left[\|(\tilde{M}_k - \bar{M}_k)\bar{s}^k\|_V + \|-H(q^k) + H(\bar{q}) + \bar{M}_k\bar{s}^k\|_V\right],$$

where $\bar{M}_k := F'(\bar{u})M_k + \hat{M}_k \in \partial H(q^k)$. Since $(\tilde{M}_k - \bar{M}_k)\bar{s}^k = E_k M_k \bar{s}^k$, the claim follows by (DMT2) and the semismoothness of H at $\bar{q}$ wrt. ∂H. □

In the remainder of this section we establish that (DMT2) is fulfilled for Algorithm SQN. As an intermediate step we show that the classical Dennis–Moré condition is satisfied for the iterates $(u^k) = (G(q^k))$. To this end, we use the approach of [KS91] to first prove that the Dennis–Moré condition holds in a weak sense and then infer strong convergence from the compactness of E_0, cf. also [HK92].

Lemma 4.12. *Let Assumption 3.1 hold and let (u^k) be generated by Algorithm SQN with $0 < \liminf_{k\to\infty} \sigma_k \le \limsup_{k\to\infty} \sigma_k < 2$. If $\sum_k \|u^k - \bar u\|_U^\eta < \infty$ holds, then*

$$\forall l \in V^* : \qquad \lim_{k\to\infty} \left\langle l, E_k \hat s_u^k \right\rangle_{V^*,V} = 0.$$

Proof. Let $l \in V^*$ and denote $e^k := E_k^* l \in U$ for all $k \in \mathbb{N}_0$, where E_k^* is the adjoint of E_k. It is readily checked that (e^k) satisfies the recursion

$$\forall k \in \mathbb{N}_0 : \qquad e^{k+1} = e^k - \sigma_k \left(\hat s_u^k, e^k\right)_U \hat s_u^k + \epsilon_k,$$

where $\epsilon_k \in U$ is given by

$$\epsilon_k := \sigma_k \left\langle l, \frac{F(u^{k+1}) - F(u^k) - F'(\bar u) s_u^k}{\|s_u^k\|_U} \right\rangle_{V^*,V} \hat s_u^k.$$

Since for sufficiently large k there holds

$$\|\epsilon_k\|_U \le 2C_F \|l\|_{V^*} \max\{\|u^{k+1} - \bar u\|_U, \|u^k - \bar u\|_U\}^\eta,$$

we infer that $\sum_k \|\epsilon_k\|_U < \infty$. The claim now follows from [HK92, Lemma 2.1]. ☐

We can pass from weak convergence to strong convergence if E_0 is compact.

Lemma 4.13. *Let Assumption 3.1 hold and let (u^k) be generated by Algorithm SQN with $0 < \liminf_{k\to\infty} \sigma_k \le \limsup_{k\to\infty} \sigma_k < 2$. Let E_0 be compact. If (u^k) satisfies $\sum_k \|u^k - \bar u\|_U^\eta < \infty$, then*

$$\lim_{k\to\infty} \|E_k \hat s_u^k\|_V = 0. \tag{10}$$

Proof. The proof can be carried out similarly to [HK92, proof of Theorem 3.3]. ☐

The Dennis–Moré condition (10) implies convergence of the updates to zero.

Lemma 4.14. *Let Assumption 3.1 hold and let (u^k) be generated by Algorithm SQN. If (u^k) converges to $\bar u$, then the following implication is true:*

$$\lim_{k\to\infty} \|E_k \hat s_u^k\|_V = 0 \qquad \Longrightarrow \qquad \lim_{k\to\infty} \|B_{k+1} - B_k\|_{\mathcal{L}(U,V)} = 0.$$

