Convergence and Applications of Newton-type Iterations

Ioannis K. Argyros

Convergence and Applications of Newton-type Iterations

Springer

Ioannis K. Argyros
Department of Mathematical Sciences
Cameron University
Lawton, OK 73505
USA
ioannisa@cameron.edu

ISBN: 978-0-387-72741-7 e-ISBN: 978-0-387-72743-1
DOI: 10.1007/978-0-387-72743-1

Library of Congress Control Number: 2007935498

Mathematics Subject Classification (2000): 65G99, 65K10, 65B05, 90C30, 90C33, 47J20, 47H04, 49M15

Printed on acid-free paper.

9 8 7 6 5 4 3 2 1

springer.com

Dedicated to my father Konstantinos

Contents

Introduction

Researchers in computational sciences are faced with the problem of solving a variety of equations. A large number of problems are solved by finding the solutions of certain equations. For example, dynamic systems are mathematically modeled by difference or differential equations, and their solutions represent usually the states of the systems. For the sake of simplicity, assume that a time-invariant system is driven by the equation $x' = f(x)$, where x is the state, then the equilibrium states are determined by solving the equations $f(x) = 0$. Similar equations are used in the case of discrete systems. The unknowns of engineering equations can be functions (difference, differential, integral equations), vectors (systems of linear or nonlinear algebraic equations), or real or complex numbers (single algebraic equations with single unknowns). Except special cases, the most commonly used solutions methods are iterative; when starting from one or several initial approximations, a sequence is constructed, which converges to a solution of the equation. Iteration methods are applied also for solving optimization problems. In such cases, the iteration sequences converge to an optimal solution of the problem in hand. Because all of these methods have the same recursive structure, they can be introduced and discussed in a general framework.

To complicate the matter further, many of these equations are nonlinear. However, all may be formulated in terms of operators mapping a linear space into another, the solutions being sought as points in the corresponding space. Consequently, computational methods that work in this general setting for the solution of equations apply to a large number of problems and lead directly to the development of suitable computer programs to obtain accurate approximate solutions to equations in the appropriate space.

This monograph is written with optimization considerations including the weakening of existing hypotheses for solving equations. It can also be used as a reference book for an advanced numerical-functional analysis course. The goal is to introduce these powerful concepts and techniques at the earliest possible stage. The reader is assumed to have had courses in numerical functional analysis and linear algebra.

We have divided the material into 11 chapters. Each chapter contains several new theoretical results and important applications in engineering, in dynamic economic systems, in input-output systems, in the solution of nonlinear and linear differential equations, and optimization problems. The applications appear in the form of Examples or Applications or Exercises or they are implied as our results improve (weaken) (extend the applicability of) earlier ones that have already been applied in concrete problems. Sections have been written as independent of each other as possible. Hence the interested reader can go directly to a certain section and understand the material without having to go back and forth in the whole textbook to find related material.

There are four basic problems connected with iterative methods.

Problem 1: Show that the iterates are well defined. For example, if the algorithm requires the evaluation of F at each x_n, it has to be guaranteed that the iterates remain in the domain of F. It is, in general, impossible to find the exact set of all initial data for which a given process is well defined, and we restrict ourselves to giving conditions that guarantee that an iteration sequence is well defined for certain specific initial guesses.

Problem 2: Concerns the convergence of the sequences generated by a process and the question of whether their limit points are, in fact, solutions of the equation. There are several types of such convergence results. The first, which we call a local convergence theorem, begins with the assumption that a particular solution x^* exists, and then asserts that there is a neighborhood U of x^* such that for all initial vectors in U the iterates generated by the process are well defined and converge to x^*. The second type of convergence theorem, which we call semilocal, does not require knowledge of the existence of a solution, but states that, starting from initial vectors for which certain—usually stringent—conditions are satisfied, convergence to some (generally nearby) solutions x^* is guaranteed. Moreover, theorems of this type usually include computable (at least in principle) estimates for the error $x_n - x^*$, a possibility not afforded by the local convergence theorems. Finally, the third and most elegant type of convergence result, the global theorem, asserts that starting anywhere in a linear space, or at least in a large part of it, convergence to a solution is ensured.

Problem 3: Concerns the economy of the entire operations and, in particular, the question of how fast a given sequence will converge. Here, there are two approaches, which correspond with the local and semilocal convergence theorems. As mentioned above, the analysis that leads to the semilocal type of theorem frequently produces error estimates, and these, in turn, may sometimes be reinterpreted as estimates of the rate of convergence of the sequence. Unfortunately, however, these are usually overly pessimistic. The second approach deals with the behavior of the sequence $\{x_n\}$ when n is large, and hence when x_n is near the solutions x^*. This behavior may then be determined, to a first approximation, by the properties of the iteration function near x^* and leads to so-called asymptotic rates of convergence.

