Mathematical Programming 35 (1986) 71-82
North-Holland

BROYDEN'S METHOD IN HILBERT SPACE

Ekkehard W. SACHS
Department of Mathematics, Box 8205, North Carolina State University, Raleigh, NC 27695-8205, USA

Received 6 July 1984
Revised manuscript received 13 September 1985

Broyden's method is formulated for the solution of nonlinear operator equations in Hilbert spaces. The algorithm is proven to be well defined and a linear rate of convergence is shown. Under an additional assumption on the initial approximation for the derivative we prove the superlinear rate of convergence.

Key words: Broyden's method, quasi-Newton methods, superlinear convergence rate.

1. Introduction

In this paper we investigate the local convergence rate properties of Broyden's method for the solution of nonlinear operator equations in a Hilbert space setting. These problems occur for example in the solution of necessary conditions of optimal control problems or in the solution of nonlinear integral equations and various other areas of applied analysis. The convergence rates for Newton's method have been studied in infinite-dimensional spaces at least since the forties and the results are meanwhile considered to be classical, see e.g. Kantorovich [14], Kantorovich and Akilov [15], Anselone [2], Moore [20].

Broyden [3] introduced a method for the solution of a finite-dimensional system of nonlinear equations which does not require the computation of any derivatives at each iteration and which retains favorable convergence rate properties such as a superlinear rate. It consists of a reasonably good approximation of the actual derivative which is updated by rank-one operators at each iteration. In contrast to many other popular Quasi-Newton updates it does not require any symmetry of the Jacobian of the system. Broyden's method has been extended in various directions. We mention nonlinear Chebyshev-approximation (Madsen [18]), nonlinear least norm problems and nonsmooth optimization problems with or without constraints (Gruver and Sachs [12], Sachs [22, 23]), and singular problems (Decker and Kelley [6]).

Unlike for Newton's method little attention has been given to the extension of Broyden's method to infinite-dimensional spaces. However, such an investigation could give more insight into the behavior of convergence rates for a discretized infinite-dimensional problem, where the discretization is rather fine. As a finite-dimensional problem it is supposed to converge superlinearly under the standard assumptions, but if for the original infinite dimensional problem Broyden's method is not superlinearly convergent, then for very fine discretizations a fast performance of Broyden's method is unlikely. Discretizations of infinite-dimensional problems occur in nonlinear integral equations, nonlinear differential equations and control problems such as nonlinear parabolic boundary control, see e.g. [12].

Todd [24] studies Broyden's method among others in abstract vector spaces of finite dimension under the viewpoint of deriving general properties of various update formulas. In this paper, we focus on the local convergence behavior of Broyden's method for the solution of the nonlinear equation

$$Fx = 0_Y,$$

where $F: X \to Y$ is a nonlinear mapping with Hilbert spaces X, Y. The basic iteration scheme is

$$x_{i+1} = x_i - B_i^{-1} Fx_i,$$

where B_i lies in $B(X, Y)$, the space of all bounded linear operators defined on X with range in Y. The operators B_i are updated according to Broyden's method.

Dennis proved in [7] for Broyden's method the linear rate of convergence of the iterates. In [12] and [23], also the superlinear rate of convergence was addressed. It was proven that the following property holds for all linear bounded functionals l on X

$$\lim_{i\to\infty} \frac{\langle l, x_{i+1} - \hat{x}\rangle}{\|x_i - \hat{x}\|} = 0, \tag{1.1}$$

i.e. the sequence $(x_{i+1} - \hat{x})\|x_i - \hat{x}\|^{-1}$ converges weakly to 0_X. If specialized to the finite-dimensional case, (1.1) gives the known result, because weak and strong convergence coincide. However, this result seems to indicate that in infinite dimensions the strong superlinear convergence rate, i.e.

$$\lim_{i\to\infty} \frac{\|x_{i+1} - \hat{x}\|}{\|x_i - \hat{x}\|} = 0, \tag{1.2}$$

does not hold. Under the usual assumptions, for another optimization algorithm, similar observations have been made by Fortuna [10], who gave an example of an infinite-dimensional quadratic minimization problem where the conjugate gradient method fails to converge superlinearly. Other Quasi-Newton methods besides Broyden's method have also been investigated in infinite-dimensional spaces: quadratic minimization problems and the DFP-algorithm by Horwitz, Sarachik [13] and Tokumaru, Adachi, and Goto [25]. A linear rate of convergence has been shown for algorithms of the Huang class in Turner and Huntley [26]. Mayorga and Quintana [19] give a sufficient condition for the superlinear rate of convergence of the BFGS-algorithm. However, in this reference, no verification of this condition for the BFGS-algorithm has been carried out. Independently, the author considered in