Proof. We compute for all $k \in \mathbb{N}_0$ with $s_u^k \ne 0$ and any $h \in U$ with $\|h\|_U = 1$

$$\begin{aligned}(B_{k+1} - B_k)\, h &= \sigma_k \left(y^k - B_k s_u^k\right) \frac{\left(s_u^k, h\right)_U}{\|s_u^k\|_U^2} \\ &= \sigma_k \left(F(u^{k+1}) - F(u^k) - F'(\bar u) s_u^k\right) \frac{\left(s_u^k, h\right)_U}{\|s_u^k\|_U^2} - \sigma_k E_k s_u^k \frac{\left(s_u^k, h\right)_U}{\|s_u^k\|_U^2}.\end{aligned}$$

Taking norms we infer that for all k with $s_u^k \ne 0$ we have

$$\|B_{k+1} - B_k\|_V \le 2C_F \max\{\|u^{k+1} - \bar u\|_U, \|u^k - \bar u\|_U\}^\eta + 2\, \|E_k \hat s_u^k\|_V.$$

Since $B_{k+1} = B_k$ if $s_u^k = 0$, cf. Lines 10 to 11 in Algorithm SQN, this estimate holds trivially if $s_u^k = 0$. The assertion follows by letting k go to infinity. ☐

Remark 4.15. In [Mar00, Theorem 6] a finite-dimensional analogue of Lemma 4.14 can be found. [ABDL14, Proof of Theorem 4.8] contains an infinite-dimensional version under stronger assumptions than the ones we use.

The proof of (DMT2) will also use the following simple lemma.

Lemma 4.16. *Let $(a_k), (b_k) \subset \mathbb{R}$ be bounded from above with $\limsup_{k\to\infty} b_k \le 0$. Moreover, let $\beta \in [0,1)$ and suppose there exists $K \in \mathbb{N}$ such that*

$$0 \le a_k \le b_k + \beta a_{k+1} \tag{11}$$

is satisfied for all $k \ge K$. Then $\lim_{k\to\infty} a_k = 0$.

Proof. Let $\bar{a} := \limsup_{k\to\infty} a_k$. From (11) and $\limsup_{k\to\infty} b_k \le 0$ we infer

$$\bar{a} = \limsup_{k\to\infty} a_k \le \limsup_{k\to\infty} (b_k + \beta a_{k+1}) \le \limsup_{k\to\infty} b_k + \limsup_{k\to\infty} \beta a_{k+1} \le \beta\bar{a},$$

hence $\bar{a} \le 0$ since $\beta < 1$. Together with (11) this implies the claim. □

As main result of this section we establish (DMT2).

Theorem 4.17. *Let Assumption 3.1 hold and let (q^k) be generated by Algorithm SQN with $0 < \liminf_{k\to\infty} \sigma_k \le \limsup_{k\to\infty} \sigma_k < 2$. Let E_0 be compact. If (q^k) converges q-linearly to $\bar{q}$, then*

$$\lim_{k\to\infty} \frac{\left\| E_k M_k \bar{s}^k \right\|_V}{\|\bar{s}^k\|_Q} = 0. \tag{12}$$

Proof. From $q^k \neq \bar{q}$ for all $k \in \mathbb{N}_0$ it follows that $\bar{s}^k \neq 0$ for all $k \in \mathbb{N}_0$, so (12) is sensible. By q-linear convergence of (q^k) there exist $K \in \mathbb{N}_0$ and $\beta \in (0,1)$ such that $\|q^{k+1} - \bar{q}\|_Q \le \beta \|q^k - \bar{q}\|_Q$ for all $k \ge K$. By increasing K if need be, we can also assume that $\|q^k - \bar{q}\|_Q \le \delta_M$ for all $k \ge K$, so that we can use the calmness of G at $\bar{q}$ for all q^k with $k \ge K$, cf. Notation 3.6. Moreover, let us introduce the sequence

$$\alpha_k := (1+\beta)\, L_G \left\| E_k \hat{s}_u^k \right\|_V + \beta L_G \left\| B_k - B_{k+1} \right\|_{\mathcal{L}(U,V)}, \quad k \in \mathbb{N}_0,$$