Problem 4: Concerns with how to best choose a method, algorithm, or software program to solve a specific type of problem and its descriptions of when a given algorithm or method succeeds or fails.

We have included a variety of new results dealing with Problems 1–4.

This monograph is an outgrowth of research work undertaken by us and complements/updates earlier works of ours focusing on in-depth treatment of convergence theory for iterative methods [7]–[43]. Such a comprehensive study of optimal iterative procedures appears to be needed and should benefit not only those working in the field but also those interested in, or in need of, information about specific results or techniques. We have endeavored to make the main text as self-contained as possible, to prove all results in full detail, and to include a number of exercises throughout the monograph. In order to make the study useful as a reference source, we have complemented each section with a set of "Remarks" in which literature citations are given, other related results are discussed, and various possible extensions of the results of the text are indicated. For completion, the monograph ends with a comprehensive list of references. Because we believe our readers come from diverse backgrounds and have varied interests, we provide "recommended reading" throughout the textbook. Often a long textbook summarizes knowledge in a field. This monograph, however, may be viewed as a report on work in progress. We provide a foundation for a scientific field that is rapidly changing. Therefore we list numerous conjectures and open problems as well as alternative models that need to be explored.

The monograph is organized as follows:

Chapter 1: The essentials on the solution of equations are provided.

Newton-type methods and their implications/applications are covered in the rest of the chapters.

The Newton-Kantorovich Theorem 2.2.4 for solving nonlinear equations is one of the most important tools in nonlinear analysis and in classic numerical analysis. This theorem has been successfully used for obtaining optimal bounds for many iterative procedures. The original paper or Kantorovich [124] contains optimal a priori bounds for the Newton-Kantorovich (NK) method (2.1.3), albeit not in explicit form. Explicit forms of those a priori bounds were obtained independently by Ostrowski [155], Gragg and Tapia [102].

The paper of Gragg and Tapia [102] also contains sharp a posteriori bounds for the NK method. By using different techniques and/or different a posteriori information, these bounds were refined by others [6], [53], [58], [59], [64], [74], [76]–[78], [128], [135], [139]–[142], [154], [162], [167], [184], [191], [209]–[212], [214]–[216], [218]–[220], and us [11]–[43]. Various extensions of the NK theorem also have been used to obtain error bounds for Newton-like (or Newton-type) methods: Inexact Newton method, the secant method, Halley's method, etc. A survey of such methods can be found in [26], [43].

The NK theorem has also been used in concrete applications for proving existence and uniqueness of solutions for nonlinear equations arising in various fields. The spectrum of applications of this theorem is immense. An Internet search seeking "Newton-Kantorovich Theorem" leads to hundreds if not thousands of works related/based on this theorem.

The list given below is therefore incomplete. However, we have included diverse problems such as the NK method on a cone, Robinson [178] (Section 2.6); the weak NK method, Tapia [188] (Section 2.4); bounds on manifolds, Argyros [39], Paardekooper [156] (Section 2.10); radius of convergence and one-parameter imbedding Meyer [139] (Section 2.11); NK method on Riemannian manifolds, Ferreira and Svaiter [94] (Section 2.12); shadowing orbits in dynamical systems, Hadeller [108] (Section 2.13); computation of continuation curves, Deuflhard, Pesh, Rentrop [77], Rheinboldt [176] (Section 2.14); Moore's theorem [143] from interval analysis, Rall [171], Neumaier and Shen [146], Zuhe and Wolfe [220] (Section 3.1); Miranda's theorem [142] for enclosing solutions of equations, Mayer [136] (Section 3.2); point-based approximation (PBA) used successfully by Robinson [179], [180] in Mathematical Programming (Section 3.3); curve tracing, Allgower [2], Chu [62], Rheinboldt [176] (Section 3.4); finite element analysis for boundary value problems, Tsuchiya [194], Pousin [168], Feinstauer-Zernicek [93] (Section 3.5); PSB updates in Hilbert spaces using quasi-NK method, Laumen [134] (Section 3.6); shadowing Lemma and chaotic behavior for nonlinear equations, Palmer [156], Stoffer [186] (Section 3.7); mesh independence principle for optimal design problems, Laumen [133], Allgower, Böhmer, Potra, Rheinboldt [2] (Section 3.8); conditioning of semidefinite programs, Nayakkankuppam [144], Alizadeh [1], Haeberly [109] (Section 3.9); analytic complexity/enlarging the set of initial guesses for the NK method, Kung [131], Traub [192] (Chapter 6, Sections 6.1, 6.2, 6.3); interior point methods, Potra [165] (Section 11.1); LP methods, Rheinboldt [177], Wang-Zhao [206], Renegar-Shub [174], Smale [184] (Section 11.2).