[12] the Dennis–Moré condition, i.e.

$$\lim_{i\to\infty} \frac{\|(B_i - F'_{\hat{x}})(x_{i+1}-x_i)\|}{\|x_{i+1}-x_i\|} = 0, \tag{1.3}$$

for optimization problems in infinite-dimensional spaces. This is a weaker condition than the one given in [19]. Moreover, similar to the finite-dimensional case proven by Dennis and Moré [8], this condition (1.3) even characterizes the superlinear convergence rate of Quasi-Newton algorithms in optimization. However, the actual verification of (1.3) for Broyden's update formula is a nontrivial task in finite dimensions and even more in infinite-dimensional Hilbert spaces. In general, the weak limit in (1.3), considered as a sequence in the space Y, is O_Y and thus (1.1) can be shown to hold. We show in the third section, that under the assumption, that $B_1 - F'_{\hat{x}}$ is not only small in the operator norm but also an operator of the Hilbert–Schmidt type, the Dennis-Moré condition (1.3) holds and the strong superlinear rate of convergence (1.2) follows for Broyden's method in infinite-dimensional spaces. Specialized to the $\mathbb{R}^n$, this covers the existing theory and hence yields a (second) extension to Hilbert spaces. The proof follows the classical Frobenius norm estimates in [4] and [5]. This theorem can also be interpreted in another way. For this rate of convergence to hold, we require that B_1 is not only close to $F'_{\hat{x}}$ in some norm, but in a particular one, the Hilbert–Schmidt-norm. In finite dimensions, both norms are equivalent, so this problem does not occur. In the last section we give an application to operator equations and two theorems on Hilbert–Schmidt operators.

After the completion of this paper, the author has become aware of related work by Winther [27] and Griewank [11].

2. Linear rate of convergence

A theorem on the linear rate of convergence of Broyden's method can be derived from existing theories such as e.g. [7, 23], where more general methods are investigated.

For given $y \in Y$, $z \in X$, X, Y Hilbert spaces we denote by $y \otimes z$ the rank one operator

$$y \otimes z = \langle z, x\rangle y \quad \text{for all } x \in X.$$

The linear convergence rate theorem can be stated as follows:

Theorem 2.1. *Let X, Y be Hilbert spaces and let $\hat{x} \in X$ be a root of $F: X \to Y$ which is Fréchet-differentiable, with a Lipschitz-continuous Fréchet-derivative $F'_{(\cdot)}$ in a ball of radius ε_L about $\hat{x}$. Suppose that, for some $\gamma > 0$,*

$$\|F'_{\hat{x}}(x)\| \geq \gamma \|x\| \quad \text{for all } x \in X \text{ and range } F'_{\hat{x}} = Y \tag{2.1}$$

holds. For each $\kappa \in (0,1)$ *there exist* $\varepsilon, \varepsilon^* > 0$ *such that if* $x_1 \in X$ *and* $B_1 \in B(X, Y)$ *satisfy*

$$\|x_1 - \hat{x}\| \leq \varepsilon^* \quad \text{and} \quad \|B_1 - F'_{\hat{x}}\| \leq \varepsilon^*, \tag{2.2}$$

then the iterates

$$x_{i+1} = x_i - B_i^{-1} F x_i, \tag{2.3}$$

$$B_{i+1} = B_i + \frac{1}{\|p_i\|^2} F x_{i+1} \otimes p_i, \tag{2.4}$$

$$p_i = x_{i+1} - x_i,$$

are well defined and converge to $\hat{x}$ *at a linear rate,*

$$\|x_{i+1} - \hat{x}\| \leq \kappa \|x_i - \hat{x}\| \tag{2.5}$$

and

$$\|B_i - F'_{\hat{x}}(x)\| \leq \varepsilon \quad \text{for all } i \in \mathbb{N}. \tag{2.6}$$

3. Superlinear rate of convergence

Whereas the theory for the linear rate of convergence gives the same results as for the finite-dimensional case, this does not seem to be true for the problem of a superlinear convergence rate. An extension to the Hilbert space problem has been proved in [23]. If it is specialized to the finite-dimensional case, it yields the known result.