and observe that $\lim_{k\to\infty} \alpha_k = 0$ because of Lemma 4.13 and Lemma 4.14 (the property $\sum_k \|u^k - \bar{u}\|_U^\eta < \infty$ holds because of $\sum_k \|q^k - \bar{q}\|_Q^\eta < \infty$). From Lemma 4.9 we further obtain that $C_B := \sup_{k\in\mathbb{N}_0} \|E_k\|_{\mathcal{L}(U,V)} < \infty$. To establish (12) we compute for all $k \in \mathbb{N}_0$

$$\begin{aligned} \frac{\left\| E_k M_k \bar{s}^k \right\|_V}{\|\bar{s}^k\|_Q} &\le \frac{\left\| E_k (M_k \bar{s}^k + G(\bar{q}) - G(q^k)) \right\|_V}{\|\bar{s}^k\|_Q} + \frac{\left\| E_k (G(q^k) - G(\bar{q})) \right\|_V}{\|\bar{s}^k\|_Q} \\ &\le C_B \frac{\left\| G(q^k) - G(\bar{q}) - M_k \bar{s}^k \right\|_U}{\|\bar{s}^k\|_Q} + \frac{\left\| E_k (G(q^k) - G(\bar{q})) \right\|_V}{\|\bar{s}^k\|_Q}. \end{aligned}$$

Due to the semismoothness of G at $\bar{q}$ it is sufficient to prove that the sequence

$$R_k := \frac{\left\| E_k (G(q^k) - G(\bar{q})) \right\|_V}{\|\bar{s}^k\|_Q}, \quad k \in \mathbb{N}_0,$$

converges to zero. We note that (R_k) is bounded from above since for all $k \geq K$

$$R_k = \frac{\left\|E_k(u^k - \bar{u})\right\|_V}{\|\bar{s}^k\|_Q} \leq \frac{C_B \left\|u^k - \bar{u}\right\|_U}{\|\bar{s}^k\|_Q} \leq \frac{C_B L_G \left\|q^k - \bar{q}\right\|_Q}{\|\bar{s}^k\|_Q} = C_B L_G.$$

Furthermore, for all $k \geq K$ we have

$$\|s_u^k\|_U \leq \|u^{k+1} - \bar{u}\|_U + \|u^k - \bar{u}\|_U \leq (\beta + 1) L_G \|q^k - \bar{q}\|_Q = (1 + \beta) L_G \|\bar{s}^k\|_Q,$$

whence

$$\frac{1}{\|\bar{s}^k\|_Q} \leq \frac{(1+\beta) L_G}{\|s_u^k\|_U}$$

for all $k \geq K$ with $s_u^k \neq 0$. Thus, using $E_k - E_{k+1} = B_k - B_{k+1}$ we obtain for these k

$$\begin{aligned} R_k &\leq \frac{\left\|E_k(u^k - u^{k+1})\right\|_V}{\|\bar{s}^k\|_Q} + \frac{\left\|E_{k+1}(u^{k+1} - \bar{u})\right\|_V}{\|\bar{s}^k\|_Q} + \frac{\left\|(E_k - E_{k+1})\left(u^{k+1} - \bar{u}\right)\right\|_V}{\|\bar{s}^k\|_Q} \\ &\leq (1+\beta) L_G \frac{\left\|E_k s_u^k\right\|_V}{\|s_u^k\|_U} + \frac{\left\|E_{k+1} \bar{s}_u^{k+1}\right\|_V}{\|\bar{s}^k\|_Q} + \beta L_G \frac{\|B_k - B_{k+1}\|_{\mathcal{L}(U,V)} \left\|q^k - \bar{q}\right\|_Q}{\|\bar{s}^k\|_Q} \\ &\leq \alpha_k + \beta \frac{\left\|E_{k+1} \bar{s}_u^{k+1}\right\|_V}{\|\bar{s}^{k+1}\|_Q}, \end{aligned}$$

where we have used the definition of α_k and the q-linear convergence of (q^k) with factor β. Therefore, we have established that for all $k \geq K$ with $s_u^k \neq 0$ there holds

