**SpezialForschungsBereich F 32**

Karl–Franzens Universität Graz
Technische Universität Graz
Medizinische Universität Graz

# A Hybrid Semismooth Quasi-Newton Method Part 1: Theory

F. Mannel A. Rund

SFB-Report No. 2018-005 Sep 2018

A–8010 GRAZ, HEINRICHSTRASSE 36, AUSTRIA

Supported by the
Austrian Science Fund (FWF)

SFB sponsors:

- **Austrian Science Fund** (FWF)
- **University of Graz**
- **Graz University of Technology**
- **Medical University of Graz**
- **Government of Styria**
- **City of Graz**

Das Land Steiermark

Stadt GRAZ Wissenschaft

# A HYBRID SEMISMOOTH QUASI-NEWTON METHOD PART 1: THEORY

FLORIAN MANNEL* AND ARMIN RUND*

**Abstract.** We present an algorithm for the solution of structured nonsmooth operator equations in Banach spaces. Specifically, we consider equations that are composed of a smooth and a semismooth part. The key feature of the new method is the following hybrid approach: While the semismooth part of the equation is linearized in the same way as in semismooth Newton methods, the smooth part is handled by a quasi-Newton method, namely a generalized Broyden's method. The resulting algorithm can be regarded as a semismooth Newton-type method that exploits the additional smoothness of the equation to reduce the computational costs.

We study the local convergence properties of the new method in an infinite-dimensional setting and find that it is q-superlinearly convergent under reasonable assumptions. For the variable on which the quasi-Newton method acts we establish convergence not only with respect to the underlying Hilbert space norm, but also with respect to Banach space norms. As a by-product, the convergence analysis shows that replacing Broyden's method by the simplified Newton method results in a hybrid algorithm that is locally q-linearly convergent.

This paper is the first of two devoted to the new hybrid approach. Here, we focus on developing the convergence theory, while applications are presented in the second, complementary paper. Among others, we demonstrate there that the new approach is capable of solving large-scale real-world problems from PDE-constrained optimal control in a fraction of the time that semismooth Newton methods take. We conclude from these two papers that the novel method has favorable theoretical properties, is widely applicable, and yields highly competitive numerical schemes.

**Key words.** Semismooth Newton methods, semismooth Newton-type methods, generalized Broyden's method, quasi-Newton methods, superlinear convergence, linear convergence, nonsmooth operator equations, semismooth equations, semismoothness, metric subregularity, calmness

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

## 1. Introduction.

Quasi-Newton and semismooth Newton methods are arguably among the most successful numerical tools for solving nonlinear equations. In this paper we combine these two cornerstones of modern computational mathematics to form a superlinearly convergent algorithm for the solution of structured nonsmooth operator equations in Banach spaces that is substantially faster than semismooth Newton methods.

Throughout this work we consider equations of the form

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

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; the precise setting is contained in section 3. We stress that 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 PDE-constrained optimal control; we detail this in the complementary paper [34]. Under mild assumptions the mapping

$$H : Q \to V, \qquad H(q) := F(G(q)) + \hat{G}(q)$$

is semismooth, which can be used to establish local q-superlinear convergence of semismooth Newton methods applied to (P); these methods require the evaluation

*University of Graz, Institute of Mathematics and Scientific Computing, Heinrichstr. 36, 8010 Graz, Austria (florian.mannel@uni-graz.at, armin.rund@uni-graz.at).

of $F'$. In this work we develop a semismooth Newton-type method that converges locally q-superlinearly, but does not require the evaluation of $F'$. In comparison to semismooth Newton methods this reduces the computational costs quite significantly whenever the evaluation of $F'$ is expensive. Such is the case, for instance, in PDE-constrained optimal control.

The key idea of the novel method is to replace the operator $F'$ in semismooth Newton methods by a quasi-Newton approximation, while the generalized derivatives for $G$ and $\hat{G}$ are left unchanged. The resulting algorithm therefore combines a quasi-Newton method with a semismooth Newton method and can be regarded as a hybrid approach. We stress that the choice to apply the quasi-Newton method only to the smooth part $F$, but not to the entire semismooth 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 very strong assumptions and do not result in widely applicable numerical algorithms.

Let us shortly comment on these two issues. Concerning the first point we mention that there are simple examples in one real variable 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; we provide details on this subject matter in subsection 4.2.1. For this reason several authors have developed modified quasi-Newton updates when dealing with nonsmooth problems. To the best of our knowledge, however, all but one of these methods are designed for finite-dimensional spaces, and it is unclear to us whether they can be extended to infinite-dimensional spaces while retaining fast local convergence. The single exception is presented in [1], 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 very particular problems, cf. [1, Remark 4]. In contrast, the hybrid method that we develop in this paper converges superlinearly and applies to the plethora of practically relevant problems that amount to solving an equation of the form (P).

Let us set our work in perspective with the literature. To begin with, we remark that this paper is concerned with *local* convergence analysis. Therefore, we ignore the large body of literature devoted to *semilocal* convergence of quasi-Newton methods.

*Quasi-Newton methods in finite-dimensional spaces.* Quasi-Newton methods for finite-dimensional problems were introduced in [13, 14, 20, 5]. By now they are well-established and introductory material on these methods is part of many textbooks, e.g., [39, 29, 28]. As starting points for a deeper treatment of quasi-Newton methods for finite-dimensional *smooth* equations we refer to the survey articles [17, 35] and the more recent historical note [22]. From the many contributions on modified and unmodified quasi-Newton methods for *nonsmooth* equations in finite dimensions, e.g., [26, 10, 41, 7, 32, 47, 23, 40, 3, 8, 33, 42, 50, 11], we highlight the papers [11, 50, 42, 23, 47] because the methods presented in these papers contain the idea to apply a quasi-Newton method to the smooth part of a structured nonsmooth equation. Among these five papers, [23] is the closest to our approach. In fact, the algorithm of [23] can be viewed as a special case of the hybrid method that we develop here. To point out differences let us stress that already in finite dimensions our approach includes important problem classes that are not covered by any of these references. A prominent example of such a class is given by $\ell^1$-regularized optimization problems, e.g., the Lasso problem [48]. These problems occur, for instance, in statistics, machine learning and image reconstruction, and have been investigated intensively during the last decade. We emphasize that the numerical study in the complementary paper [34] contains $L^1$-regularized optimal control problems, which fall into this category.

*Quasi-Newton methods in infinite-dimensional spaces.* The superlinear convergence of quasi-Newton methods for *smooth* equations in infinite-dimensional spaces has been established in [44, 21, 25]. A simplified proof of the results in [44] is obtained in [30]. We point out that the convergence analysis of the hybrid method allows to recover the results from [44, 30] under weaker assumptions. Recently, quasi-Newton methods have been investigated for generalized equations with smooth base mapping, cf. [4]. For *nonsmooth* equations in infinite-dimensional spaces, references are few. In fact, we found only the two recent papers [38, 1], one of which we already discussed above. The other one, [38], contains the idea to use a quasi-Newton method on the smooth part of a structured nonsmooth equation. However, the structure of the equation is different from and less general than the one we use, and only linear convergence is established. For completeness we mention that the results on linear convergence in [21] also allow a certain degree of nonsmoothness, cf. [21, (1.12) and (1.13)]. In conclusion, the approach to study quasi-Newton methods in infinite-dimensional spaces for structured nonsmooth equations of the form (P) is new.

The main contributions of this and the complementary paper [34] are as follows. Above all, we provide an approach through which quasi-Newton methods can be used effectively in the context of nonsmooth operator equations in infinite-dimensional spaces. In particular, the hybrid method presented in this paper

- converges locally q-superlinearly under mild assumptions, cf. Theorem 4.18;
- is widely applicable, for instance to variational inequalities and structured nonsmooth optimization problems, cf. [34];
- allows to derive practical algorithms that are significantly faster than their semismooth Newton counterparts, cf. [34].

This paper is organized as follows. In section 2 we fix the notation, introduce necessary concepts, and establish results that are required for the convergence analysis. In section 3 we provide the precise problem setting and present the hybrid method along with the assumptions that we use for its convergence analysis. In addition, we establish fundamental properties of the mapping $H$. Section 4 is concerned with the convergence analysis. Besides proving local linear and superlinear convergence of the hybrid method, this section includes a short synopsis on superlinear convergence of quasi-Newton methods in smooth and nonsmooth settings, where we work out, in particular, that superlinear convergence cannot be ensured on semismooth equations. The paper ends with a summary and an outlook on the complementary paper [34].

## 2. Preliminaries.

This section contains definitions and results that are required for the convergence analysis. We begin with the following convention.

Throughout section 2, $X, Y, Z$ are normed linear spaces and $U$ is a Hilbert space.

We point out that $X$, $Y$ and $Z$ are not assumed to be complete in section 2, except in Lemma 2.13 and Lemma 2.18, where this is explicitly stated.

### 2.1. Notation.

We employ the following notation.

For the natural numbers we use $\mathbb{N} := \{1, 2, 3, \dots\}$ and $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$. For a mapping $f : A \to B$ between sets $A$ and $B$ we write $f \equiv b$, where $b \in B$, to indicate that $f(a) = b$ for all $a \in A$. Furthermore, all linear spaces are real linear spaces. In normed linear spaces $X$ and $Y$ we denote

- $\mathrm{id} : X \to X$ for the identity mapping, i.e., $\mathrm{id}(x) = x$ for all $x \in X$;
- $\mathbb{B}_\delta(\bar{x}) := \{x \in X : \|x - \bar{x}\|_X < \delta\}$ for $\delta > 0$ and $\bar{x} \in X$;
- $\mathbb{B}_\delta'(\bar{x}) := \mathbb{B}_\delta(\bar{x}) \setminus \{\bar{x}\}$ for $\delta > 0$ and $\bar{x} \in X$;
- $\overline{\mathbb{B}}_\delta(\bar{x}) := \{x \in X : \|x - \bar{x}\|_X \le \delta\}$ for $\delta > 0$ and $\bar{x} \in X$;
- $\mathcal{L}(X, Y)$ for the continuous linear functionals from $X$ to $Y$;
- $A + B := \{a + b : a \in A, b \in B\}$ for nonempty sets $A, B \subset X$.

Moreover, in the Hilbert space $U$ we write
- $(v, w)_U$ for the scalar product of $v, w \in U$;
- $(v, \cdot)_U$ for the linear operator $w \mapsto (v, w)_U$ from $U$ to $\mathbb{R}$.

For clarity let us also mention that
- phrases such as "Let the sequence $(q^k)$ be generated by Algorithm 1." are understood in the sense that infinitely many iterates have been generated, i.e., Algorithm 1 has not terminated after finitely many steps.

**2.2. Semismoothness and strict differentiability with radial rate.** We use the following definition of semismoothness in this paper.

Definition 2.1. *Let $\bar{x} \in X$ and let $Q : X \to Y$ be continuous in an open neighborhood of $\bar{x} \in X$. Moreover, let $\partial Q : X \rightrightarrows \mathcal{L}(X, Y)$ satisfy $\partial Q(x) \neq \emptyset$ for all $x \in X$. We say that $Q$ is* semismooth at $\bar{x}$ with respect to $\partial Q$ *iff there holds*

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

*The set-valued mapping $\partial Q : X \rightrightarrows \mathcal{L}(X, Y)$ is called a* generalized derivative of $Q$*. For $x \in X$ every $M \in \partial Q(x)$ is called a* generalized differential of $Q$ at $x$*.*

*Remark* 2.2. This definition can be found in [49, Definition 3.1] and is analyzed thoroughly in [49]. It includes, as special cases, Newton differentiability, cf. [27, Definition 8.10], semismoothness based on slant derivatives, cf. [9], as well as most finite-dimensional notions of semismoothness. Let us, however, mention that under additional technicalities it would be possible to extend the results of this paper to the more general notion of semismoothness introduced in [31, Definition 3]; in particular, the continuity assumption on $Q$ is not necessary to develop the theory in this paper.

We will use the following differentiability concept that is inspired by [41].

Definition 2.3. *We call $Q : X \to Y$* strictly differentiable at $\bar{x} \in X$ with radial rate $\eta > 0$ *iff there exists a bounded linear operator $Q'(\bar{x}) \in \mathcal{L}(X, Y)$ and constants $C_Q, \delta_Q > 0$ such that*

$$\|Q(y) - Q(x) - Q'(\bar{x})(y - x)\|_Y \le C_Q \|y - x\|_X \max\big\{\|y - \bar{x}\|_X, \|x - \bar{x}\|_X\big\}^\eta$$

*is satisfied for all $x, y \in \mathbb{B}_{\delta_Q}(\bar{x})$. We also say that $Q$ is* $\eta$-strictly differentiable at $\bar{x}$*.*

We collect elementary facts about $\eta$-strictly differentiable functions. For the concept of strict differentiability we refer to [36, Definition 1.13] and [19, Section 1.4].

Lemma 2.4. *1) If $Q$ satisfies Definition 2.3, then it is continuous in $\mathbb{B}_{\delta_Q}(\bar{x})$, (strongly) semismooth at $\bar{x}$ wrt. $\partial Q(x) := \{Q'(\bar{x})\}$ for all $x \in X$ (with rate $\eta$), strictly differentiable at $\bar{x}$, and Fréchet differentiable at $\bar{x}$.*

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

*Proof.* All claims follow from the respective definitions. ∎

**2.3. Local calmness and local metric subregularity, uniform' boundedness and uniform' invertibility.**

DEFINITION 2.5. *Let $D \subset X$, $Q : D \to Y$ and $\bar{x} \in D$.*

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

$$\|Q(x) - Q(\bar{x})\|_Y \le L_Q \|x - \bar{x}\|_X .$$

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

$$\|x - \bar{x}\|_X \le \kappa_Q \|Q(x) - Q(\bar{x})\|_Y .$$

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

*Remark* 2.6. Clearly, (locally) Lipschitz continuous functions are (locally) calm.