Theorem 3.1. *Let* X, Y *be Hilbert spaces and let* $F: X \to Y$ *be Fréchet-differentiable with Lipschitz-continuous Fréchet-derivative* F'_x *in a ball of radius* ε_L *about the root* $\hat{x}$ *of* F. *Suppose that* $F'_{\hat{x}}$ *satisfies the regularity condition* (2.1). *There exists* $\varepsilon > 0$ *such that if* $x_1 \in X$ *and* $B_1 \in B(X, Y)$ *satisfy*

$$B_1^{-1} \in B(Y, X), \|x_1 - \hat{x}\| \leq \varepsilon \quad \text{and} \quad \|B_1 - F'_{\hat{x}}\| \leq \varepsilon,$$

then the iterates $\{x_i\}$ *defined by* (2.3) *and* (2.4) *converge linearly and*

$$\lim_{i \to \infty} \frac{\langle h, x_{i+1} - \hat{x} \rangle}{\|x_i - \hat{x}\|} = 0 \tag{3.1}$$

for all $h \in X$.

Proof. The linear rate of convergence follows from Theorem 2.1. Since the operators $\{B_i\}$ stay in a neighborhood of $F'_{\hat{x}}$, Theorem 6.3 in [23] is applicable and with the specialization to the case $\phi = \|\cdot\|$ and $G = F$ we deduce

$$\lim_{i \to \infty} \frac{\langle l, (B_i - F'_{\hat{x}}) p_i \rangle}{\|p_i\|} = 0 \quad \text{for all } l \in Y. \tag{3.2}$$

The iteration scheme (2.3) and $F\hat{x}=0$ yield the equality

$$\begin{aligned} F'_{\hat{x}}(x_{i+1}-\hat{x}) &= (F'_{\hat{x}}-B_i)(x_{i+1}-x_i)+F'_{\hat{x}}(x_i-\hat{x})+B_i(x_{i+1}-x_i) \\ &= (F'_{\hat{x}}-B_i)(x_{i+1}-x_i)+F'_{\hat{x}}(x_i-\hat{x})-F\hat{x}-Fx_i \end{aligned}$$

Then, from an application of the mean value theorem [9, (8.6.2)], we obtain, for some Lipschitz-constant $\nu>0$ and all $l\in Y$,

$$\langle l, F'_{\hat{x}}(x_{i+1}-\hat{x})\rangle \leq \langle l, (F'_{\hat{x}}-B_i)(x_{i+1}-x_i)\rangle + \nu\|l\|\,\|x_i-\hat{x}\|^2. \tag{3.3}$$

The linear rate of convergence implies

$$\|x_{i+1}-x_i\| \leq 2\|x_i-\hat{x}\|. \tag{3.4}$$

From (3.3) and (3.4) we deduce

$$\frac{\langle l, F'_{\hat{x}}(x_{i+1}-\hat{x})\rangle}{\|x_i-\hat{x}\|} \leq \frac{2\langle l, (F'_{\hat{x}}-B_i)p_i\rangle}{\|p_i\|} + \nu\|l\|\,\|x_i-\hat{x}\|^2 \tag{3.5}$$

and therefore

$$\lim_{i\to\infty} \frac{\langle l, F'_{\hat{x}}(x_{i+1}-\hat{x})\rangle}{\|x_i-\hat{x}\|} = 0 \quad \text{for all } l\in Y. \tag{3.6}$$

Equation (3.6) and assumption (2.1) imply that (3.1) holds.

Note that the 'weak' superlinear convergence rate (3.1) is identical with the usual statement on a superlinear rate

$$\lim_{i\to\infty} \frac{\|x_{i+1}-\hat{x}\|}{\|x_i-\hat{x}\|} = 0, \tag{3.7}$$

if X is finite-dimensional. The same statement (3.1) has been investigated in [12] for Quasi-Newton methods in Hilbert spaces which are applied to optimization problems.

However, despite the fact that Theorem 3.1 is an extension of the finite-dimensional theory, one would like to know under which conditions the 'strong' superlinear rate of convergence (3.7) holds. As it will be shown in the following theorem, this is the case if the initial approximation B_1 of the derivative $F'_{\hat{x}}$ is close enough in a sense that is stronger than the norm topology (2.2). Before we formulate the theorem we state a few facts from the operator theory in Hilbert spaces.

Let us review some properties of rank-one operators and their composition with other operators which will be used in the proof.

Lemma 3.2. *Let X, Y be Hilbert spaces and $x_i\in X$, $y_i\in Y$, $i=1,2$, $T\in B(X,Y)$. Then the following properties hold, where T^* denotes the adjoint operator:*

(i) $(y_1\otimes x_1)^* = x_1\otimes y_1 \in B(Y,X)$,

(ii) $(y_1\otimes x_1)(x_2\otimes y_2) = \langle x_1, x_2\rangle (y_1\otimes y_2)\in B(Y,Y)$,

(iii) $T(x_1\otimes y_1) = Tx_1\otimes y_1 \in B(Y,Y)$,

(iv) $(y_1\otimes x_1)T^* = y_1\otimes Tx_1 \in B(Y,Y)$,

(v) $\|y_1\otimes x_1\| = \|y_1\|\,\|x_1\|$.