$$0 \leq R_k \leq \alpha_k + \beta R_{k+1}. \tag{13}$$

In the case $s_u^k = 0$ we have $u^{k+1} = u^k$ and $B_{k+1} = B_k$, thus $E_{k+1} = E_k$ and $G(q^{k+1}) = G(q^k)$, so that the definition of R_k implies $0 \leq R_k \leq \beta R_{k+1}$ in this case. As $\alpha_k \geq 0$ we conclude that (13) holds for all $k \geq K$. Since (R_k) is bounded from above and since $\lim_{k\to\infty} \alpha_k = 0$, Lemma 4.16 implies $\lim_{k\to\infty} R_k = 0$. $\square$

Remark 4.18. The proof of Theorem 4.17 shows that the compactness of E_0 and $0 < \liminf_{k\to\infty} \sigma_k \leq \limsup_{k\to\infty} \sigma_k < 2$ are only needed to ensure that

$$\lim_{k\to\infty} \|E_k \hat{s}_u^k\|_V = 0. \tag{14}$$

4.2.2 Results

The following two theorems on q-superlinear convergence are the main results of this paper. We are not aware of other works that show superlinear convergence of semismooth quasi-Newton methods in infinite dimensions.

In the first result we *assume* that (q^k) exists and converges q-linearly to $\bar{q}$. This allows to show q-superlinear convergence without supposing that $\|E_0\|_{\mathcal{L}(U,V)}$ is small.

Theorem 4.19. *Let Assumption 3.1 hold and let (q^k) be generated by Algorithm SQN with $0 < \liminf_{k\to\infty} \sigma_k \le \limsup_{k\to\infty} \sigma_k < 2$. Let E_0 be compact. If (q^k) converges q-linearly to $\bar{q}$, then the convergence is q-superlinear if a) or b) is satisfied:*

a) The sequence $(\|\tilde{M}_k^{-1}\|_{\mathcal{L}(V,Q)})$ is bounded.

b) Assumption 3.3 is fulfilled and there holds

$$\limsup_{k\to\infty} \frac{\left\|E_k(M_k - M_{k+1})\bar{s}^{k+1}\right\|_V}{\|\bar{s}^{k+1}\|_Q} < \frac{1}{C_{\bar{M}^{-1}}}.$$

Proof. If a) is satisfied, then the q-superlinear convergence of (q^k) follows from Lemma 4.11, whose prerequisites are fulfilled due to a) and Theorem 4.17. For the q-superlinear convergence under condition b) we deduce from Lemma 4.10 and Theorem 4.17 that it suffices to establish

$$\limsup_{k\to\infty} \frac{\left\|E_k M_k \bar{s}^{k+1}\right\|_V}{\|\bar{s}^{k+1}\|_Q} < \frac{1}{C_{\bar{M}^{-1}}}. \tag{15}$$

Since $\|M_k\|_{\mathcal{L}(Q,U)} \le C_M$ for all sufficiently large k by Assumption 3.1, we have

$$\begin{aligned}\frac{\left\|E_k M_k \bar{s}^{k+1}\right\|_V}{\|\bar{s}^{k+1}\|_Q} &\le \frac{\left\|E_k(M_k - M_{k+1})\bar{s}^{k+1}\right\|_V}{\|\bar{s}^{k+1}\|_Q} + \frac{\left\|E_k M_{k+1}\bar{s}^{k+1}\right\|_V}{\|\bar{s}^{k+1}\|_Q} \\ &\le \frac{\left\|E_k(M_k - M_{k+1})\bar{s}^{k+1}\right\|_V}{\|\bar{s}^{k+1}\|_Q} + \frac{\left\|E_{k+1} M_{k+1}\bar{s}^{k+1}\right\|_V}{\|\bar{s}^{k+1}\|_Q} + C_M \|E_k - E_{k+1}\|_{\mathcal{L}(U,V)}\end{aligned}$$

for these k. The second term on the right-hand side converges to zero by Theorem 4.17. The third term on the right-hand side converges to zero by Lemma 4.14, whose prerequisites are fulfilled due to Lemma 4.13. Thus, (15) follows by b). □

In the second result we require Assumption 3.3 and that (q^0, B_0) be close to $(\bar{q}, F'(\bar{u}))$, but neither q-linear convergence of (q^k) nor a) or b) from Theorem 4.19. Instead, we show in the proof that all these properties follow from the assumptions.