To derive sufficient conditions for local calmness and local metric subregularity that are easy to verify, we use the following concepts.

DEFINITION 2.7. *Let $Q : X \to Y$ be semismooth at $\bar{x} \in X$.*

*1) We say that $\partial Q$ has a* uniformly' (uniformly) bounded selection near $\bar{x}$ *iff there are $C_M, \delta_M > 0$ such that for every $x \in \mathbb{B}'_{\delta_M}(\bar{x})$ ($x \in \mathbb{B}_{\delta_M}(\bar{x})$) there exists $M \in \partial Q(x)$ with $\|M\|_{\mathcal{L}(X,Y)} \le C_M$.*

*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 $M \in \partial Q(x)$, then $\partial Q$ is called* uniformly' (uniformly) bounded near $\bar{x}$.

*3) We say that $\partial Q$ 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 exists an invertible $M \in \partial Q(x)$ with $\|M^{-1}\|_{\mathcal{L}(Y,X)} \le C_{M^{-1}}$.*

*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 $M \in \partial Q(x)$, then $\partial Q$ is called* uniformly' (uniformly) invertible near $\bar{x}$.

LEMMA 2.8. *Let $Q : X \to Y$ be semismooth at $\bar{x} \in X$. Then:*

*1) $Q$ is locally calm at $\bar{x}$ if $\partial Q$ admits a uniformly' bounded selection near $\bar{x}$.*

*2) $Q$ is locally metrically subregular at $\bar{x}$ if $\partial Q$ admits a uniformly' invertible selection near $\bar{x}$.*

*Proof.* **Proof of 1):** By the selection property there are $C_M, \delta_M > 0$ such that for every $x \in \mathbb{B}'_{\delta_M}(\bar{x})$ there is at least one $M = M_x \in \partial Q(x)$ with $\|M_x\|_{\mathcal{L}(X,Y)} \le C_M$. The semismoothness implies that there exists $\hat{\delta} \in (0, \delta_M]$ such that

$$\sup_{M \in \partial Q(x)} \|Q(x) - Q(\bar{x}) - M(x - \bar{x})\|_Y \le \|x - \bar{x}\|_X$$

is satisfied for all $x \in \mathbb{B}_{\hat{\delta}}(\bar{x})$. Hence, by the reverse triangle inequality, we have

$$\|Q(x) - Q(\bar{x})\|_Y - \|M(x - \bar{x})\|_Y \le \|Q(x) - Q(\bar{x}) - M(x - \bar{x})\|_Y \le \|x - \bar{x}\|_X$$

for all these $x$ and all $M \in \partial Q(x)$. Choosing for every $x \in \mathbb{B}'_{\hat{\delta}}(\bar{x})$ the corresponding $M = M_x \in \partial Q(x)$ from the prerequisite provides for all $x \in \mathbb{B}'_{\hat{\delta}}(\bar{x})$ the inequality

$$\|Q(x) - Q(\bar{x})\|_Y \le \|M(x - \bar{x})\|_Y + \|x - \bar{x}\|_X \le (C_M + 1) \|x - \bar{x}\|_X .$$

This establishes that $Q$ is locally calm at $x$ since there is nothing to prove for $x = \bar{x}$.
**Proof of 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 $M = M_x \in \partial Q(x)$ that is invertible with $\|M_x^{-1}\|_{\mathcal{L}(Y,X)} \le C_{M^{-1}}$. The semismoothness implies that there exists $\hat{\delta} \in (0, \delta_{M^{-1}}]$ such that

$$\sup_{M \in \partial Q(x)} \|Q(x) - Q(\bar{x}) - M(x - \bar{x})\|_Y \le \frac{1}{2C_{M^{-1}}} \|x - \bar{x}\|_X$$

is satisfied for all $x \in \mathbb{B}_{\hat{\delta}}(\bar{x})$. Hence, by the reverse triangle inequality, we have

$$\|M(x - \bar{x})\|_Y - \|Q(x) - Q(\bar{x})\|_Y \le \|Q(x) - Q(\bar{x}) - M(x - \bar{x})\|_Y \le \frac{1}{2C_{M^{-1}}} \|x - \bar{x}\|_X$$

for all these $x$ and all $M \in \partial Q(x)$. This yields

$$\|M(x - \bar{x})\|_Y - \frac{1}{2C_{M^{-1}}} \|x - \bar{x}\|_X \le \|Q(x) - Q(\bar{x})\|_Y \,. \tag{2.1}$$

Choosing for every $x \in \mathbb{B}'_{\hat{\delta}}(\bar{x})$ the corresponding $M = M_x \in \partial Q(x)$ from the prerequisite provides for all $x \in \mathbb{B}'_{\hat{\delta}}(\bar{x})$ the inequality

$$\|x - \bar{x}\|_X = \left\|M_x^{-1} M_x(x - \bar{x})\right\|_X \le C_{M^{-1}} \|M_x(x - \bar{x})\|_Y \,,$$

which implies for all these $x$ that

$$\frac{1}{C_{M^{-1}}} \|x - \bar{x}\|_X - \frac{1}{2C_{M^{-1}}} \|x - \bar{x}\|_X \le \|M_x(x - \bar{x})\|_Y - \frac{1}{2C_{M^{-1}}} \|x - \bar{x}\|_X \,.$$

The assertion now follows by use of (2.1) since there is nothing to prove for $x = \bar{x}$. ∎

Corollary 2.9. *Let $Q : X \to Y$ be $\eta$-strictly differentiable at $\bar{x} \in X$. Then:*
*1) $Q$ is locally calm at $\bar{x}$.*
*2) $Q$ is locally metrically subregular at $\bar{x}$ if $Q'(\bar{x})$ is invertible.*

*Proof.* Apply Lemma 2.8 ($Q$ is semismooth at $\bar{x}$ wrt. $\partial Q(x) := \{Q'(\bar{x})\}$, $x \in X$). ∎

**2.4. A chain rule for semismooth mappings.** The following chain rule for semismooth mappings generalizes [49, Proposition 3.8].

Lemma 2.10. *Let $Q : X \to Y$ be semismooth at $\bar{x} \in X$ and locally calm at $\bar{x}$. Let $R : Y \to Z$ be semismooth at $\bar{y} := Q(\bar{x})$ with uniformly bounded generalized derivative near $\bar{y}$. Then $S : X \to Z$, $S := R(Q(x))$ is semismooth at $\bar{x}$ with respect to*

$$\partial S : X \rightrightarrows \mathcal{L}(X, Z), \qquad \partial S(x) := \big\{M_R \circ M_Q : \; M_R \in \partial R(Q(x)),\, M_Q \in \partial Q(x)\big\}.$$

*Also, $S$ is locally calm at $\bar{x}$.*

*Proof.* The continuity of $S$ in an open neighborhood of $\bar{x}$ is clear. To prove the remainder property of $S$ at $\bar{x}$, we define $\varphi : X \to Y$, $\varphi(h) := Q(\bar{x} + h) - Q(\bar{x}) = Q(\bar{x} + h) - \bar{y}$. For all $h \in X$ and all $M_S = M_R \circ M_Q \in \partial S(\bar{x} + h)$ there holds

$$\begin{aligned} &S(\bar{x} + h) - S(\bar{x}) - M_S h \\ &\quad = \big[R(\bar{y} + \varphi(h)) - R(\bar{y}) - M_R \varphi(h)\big] + \big[M_R\big(Q(\bar{x} + h) - Q(\bar{x}) - M_Q h\big)\big]. \end{aligned} \tag{2.2}$$

The continuity of $Q$ at $\bar{x}$ implies $\varphi(h) \to 0$ for $h \to 0$, hence

$$\sup_{M_R \in \partial R(\bar{y} + \varphi(h))} \|R(\bar{y} + \varphi(h)) - R(\bar{y}) - M_R \varphi(h)\|_Z = o(\|\varphi(h)\|_Y) \quad \text{for } h \to 0$$

by the semismoothness of $R$ at $\bar{y}$. Invoking the local calmness of $Q$ at $\bar{x}$ yields

$$\sup_{M_R \in \partial R(\bar{y}+\varphi(h))} \|R(\bar{y}+\varphi(h)) - R(\bar{y}) - M_R\varphi(h)\|_Z = o(\|h\|_X) \quad \text{for } h \to 0. \tag{2.3}$$

Since $\partial R$ is uniformly bounded near $\bar{y} = Q(\bar{x})$ and since $Q$ is continuous at $\bar{x}$, there are $C_R, \delta_R > 0$ such that $\|M_R\|_{\mathcal{L}(Y,Z)} \le C_R$ for every $M_R \in \partial R(Q(\bar{x}+h))$ and all $h \in \mathbb{B}_{\delta_R}(0)$. Hence,

$$\sup_{M_S \in \partial S(\bar{x}+h)} \big\|M_R\big(Q(\bar{x}+h) - Q(\bar{x}) - M_Q h\big)\big\|_Z = o(\|h\|_X) \quad \text{for } h \to 0 \tag{2.4}$$

by the semismoothness of $Q$ at $\bar{x}$. The semismoothness of $S$ follows by combining (2.2)–(2.4) in an obvious way. Lemma 2.8 implies that $R$ is locally calm at $\bar{y}$. Since $Q$ and $R$ are locally calm at $\bar{x}$ and $\bar{y}$, respectively, $S = R \circ Q$ is locally calm at $\bar{x}$. ◻

**2.5. Rates of convergence.** We recall well-known concepts that measure the speed of convergence of sequences and establish connections between them.

Definition 2.11. *Let $(x^k) \subset X$.*

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

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

*is satisfied for all $k \ge K$. Instead of "1-step q-linear convergence" we speak of q-linear convergence.*

2) *We say that $(x^k)$* converges q-superlinearly *to $\bar{x} \in X$ iff there is a null sequence $(\varepsilon_k) \subset [0,\infty)$ such that*

$$\|x^{k+1} - \bar{x}\|_X \le \varepsilon_k\|x^k - \bar{x}\|_X$$

*is satisfied for all $k \in \mathbb{N}_0$.*

3) *We say that $(x^k)$* converges r-linearly (r-superlinearly) *to $\bar{x} \in X$ iff there is a q-linearly (q-superlinearly) convergent null sequence $(\alpha_k) \subset [0,\infty)$ such that*

$$\|x^k - \bar{x}\|_X \le \alpha_k$$

*is satisfied for all $k \in \mathbb{N}_0$.*

Local calmness and local metric subregularity are key to deduce from the convergence rate of a sequence $(x^k)$ the convergence rate of $(Q(x^k))$, and vice versa.

Lemma 2.12. *Let $(x^k) \subset X$, $\bar{x} \in X$, and $Q : X \to Y$.*

1) *If $(x^k)$ converges q-linearly (q-superlinearly) to $\bar{x}$, then $(Q(x^k))$ converges r-linearly (r-superlinearly) to $Q(\bar{x})$ provided that $Q$ is locally calm at $\bar{x}$.*
2) *If $(x^k)$ converges q-linearly with rate $\beta \in [0,1)$ to $\bar{x}$, then $(Q(x^k))$ converges $j$-step q-linearly to $Q(\bar{x})$ provided that $Q$ is locally calm at $\bar{x}$ and locally metrically subregular at $\bar{x}$ and provided that $j \in \mathbb{N}$ is such that $\hat{\beta} := \beta^j L_Q \kappa_Q < 1$ is satisfied. The rate of convergence of $(Q(x^k))$ is $\hat{\beta}$.*
3) *If $(x^k)$ converges q-superlinearly to $\bar{x}$, then $(Q(x^k))$ converges q-superlinearly to $Q(\bar{x})$ provided that $Q$ is locally calm at $\bar{x}$ and locally metrically subregular at $\bar{x}$.*

*4) If $(Q(x^k))$ converges q-linearly (q-superlinearly) to $Q(\bar{x})$, then $(x^k)$ converges r-linearly (r-superlinearly) to $\bar{x}$ provided that $Q$ is locally metrically subregular at $\bar{x}$.*

*5) If $(Q(x^k))$ converges q-linearly with rate $\hat{\beta} \in [0,1)$ to $Q(\bar{x})$, then $(x^k)$ converges $j$-step q-linearly to $\bar{x}$ provided that $Q$ is locally calm at $\bar{x}$ and locally metrically subregular at $\bar{x}$ and provided that $j \in \mathbb{N}$ is such that $\beta := \hat{\beta}^j L_Q \kappa_Q < 1$ is satisfied. The rate of convergence of $(x^k)$ is $\beta$.*

*6) If $(Q(x^k))$ converges q-superlinearly to $Q(\bar{x})$, then $(x^k)$ converges q-superlinearly to $\bar{x}$ provided that $Q$ is locally metrically subregular at $\bar{x}$ and locally calm at $\bar{x}$.*

*Proof.* We prove 1)-3). The remaining claims can be established similarly.