The proofs of these well known properties can be carried out by inspection.

We also need to cite several statements on traces of operators and Hilbert–Schmidt-operators, see e.g. [16, 263].

Definition 3.3. *Let X, Y be Hilbert spaces. An operator $D \in B(X, Y)$ is a Hilbert–Schmidt operator if there exists an orthonormal basis $\{a_\alpha : \alpha \in A\}$ of X such that*

$$\sum_{\alpha \in A} \|Da_\alpha\|_Y^2 \quad \textit{is finite.}$$

Lemma 3.4. *Let X, Y be Hilbert spaces, $x_1, x_2 \in X$, and D_1 a Hilbert–Schmidt operator. Then for the orthonormal basis $\{a_\alpha : \alpha \in A\}$ from Definition 3.3 we define, for each operator $T \in B(X, X)$,*

$$\operatorname{tr} T = \sum_{\alpha \in A} \langle a_\alpha, Ta_\alpha \rangle$$

as long as the sum is finite. Then

(i) $\operatorname{tr} T_i < \infty$ *for* $i = 1, 2$ *implies* $tr(T_1 + T_2) = \operatorname{tr} T_1 + \operatorname{tr} T_2$,

(ii) $\operatorname{tr} x_1 \otimes x_2 = \langle x_1, x_2 \rangle$.

The proof of (i) follows by inspection, whereas for (ii) one only needs to use the definition of rank-one operator. The linear functional tr on $B(X, Y)$ is called the trace of an operator and several more properties can be shown. Since they are not relevant for our application we confine ourselves to the observations mentioned above.

Theorem 3.5. *Let all the assumptions of Theorem 3.1 be satisfied and suppose in addition that $B_1 \in B(X, Y)$ is such that*

$$B_1 - F'_{\hat{x}} \quad \textit{is a Hilbert–Schmidt-operator.}$$

Then the iterates $\{x_i\}_{\mathbb{N}}$ defined by (2.3) and (2.4) converge superlinearly to $\hat{x}$,

$$\lim_{i \to \infty} \frac{\|x_{i+1} - \hat{x}\|}{\|x_i - \hat{x}\|} = 0.$$

Proof. The proof is closely related to the one given by Broyden [4] and Broyden, Dennis and Moré [5] in the finite-dimensional case. Observe that the definition of the update and $y_i = Fx_{i+1} - Fx_i$ imply

$$B_{i+1} - F'_{\hat{x}} = (B_i - F'_{\hat{x}})\left(I - \frac{p_i \otimes p_i}{\langle p_i, p_i \rangle} \right) + \frac{(y_i - F'_{\hat{x}} p_i) \otimes p_i}{\langle p_i, p_i \rangle} \tag{3.8}$$

or, with

$$D_i = B_i - F'_{\hat{x}} \in B(X, Y), \quad e_i = \frac{p_i}{\|p_i\|} \in X, \quad d_i = \frac{y_i - F'_{\hat{x}} p_i}{\|p_i\|} \in Y,$$

we can write (3.8) shorter as

$$D_{i+1}=D_i(I-e_i\otimes e_i)+d_i\otimes e_i.$$

Using properties (i)–(iv) in Lemma 3.2 and $\|e_i\|=1$ we can simplify $D_{i+1}^*D_{i+1}\in B(Y, Y)$ as follows:

$$\begin{aligned}D_{i+1}^*D_{i+1}&=((I-e_i\otimes e_i)D_i^*+e_i\otimes d_i)(D_i(I-e_i\otimes e_i)+d_i\otimes e_i)\\&=(I-e_i\otimes e_i)D_i^*D_i(I-e_i\otimes e_i)+(e_i\otimes d_i)(D_i-D_ie_i\otimes e_i)\\&\quad+(I-e_i\otimes e_i)D_i^*d_i\otimes e_i+\|d_i\|^2e_i\otimes e_i\\&=D_i^*D_i-(e_i\otimes D_i^*D_ie_i+D_i^*D_ie_i\otimes e_i)+\|D_ie_i\|^2e_i\otimes e_i\\&\quad+e_i\otimes D_i^*d_i+D_i^*d_i\otimes e_i+(\|d_i\|^2-2\langle d_i, D_ie_i\rangle)e_i\otimes e_i.\end{aligned}\tag{3.9}$$