The foundation of the NK theorem is famous for its simplicity and clarity of NK hypothesis (2.2.17) (or (2.2.37) in affine invariant form).

This hypothesis is the crucial sufficient condition for the convergence of Newton's method. However, convergence of Newton's method can be obtained even if the NK hypothesis is violated (see, e.g., Example 2.2.14). Therefore weakening this condition is of extreme importance because the applicability of this powerful method will be extended. Recently we showed [39] by considering more precise majorizing sequences that the NK hypothesis can always be replaced by the weaker (2.2.52) (if $\ell_0 \neq \ell$) (see also Theorem 2.2.11) which doubles (at most if $\ell_0 = 0$) the applicability of this theorem. Note that the verification of condition (2.2.56) requires the same information and computational cost as (2.2.37) because in practice the computation of Lipschitz constant ℓ requires the evaluation of center-Lipschitz constant ℓ_0 too.

Moreover the following advantages hold (see Theorem 2.2.11 and the Remarks that follow): <u>semilocal case</u>: finer error estimates on the distances involved and an at least as precise information on the location of the solution; <u>local case</u>: finer error bounds and larger trust regions (radius of convergence).

The following advantages carry over if our approach is extended to related methods/hypotheses: Below we provide a list: secant method, Argyros [12], [43], Dennis [74], Potra [162], Hernandez [116], [117] (Section 2.3); "Terra Incognita" and Hölder continuity, Argyros [32], [35], Lysenko [135], Ciancarruso, De Pascale [64] (Section 2.4); NK method under regular smoothness conditions, Galperin [98], Galperin and Waksman [99] (Section 2.5); enlarging the radius of convergence for the NK method using hypotheses on the m ($m > 1$ an integer) Fréchet-differentiable operators, Argyros [27], [43], Ypma [216] (Section 2.8); Gauss-Newton method, Ben-Israel [46], Häussler [110] (Section 2.15); Broyden's method [52], Dennis [75]

(Section 4.1); Stirling's method [185], Rall [170] (Section 4.2); Steffenssen-Aitken method, Catinas [54], Pavaloiu [158], [159]; method of tangent hyperbolas, Kanno [123], Yamamoto [211] (Section 4.5); modified secant method with applications in function optimization, Amat, Busquier, Gutierrez [4], Bi, Ren, Wu [47], Ren [172] (Section 4.6); the King-Werner method, Ren [172]; Newton methods (including two-point), Argyros [34], [35], [43] Dennis [74], [75], Chen, Yamamoto, [58], [59], [60], (in Chapters 5 and 8); variational inequalities in Chapter 7, K-theory and convergence on generalized Banach spaces with a convergence structure, Caponetti, De Pascale, Zabrejko [53], Meyer [139]–[141] in Chapter 9, and extensions to set-to-set mappings in Chapter 10.

Earlier results by us or others are included in sections mentioned above directly or indirectly as special cases of our results. Note that revisiting all results to date that have used the NK hypothesis (2.2.37) and replacing (2.2.37) with our weaker hypothesis (2.2.56) is worth it for the reasons/benefits mentioned above. However, this will be an enormous or even impossible task. That is why in this monograph we decided to include only the above chapters and leave the rest for the motivated reader. Note that some results are also listed as exercises to reduce the size of the book.

Finally we state that although the refinement of majorizing sequences technique inaugurated by us in [39] is very recent, several authors have already succesfully used it: Amat, Busquier, Gutierrez [4] (see Section 4.8), Bi, Ren, Wu [47] (see Section 4.6), and Ren [172] (see Section 4.7).

1

Operators and Equations

The basic background for solving equations is introduced here.

1.1 Operators on linear spaces

Some mathematical operations have certain properties in common. These properties are given in the following definition.

Definition 1.1.1. *An operator T that maps a linear space X into a linear space Y over the same scalar field S is said to be additive if*

$$T(x+y) = T(x) + T(y), \qquad \textit{for all } x, y \in X,$$

and homogeneous if

$$T(sx) = sT(x), \qquad \textit{for all } x \in X, s \in S.$$

An operator that is additive and homogeneous is called a linear operator.