**Proof of 1):** The local calmness of $Q$ at $\bar{x}$ implies

$$\|Q(x^k) - Q(\bar{x})\|_Y \le L_Q \|x^k - \bar{x}\|_X$$

for all $k$ sufficiently large. Since the sequence $(L_Q\|x^k - \bar{x}\|_X)$ converges q-linearly (q-superlinearly) to zero, we obtain r-linear (r-superlinear) convergence of $(Q(x^k))$.

**Proof of 2):** Let $j \in \mathbb{N}$ be such that $\beta^j L_Q \kappa_Q < 1$ is satisfied. The local calmness of $Q$ at $\bar{x}$ and the local metric subregularity at $\bar{x}$ imply for all $k$ sufficiently large

$$\|Q(x^{k+j}) - Q(\bar{x})\|_Y \le L_Q\|x^{k+j} - \bar{x}\|_X \le L_Q \beta^j \|x^k - \bar{x}\|_X \le \beta^j L_Q \kappa_Q \|Q(x^k) - Q(\bar{x})\|_Y,$$

which establishes the assertion.

**Proof of 3):** The local calmness of $Q$ at $\bar{x}$ in combination with the local metric subregularity at $\bar{x}$ implies

$$\|Q(x^{k+1}) - Q(\bar{x})\|_Y \le L_Q\|x^{k+1} - \bar{x}\|_X \le L_Q \varepsilon_k \|x^k - \bar{x}\|_X \le L_Q \kappa_Q \varepsilon_k \|Q(x^k) - Q(\bar{x})\|_Y$$

for all $k$ sufficiently large and a null sequence $(\varepsilon_k) \subset [0, \infty)$. This proves the claim. ◻

**2.6. Results on linear operators.** We use the following consequence of Banach's lemma on the invertibility of perturbed linear operators.

LEMMA 2.13. *Let $X$ and $Y$ be Banach spaces and $A, B \in \mathcal{L}(X,Y)$. If $A$ is invertible with $\rho := \|I - A^{-1}B\|_{\mathcal{L}(X,X)} < 1$, then $B$ is invertible, too, and there holds*

$$\left\|B^{-1}\right\|_{\mathcal{L}(Y,X)} \le \frac{\left\|A^{-1}\right\|_{\mathcal{L}(Y,X)}}{1-\rho}.$$

*In particular, if $A$ is invertible with $\|A^{-1}\|_{\mathcal{L}(Y,X)} \le C$ for some $C > 0$ and $B$ is such that $\|A - B\|_{\mathcal{L}(X,Y)} \le 1/(2C)$, then $B$ is invertible with*

$$\left\|B^{-1}\right\|_{\mathcal{L}(Y,X)} \le 2C.$$

*Proof.* Let $T := I - A^{-1}B$. By assumption, $T \in \mathcal{L}(X,X)$ and $\|T\|_{\mathcal{L}(X,X)} = \rho < 1$. Thus, we can apply Banach's lemma, cf., e.g., [12, Theorem 3.6-2]. This yields the invertibility of $I - T = A^{-1}B$ and $\|(A^{-1}B)^{-1}\|_{\mathcal{L}(X,X)} \le \frac{1}{1-\rho}$. Since $A$ is invertible, $B = AA^{-1}B$ is invertible, too, and there holds

$$\left\|B^{-1}\right\|_{\mathcal{L}(Y,X)} \le \left\|(A^{-1}B)^{-1}\right\|_{\mathcal{L}(X,X)} \left\|A^{-1}\right\|_{\mathcal{L}(Y,X)} \le \frac{\left\|A^{-1}\right\|_{\mathcal{L}(Y,X)}}{1-\rho}.$$

The second assertion follows from the first since

$$\rho := \left\|I - A^{-1}B\right\|_{\mathcal{L}(X,X)} \leq \left\|A^{-1}\right\|_{\mathcal{L}(Y,X)} \|A - B\|_{\mathcal{L}(X,Y)} \leq \frac{1}{2} < 1$$

under the stated assumptions, which implies

$$\left\|B^{-1}\right\|_{\mathcal{L}(Y,X)} \leq \frac{\left\|A^{-1}\right\|_{\mathcal{L}(Y,X)}}{1-\rho} \leq \frac{\left\|A^{-1}\right\|_{\mathcal{L}(Y,X)}}{1-\frac{1}{2}} \leq 2C.$$

□

The next lemma helps to analyze the generalized Broyden update.

LEMMA 2.14. *Let $s \in U$ with $\|s\|_U \leq \sqrt{2}$. Then the operator*

$$A_s : U \to U, \qquad A_s := I - s\,(s, \cdot)_U$$

*is linear and continuous with $\|A_s\|_{\mathcal{L}(U,U)} \leq 1$.*

*Proof.* Clearly, $A_s$ is linear and continuous. Moreover, it is easy to check that $A_s$ is self-adjoint, i.e., $(A_s v, w)_U = (v, A_s w)_U$ for all $v, w \in U$. Therefore, we have

$$\begin{aligned}\|A_s\|_{\mathcal{L}(U,U)} = \sup_{\|v\|_U=1} \left|(A_s v, v)_U\right| &= \sup_{\|v\|_U=1} \left|(v,v)_U - (s,v)_U^2\right| \\ &= \sup_{\|v\|_U=1} \left|1 - (s,v)_U^2\right| \leq 1,\end{aligned}$$

where we used that $0 \leq (s,v)_U^2 \leq \|s\|_U^2 \|v\|_U^2 \leq 2$ for all $v \in U$ with $\|v\|_U = 1$. □

The following auxiliary result is needed to prove weak superlinear convergence.

LEMMA 2.15. *For $e, s \in U$ there holds*

$$\sup_{\|h\|_U \leq 1} \left|\left(e, h - s\,(s,h)_U\right)_U\right| = \sqrt{\|e\|_U^2 - (2 - \|s\|_U^2)\,(e,s)_U^2}.$$

*Proof.* For any $h \in U$ we have

$$\left(e, h - s\,(s,h)_U\right)_U = (e,h)_U - (e,s)_U\,(s,h)_U = \left(e - s\,(e,s)_U\,, h\right)_U. \tag{2.5}$$

Defining $v := e - s(e,s)_U$ we observe that the assertion is trivially fulfilled if $v = 0$. If $v \neq 0$, then (2.5) implies that the supremum in question is attained for $h = \frac{v}{\|v\|_U}$ and has the value $\|v\|_U = \sqrt{(v,v)_U}$. Thus, the claim follows from

$$(v,v)_U = (e,e)_U - 2\,(e,s)_U^2 + (e,s)_U^2\,(s,s)_U = \|e\|_U^2 - \left(2 - \|s\|_U^2\right)(e,s)_U^2$$

after taking square roots (which is possible, as we observe). □

**2.7. Bounded deterioration.** We provide an error estimate for the linear operator that is used in place of $F'$ in Broyden's method. It is a crucial ingredient for the proof of liner convergence of the new hybrid method.

LEMMA 2.16. *Let $F : U \to Y$ be $\eta$-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 $B \in \mathcal{L}(U,Y)$ 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*

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

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

$$\|F(u_+) - F(u) - F'(\bar u)(u_+ - u)\|_Y \le C_F \|u_+ - u\|_U \max\big\{\|u_+ - \bar u\|_U, \|u - \bar u\|_U\big\}^\eta. \tag{2.6}$$

Let $u, u_+ \in \mathbb{B}_{\delta_F}(\bar u)$, $\sigma \in [0,2]$, and $B \in \mathcal{L}(U,Y)$. Set $s := u_+ - u$. For $s = 0$ the assertion is trivial. Hence, we assume $s \neq 0$, i.e., $u_+ \neq u$, for the rest of the proof. Defining $\hat s := \frac{s}{\|s\|_U}$ we compute

$$\begin{aligned} &B_+ - F'(\bar u) \\ &\quad = [B - F'(\bar u)] + \sigma\,[F(u_+) - F(u) - F'(\bar u)s]\,\frac{(s,\cdot)_U}{\|s\|_U^2} - \sigma\,[B - F'(\bar u)]\,\frac{s\,(s,\cdot)_U}{\|s\|_U^2} \\ &\quad = [B - F'(\bar u)]\left[I - \sqrt{\sigma}\hat s\left(\sqrt{\sigma}\hat s,\cdot\right)_U\right] + \sigma\,[F(u_+) - F(u) - F'(\bar u)s]\,\frac{(s,\cdot)_U}{\|s\|_U^2}. \end{aligned}$$

Using Lemma 2.14, $\|(s,\cdot)_U\|_{\mathcal{L}(U,\mathbb{R})} \le \|s\|_U$, and (2.6) we deduce that

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

*Remark* 2.17.
1) For $s \neq 0$ and $\sigma = 1$ the definition of $B_+$ in Lemma 2.16 is Broyden's update of $B$ at $u_+$. Note, however, that $u_+ = u + s$ does not have to be the quasi-Newton iterate succeeding $u$, i.e., the relation $Bs = -F(u)$ is not required. When we apply Lemma 2.16, we will need this generality.
2) Roughly speaking, Lemma 2.16 shows that the approximation quality of $B$ is kept to a certain extent by $B_+$. For the classical Broyden's method this property of the Broyden update is referred to as *bounded deterioration.*

**2.8. Relative compactness of infinitely many rank-one updates.** We reformulate and simplify an argument contained in [30, Proof of Theorem 2.5].

LEMMA 2.18. *Let $X$ be a Banach space, let $C_W > 0$, and let $(w^k)_{k\in\mathbb{N}_0} \subset X$ satisfy $\|w^k\|_X \le C_W$ for all $k \in \mathbb{N}_0$. Let $\beta \in [0,1)$ and define for all $k \in \mathbb{N}_0$ the sets*

$$I_k := \big\{\alpha\beta^k w^k : \ \alpha \in [-1,1]\big\} \subset X \quad \text{and} \quad I := \bigcup_{l=0}^{\infty}\left\{\sum_{k=0}^{l} x^k : \ x^k \in I_k \ \ \forall k\right\} \subset X.$$

*Then $I$ is relatively compact, i.e., its closure is compact.*

*Proof.* We recall, e.g. from [2, 4.7 (5)], that in a Banach space a set is relatively compact iff it is totally bounded. Therefore, it is enough to show that for every $\varepsilon > 0$ there are $N \in \mathbb{N}_0$ and $v^0, v^1, \ldots, v^N \in X$ such that $I \subset \cup_{j=0}^N \mathbb{B}_\varepsilon(v^j)$ is satisfied. Let $\varepsilon > 0$ and set $\hat\varepsilon := \frac{\varepsilon}{2}$. Choose $M \in \mathbb{N}_0$ so large that $\frac{\beta^M}{1-\beta} < \frac{\hat\varepsilon}{C_W}$. For every sequence $(x^k)_{k\ge M}$ with $x^k \in I_k$ for all $k \ge M$, we have

$$\left\|\sum_{k=M}^{l} x^k\right\|_X \le C_W \sum_{k=M}^{l} \beta^k \le C_W \frac{\beta^M}{1-\beta} < \hat\varepsilon$$

for all $l \geq M$. Hence,

$$\bigcup_{l=M}^{\infty} \left\{ \sum_{k=M}^{l} x^k : \;\; x^k \in I_k \;\; \forall k \geq M \right\} \subset \mathbb{B}_{\hat{\varepsilon}}(0). \tag{2.7}$$

Since every $I_k$ is sequentially compact, thus compact, it is easy to see that the set

$$\bigcup_{l=0}^{M-1} \left\{ \sum_{k=0}^{l} x^k : \;\; x^k \in I_k \;\; \forall k < M \right\} = \left\{ \sum_{k=0}^{M-1} x^k : \;\; x^k \in I_k \;\; \forall k < M \right\}$$

is relatively compact (in fact compact, as it is closed). This implies that it is totally bounded, thus there are $N \in \mathbb{N}_0$ and $v^0, v^1, \ldots, v^N \in X$ such that

$$\bigcup_{l=0}^{M-1} \left\{ \sum_{k=0}^{l} x^k : \;\; x^k \in I_k \;\; \forall k < M \right\} \subset \bigcup_{j=0}^{N} \mathbb{B}_{\hat{\varepsilon}}(v^j).$$

By the triangle inequality this yields in combination with (2.7) that

$$I = \bigcup_{l=0}^{\infty} \left\{ \sum_{k=0}^{l} x^k : \;\; x^k \in I_k \;\; \forall k \right\} \subset \bigcup_{j=0}^{N} \mathbb{B}_{\varepsilon}(v^j).$$

∎

**2.9. A result on null sequences.** The following lemma is used in the proof of q-superlinear convergence of the hybrid method.

LEMMA 2.19. *Let $(a_k), (b_k) \subset \mathbb{R}$ be bounded from above with $\limsup_{k\to\infty} b_k \leq 0$. Moreover, let $\beta < 1$ and suppose there exists $K \in \mathbb{N}$ such that*

$$0 \leq a_k \leq b_k + \beta a_{k+1} \tag{2.8}$$

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

*Proof.* Let $\bar{a} := \limsup_{k\to\infty} a_k$. From (2.8) and $\limsup_{k\to\infty} b_k \leq 0$ we infer

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

hence $\bar{a} \leq 0$ since $\beta < 1$. It also follows from (2.8) that $\liminf_{k\to\infty} a_k \geq 0$. Together, we have $0 \leq \liminf_{k\to\infty} a_k \leq \limsup_{k\to\infty} a_k = \bar{a} \leq 0$, which implies the assertion. ∎

## 3. Problem setting, hybrid method, main assumptions, consequences.

**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 $U$,
- mappings $G : Q \to U$, $F : U \to V$ and $\hat{G} : Q \to V$,
- $H : Q \to V, \quad H(q) := F(G(q)) + \hat{G}(q)$,

our goal is to

$$\text{find } \bar{q} \in Q \text{ such that } H(\bar{q}) = 0. \tag{P}$$

For concrete problem classes that are covered by (P) we refer to the complementary paper [34], where we also explicate why we use the three different spaces $Q$, $U$, and $V$.