By assumption, D_1 is a Hilbert–Schmidt operator, i.e. for some orthonormal system $\{a_\alpha: \alpha\in A\}$ of X we have

$$\sum_{\alpha\in A}\langle a_\alpha, D_1^*D_1a_\alpha\rangle<\infty.$$

Each correction from $D_i^*D_i$ to $D_{i+1}^*D_{i+1}$ in (3.9) is given by a finite sum of rank one operators. Hence by Lemma 3.4 all operators $D_i^*D_i$, $i\in\mathbb{N}$, have a finite trace $\operatorname{tr}(D_i^*D_i)$. Using the linearity of tr and $\operatorname{tr}(x_1\otimes x_2)=\langle x_1, x_2\rangle$, we can simplify (3.9) considerably by taking the trace on both sides:

$$\operatorname{tr}(D_{i+1}^*D_{i+1})=\operatorname{tr}(D_i^*D_i)-\|D_ie_i\|^2+\|d_i\|^2.\tag{3.10}$$

We use (3.10) consecutively and obtain

$$\operatorname{tr}(D_{i+1}^*D_{i+1})=\operatorname{tr}(D_1^*D_1)+\sum_{k=1}^{i}(\|d_k\|^2-\|D_ke_k\|^2.\tag{3.11}$$

By definition $\operatorname{tr}(D_{i+1}^*D_{i+1})$ is a sum of norms and hence nonnegative so that (3.11) yields

$$\sum_{k=1}^{i}\|D_ke_k\|^2\leq\operatorname{tr}(D_1^*D_1)+\sum_{k=1}^{i}\|d_k\|^2.\tag{3.12}$$

With the differentiability requirements imposed on F we can estimate $\|d_i\|$ using a Lipschitz constant $\nu>0$ and an application of the mean value theorem [9, (8.6.2)]

$$\begin{aligned}\|d_i\|&=\|Fx_{i+1}-Fx_i-F'_{\hat{x}}(x_{i+1}-x_i)\|\,\|x_{i+1}-x_i\|^{-1}\\&\leq\nu\|x_i-\hat{x}\|\end{aligned}$$

With the linear convergence rate we obtain then for some $\kappa\in(0,1)$

$$\begin{aligned}\sum_{k=1}^{i}\|d_k\|^2&\leq\sum_{k=1}^{i}\nu^2\|x_k-\hat{x}\|^2\leq\nu^2\|x_1-\hat{x}\|^2\sum_{k=1}^{1}(\kappa^2)^{k-1}\\&\leq\nu^2\|x_1-\hat{x}\|(1-\kappa^2)^{-1}\quad\text{for all } i\in\mathbb{N}.\end{aligned}$$

Hence the sum of the left inequality in (3.12) is also finite

$$\sum_{k=1}^{\infty} \|D_k e_k\|^2 < \infty$$

and in particular

$$\lim_{i\to\infty} \|D_i e_i\| = \frac{\|(B_i - F'_{\hat{x}})p_i\|}{\|p_i\|} = 0. \tag{3.13}$$

This is the Dennis–Moré condition [8] which is known to characterize the superlinear convergence in the finite-dimensional case. For our purpose we can use (3.5) and take the supremum over all elements l in the unit ball on both sides of the inequality to obtain will (2.1)

$$\gamma \frac{\|x_{i+1} - \hat{x}\|}{\|x_i - \hat{x}\|} \leqslant \frac{\|F'_{\hat{x}}(x_{i+1} - \hat{x})\|}{\|x_i - \hat{x}\|}$$

$$\leqslant 2 \frac{\|(B_i - F'_{\hat{x}})p_i\|}{\|p_i\|} + \nu \|x_i - \hat{x}\|^2.$$

Hence (3.13) and the convergence of x_i to $\hat{x}$ yield (3.10).

Remark 3.6. In the proof of Theorem 3.1 we used a result in [23] to state line (3.2). This equality can be proved similar to the proof of Theorem 3.5 by deriving a recursion formula for $D_{i+1}D^*_{i+1}$ and applying $l \in Y$ on both sides of $D_{i+1}D^*_{i+1}$. It should be noted that expression $D_{i+1}D^*_{i+1}$ in contrast to (3.9) is rather short:

$$D_{i+1}D^*_{i+1} = D_i D^*_i - D_i e_i \otimes D_i e_i + d_i \otimes d_i. \tag{3.14}$$

Remark 3.7. In the case that F is an affine operator, i.e. $Fx = Ax + b$ ($A \in B(X, Y)$, $b \in X$), the elements d_i equal θ_x and (3.14) simplifies to

$$D_{i+1}D^*_{i+1} = D_i D^*_i - D_i e_i \otimes D_i e_i. \tag{3.15}$$

The sequence of positive operators $\{D_i D^*_i\}_{\mathbb{N}}$ is a decreasing sequence which is bounded by $\theta_{B(Y,Y)}$ and $D_1 D^*_1$ and hence there exists operator $S \in B(Y, Y)$ which is the strongly convergent limit of the sequence, see Martin [17, p. 100]. This observation is independent on the convergence of the iterates. In the case of a Hilbert-Schmidt operator $D_1 = B_1 - A$, the statement in Theorem 3.5 can be sharpened to obtain the infinite dimensional analogue of a theorem of Moré and Trangenstein [21]. We state a version where the relaxation factor θ_i in [21] is set equal to 1.