Theorem 4.20. *Let Assumption 3.3 hold and let E_0 be compact. Then:*

1) There exist $\delta, \varepsilon > 0$ with the following property: For every initial $(q^0, B_0) \in Q \times \mathcal{L}(U, V)$ with $\|q^0 - \bar{q}\|_Q < \delta$ and $\|E_0\|_{\mathcal{L}(U,V)} < \varepsilon$, Algorithm SQN is well-defined and either terminates with output $q^* = \bar{q}$ or it generates (q^k) that converges q-superlinearly to $\bar{q}$ if $0 < \liminf_{k\to\infty} \sigma_k \le \limsup_{k\to\infty} \sigma_k < 2$.

2) If additionally F is Gâteaux differentiable in a neighborhood of $\bar{u}$ and the Gâteaux derivative is continuous at $\bar{u}$, then the condition $\|E_0\|_{\mathcal{L}(U,V)} < \varepsilon$ in 1) can be replaced by $\|B_0 - F'(u^0)\|_{\mathcal{L}(U,V)} < \varepsilon$. In particular, this replacement is possible if F is Hölder continuously Fréchet differentiable near $\bar{u}$.

Proof. **Proof of 1):** Theorem 4.4 provides $\delta, \varepsilon > 0$ such that well-definition, q-linear convergence of (q^k), and a) of Theorem 4.19 are ensured. The q-superlinear convergence of (q^k) thus follows from Theorem 4.19. (We remark that due to (4) and the uniform* boundedness of ∂G near $\bar{q}$, cf. Assumption 3.1, the numbers δ, ε can be chosen such that b) of Theorem 4.19 is satisfied, too.)
Proof of 2): Verbatim as in part 2) of Theorem 4.4. □

Remark 4.21.

1) If V is finite-dimensional, then E_0 is compact regardless of the choice of B_0.
2) Theorem 4.19 and Theorem 4.20 stay valid if (14) replaces the condition $0 < \liminf_{k\to\infty} \sigma_k \le \limsup_{k\to\infty} \sigma_k < 2$ and the compactness of E_0.
3) Theorem 4.19 and Theorem 4.20 allow to recover well-known superlinear convergence results for semismooth Newton methods ($F \equiv G \equiv 0$) under standard assumptions, and they also generalize superlinear convergence results for Broyden–like methods ($G = \mathrm{id}, \hat{G} \equiv 0$), e.g. [MT76, Theorem 3.4] and [Sac85, Section 6], for instance since F' is not required to be locally Lipschitz. Remarkably, no additional assumptions are needed to obtain superlinear convergence when the two methods are combined in the hybrid approach. Moreover, Theorem 4.19 seems to be the first result on superlinear convergence for semismooth quasi-Newton methods that does not require $\|E_0\|_{\mathcal{L}(U,V)}$ to be sufficiently small. Theorem 4.20 also generalizes finite-dimensional results for semismooth quasi-Newton methods such as [HS97, Theorem 4.1].
4) As remarked by a referee, the condition $\|E_0\|_{\mathcal{L}(U,V)} < \varepsilon$ appears quite restrictive from a practical point of view. To the best of our knowledge, however, all available local convergence results for semismooth quasi-Newton methods for equations use an assumption of this type, except for Theorem 4.19. In contrast, for *smooth* operator equations between separable Hilbert spaces there are the strong results of Griewank [Gri87] who considers a modified version of Broyden's method and proves convergence results in which the initial operator error needs to be small only in the essential norm. In finite dimensions this means that there is no restriction on this error. In our framework, however, F generally maps from a Hilbert space into a Banach space.