Moreover, note that (P) includes smooth equations (take $Q = U$, $G = \mathrm{id}$, $\hat{G} \equiv 0$) and semismooth equations ($F \equiv 0$, $G \equiv 0$) as special instances. Finally, let us emphasize that the assumption that $F$, $G$ and $\hat{G}$ are defined on the whole space is only made to simplify the presentation — the results established in this paper remain true if these mappings are defined only locally around $\bar{u} := G(\bar{q})$ and $\bar{q}$, respectively.

We now present the hybrid semismooth quasi-Newton method. Roughly speaking, we propose to apply a semismooth Newton method to the nonsmooth parts $G$ and $\hat{G}$ of $H$, while the smooth part $F$ is dealt with using a quasi-Newton method. Specifically, the algorithm reads as follows.

**Algorithm 1:** Hybrid semismooth quasi-Newton method

**Input:** $q_0 \in Q, \quad B_0 \in \mathcal{L}(U, V), \quad 0 \le \sigma_{\min} \le \sigma_{\max} \le 2.$

1 Let $u^0 := G(q^0)$.
2 **for** $k = 0, 1, 2, \dots$ **do**
3 **if** $H(q^k) = 0$ **then** let $\bar{q} := q^k$; STOP.
4 Choose $M_k \in \partial G(q^k)$ and $\hat{M}_k \in \partial \hat{G}(q^k)$.
5 Let $\tilde{M}_k := B_k M_k + \hat{M}_k$.
6 Solve $\tilde{M}_k s^k = -H(q^k)$ 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 $\sigma_k \in [\sigma_{\min}, \sigma_{\max}]$.
10 **if** $s_u^k \neq 0$ **then** let $B_{k+1} := B_k + \sigma_k (y^k - B_k s_u^k) \frac{(s_u^k, \cdot)_U}{\|s_u^k\|_U^2}$;
11 **else** let $B_{k+1} := B_k$.
12 **end**

**Output:** $\bar{q}$

Extensions of Algorithm 1 that converge globally are examined in [34].

Before we dive into the details of Algorithm 1, let us point out that it contains several well-known methods: In fact, it contains Broyden's method for $F$ (take $Q = U$, $G = \mathrm{id}$, $\hat{G} \equiv 0$, $(\sigma_k) \equiv 1$), the simplified Newton method for $F$ ($Q = U$, $G = \mathrm{id}$, $\hat{G} \equiv 0$, $(\sigma_k) \equiv 0$), and the semismooth Newton method for $\hat{G}$ ($F \equiv 0$, $G \equiv 0$, $(\sigma_k)$ arbitrary). Except for the simplified Newton method these methods are all locally q-superlinearly convergent under suitable assumptions, and it is therefore reasonable to expect that superlinear convergence of Algorithm 1 will require all these assumptions to be satisfied. Remarkably, no additional assumptions are needed, and this is true for a very flexible choice of $(\sigma_k)$. Also, this clarifies why we can recover results for these methods from results on the hybrid method. In this respect, let us also mention that if $F$, $G$ and $\hat{G}$ are affine and the choice $B_0 = F'(u_0) = F'(\bar{u})$ is made, then Algorithm 1 converges in one iteration; if only $F$ is affine and the choice $B_0 = F'(u_0) = F'(\bar{u})$ is made, then it holds by induction that

$$y^k - B_k s_u^k = F(u^{k+1}) - F(u^k) - F'(\bar{u}) s_u^k = F'(\bar{u}) s_u^k - F'(\bar{u}) s_u^k = 0.$$

This implies $B_k = F'(u^k)$ for all $k \in \mathbb{N}_0$, whence Algorithm 1 coincides with a semismooth Newton method for $H$.

We now make some more specific comments on Algorithm 1. To begin with, we emphasize that we will provide assumptions that guarantee the unique solvability of the linear system in Line 6, thereby ensuring that the algorithm is well-defined. Next we point out that Algorithm 1 involves a generalized version of the update formula of Broyden's method. The classical Broyden update is obtained for $\sigma_k = 1$ in Line 10. The idea to generalize the update through the parameter $\sigma_k \in [\sigma_{\min}, \sigma_{\max}]$ is well-known; cf., e.g., [37], [43, Section 6] and [32, Algorithm 1]. We will find that for $\sigma_{\min} = 0$ and $\sigma_{\max} = 2$ the hybrid method is locally q-linearly convergent, while the combination of $\sigma_{\min} > 0$ and $\sigma_{\max} = 2$ ensures q-superlinear convergence. Note that the result on linear convergence includes, for instance, the choice $(\sigma_k) \equiv 0$, i.e., $B_k = B_0$ during the entire algorithm. That is, if we replace Broyden's method by the simplified Newton method, we obtain a hybrid algorithm that converges q-linearly. Finally, the Sherman–Morrison formula implies that if $B_k$ is invertible, then $B_{k+1}$ is invertible, too, except for at most one value of $\sigma_k$ that can be computed explicitly.

**3.2. Main assumptions and consequences.** The convergence analysis rests upon the following assumption. We recall the notions of uniform' boundedness, uniform' invertibility and $\eta$-strict differentiability from Definition 2.7 and Definition 2.3.

*Assumption* 3.1. Suppose that
- there is $\bar{q} \in Q$ with $H(\bar{q}) = 0$;
- $G$ is semismooth at $\bar{q}$;
- $\partial G$ is uniformly' bounded near $\bar{q}$ with constants $C_M, \delta_M > 0$;
- $\hat{G}$ is semismooth at $\bar{q}$;
- $F$ is $\eta$-strictly differentiable at $\bar{u} := G(\bar{q})$ with constants $C_F, \delta_F > 0$;
- the generalized derivative $\partial H : Q \rightrightarrows \mathcal{L}(Q, V)$ given by

$$\partial H(q) := \left\{ F'(\bar{u}) \circ M + \hat{M} : \;\; M \in \partial G(q),\, \hat{M} \in \partial \hat{G}(q) \right\} \tag{3.1}$$

  is uniformly' invertible near $\bar{q}$ with constants $C_{\bar{M}^{-1}}, \delta_{\bar{M}^{-1}} > 0$.

*Remark* 3.2.
1) If $F$ is Hölder continuously Fréchet differentiable in a neighborhood of $\bar{u}$, then it is also $\eta$-strictly differentiable at $\bar{u}$, cf. Lemma 2.4 2).
2) Since the generalized derivative $\partial H$ contains the unknown point $\bar{u}$, we stress that it is not used in the algorithm, but for theoretical purposes only.
3) If $q \mapsto \partial H(q)$ is upper semicontinuous at $\bar{q}$, then $\partial H$ 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 $\|\bar{M}^{-1}\|_{\mathcal{L}(V,Q)} \leq C_{\bar{M}^{-1}}$. This can be argued using Lemma 2.13.

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

LEMMA 3.3. *Let Assumption 3.1 hold. Then $H$ is semismooth at $\bar{q}$ with respect to $\partial H$ defined in (3.1), and $F$, $G$, and $F \circ G$ are locally calm at $\bar{u}$ and $\bar{q}$, respectively.*

*Proof.* The uniform' boundedness of $\partial G$ near $\bar{q}$ implies by Lemma 2.8, part 1), that $G$ is locally calm at $\bar{q}$. Next we demonstrate the semismoothness of $H$. To this end, note that $\hat{G}$ is semismooth at $\bar{q}$ by assumption, hence it is sufficient to show that $F \circ G$ is semismooth at $\bar{q}$. Since $F$ is $\eta$-strictly differentiable at $\bar{u}$ by Assumption 3.1, we infer from Lemma 2.4 1) that $F$ is semismooth at $\bar{u}$ wrt. the generalized derivative $\partial F(u) := \{F'(\bar{u})\}$, $u \in U$. Evidently, $\partial F$ is uniformly bounded near $\bar{u}$. Moreover, $G$ is semismooth at $\bar{q}$ by assumption and locally calm at $\bar{q}$ as we have already established. Therefore, the semismoothness of $F \circ G$ at $\bar{q}$ follows from Lemma 2.10. In addition, this lemma yields that $F \circ G$ is locally calm at $\bar{q}$. Since $\partial F$ is uniformly bounded near $\bar{u}$, Lemma 2.8, part 1), yields that $F$ is locally calm at $\bar{u}$. ∎

*Remark* 3.4. The uniform' invertibility of $\partial H$ is not required for Lemma 3.3.

We have just established that $F$ and $G$ are locally calm at $\bar{u}$, respectively, $\bar{q}$, if Assumption 3.1 holds. In particular, we can assume without loss of generality that the constant $\delta_M$ in Assumption 3.1 is so small that $G$ is locally calm at $\bar{q}$ in $\mathbb{B}_{\delta_M}(\bar{q})$.

DEFINITION 3.5. *Under Assumption 3.1, let $L_F > 0$ be the constant of local calmness of $F$ at $\bar{u}$ and $L_G > 0$ be 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 3.6. *Let Assumption 3.1 hold. Then $H$ is locally metrically subregular at $\bar{q}$. If, in addition, $\hat{G}$ is locally calm at $\bar{q}$, then $H$ is also locally calm at $\bar{q}$.*

*Proof.* By Assumption 3.1, $\partial H$ is uniformly' invertible near $\bar{q}$. Hence, it follows from Lemma 2.8, part 2), that $H$ is locally metrically subregular at $\bar{q}$.

Since $F \circ G$ is locally calm at $\bar{q}$ by Lemma 3.3, the local calmness of $\hat{G}$ at $\bar{q}$ yields that $H = F \circ G + \hat{G}$ is locally calm at $\bar{q}$. ◻

COROLLARY 3.7. *Let Assumption 3.1 hold and suppose that $\hat{G}$ is locally calm at $\bar{q}$. Then there are constants $L_H, \kappa_H, \delta_H > 0$ such that*

$$\kappa_H \|q - \bar{q}\|_Q \leq \|H(q) - H(\bar{q})\|_V \leq L_H \|q - \bar{q}\|_Q \tag{3.2}$$

*holds for all $q \in \mathbb{B}_{\delta_H}(\bar{q})$. Moreover, $(q^k)$ converges to $\bar{q}$ (q-superlinearly) if and only if $(H(q^k))$ converges to $H(\bar{q})$ (q-superlinearly).*

*Proof.* By Lemma 3.6, $H$ is locally metrically subregular at $\bar{q}$ and locally calm at $\bar{q}$. The inequalities (3.2) follow from the definition of local calmness and local metric subregularity. Also, (3.2) evidently implies the claim for the case of mere convergence. The case of q-superlinear convergence follows from Lemma 2.12, parts 3) and 6). ◻

*Remark* 3.8. Corollary 3.7 and particularly (3.2) provide a certain justification for the numerical experience that it is often successful to use $\|H(\cdot)\|_V$ as a merit function. Note that via $F \equiv 0$ and $G \equiv 0$, Corollary 3.7 also applies to the semismooth case

## 4. Convergence analysis.

In this section we establish local convergence results for Algorithm 1. Throughout, we will use the notation introduced in section 3 and the following definition.

DEFINITION 4.1. *Assume that Algorithm 1 has generated iterates $q^{k+1}, q^k$ and $u^{k+1}, u^k$, along with an operator $B_k$. Then we denote*

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

*as well as*

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

*and finally*

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

### 4.1. Linear convergence.

We establish local q-linear convergence of Algorithm 1. The key for proving this is the bounded deterioration property discussed in Lemma 2.16. We recall that $C_M$ and $C_{\bar{M}^{-1}}$ are introduced in Assumption 3.1.

THEOREM 4.2. *Let Assumption 3.1 hold and let $\beta \in (0, 1)$. Then:*

*1) There exist $\delta, \varepsilon > 0$ such that for every pair of starting values $(q^0, B_0) \in Q \times \mathcal{L}(U,V)$ with $\|q^0 - \bar{q}\|_Q < \delta$ and $\|E_0\|_{\mathcal{L}(U,V)} < \varepsilon$, Algorithm 1 is well-defined and either terminates after finitely many iterations or generates a sequence of iterates $(q^k)$ that satisfies for all $k \in \mathbb{N}_0$ the inequalities*

$$\|q^{k+1} - \bar{q}\|_Q \leq \beta \|q^k - \bar{q}\|_Q \quad \text{and} \quad \|E_k\|_{\mathcal{L}(U,V)} \leq \frac{\beta}{4C_M C_{\bar{M}^{-1}}}. \tag{4.1}$$

*2) If, in addition to Assumption 3.1, $F$ 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 in a neighborhood of $\bar{u}$.*

*Proof.* **Proof of 1):** The proof of this part requires some preparations. To begin with, we mention that in the following, $\tilde{M}_k$, $M_k$ and $\hat{M}_k$ are the quantities from Algorithm 1. By shrinking $\beta$ if necessary, we can assume without loss of generality that $\beta \leq \frac{1}{2^{1/\eta}}$, where $\eta > 0$ is the constant from Assumption 3.1. Thus, we have $\hat{\beta} := \beta^\eta \leq \frac{1}{2}$. Since $H$ is semismooth at $\bar{q}$, cf. Lemma 3.3, there is $\delta_H > 0$ such that

$$\sup_{\bar{M} \in \partial H(q)} \left\| H(q) - H(\bar{q}) - \bar{M}(q - \bar{q}) \right\|_V \leq \frac{\beta}{4C_{\bar{M}^{-1}}} \|q - \bar{q}\|_Q \tag{4.2}$$

holds for all $q \in \mathbb{B}_{\delta_H}(\bar{q})$. Also, we recall from Definition 3.5 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 $\eta$-strictly differentiable at $\bar{u}$ with constants $C_F, \delta_F > 0$. The definitions $\breve{C} := 8C_M C_{\bar{M}^{-1}}$ and $\mathring{C} := 2C_F L_G^\eta > 0$ conclude the preparations. We now claim that the values

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

ensure that Algorithm 1 is well-defined and either terminates after finitely many iterations or generates a sequence of iterates $(q^k)$ that satisfies (4.1). We prove this by induction. 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. For the induction argument we consider Line 2 to Line 11 in Algorithm 1 with an arbitrary $k \in \mathbb{N}_0$. We will show that either the algorithm terminates in Line 3 or the operator $\tilde{M}_k$ is boundedly invertible—which implies that $q^{k+1}$ and $E_{k+1}$ exist—and there hold

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

Due to $\|q^0 - \bar{q}\|_Q < \delta$, $\|E_0\|_{\mathcal{L}(U,V)} < \varepsilon$, and the definition of $\delta$ and $\varepsilon$, this yields (4.1).