Corollary 3.8. *Let X, Y be Hilbert spaces and $\hat{x}$ a root of $Fx = Ax + b = 0$, where $A \in B(X, Y)$ and $b \in X$. Suppose that $F'_{\hat{x}} = A$ is regular, i.e. (2.1) holds. If $x_1 \in X$ and $B_1 \in B(X, Y)$ is such that $B_1 - A$ is a Hilbert-Schmidt operator and if the iterates are well defined, then x_i converges to $\hat{x}$ at a superlinear rate.*

Proof. We specialize the proof of Theorem 3.5 to the case $d_i = \theta_Y$. Then (3.10) reduces to

$$\operatorname{tr}(D_{i+1}^* D_{i+1}) = \operatorname{tr}(D_1^* D_1) - \sum_{k=1}^{i} \|D_k e_k\|^2$$

and we obtain (3.13). This implies with the definition of the iteration rule (2.3) that

$$\lim_{i\to\infty} \frac{\|Fx_{i+1}\|}{\|x_{i+1}-x_i\|} = 0.$$

Hence (2.1) yields

$$0 = \lim_{i\to\infty} \frac{\|Ax_{i+1}+b\|}{\|x_{i+1}-x_i\|} = \lim_{i\to\infty} \frac{\|A(x_{i+1}-\hat{x})\|}{\|x_{i+1}-x_i\|}$$
$$\geqslant \gamma \lim_{i\to\infty} \frac{\|x_{i+1}-\hat{x}\|}{\|x_{i+1}-x_i\|} \geqslant \gamma \lim_{i\to\infty} \left(1 + \frac{\|x_i-\hat{x}\|}{\|x_{i+1}-\hat{x}\|}\right)^{-1}$$

and therefore

$$\lim_{i\to\infty} \frac{\|x_{i+1}-\hat{x}\|}{\|x_i-\hat{x}\|} = 0.$$

Remark 3.9. As one can see from this section, the conditions (3.2) and (3.13) carry the information how well the operators B_i approximate the Fréchet-derivative $F'_{\hat{x}}$ of F at the solution $\hat{x}$. In order to see that (3.2) is in general not sufficient for a superlinear rate of convergence, we state the infinite-dimensional extension of the characterization of the superlinear convergence rate given in the finite-dimensional case by Dennis and Moré [8].

Theorem 3.10. *Let X, Y be Hilbert spaces, $\hat{x} \in X$ a root of $Fx = \theta_Y$ where $F: X \to Y$ is continuously Fréchet-differentiable on an open convex set V containing $\hat{x}$, $\{B_i\}_{\mathbb{N}} \subset B(X, Y)$ a sequence of operators which have inverse operators defined on the range of F, and assume that, for some $\gamma > 0$,*

$$\|F'_{\hat{x}} x\| \geqslant \gamma \|x\| \quad \text{for all } x \in X.$$

Suppose that the sequence of $\{x_i\}_{\mathbb{N}}$ defined by

$$x_1 \in V, \qquad x_{i+1} = x_i - B_i^{-1} F x_i \quad \text{for } i \in \mathbb{N},$$

remains in V and converges to $\hat{x}$. Then

$$\lim_{i\to\infty} \frac{\|x_{i+1}-\hat{x}\|}{\|x_i-\hat{x}\|} = 0$$

holds if and only if

$$\lim_{i\to\infty} \frac{\|(B_i - F'_{\hat{x}})(x_{i+1}-x_i)\|}{\|x_{i+1}-x_i\|} = 0.$$

The proof follows exactly the one given in [8, Theorem 2.2] and is omitted.

4. Applications

Let us consider one application to operator equations in connection with Broyden's method. Obviously, in order to obtain the (strong) superlinear rate of convergence, the initial update B_1 for the derivative of F at the solution needs not only to be close in the operator norm, but it is assumed to differ only by a Hilbert–Schmidt operator. Nonlinear integral equations often can be written in the following way, see e.g. Anselone [2].