Regarding convergence of (u^k), $(F(u^k))$ and $(H(q^k))$, we note the following.

Lemma 4.22. *Let Assumption 3.1 hold and let (q^k) be generated by Algorithm SQN. If (q^k) converges r-superlinearly to $\bar{q}$, then:*

1) *(u^k) and $(F(u^k))$ converge r-superlinearly to $\bar{u}$, respectively, $F(\bar{u})$, and there hold $\|u^k - \bar{u}\|_U \le L_G\|q^k - \bar{q}\|_Q$, $\|F(u^k) - F(\bar{u})\|_V \le L_F\|u^k - \bar{u}\|_U$, and $\|F(u^k) - F(\bar{u})\|_V \le L_F L_G\|q^k - \bar{q}\|_Q$ for all k sufficiently large.*
2) *$(H(q^k))$ converges to zero. If $\hat{G}$ is locally calm at $\bar{q}$, then $(H(q^k))$ converges r-superlinearly. If $\hat{G}$ is locally calm at $\bar{q}$, Assumption 3.3 holds, and (q^k) converges q-superlinearly to $\bar{q}$, then $(H(q^k))$ converges q-superlinearly.*
3) *If there is $\kappa_G > 0$, respectively, $\kappa_F > 0$ such that $\|q^k - \bar{q}\|_Q \le \kappa_G\|G(q^k) - G(\bar{q})\|_U$, respectively, $\|q^k - \bar{q}\|_Q \le \kappa_F\|F(G(q^k)) - F(G(\bar{q}))\|_V$ is valid for all k sufficiently large, then (u^k), respectively, $(F(u^k))$ converges q-superlinearly.*

Proof. The proof of 1) and 2) is analogue to the proof of 1) and 2) in Lemma 4.7. The proof of 3) is also similar to 2) of Lemma 4.7. □

Remark 4.23. The inequalities for G and $F \circ G$ in 3) hold, for instance, under a uniform invertibility assumption similar to the one of Lemma 2.7 2).

5 Summary and outlook

We have presented an algorithm that combines a Broyden–like method with semismooth Newton methods. This hybrid approach enables the use of quasi-Newton methods in infinite-dimensional nonsmooth regimes and yields local superlinear convergence under standard assumption, as we proved in Theorem 4.19 and Theorem 4.20. In the complementary paper [MR20] we apply the approach to nonsmooth optimal control problems and show in an extensive numerical study that it is significantly faster in runtime than semismooth Newton methods.

Acknowledgments

- Armin Rund was supported by the Austrian Science Fund (FWF) under grant SFB F32-N18 (SFB "Mathematical Optimization and Applications in Biomedical Sciences").
- We would like to thank an anonymous referee who helped to improve the quality of the paper substantially.

References

[AB03] S. Amat and S. Busquier. A modified secant method for semismooth equations. *Appl. Math. Lett.*, 16(6):877–881, 2003. doi:10.1016/S0893-9659(03)90011-5.

[ABDL14] F.J. Aragón Artacho, A. Belyakov, A.L. Dontchev, and M. López. Local convergence of quasi-Newton methods under metric regularity. *Comput. Optim. Appl.*, 58(1):225–247, 2014. doi:10.1007/s10589-013-9615-y.

[AN18] S. Adly and H.V. Ngai. Quasi-Newton methods for solving nonsmooth equations: Generalized Dennis-Moré theorem and Broyden's update. *J. Convex Anal.*, 25(4):1075–1104, 2018.

[BDM73] C.G. Broyden, J.E.jun. Dennis, and J.J. More. On the local and superlinear convergence of quasi-Newton methods. *J. Inst. Math. Appl.*, 12:223–245, 1973. doi:10.1093/imamat/12.3.223.