If the algorithm terminates in Line 3, then there is nothing to prove. Otherwise, we have $H(q^k) \neq 0$, hence $q^k \neq \bar{q}$. We will first demonstrate that $\tilde{M}_k$ is boundedly invertible. Apparently, there hold both $\bar{M}_k := F'(\bar{u}) M_k + \hat{M}_k \in \partial H(q^k)$ and

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

The induction assumption provides $\|q^{j+1} - \bar{q}\|_Q \leq \beta \|q^j - \bar{q}\|_Q$ for all $0 \leq j \leq k-1$, thus $q^k \in \mathbb{B}_\delta'(\bar{q})$. In particular, we have $q^k \in \mathbb{B}_{\delta_M}'(\bar{q})$, which implies $\|M_k\|_{\mathcal{L}(Q,U)} \leq C_M$

by Assumption 3.1. The induction assumption yields $\|E_k\|_{\mathcal{L}(U,V)} < \varepsilon + 2\mathring{C}\delta^\eta \le \frac{2\beta}{\breve{C}}$. Together, we infer that

$$\|\bar{M}_k - \tilde{M}_k\|_{\mathcal{L}(Q,V)} \le \frac{\beta}{4C_{\bar{M}^{-1}}} \le \frac{1}{2C_{\bar{M}^{-1}}}. \tag{4.3}$$

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 3.1. This proves that $\tilde{M}_k$ is invertible with $\|\tilde{M}_k^{-1}\|_{\mathcal{L}(V,Q)} \le 2C_{\bar{M}^{-1}}$, cf. Lemma 2.13.

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 (4.2). Using $H(\bar{q}) = 0$ and (4.2) we compute

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

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 $q^0, q^1, \ldots, q^{k+1}$. To complete the induction it is left to show the validity of

$$\|E_{k+1}\|_{\mathcal{L}(U,V)} \le \|E_0\|_{\mathcal{L}(U,V)} + \mathring{C}(2 - \hat{\beta}^k)\|q^0 - \bar{q}\|_Q^\eta. \tag{4.4}$$

If $s_u^k = 0$, then $E_{k+1} = E_k$, hence (4.4) follows from the induction assumption in this case. 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 2.16. This produces

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

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

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

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 that

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

thereby concluding the induction as well as the proof of 1).

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

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

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 shrinking $\hat\varepsilon$ if necessary, we can assume that $\hat\varepsilon \le \frac{\varepsilon}{2}$ is satisfied. Recall from Definition 3.5 that $G$ is locally calm at $\bar q$ in $\mathbb{B}_{\delta_M}(\bar q)$ with constant $L_G > 0$. Defining $\hat\delta := \min\{\delta, \delta_M, \frac{\hat\varepsilon}{L_G}\}$ we infer that $\|q^0-\bar q\|_Q < \hat\delta$ yields $\|q^0-\bar q\|_Q < \delta$. It remains to establish $\|B_0-F'(\bar u)\|_{\mathcal{L}(U,V)} < \varepsilon$. From $\|q^0-\bar q\|_Q < \hat\delta \le \frac{\hat\varepsilon}{L_G}$ it follows that $\|u^0-\bar u\|_U = \|G(q^0)-G(\bar q)\|_Q \le L_G\|q^0-\bar q\|_Q < \hat\varepsilon$, where the calmness of $G$ at $\bar q$ applies due to $\|q^0-\bar q\|_Q < \hat\delta \le \delta_M$. 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 infer that

$$\|B_0-F'(\bar u)\|_{\mathcal{L}(U,V)} \le \|B_0-F'(u^0)\|_{\mathcal{L}(U,V)} + \|F'(u^0)-F(\bar u)\|_{\mathcal{L}(U,V)} < \varepsilon. \qquad \square$$

*Remark* 4.3.

1) After rather small modifications in the previous proof it follows that Theorem 4.2 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 (4.2).
2) It is possible to include a step length $\alpha_k > 0$ in Algorithm 1, 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 4.2 is preserved if $\alpha_k \in [\alpha_-, \alpha_+]$ holds for all $k$. The proof requires in addition to Assumption 3.1 only that $\hat G$ is locally calm at $\bar q$. However, our main focus is on superlinear convergence, which requires $\alpha_k \to 1$ anyway, so we choose to work with unit step length for simplicity.
3) In the special case $Q = U$, $G = \mathrm{id}$, $\hat G \equiv 0$, and $(\sigma_k) \equiv 1$, Algorithm 1 reduces to the classical Broyden's method for $F$. It is noteworthy that in this case Assumption 3.1 is identical to the assumptions needed for q-linear convergence of the classical Broyden's method in infinite-dimensional spaces. Moreover, since $(q^k) \equiv (u^k)$ in this case, the standard result on local q-linear convergence of Broyden's method is reproduced, compare [15, Theorem 5].

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

COROLLARY 4.4. *Let Assumption 3.1 hold and suppose that Algorithm 1 generates a sequence $(q^k)$ that converges q-linearly to $\bar q$. Then $(u^k)$ and $(F(u^k))$ converge r-linearly to $\bar u$ and $F(\bar u)$, respectively, 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. If $\hat G$ is calm at $\bar q$, then $(H(q^k))$ converges both r-linearly and $j$-step q-linearly for an appropriate $j \in \mathbb{N}$.*

*Proof.* Since $G$ and $F \circ G$ are locally calm at $\bar q$ by Lemma 3.3, the r-linear convergence of $(u^k) = (G(q^k))$ and $(F(u^k)) = ((F\circ G)(q^k))$ is an immediate consequence of Lemma 2.12, part 1). The asserted estimates follow from the local calmness of $G$ and $F \circ G$ at $\bar{q}$, respectively, $F$ at $\bar{u}$; cf. Lemma 3.3. Since the local calmness only applies if $q^k$ is sufficiently close to $\bar{q}$, $k$ has to be sufficiently large. The r-linear and multi-step q-linear convergence of $(H(q^k))$ follow from part 1) and 2) of Lemma 2.12, the prerequisites of which are satisfied due to Lemma 3.6. ∎

*Remark* 4.5. Let us comment further on the connection between convergence of $(q^k)$ and convergence of $(u^k)$ and $(F(u^k))$. A sufficient criterion to infer (possibly multi-step) q-linear convergence of $(u^k)$ from q-linear convergence of $(q^k)$ is contained in Lemma 2.12. It requires $G$ to be locally metrically subregular at $\bar{q}$. In [34] we observe that in many applications, $G$ is a projection or, more generally, a proximal mapping. Hence, local metrical subregularity of $G$ at $\bar{q}$ cannot be expected in these applications. Moreover, it is well-known that if $G$ is a superposition operator between the Lebesgue spaces $Q = L^p$ and $U = L^2$, then the semismoothness of $G$ necessitates $p > 2$ except if $G$ is affine, cf. [24, Lemma 2.7 in Section 2.5.2] or [45, Theorem 3.3]. Thus, also in this setting we cannot expect $G$ to be locally metrically subregular. On the other hand, for $F \equiv G \equiv 0$—which is included in our convergence analysis—it is evident that convergence of $(u^k)$ and $(F(u^k))$ does not imply convergence of $(q^k)$.

An inspection of the proof of Theorem 4.2, part 1), produces the following result.

Corollary 4.6. *Let Assumption 3.1 hold. Then:*

1) *If $(q^0, B_0)$ is chosen as in Theorem 4.2 1), then $(\|\tilde{M}_k^{-1}\|_{\mathcal{L}(V,Q)})$ is bounded.*
2) *If Algorithm 1 generates a sequence of iterates $(q^k)$ that converges r-linearly to $\bar{q}$, then $(\|E_k\|_{\mathcal{L}(U,V)})$ is bounded.*

*Proof.* **Proof of 1):** If finite termination occurs, then there is nothing to prove. Otherwise, the asserted boundedness of $(\|\tilde{M}_k^{-1}\|_{\mathcal{L}(V,Q)})$ follows from the proof of Theorem 4.2 1). In fact, note that below (4.3) we showed that $\tilde{M}_k$ is invertible with $\|\tilde{M}_k^{-1}\|_{\mathcal{L}(V,Q)} \leq 2C_{\bar{M}^{-1}}$. Since this holds for all $k \in \mathbb{N}_0$, $(\|\tilde{M}_k^{-1}\|_{\mathcal{L}(V,Q)})$ is bounded.
**Proof of 2):** By the r-linear convergence of $(q^k)$ there exist $\beta \in (0,1)$, $(\alpha_k) \subset [0,\infty)$ and $K \in \mathbb{N}_0$ such that $\|q^k - \bar{q}\|_Q \leq \alpha_k \leq \beta^{k-K}\alpha_K$ is valid for all $k \geq K$. Since $(q^k)$ converges to $\bar{q}$, we may assume without loss of generality that $(q^k)_{k\geq K} \subset \mathbb{B}_{\delta_M}(\bar{q})$, so that $G$ is locally calm at $\bar{q}$ on these iterates. In turn, this yields that Lemma 2.16 is applicable to, without loss of generality, $(u^k)_{k\geq K}$. Thus, for all $k \geq K$ there holds

$$\|E_{k+1}\|_{\mathcal{L}(U,V)} \leq \|E_K\|_{\mathcal{L}(U,V)} + \mathring{C} \sum_{j=K}^{k} \|q^j - \bar{q}\|_Q^\eta \leq \|E_K\|_{\mathcal{L}(U,V)} + \mathring{C} \frac{\alpha_K^\eta}{1-\beta^\eta},$$

where $\mathring{C} := 2C_F L_G^\eta$. As the right-hand side is independent of $k$, the claim follows. ∎

*Remark* 4.7.

1) In the situation of Corollary 4.6, part 2), it is easy to see that $(\|\tilde{M}_k\|_{\mathcal{L}(Q,V)})$ is bounded if $(\|\hat{M}_k\|_{\mathcal{L}(Q,V)})$ is bounded.
2) In several of the following results the sequence $(q^k)$ is required to converge q-linearly. Most of these results remain valid if the q-linear convergence is replaced by r-linear convergence. This can be inferred from the respective proofs by minor changes that are similar to the argument used in the proof of Corollary 4.6, part 2), to pass from $(q^k)$ to $(\alpha_k)$. However, the proof of the fundamental result Theorem 4.16 requires q-linear convergence of $(q^k)$. For convenience we therefore work with q-linear convergence of $(q^k)$ from now on.

**4.2. Superlinear convergence.** In the first part of this section we give an overview of superlinear convergence results for quasi-Newton methods applied to smooth and to semismooth mappings. In the second part we prepare the results on superlinear convergence, and in the third part we collect and prove them.

**4.2.1. On superlinear convergence of quasi-Newton methods.** In this section we summarize results regarding superlinear convergence of quasi-Newton methods to work out that, in contrast to the smooth case, unmodified quasi-Newton methods applied to semismooth mappings do not achieve superlinear convergence, in general.