Let X be a Hilbert space, Y a Banach space, $G: X \to Y$ a nonlinear operator, $K: Y \to X$ a linear operator. For given $g \in Y$ find $\hat{x} \in X$ such that

$$\hat{x} - KG\hat{x} = g. \tag{4.1}$$

Theorem 4.1. *Let X be a Hilbert space, Y a Banach space, $G: X \to Y$ Fréchet-differentiable with Lipschitz-continuous derivative $G'_{\hat{x}}$ in a ball around the solution $\hat{x}$ of (4.1), $K \in B(Y, X)$ such that $K \circ G'_{\hat{x}}: X \to X$ is a Hilbert–Schmidt operator for all $x \in X$ and 1 is not an eigenvalue of $K \circ G'_{\hat{x}}$. Then there exists $\varepsilon > 0$ such that if $x_1 \in X$ satisfies*

$$\|x_1 - \hat{x}\| \leq \varepsilon$$

and $B_1 \in B(X, X)$ is chosen to be

$$B_1 = I - KG'_{x_1}, \tag{4.2}$$

then the sequence of iterates defined by (2.3) and (2.4) converges at a superlinear rate to $\hat{x}$.

Proof. We define $F: X \to X$ by

$$Fx = x - KGx - g.$$

By the assumption F is Fréchet-differentiable,

$$F'_x = I - KG'_x$$

and the continuity requirements in Theorem 3.1 are met. Furthermore, since 1 is not an eigenvalue of $KG'_{\hat{x}}$ there exists $m > 0$ with

$$\|F'_{\hat{x}}(u)\| = \|(I - KG'_{\hat{x}})u\| \geq m\|u\|$$

for all $u \in U$. By the choice (4.2)

$$B_1 - F'_{\hat{x}} = I - KG'_{x_1} - (I - KG'_{\hat{x}}) = K(G'_{\hat{x}} - G'_{x_1}) \tag{4.3}$$

and

$$\|B_1 - F'_{\hat{x}}\| \leq \|K\| \, \|G'_{\hat{x}} - G'_{x_1}\| \leq \|K\| \kappa \|\hat{x} - x_1\|$$

for some constant $\kappa > 0$. Hence $B_1 - F'_{\hat{x}}$ becomes small if $\|\hat{x} - x_1\|$ is small. Also, by assumption, (4.3) shows that $B_1 - F'_{\hat{x}}$ is a Hilbert-Schmidt operator because of the additivity of this class of operators. Hence, Theorem 3.5 yields the superlinear rate of convergence.

Remark 4.2. If $\|KG'_{\hat{x}}\|$ is small enough, it is sufficient to choose

$$B_1 = I, \tag{4.4}$$

since $B_1 - F'_{\hat{x}} = -KG'_{\hat{x}}$ which is a Hilbert-Schmidt operator. The choice (4.4) is always feasible, if G is linear.

Remark 4.3. Bellman proposed a method for the solution of operator equation which is called 'Quasilinearization': Given $x_i \in X$, solve

$$Fx_i + F'_{x_i}(x_{i+1} - x_i) = 0$$

which is a linear equation in x_{i+1}. For the problem (4.1) this reduces to

$$x_{i+1} - KG'_{x_i}(x_{i+1}) = KGx_i - KG'_{x_i(x_i)} + g \tag{4.5}$$

which is a linear equation. Under similar smoothness assumptions one can show a quadratic rate of convergence for this method. The property that certain mappings are Hilbert-Schmidt operators is not needed here. However, at each iteration step a linear equation (4.5) has to be solved and G'_{x_i} needs to computed. This is not necessary if Broyden's method is used instead and the fast superlinear rate of convergence is maintained if $B_1 - F'_{\hat{x}}$ is of the Hilbert-Schmidt class.

The importance of Hilbert-Schmidt operators for Broyden's method in Hilbert spaces has become evident so that we cite a characterization of these operators on L_2-spaces.

Theorem 4.4 [1, p. 264]. *Let M_1 and M_2 be measurable subsets of $\mathbb{R}^p$ and $\mathbb{R}^q$, resp. The operator $T \in B(L_2(M_1), L_2(M_2))$ is a Hilbert-Schmidt operator if and only if there exists a kernel $k \in L_2(M_2 \times M_1)$ such that, for all $f \in L_2(M_1)$,*

$$Tf(x) = \int_{M_1} k(x, y) f(y)\, \mathrm{d}y \quad \textit{a.e. in } M_2.$$

Another sufficient condition can be found in Agmon [1, p. 211]:

Theorem 4.5. *Let M be a bounded and open subset of $\mathbb{R}^n$ having the segment property and the ordinary cone property (see [1, p. 11]). If $T \in B(L_2(M), L_2(M))$ has the property that*

$$Ty \in H_m(M) \quad \textit{for all } y \in L_2(M),$$

where $H_m(M)$ is the subspace of L_2 where all functions have strong L_2-derivatives of order up to m, and if $2m > n$, then T is a Hilbert-Schmidt operator.