[Che90] X. Chen. On the convergence of Broyden-like methods for nonlinear equations with nondifferentiable terms. *Ann. Inst. Stat. Math.*, 42(2):387–401, 1990. doi:10.1007/BF00050844.

[Che97] X. Chen. Superlinear convergence of smoothing quasi-Newton methods for nonsmooth equations. *J. Comput. Appl. Math.*, 80(1):105–126, 1997. doi:10.1016/S0377-0427(97)80133-1.

[Cia13] P.G. Ciarlet. *Linear and nonlinear functional analysis with applications*. SIAM, 2013.

[CNQ00] X. Chen, Z. Nashed, and L. Qi. Smoothing methods and semismooth methods for nondifferentiable operator equations. *SIAM J. Numer. Anal.*, 38(4):1200–1216, 2000. doi:10.1137/S0036142999356719.

[CQ94] X. Chen and L. Qi. A parameterized Newton method and a quasi-Newton method for nonsmooth equations. *Comput. Optim. Appl.*, 3(2):157–179, 1994. doi:10.1007/BF01300972.

[CQ10] J. Chen and L. Qi. Globally and superlinearly convergent inexact Newton-Krylov algorithms for solving nonsmooth equations. *Numer. Linear Algebra Appl.*, 17(1):155–174, 2010. doi:10.1002/nla.673.

[CY92] X. Chen and T. Yamamoto. On the convergence of some quasi-Newton methods for nonlinear equations with nondifferentiable operators. *Computing*, 49(1):87–94, 1992. doi:10.1007/BF02238652.

[Den71] J.E.jun. Dennis. On the convergence of Broyden's method for nonlinear systems of equations. *Math. Comput.*, 25:559–567, 1971. doi:10.2307/2005218.

[DM74] J.E.jun. Dennis and J.J. More. A characterization of superlinear convergence and its application to quasi-Newton methods. *Math. Comput.*, 28:549–560, 1974. doi:10.2307/2005926.

[Don12] A.L. Dontchev. Generalizations of the Dennis-Moré theorem. *SIAM J. Optim.*, 22(3):821–830, 2012. doi:10.1137/110833567.

[DR14] A.L. Dontchev and R.T. Rockafellar. *Implicit functions and solution mappings. A view from variational analysis*. Springer, 2nd updated edition, 2014. doi:10.1007/978-1-4939-1037-3.

[Gri87] A. Griewank. The local convergence of Broyden–like methods on lipschitzian problems in Hilbert spaces. *SIAM J. Numer. Anal.*, 24(3):684–705, 1987. doi:10.1137/0724045.

[HK92] D.M. Hwang and C.T. Kelley. Convergence of Broyden's method in Banach spaces. *SIAM J. Optim.*, 2(3):505–532, 1992. doi:10.1137/0802025.

[HS97] J. Han and D. Sun. Newton and quasi-Newton methods for normal maps with polyhedral sets. *J. Optim. Theory Appl.*, 94(3):659–676, 1997. doi:10.1023/A:1022653001160.

[IK92] C.-M. Ip and J. Kyparisis. Local convergence of quasi-Newton methods for B-differentiable equations. *Math. Program.*, 56(1 (A)):71–89, 1992. doi:10.1007/BF01580895.

[IK08] K. Ito and K. Kunisch. *Lagrange multiplier approach to variational problems and applications*. SIAM, 2008. doi:10.1137/1.9780898718614.

[IS14] A.F. Izmailov and M.V. Solodov. *Newton-type methods for optimization and variational problems*. Springer, 2014. doi:10.1007/978-3-319-04247-3.

[Kel95] C.T. Kelley. *Iterative methods for linear and nonlinear equations*. SIAM, 1995. doi:10.1137/1.9781611970944.

[Kru18] F. Kruse. Semismooth implicit functions. *J. Convex Anal.*, 25(2):595–622, 2018.

[KS91] C.T. Kelley and E.W. Sachs. A new proof of superlinear convergence for Broyden's method in Hilbert space. *SIAM J. Optim.*, 1(1):146–150, 1991. doi:10.1137/0801011.