We consider first the problem of finding $\bar{x}$ with $F(\bar{x}) = 0$, where $F : X \to Y$ is a mapping between Banach spaces $X$ and $Y$ that is Lipschitz continuously Fréchet differentiable. Let $(x^k)$ be generated by a Newton-type method and let $(x^k)$ converge to $\bar{x}$. If $F'(\bar{x}) \in \mathcal{L}(X,Y)$ is invertible, then it is both necessary and sufficient for q-superlinear convergence of $(x^k)$ to $\bar{x}$ that one of the two equivalent conditions

$$\text{(DM)} \qquad \lim_{k\to\infty} \frac{\left\|(B_k - F'(\bar{x}))(s_x^k)\right\|_Y}{\|s_x^k\|_X} = 0 \qquad \text{or} \qquad \lim_{k\to\infty} \frac{\left\|(B_k - F'(x^k))(s_x^k)\right\|_Y}{\|s_x^k\|_X} = 0$$

is satisfied, where $B_k$ is the linear operator that is used in place of $F'(x^k)$ and $s_x^k = x^{k+1} - x^k$ is the resulting step, i.e., $B_k s_x^k = -F(x^k)$. The first condition is called Dennis–Moré condition after their discoverers, who introduced it in [16] and showed there that it is satisfied for the DFP method. Obviously, the two conditions are fulfilled for Newton's method and, more generally, if $\|B_k - F'(\bar{x})\|_{\mathcal{L}(X,Y)} \to 0$ for $k \to \infty$. However, the ground-breaking achievement of (DM) is that it shows that convergence in operator norm is *not* required for superlinear convergence. This observation is especially important for quasi-Newton methods because, in general, the quasi-Newton approximations $(B_k)$ do *not* converge to $F'(\bar{x})$ in operator norm, not even if $X$ and $Y$ are finite-dimensional, cf., e.g., [17, Example 5.3] and [18, Lemma 8.2.7]. Despite, several quasi-Newton methods satisfy the Dennis–Moré condition, as Broyden, Dennis and Moré showed in finite dimensions in [6], thereby proving q-superlinear convergence of these methods. Subsequently, Stoer [46] presented two examples that demonstrated that the assumptions of the finite-dimensional case are not sufficient for q-superlinear convergence of quasi-Newton methods in the infinite-dimensional case; another example is given by Griewank in [21, p.688–689]. This implies that the Dennis–Moré condition (DM) is not satisfied in this case. Subsequently, Sachs and, independently, Griewank proved q-superlinear convergence for quasi-Newton methods in infinite dimensions, cf. [44] and [21]. They realized that the missing ingredient is to require the initial approximation $B_0 - F'(\bar{x})$ to be compact. Compactness comes into play because, as Sachs proved, quasi-Newton methods fulfill (DM) if the strong convergence in $Y$ is replaced by weak convergence, i.e., they satisfy

$$\text{(DMW)} \qquad \forall l \in Y^* : \quad \lim_{k\to\infty} \frac{\left\langle l, B_k - F'(\bar{x})(s_x^k)\right\rangle_{Y^*,Y}}{\|s_x^k\|_X} = 0.$$

If the difference $B_0 - F'(\bar{x})$ is a compact operator, then (DM) follows from (DMW), yielding q-superlinear convergence; cf. [30]. The relationship between the requirements in the finite-dimensional and the infinite-dimensional case becomes clear by realizing that in the finite-dimensional case (DM) and (DMW) are equivalent and, moreover, all linear operators are compact so the requirement on $B_0 - F'(\bar{x})$ is void.

We now consider the problem of finding $\bar{x}$ with $G(\bar{x}) = 0$, where $G : X \to Y$ is a mapping between Banach spaces $X$ and $Y$ that is semismooth at $\bar{x}$. At first glance, the situation seems to be similar to the smooth case just discussed: A semismooth Newton-type method for $G(\bar{x}) = 0$, where $G$ satisfies the conditions needed for q-superlinear convergence of the semismooth Newton method, that generates a converging sequence of iterates converges q-superlinearly if $G$ is calm at $\bar{x}$ and the second condition of (DM) is satisfied with $F'(x^k)$ replaced by an arbitrary element of $\partial G(x^k)$. This statement is true in infinite dimensions, cf. [49, Theorem 3.18(a)]. However, there are simple examples for $X = Y = \mathbb{R}$ in which quasi-Newton methods converge only q-linearly for semismooth functions, even if the uniform invertibility property of Definition 2.7 holds (which guarantees local q-superlinear convergence of the semismooth Newton method in these examples); cf. [21, Introduction] and [28, Example 2.40]. In fact, in this very situation the convergence may not even be q-linear, cf. [1, Example 1]. In conclusion, we obtain, firstly, that Algorithm 1 could not be superlinearly convergent if the generalized Broyden method in this algorithm were applied to $H$. Secondly, the semismooth analogue of the Dennis–Moré condition is generally not satisfied for *unmodified* quasi-Newton methods applied to semismooth mappings, not even in finite dimensions. Concerning *modified* quasi-Newton methods that achieve superlinear convergence in infinite dimensions we are only aware of [1, Theorem 5], which requires strong assumptions and the computation of a quantity that is difficult to compute for many problems, as the authors note in [1, Remark 4].

**4.2.2. Superlinear convergence of the hybrid method: Preliminaries.** To prove q-superlinear convergence of $(q^k)$ generated by Algorithm 1 we view this algorithm as a semismooth Newton-type method for $H$ with generalized derivative

$$\partial H : Q \rightrightarrows \mathcal{L}(Q, V), \qquad \partial H(q) := \left\{ F'(\bar{u}) \circ M + \hat{M} : \; M \in \partial G(q), \, \hat{M} \in \partial \hat{G}(q) \right\},$$

cf. (3.1). We now provide two Dennis–Moré-type conditions that will enable us to prove two somewhat different results on q-superlinear convergence of the hybrid method. We recall that the constant $C_{\bar{M}^{-1}}$ is introduced in Assumption 3.1, that $M_k$ and $B_k$ are generated by Algorithm 1 in Line 4, respectively, Lines 10 and 11, and that we use the notation $\bar{s}^k = q^k - \bar{q}$, $s^k = q^{k+1} - q^k$, and $E_k = B_k - F'(\bar{u})$.

LEMMA 4.8. *Let Assumption 3.1 hold and let $(q^k)$ be generated by Algorithm 1. If $(q^k)$ converges to $\bar{q}$ and satisfies both*

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

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

*Proof.* Since Algorithm 1 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 $(\varepsilon_k) \subset [0, \infty)$ and constants $c < 1$ and $K \in \mathbb{N}_0$ such that

$$(4.5) \quad C_{\bar{M}^{-1}} \|E_k M_k \bar{s}^k\|_V \leq \varepsilon_k \|\bar{s}^k\|_Q \qquad \text{and} \qquad C_{\bar{M}^{-1}} \|E_k M_k \bar{s}^{k+1}\|_V \leq c \, \|\bar{s}^{k+1}\|_Q$$

are satisfied for all $k \geq K$. Moreover, since $H = F \circ G + \hat{G}$ is semismooth at $\bar{q}$ by Lemma 3.3 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. $(\varepsilon_k)$, such that for every $k \in \mathbb{N}_0$

$$(4.6) \qquad C_{\bar{M}^{-1}} \|H(q^k) - H(\bar{q}) - \bar{M}_k \bar{s}^k\|_V \leq \varepsilon_k \|\bar{s}^k\|_Q$$

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 1 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 (4.5) and (4.6) this implies

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

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

In the second Dennis–Moré-type condition, the lim sup condition of (DMT1) is replaced by uniform invertibility of $(\tilde{M}_k)$ (which is generated by Algorithm 1 in Line 5).

LEMMA 4.9. *Let Assumption 3.1 hold and let $(q^k)$ be generated by Algorithm 1. 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 1 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

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

Since $(\tilde{M}_k - \bar{M}_k)\bar{s}^k = E_k M_k \bar{s}^k$, the claim follows by (DMT2) and semismoothness. ∎

*Remark* 4.10. Neither of the Dennis–Moré-type conditions (DMT1) and (DMT2) reproduces the classical Dennis–Moré condition (DM) for $Q = U$ and $G = \mathrm{id}$.

In the remainder of this section we establish that (DMT1) and (DMT2) are fulfilled for Algorithm 1. A crucial part of the proof is contained in the next three lemmas, whose main goal it is to show that the Dennis–Moré condition (DM) is satisfied for the iterates $(u^k) = (G(q^k))$ of Algorithm 1; this is achieved in Lemma 4.13. The main difficulty is that for the hybrid method (in contrast to classical quasi-Newton methods), the direction $s_u^k = u^{k+1} - u^k = G(q^{k+1}) - G(q^k)$ is not the solution of the quasi-Newton equation, i.e., $B_k s_u^k = -F(u^k)$ is not satisfied, in general.

Let us begin by demonstrating that the iterates $(u^k)$ of Algorithm 1 satisfy the Dennis–Moré condition in the sense of weak convergence. To this end, we recall the definition $\hat{s}_u^k = \frac{s_u^k}{\|s_u^k\|_U}$ for all $k \in \mathbb{N}_0$ with $s_u^k \neq 0$ and $\hat{s}_u^k = 0$ for all $k$ with $s_u^k = 0$.

LEMMA 4.11. *Let Assumption 3.1 hold and let $(q^k)$ be generated by Algorithm 1 with $0 < \sigma_{\min} \le \sigma_{\max} < 2$. If $(q^k)$ converges q-linearly to $\bar{q}$, then*

$$\forall l \in U : \quad \lim_{k\to\infty} \left(l, E_k \hat{s}_u^k\right)_U = 0.$$

*Proof.* We start with some preparations. By the q-linear convergence of $(q^k)$ there exist constants $K \in \mathbb{N}_0$ and $\beta \in (0,1)$ such that $\|q^{k+1} - \bar{q}\|_Q \le \beta \|q^k - \bar{q}\|_Q$ holds for all $k \ge K$. Since $G$ is locally calm at $\bar{q}$, wlog. $\|G(q^k) - G(\bar{q})\|_U \le L_G \|q^k - \bar{q}\|_Q$ for all $k \ge K$, we obtain for all these $k$

$$\begin{aligned}
&\max\left\{\|u^{k+1} - \bar{u}\|_U, \|u^k - \bar{u}\|_U\right\} \\
&\qquad = \max\left\{\|G(q^{k+1}) - G(\bar{q})\|_U, \|G(q^k) - G(\bar{q})\|_U\right\} \\
&\qquad \le \max\left\{L_G\|q^{k+1} - \bar{q}\|_Q, L_G\|q^k - \bar{q}\|_Q\right\} \\
&\qquad \le \max\left\{L_G\beta\|q^k - \bar{q}\|_Q, L_G\|q^k - \bar{q}\|_Q\right\} \le L_G\|q^k - \bar{q}\|_Q.
\end{aligned} \tag{4.7}$$

This estimate implies, in particular, $u^k \to \bar{u}$ for $k \to \infty$. Due to the $\eta$-strict differentiability of $F$ at $\bar{u}$, cf. Assumption 3.1, we infer from $u^k \to \bar{u}$ that $\|F(u^{k+1}) - F(u^k) - F'(\bar{u})s_u^k\|_U \le C_F \|s_u^k\|_U \max\{\|u^{k+1} - \bar{u}\|_U, \|u^k - \bar{u}\|_U\}^\eta$ holds for all $k$ sufficiently large, wlog. $k \ge K$. Using (4.7) and $\mathring{C} := 2C_F L_G^\eta$ this yields

$$\|F(u^{k+1}) - F(u^k) - F'(\bar{u})s_u^k\|_U \le \frac{\mathring{C}}{2}\|s_u^k\|_U \|q^k - \bar{q}\|_Q^\eta \tag{4.8}$$

for all $k \ge K$. We conclude the preparations by noting that Corollary 4.6, part 2), implies $C_B := \max\{1, \sup_{k\in\mathbb{N}_0} \|E_k\|_{\mathcal{L}(U,V)}\} < \infty$.

To establish the assertion, fix $l \in U$ and denote $e^k := E_k^* l \in U$ for all $k \in \mathbb{N}_0$. For $l = 0$ the claim is obviously true, so we may assume $l \ne 0$. For $h \in U$ there holds

$$\begin{aligned}
&\left(e^{k+1}, h\right)_U = (l, E_{k+1}h)_U \\
&= \left(l, E_k\left(h - \sigma_k s_u^k \frac{\left(s_u^k, h\right)_U}{\|s_u^k\|_U^2}\right)\right)_U + \sigma_k \left(l, (y^k - F'(\bar{u})s_u^k)\frac{\left(s_u^k, h\right)_U}{\|s_u^k\|_U^2}\right)_U \\
&= \left(e^k, h - \sqrt{\sigma_k}\hat{s}_u^k\left(\sqrt{\sigma_k}\hat{s}_u^k, h\right)_U\right)_U + \sigma_k\left(l, (F(u^{k+1}) - F(u^k) - F'(\bar{u})s_u^k)\frac{\left(s_u^k, h\right)_U}{\|s_u^k\|_U^2}\right)_U.
\end{aligned}$$

Introducing the constant $\mathring{C}_l := \|l\|_U \mathring{C}$ we infer from this by Lemma 2.15 and (4.8) that for all $k \ge K$ we have

$$\begin{aligned}
\sup_{\|h\|_U \le 1} \left(e^{k+1}, h\right)_U &\le \sup_{\|h\|_U \le 1} \left(e^k, h - \sqrt{\sigma_k}\hat{s}_u^k\left(\sqrt{\sigma_k}\hat{s}_u^k, h\right)_U\right)_U \\
&\quad + \sigma_{\max} \sup_{\|h\|_U \le 1}\left(l, (F(u^{k+1}) - F(u^k) - F'(\bar{u})s_u^k)\frac{\left(s_u^k, h\right)_U}{\|s_u^k\|_U^2}\right)_U \\
&\le \sqrt{\|e^k\|_U^2 - (2 - |\sigma_k|)\,|\sigma_k|\,(e^k, \hat{s}_u^k)_U^2} + \mathring{C}_l \|q^k - \bar{q}\|_Q^\eta.
\end{aligned} \tag{4.9}$$