References

[1] S. Agmon, *Lectures on elliptic boundary value problems* (Van Nostrand, Princeton, 1965).

[2] P.M. Anselone, *Collectively compact operator approximation theory and applications to integral equations* (Prentice-Hall, Englewood Cliffs, 1971).

[3] C.G. Broyden, "A class of methods for solving nonlinear simultaneous equations", *Mathematics of Computation* 19(1965) 577–593.

[4] C.G. Broyden, "The convergence of single-rank quasi-Newton methods", *Mathematics of Computation* 24(1970) 365–382.

[5] C.G. Broyden, J.E. Dennis, Jr., and J.J. Moré, "On the local and superlinear convergence of quasi-Newton methods", *Journal of Institute of Mathematics and Applications* 12(1973) 223–246.

[6] D.W. Decker and C.T. Kelley, "Broyden's method for a class of problems having singular Jacobian at the root", *SIAM Journal Control and Optimization* 22(1985) 566–574.

[7] J.E. Dennis, Jr., "Toward a unified convergence theory for Newton-like methods", in: L.B. Rall, ed., *Nonlinear functional analysis and applications* (Academic Press, New York, London, 1971) pp. 425–472.

[8] J.E. Dennis, Jr. and J.J. Moré, "A characterization of superlinear convergence and its application to quasi-Newton methods", *Mathematics of Computation* 28(1974) 543–560.

[9] J. Dieudonné, *Foundations of modern analysis* (Academic Press, New York and London, 1960).

[10] Z. Fortuna, "Some convergence properties of the conjugate gradient method in Hilbert space", *SIAM Journal on Numerical Analysis* 16(1979) 380–384.

[11] A Griewank, "The superlinear convergence of secant method on mildly non-linear problems in Hilbert space", preprint, Southern Methodist University (Dallas, August 1984).

[12] W.A. Gruver and E. Sachs, *Algorithmic methods in optimal control* (Pitman, Boston, London, Melbourne, 1981).

[13] L.B. Horwitz and P.E. Sarachik, "Davidon's methods in Hilbert space", *SIAM Journal on Applied Mathematics* 16(1968) 676–695.

[14] *L.V. Kantorovich*, "Functional analysis and applied mathematics", *Upspekhi Matematiki Nauk* 3(1948) 89–185.

[15] L.V. Kantorovich and G.P. Akilov, *Functional analysis in normed spaces* (MacMillan, New York, 1964).

[16] T. Kato, *Perturbation theory for linear operators* (Springer-Verlag, New York, 1966).

[17] R.H. Martin, *Nonlinear operators and differential equations in Banach spaces* (Wiley, New York-Chichester-Brisbane-Toronto, 1976).

[18] K. Madsen, "Minimax solutions of nonlinear equations without calculating derivatives", *Mathematical Programming Study* 3(1975) 110–126.

[19] R.V. Mayorga and V.H. Quintana, "A family of variable metric methods in function space, without exact line searches", *Journal of Optimization Theory and Applications* 31 (1980) 303–329.

[20] R.H. Moore, "Approximations to nonlinear operator equations and Newton's methods", *Numerische Mathematik* 12(1968) 23–34.

[21] J.J. Moré and J.A. Trangenstein, "On the global convergence of Broyden's method", *Mathematics of Computation* 30(1976) 523–540.

[22] E. Sachs, "Global convergence of quasi-Newton-type algorithms for some nonsmooth optimization problems", *Journal of Optimization Theory and Applications* 40(1983) 201–219.

[23] E. Sachs, "Convergence rates of quasi-Newton algorithms for some nonsmooth optimization problems", *SIAM Journal on Control and Optimization* 23(1985) 401–418.

[24] M.J. Todd, "Quasi-Newton updates in abstract vector spaces", *SIAM Review* 26(1984) 367–377.

[25] H. Tokumaru, N. Adachi, and K. Goto, "Davidon's method for minimization problems in Hilbert space with an application to control problems", *SIAM Journal on Control and Optimization* 8(1970) 163–178.

[26] P.R. Turner and E. Huntley, "Variable metric methods in Hilbert space with applications to control problems", *Journal of Optimization Theory and Applications* 19(1976) 381–400.

[27] R. Winther, "A Numerical Galerkin Method for a Parabolic Problem", Ph.D. Thesis, Cornell University (Ithaca, NY, 1977).