[LF00] D. Li and M. Fukushima. A derivative-free line search and global convergence of Broyden-like method for nonlinear equations. *Optim. Methods Softw.*, 13(3):181–201, 2000. doi:10.1080/10556780008805782.

[LF01] D. Li and M. Fukushima. Globally convergent Broyden-like methods for semismooth equations and applications to VIP, NCP and MCP. *Ann. Oper. Res.*, 103:71–97, 2001. doi:10.1023/A:1012996232707.

[LG06] J. Liu and Y. Gao. Inexact-Newton method for solving operator equations in infinite-dimensional spaces. *J. Appl. Math. Comput.*, 22(1-2):351–360, 2006. doi:10.1007/BF02896484.

[Mar00] J.M. Martínez. Practical quasi-Newton methods for solving nonlinear systems. *J. Comput. Appl. Math.*, 124(1-2):97–121, 2000. doi:10.1016/S0377-0427(00)00434-9.

[MHMP13] P.Q. Muoi, D.N. Hào, P. Maass, and M. Pidcock. Semismooth Newton and quasi-Newton methods in weighted ℓ^1-regularization. *J. Inverse Ill-Posed Probl.*, 21(5):665–693, 2013. doi:10.1515/jip-2013-0031.

[Mor05] B.S. Mordukhovich. *Variational analysis and generalized differentiation. I: Basic theory*. Springer, 2005. doi:10.1007/3-540-31247-1.

[MQ95] J.M. Martínez and L. Qi. Inexact Newton methods for solving nonsmooth equations. *J. Comput. Appl. Math.*, 60(1-2):127–145, 1995. doi:10.1016/0377-0427(94)00088-I.

[MR20] F. Mannel and A. Rund. A hybrid semismooth quasi-Newton method for nonsmooth optimal control with PDEs. *Optim. Eng.*, Online First:1–39, 2020. doi:10.1007/s11081-020-09523-w.

[MT76] J.J. More and J.A. Trangenstein. On the global convergence of Broyden's method. *Math. Comput.*, 30:523–540, 1976. doi:10.2307/2005323.

[PQS98] F.A. Potra, L. Qi, and D. Sun. Secant methods for semismooth equations. *Numer. Math.*, 80(2):305–324, 1998. doi:10.1007/s002110050369.

[Qi97] L. Qi. On superlinear convergence of quasi-Newton methods for nonsmooth equations. *Oper. Res. Lett.*, 20(5):223–228, 1997. doi:10.1016/S0167-6377(97)00012-6.

[QJ97] 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(2):301–325, 1997. doi:10.1287/moor.22.2.301.

[Rob92] S.M. Robinson. Normal maps induced by linear transformations. *Math. Oper. Res.*, 17(3):691–714, 1992. doi:10.1287/moor.17.3.691.

[Sac85] E.W. Sachs. Convergence rates of quasi-Newton algorithms for some nonsmooth optimization problems. *SIAM J. Control Optim.*, 23:401–418, 1985. doi:10.1137/0323026.

[Sac86] E.W. Sachs. Broyden's method in Hilbert space. *Math. Program.*, 35:71–82, 1986. doi:10.1007/BF01589442.

[SH97] D. Sun and J. Han. Newton and quasi-Newton methods for a class of nonsmooth equations and related problems. *SIAM J. Optim.*, 7(2):463–480, 1997. doi:10.1137/S1052623494274970.

[Ulb11] M. Ulbrich. *Semismooth Newton methods for variational inequalities and constrained optimization problems in function spaces*. MOS-SIAM Series on Optimization. SIAM, 2011. doi:10.1137/1.9781611970692.

[WML11] X. Wang, C. Ma, and M. Li. A globally and superlinearly convergent quasi-Newton method for general box constrained variational inequalities without smoothing approximation. *J. Glob. Optim.*, 50(4):675–694, 2011. doi:10.1007/s10898-010-9629-2.