In the case $e^k \ne 0$ we continue by

$$\|e^{k+1}\|_U = \sup_{\|h\|_U \le 1} \left(e^{k+1}, h\right)_U \le \|e^k\|_U - (2 - \sigma_k)\,\sigma_k \frac{\left(e^k, \hat{s}_u^k\right)_U^2}{2\|e^k\|_U} + \mathring{C}_l \|q^k - \bar{q}\|_Q^\eta,$$

where we used that $\sqrt{a^2 - b^2} \le a - \frac{b^2}{2a}$ holds for $a > b \ge 0$ as well as for $a = b > 0$. As $\|e^k\|_U = \|E_k^* l\|_U \le C_B \|l\|_U$ for all $k$, it follows that for $k \ge K$ with $e^k \ne 0$ we have

$$(2-\sigma_{\max})\,\sigma_{\min} \frac{\left(e^k, \hat{s}_u^k\right)_U^2}{2C_B\|l\|_U} \le (2-\sigma_k)\,\sigma_k \frac{\left(e^k, \hat{s}_u^k\right)_U^2}{2C_B\|l\|_U} \le \|e^k\|_U - \|e^{k+1}\|_U + \mathring{C}_l \|q^k - \bar{q}\|_Q^\eta.$$

This estimate is also valid for $k \ge K$ with $e^k = 0$, since for $e^k = 0$ it can be inferred directly from (4.9). Letting $\sigma := (2-\sigma_{\max})\sigma_{\min}$ we conclude via summation that

(4.10)
$$\begin{aligned}
\frac{\sigma}{2C_B\,\|l\|_U} \sum_{k=K}^{\infty} \left(e^k, \hat{s}_u^k\right)_U^2 &= \sup_{n\ge K} \left[ \frac{\sigma}{2C_B\,\|l\|_U} \sum_{k=K}^{n} \left(e^k, \hat{s}_u^k\right)_U^2 \right] \\
&\le \sup_{n\ge K} \left[ \sum_{k=K}^{n} \left(\|e^k\|_U - \|e^{k+1}\|_U\right) \right] + \sup_{n\ge K} \left[ \sum_{k=K}^{n} \mathring{C}_l \|q^k - \bar{q}\|_Q^\eta \right] \\
&\le \sup_{n\ge K} \left[ \|e^K\|_U - \|e^{n+1}\|_U \right] + \mathring{C}_l \sum_{k=K}^{\infty} \|q^k - \bar{q}\|_Q^\eta \\
&\le \|e^K\|_U + \mathring{C}_l \frac{\|q^K - \bar{q}\|_Q^\eta}{1-\beta^\eta},
\end{aligned}$$

where we used that $\|q^{k+1} - \bar{q}\|_Q \le \beta \|q^k - \bar{q}\|_Q$ for all $k \ge K$. Since the right-hand side of (4.10) is bounded, the series $\sum_{k=K}^{\infty} (e^k, \hat{s}_u^k)_U^2$ converges, yielding $(e^k, \hat{s}_u^k)_U \to 0$ for $k \to \infty$. As $(e^k, \hat{s}_u^k)_U = (l, E_k \hat{s}_u^k)_U$ by definition, the assertion follows. ∎

To facilitate the proof that the weak limit in Lemma 4.11 is, in fact, a strong limit if $E_0$ is compact, we note the following.

LEMMA 4.12. *If Algorithm 1 generates a sequence of iterates $(q^k)$, then for all $K \in \mathbb{N}_0$ the operator $B_K - B_0$ is compact.*

*Proof.* For $K = 0$ the claim is obviously true. For $K \ge 1$ we have

$$B_K - B_0 = \sum_{k=0}^{K-1} \left(B_{k+1} - B_k\right).$$

By definition of the update, cf. lines 10–11 in Algorithm 1, each of the differences on the right-hand side is either zero or a rank-one operator. Hence, the maximal rank of $B_K - B_0$ is $K$, which implies that this operator is compact. ∎

Similarly as in the smooth case, cf. subsection 4.2.1, the compactness of $E_0$ enables us to pass from weak convergence to strong convergence, even though the direction $\hat{s}_u^k$ in the following result is not the renormed solution of $B_k s^k = -F(u^k)$.

LEMMA 4.13. *Let Assumption 3.1 hold and let $(q^k)$ be generated by Algorithm 1 with $0 < \sigma_{\min} \le \sigma_{\max} < 2$. In addition, let $E_0$ be compact. If $(q^k)$ converges q-linearly to $\bar{q}$, then*

$$\lim_{k\to\infty} \|E_k \hat{s}_u^k\|_V = 0. \tag{4.11}$$

*Proof.* The following proof is a generalized version of [30, Proof of Theorem 2.5]. We have to show that the sequence $v^k := E_k \hat{s}_u^k$, $k \in \mathbb{N}_0$, converges strongly to zero. Recalling from Lemma 4.11 that $(v^k)$ converges weakly to zero, it suffices to demonstrate that any subsequence of $(v^k)$ contains a strongly convergent subsequence. We will accomplish this by proving that there is $K \in \mathbb{N}_0$ such that $(v^k)_{k>K}$ belongs to a relatively compact set. We start by following the arguments of the proof of Lemma 4.11 up until (4.8). This shows that there are $K \in \mathbb{N}_0$ and $\beta \in (0,1)$ such that for all $k \geq K$ there hold $\|q^{k+1} - \bar{q}\|_Q \leq \beta \|q^k - \bar{q}\|_Q$ and, with $\mathring{C} := 2C_F L_G^\eta$,

$$\|F(u^{k+1}) - F(u^k) - F'(\bar{u}) s_u^k\|_V \leq \frac{C^\circ}{2} \|s_u^k\|_U \|q^k - \bar{q}\|_Q^\eta. \tag{4.12}$$

Letting $w^k := F(u^{k+1}) - F(u^k) - F'(\bar{u}) s_u^k$ for all $k \in \mathbb{N}_0$ we obtain from the definition of $B_{k+1}$ that for all $k \in \mathbb{N}_0$ we have

$$E_{k+1} = E_k \left( I - \sigma_k s_u^k \frac{(s_u^k, \cdot)_U}{\|s_u^k\|_U^2} \right) + \sigma_k w^k \frac{(s_u^k, \cdot)_U}{\|s_u^k\|_U^2}.$$

In particular, this is true for all $k \geq K$. Since $\|I - \sqrt{\sigma_k} \hat{s}_u^k \left(\sqrt{\sigma_k} \hat{s}_u^k, \cdot\right)_U\|_{\mathcal{L}(U,U)} \leq 1$ by Lemma 2.14, we deduce using (4.12) that for all $k \geq K$ there holds

$$E_{k+1} \overline{\mathbb{B}}_1(0) \subset E_k \overline{\mathbb{B}}_1(0) + \left\{ \alpha \mathring{C} \|q^k - \bar{q}\|_Q^\eta \hat{w}^k : \ \alpha \in [-1,1] \right\},$$

where $\hat{w}^k = 0$ if $w^k = 0$ and $\hat{w}^k = \frac{w^k}{\|w^k\|_V}$ if $w^k \neq 0$. Introducing $\hat{\beta} := \beta^\eta$ and $C_K := \mathring{C} \|q^K - \bar{q}\|_Q^\eta$ we infer that for all $j \in \mathbb{N}$

$$E_{K+j} \overline{\mathbb{B}}_1(0) \subset E_K \overline{\mathbb{B}}_1(0) + \left\{ \sum_{k=0}^{j-1} \alpha_k \hat{\beta}^k C_K \hat{w}^{K+k} : \ \alpha_k \in [-1,1] \right\}.$$

Defining for all $k \in \mathbb{N}_0$ the sets

$$I_k := \left\{ \alpha \hat{\beta}^k C_K \hat{w}^{K+k} : \ \alpha \in [-1,1] \right\} \subset V$$

it follows that for every $j \in \mathbb{N}$ there holds

$$\begin{aligned} v^{K+j} \in E_{K+j} \overline{\mathbb{B}}_1(0) \subset E_K \overline{\mathbb{B}}_1(0) &+ \left\{ \sum_{k=0}^{j-1} x^k : \ x^k \in I_k \ \forall k \right\} \\ &\subset E_0 \overline{\mathbb{B}}_1(0) + (E_K - E_0) \overline{\mathbb{B}}_1(0) + \left\{ \sum_{k=0}^{j-1} x^k : \ x^k \in I_k \ \forall k \right\} \\ &\subset E_0 \overline{\mathbb{B}}_1(0) + (B_K - B_0) \overline{\mathbb{B}}_1(0) + \bigcup_{l=0}^{\infty} \left\{ \sum_{k=0}^{l} x^k : \ x^k \in I_k \ \forall k \right\}. \end{aligned}$$

The three sets on the right-hand side are each relatively compact; for the first set this is true by assumption, while for the second and third this follows from Lemma 4.12 and Lemma 2.18, respectively. Thus, $(v^k)_{k>K}$ is part of a relatively compact set. ∎

The final ingredient to prove that Algorithm 1 converges q-superlinearly is the observation that under the Dennis–Moré condition (4.11) the Broyden updates $B_{k+1} - B_k$ converge to zero in operator norm. As we pointed out in subsection 4.2.1, this is generally not true for $B_k - F'(\bar{u})$.

LEMMA 4.14. *Let Assumption 3.1 hold and let $(q^k)$ be generated by Algorithm 1. If $(q^k)$ converges to $\bar{q}$, 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.* As in (4.8) we can derive that for all $k$ sufficiently large it holds that

$$\|F(u^{k+1}) - F(u^k) - F'(\bar{u})s_u^k\|_V \le \frac{\mathring{C}}{2}\|s_u^k\|_U \|q^k - \bar{q}\|_Q^\eta,$$

where $\mathring{C} := 2C_F L_G^\eta$. We compute for all $k \in \mathbb{N}_0$ with $s_u^k \ne 0$ and any $h \in U$

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

By taking norms in $V$ we infer that for all sufficiently large $k$ with $s_u^k \ne 0$ we have

$$\|(B_{k+1} - B_k)\, h\|_V \le \mathring{C}\|q^k - \bar{q}\|_Q^\eta \,\|h\|_U + 2\,\|E_k \hat{s}_u^k\|_V \,\|h\|_U\,.$$

Since $B_{k+1} = B_k$ if $s_u^k = 0$, cf. lines 10–11 in Algorithm 1, this estimate is also valid in the case $s_u^k = 0$. Thus, for all $k$ sufficiently large it is true that

$$\|B_{k+1} - B_k\|_{\mathcal{L}(U,V)} \le \mathring{C}\|q^k - \bar{q}\|_Q^\eta + 2\,\|E_k \hat{s}_u^k\|_V.$$

This implies the assertion by letting $k$ go to infinity. ∎

*Remark* 4.15. We discovered Lemma 4.14 independently, but later learned that [35, Theorem 6] contains the finite-dimensional analogue of Lemma 4.14 and [4, Proof of Theorem 4.8] contains an infinite-dimensional version under stronger assumptions.

In view of Lemma 4.8 and Lemma 4.9 it is clear that the following theorem is most valuable for establishing superlinear convergence of Algorithm 1.

THEOREM 4.16. *Let Assumption 3.1 hold and let $(q^k)$ be generated by Algorithm 1 with $0 < \sigma_{\min} \le \sigma_{\max} < 2$. In addition, let $E_0$ be compact. If $(q^k)$ converges q-linearly to $\bar{q}$, then*

$$\lim_{k\to\infty} \frac{\left\|E_k M_k \bar{s}^k\right\|_V}{\|\bar{s}^k\|_Q} = 0. \tag{4.13}$$

*Proof.* We start with some preparations. 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. Definition 3.5. Moreover, let us introduce the sequence

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

and observe that $\lim_{k\to\infty} \varepsilon_k = 0$, as follows from Lemma 4.13 and Lemma 4.14. From Corollary 4.6, part 2), we obtain $C_B := \sup_{k\in\mathbb{N}_0} \|E_k\|_{\mathcal{L}(U,V)} < \infty$.

We now start the actual proof. To establish (4.13) we compute for all $k \in \mathbb{N}_0$

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

Hence, by 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. For later use we note that $(R_k)$ is bounded from above; this follows since for all $k \geq K$ there holds

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

Furthermore, for all $k \geq K$

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

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 $k$

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

where we have used the q-linear convergence of $(q^k)$ with rate $\beta$ to obtain the last inequality. Therefore, we have established that for all $k \geq K$ with $s_u^k \neq 0$ there holds

$$0 \leq R_k \leq \varepsilon_k + \beta R_{k+1}. \tag{4.14}$$

In the case $s_u^k = 0$ we have $u^{k+1} = u^k$ and $B_{k+1} = B_k$, the latter by definition of the Broyden update in Line 10–11 of Algorithm 1. Thus, the definition of $R_k$ shows that $0 \leq R_k \leq \beta R_{k+1}$ in this case, which implies by $\varepsilon_k \geq 0$ that (4.14) holds for all $k \geq K$, regardless of whether $s_u^k \neq 0$ or $s_u^k = 0$. Since $(R^k)$ is bounded from above and since there holds $\lim_{k\to\infty} \varepsilon_k = 0$, we can apply Lemma 2.19 with $a_k := R_k$ and $b_k := \varepsilon_k$. This shows that $\lim_{k\to\infty} R_k = 0$, as desired. ∎

*Remark* 4.17. The proof of Theorem 4.16 shows that $\sigma_{\min} > 0$, $\sigma_{\max} < 2$, and the compactness of $E_0$ are only needed to ensure that

$$\lim_{k\to\infty} \|E_k \hat{s}_u^k\|_V = 0. \tag{4.15}$$

In other words, if $(q^k)$ converges q-linearly to $\bar{q}$ and (4.15) holds, then (4.13) follows.

**4.2.3. Superlinear convergence of the hybrid method: Results.** Our main result is the following theorem on the q-superlinear convergence of Algorithm 1.

THEOREM 4.18. *Let Assumption 3.1 hold and consider Algorithm 1 with* $0 < \sigma_{\min} \le \sigma_{\max} < 2$. *In addition, let* $E_0$ *be compact. Then:*

1) *If* $(q^k)$ *is generated by Algorithm 1 and converges q-linearly to* $\bar{q}$*, then the convergence is q-superlinear if any of the following two conditions is satisfied:*
   a) *The sequence* $(\|\tilde{M}_k^{-1}\|_{\mathcal{L}(V,Q)})$ *is bounded.*
   b) *There holds*

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

2) *There exist* $\delta, \varepsilon > 0$ *such that for every pair of starting values* $(q^0, B_0) \in Q \times \mathcal{L}(U,V)$ *with* $\|q^0 - \bar{q}\|_Q < \delta$ *and* $\|E_0\|_{\mathcal{L}(U,V)} < \varepsilon$*, Algorithm 1 is well-defined and either terminates after finitely many iterations or generates a sequence* $(q^k)$ *that converges q-linearly to* $\bar{q}$ *and satisfies both a) and b). In particular,* $(q^k)$ *converges q-superlinearly.*
3) *If, in addition to Assumption 3.1,* $F$ *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 2) 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 in a neighborhood of* $\bar{u}$*.*

*Proof.* **Proof of 1):** If a) is satisfied, then the q-superlinear convergence of $(q^k)$ follows from Lemma 4.9, whose prerequisites are fulfilled due to a) and Theorem 4.16. For the q-superlinear convergence under condition b) we deduce from Lemma 4.8 and Theorem 4.16 that it suffices to establish

$$\limsup_{k\to\infty} \frac{\left\|E_k M_k \bar{s}^{k+1}\right\|_V}{\|\bar{s}^{k+1}\|_Q} < \frac{1}{C_{\bar{M}^{-1}}}. \tag{4.16}$$

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

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

for these $k$. The second term on the right-hand side converges to zero by Theorem 4.16. 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, (4.16) follows by b).
**Proof of 2):** Theorem 4.2 and Corollary 4.6, part 1), provide $\delta, \varepsilon > 0$ such that q-linear convergence of $(q^k)$ and a) are ensured. By (4.1) and the uniform' boundedness of $\partial G$ near $\bar{q}$, cf. Assumption 3.1, $\delta, \varepsilon$ can be chosen such that b) is satisfied, too.
**Proof of 3):** Verbatim as in part 2) of Theorem 4.2. ∎

*Remark* 4.19. 1) If $V$ is finite-dimensional, then $E_0 = B_0 - F'(\bar{u})$ is always compact. Furthermore, in PDE-constrained optimization $F'(\bar{u})$ is often compact, so that $E_0$ is compact if, e.g., $B_0 = 0$; cf. [34] for details.
2) Remark 4.17 still applies. That is, $\sigma_{\min} > 0$, $\sigma_{\max} < 2$, and the compactness of $E_0$ can be replaced by (4.15).

3) For $Q = U$, $G = \mathrm{id}$, $\hat{G} \equiv 0$ and the choice $(\sigma_k) \equiv 1$ we recover from Theorem 4.18 the classical superlinear convergence result [30, Theorem 2.5] of Broyden's method for the infinite-dimensional smooth case under weaker differentiability assumptions. Furthermore, for $F \equiv 0$ and $G \equiv 0$ we recover from Theorem 4.18 the standard superlinear convergence result for infinite-dimensional semismooth Newton methods under the standard assumptions.
4) The setting $F \equiv 0$ and $G \equiv 0$ shows that the compactness of $E_0$ is sufficient but not necessary to obtain q-superlinear convergence of Algorithm 1. Note that $(s_u^k) \equiv 0$ in this setting and consider 3). Also, it follows that q-superlinear convergence of $(u^k)$ and $(F(u^k))$ does not imply convergence of $(q^k)$.

After recalling that $L_F$ is introduced in Definition 3.5, we note the following for the convergence of $(u^k)$, $(F(u^k))$, and $(H(q^k))$.

COROLLARY 4.20. *Let Assumption 3.1 hold and let $(q^k)$ be generated by Algorithm 1. If $(q^k)$ converges q-superlinearly to $\bar{q}$, then:*

*1) $(u^k)$ converges r-superlinearly to $\bar{u}$ and satisfies $\|u^k - \bar{u}\|_U \le L_G \|q^k - \bar{q}\|_Q$ for all $k$ sufficiently large.*
*2) $(F(u^k))$ converges r-superlinearly to $F(\bar{u})$ and satisfies, for all $k$ large enough, $\|F(u^k) - F(\bar{u})\|_V \le L_F \|u^k - \bar{u}\|_U$ and $\|F(u^k) - F(\bar{u})\|_V \le L_F L_G \|q^k - \bar{q}\|_U$.*
*3) If $\hat{G}$ is locally calm at $\bar{q}$, then $(H(q^k))$ converges q-superlinearly to zero.*
*4) Define for all $K \in \mathbb{N}_0$ the set $S_q^K := \{\bar{q}\} \cup (q^k)_{k \ge K}$. If there is $\bar{K} \in \mathbb{N}_0$ such that $G|_{S_q^{\bar{K}}}$, respectively, $(F \circ G)|_{S_q^{\bar{K}}}$ is metrically subregular at $\bar{q}$, then the convergence of $(u^k)$, respectively, $(F(u^k))$ happens q-superlinearly.*

*Proof.* **Proof of 1):** Using $(u^k) = (G(q^k))$ the r-superlinear convergence follows from Lemma 2.12, part 1). The inequality is implied by the local calmness of $G$ at $\bar{q}$.
**Proof of 2):** Note that $F$ and $F \circ G$ are locally calm at $\bar{u}$, respectively, $\bar{q}$, cf. Lemma 3.3. Using $(F(u^k)) = ((F \circ G)(q^k))$ the r-superlinear convergence follows from Lemma 2.12, part 1). The inequalities are implied by local calmness.
**Proof of 3):** The assertion is contained in Corollary 3.7.
**Proof of 4):** By the local calmness of $G$ and $q^k \to \bar{q}$, there is $\bar{K} \in \mathbb{N}_0$ such that $G|_{S_q^{\bar{K}}}$ is calm and metrically subregular at $\bar{q}$. By Lemma 2.12 3) this implies q-superlinear convergence of $(G(q^k)) = (u^k)$. The same reasoning applies to $(F \circ G)|_{S_q^{\bar{K}}}$. ∎

*Remark* 4.21. Concerning 3) we mention that a strong connection between $(q^k)$ and $(H(q^k))$ is contained in Corollary 3.7. Regarding the metric subregularity in 4) we recall that $G$ is metrically subregular on $S_q^{\bar{K}}$ for a sufficiently large $K$ if $\partial G$ admits a uniformly' invertible selection near $\bar{q}$, cf. Lemma 2.8 2). Analogously for $F \circ G$.

**5. Conclusion and outlook.** We have presented an algorithm that combines quasi-Newton methods with semismooth Newton methods. This approach enables the use of quasi-Newton methods in infinite-dimensional nonsmooth regimes without sacrificing the local superlinear convergence. In contrast, quasi-Newton methods alone cannot achieve superlinear convergence on nonsmooth problems, in general.

In the complimentary paper [34] we examine the practical properties of the hybrid approach. Among others, we

- show that the new approach is widely applicable, for instance to generalized variational inequalities and structured nonsmooth optimization problems;
- globalize Algorithm 1 by a matrix-free limited-memory truncated trust-region method that proves capable of solving a nonconvex and nonsmooth large-scale real-world optimization problem involving the Bloch equations;

- conduct an extensive numerical study on problems from nonsmooth PDE-constrained optimal control and find that the new approach is at least an order of magnitude faster 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").

## REFERENCES

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

[2] H. W. Alt, *Linear Functional Analysis. An Application-Oriented Introduction*, Universitext, Springer, 2016, https://doi.org/10.1007/978-1-4471-7280-2.

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

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

[5] C. Broyden, *A class of methods for solving nonlinear simultaneous equations*, Math. Comput., 19 (1965), pp. 577–593, https://doi.org/10.2307/2003941.

[6] C. Broyden, J. Dennis, and J. J. More, *On the local and superlinear convergence of quasi-Newton methods*, J. Inst. Math. Appl., 12 (1973), pp. 223–245, https://doi.org/10.1093/imamat/12.3.223.

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

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

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

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

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

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

[13] W. C. Davidon, *Variable metric method for minimization*, Technical Report ANLC5990 (Revised), (1959), https://www.osti.gov/servlets/purl/4222000.

[14] W. C. Davidon, *Variable metric method for minimization*, SIAM Journal on Optimization, 1 (1991), pp. 1–17, https://doi.org/10.1137/0801001.

[15] J. Dennis, *On the convergence of Broyden's method for nonlinear systems of equations*, Math. Comput., 25 (1971), pp. 559–567, https://doi.org/10.2307/2005218.

[16] J. Dennis and J. J. More, *A characterization of superlinear convergence and its application to quasi-Newton methods*, Math. Comput., 28 (1974), pp. 549–560, https://doi.org/10.2307/2005926.

[17] J. Dennis and J. J. More, *Quasi-Newton methods, motivation and theory*, SIAM Rev., 19 (1977), pp. 46–89, https://doi.org/10.1137/1019005.

[18] J. Dennis and R. B. Schnabel, *Numerical methods for unconstrained optimization and nonlinear equations*, SIAM, repr. ed., 1996, https://doi.org/10.1137/1.9781611971200.

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

[20] R. Fletcher and M. Powell, *A rapidly convergent descent method for minimization*, Comput. J., 6 (1963), pp. 163–168, https://doi.org/10.1093/comjnl/6.2.163.

[21] A. Griewank, *The local convergence of broyden-like methods on lipschitzian problems in hilbert spaces*, SIAM Journal on Numerical Analysis, 24 (1987), pp. 684–705, https://doi.org/10.

1137/0724045.

[22] A. Griewank, *Broyden updating, the good and the bad!*, Doc. Math. (Bielefeld), (2012), pp. 301–315, https://www.emis.de/journals/DMJDMV/vol-ismp/45_griewank-andreas-broyden.pdf.

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

[24] M. Hinze, R. Pinnau, M. Ulbrich, and S. Ulbrich, *Optimization with PDE constraints*, vol. 23 of Mathematical Modelling: Theory and Applications, Springer, 2009, https://doi.org/10.1007/978-1-4020-8839-1.

[25] D. Hwang and C. Kelley, *Convergence of Broyden's method in Banach spaces*, SIAM J. Optim., 2 (1992), pp. 505–532, https://doi.org/10.1137/0802025.

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

[27] K. Ito and K. Kunisch, *Lagrange Multiplier Approach to Variational Problems and Applications*, SIAM, 2008, https://doi.org/10.1137/1.9780898718614.

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

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

[30] C. T. Kelley and E. W. Sachs, *A new proof of superlinear convergence for broyden's method in hilbert space*, SIAM Journal on Optimization, 1 (1991), pp. 146–150, https://doi.org/10.1137/0801011.

[31] F. Kruse, *Semismooth implicit functions*, Journal of Convex Analysis, 25 (2018), pp. 595–622.

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

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

[34] F. Mannel and A. Rund, *A hybrid semismooth quasi-Newton method. Part 2: Applications. Submitted (2018); preprint available at https://imsc.uni-graz.at/mannel/sqn2.pdf*.

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

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

[37] J. More and J. Trangenstein, *On the global convergence of Broyden's method*, Math. Comput., 30 (1976), pp. 523–540, https://doi.org/10.2307/2005323.

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

[39] J. Nocedal and S. J. Wright, *Numerical Optimization*, Springer Series in Operations Research and Financial Engineering, Springer, 2nd ed., 2006, https://doi.org/10.1007/978-0-387-40065-5.

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

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

[42] 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), pp. 301–325, https://doi.org/10.1287/moor.22.2.301.

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

[44] E. W. Sachs, *Broyden's method in Hilbert space*, Math. Program., 35 (1986), pp. 71–82, https://doi.org/10.1007/BF01589442.

[45] A. Schiela, *A simplified approach to semismooth Newton methods in function space*, SIAM J. Optim., 19 (2008), pp. 1417–1432, https://doi.org/10.1137/060674375.

[46] J. Stoer, *Two examples on the convergence of certain rank-2 minimization methods for quadratic functionals in Hilbert space*, Linear Algebra Appl., 28 (1979), pp. 217–222, https://doi.org/10.1016/0024-3795(79)90135-6.

[47] D. Sun and J. Han, *Newton and quasi-Newton methods for a class of nonsmooth equations*

*and related problems*, SIAM J. Optim., 7 (1997), pp. 463–480, https://doi.org/10.1137/S1052623494274970.

[48] R. Tibshirani, *Regression shrinkage and selection via the lasso*, J. R. Stat. Soc., Ser. B, 58 (1996), pp. 267–288, http://www.jstor.org/stable/2346178.

[49] M. Ulbrich, *Semismooth Newton Methods for Variational Inequalities and Constrained Optimization Problems in Function Spaces*, MOS-SIAM Series on Optimization, SIAM, 2011, https://doi.org/10.1137/1.9781611970692.

[50] 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 (2011), pp. 675–694, https://doi.org/10.1007/s10898-010-9629-2.