Many examples of linear operators exist.

Example 1.1.2. Define an operator T from a linear space X into itself by $T(x) = sx$, $s \in S$. Then T is a linear operator.

Example 1.1.3. The operator $D = \frac{d}{dt}$ mapping $X = C^1[0,1]$ into $Y = C[0,1]$ given by

$$D(x) = \frac{dx}{dt} = y(t), \ 0 \leq t \leq 1,$$

is linear.

If X and Y are linear spaces over the same scalar field S, then the set $L(X, Y)$ containing all linear operators from X into Y is a linear space over S if addition is defined by

I.K. Argyros, *Convergence and Applications of Newton-type Iterations*,
DOI: 10.1007/978-0-387-72743-1_1, © Springer Science+Business Media, LLC 2008

$$(T_1 + T_2)(x) = T_1(x) + T_2(x), \qquad \text{for all } x \in X,$$

and scalar multiplication by

$$(sT)(x) = s(T(x)), \qquad \text{for all } x \in X, \ s \in S.$$

We may also consider linear operators B mapping X into $L(X, Y)$. For an $x \in X$ we have

$$B(x) = T,$$

a linear operator from X into Y. Hence, we have

$$B(x_1, x_2) = (B(x_1))(x_2) = y \in Y.$$

B is called a bilinear operator from X into Y. The linear operators B from X into $L(X, Y)$ form a linear space $L(X, L(X, Y))$. This process can be repeated to generate j-linear operators ($j > 1$ an integer).

Definition 1.1.4. *A linear operator mapping a linear space X into its scalar S is called a linear functional in X.*

Definition 1.1.5. *An operator Q mapping a linear space X into a linear space Y is said to be nonlinear if it is not a linear operator from X into Y.*

Some metric concepts of importance are now introduced.

Definition 1.1.6. *An operator F from a Banach space X into a Banach space Y is continuous at $x = x^*$ if*

$$\lim_{n\to\infty} \left\| x_n - x^* \right\|_X = 0 \implies \lim_{n\to\infty} \left\| F(x_n) - F\left(x^*\right) \right\|_Y = 0$$

Theorem 1.1.7. *If a linear operator T from a Banach space X into a Banach space Y is continuous at $x^* = 0$, then it is continuous at every point x of space X.*

Proof. We have $T(0) = 0$, and from $\lim_{n\to\infty} \|x_n\| = 0$ we get $\lim_{n\to\infty} \|T(x_n)\| = 0$. If sequence $\{x_n\}$ $(n \geq 0)$ converges to x^* in X, by setting $y_n = x_n - x^*$ we obtain $\lim_{n\to\infty} \|y_n\| = 0$. By hypothesis this implies that

$$\lim_{n\to\infty} \|T(x_n)\| = \lim_{n\to\infty} \left\|T\left(x_n - x^*\right)\right\| = \lim_{n\to\infty} \left\|T(x_n) - T\left(x^*\right)\right\| = 0.$$

Definition 1.1.8. *An operator F from a Banach space X into a Banach space Y is Lipschitz continuous on the set A in X if there exists a constant $c < \infty$ such that*

$$\|F(x) - F(y)\| \leq c\,\|x - y\|, \qquad \textit{for all } x, y \in A.$$

The greatest lower bound (infimum) of numbers c satisfying the above inequality for $x \neq y$ is called the bound of F on A. An operator that is bounded on a ball (open) $U(z, r) = \{x \in X \mid \|x - z\| < r\}$ is continuous at z. It turns out that for linear operators, the converse is also true.

Theorem 1.1.9. *A continuous linear operator T from a Banach space X into a Banach space Y is bounded on X.*

Proof. By the continuity of T there exists $\varepsilon > 0$ such that $\|T(z)\| < 1$, if $\|z\| < \varepsilon$. For $0 \neq z \in X$

$$\|T(z)\| \leq \frac{1}{\varepsilon}\|z\|, \tag{1.1.1}$$

because $\|cz\| < \varepsilon$ for $|c| < \frac{\varepsilon}{\|z\|}$, and $\|T(cz)\| = |c| \cdot \|T(z)\| < 1$. Letting $z = x - y$ and $c = \varepsilon^{-1}$ in (1.1.1), we conclude that operator T is bounded on X.

The bound on X of a linear operator T denoted by $\|T\|_X$ or simply $\|T\|$ is called the norm of T. As in Theorem 1.1.9 we get

$$\|T\| = \sup_{\|x\|=1} \|T(x)\|. \tag{1.1.2}$$

Hence, for any bounded linear operator T

$$\|T(x)\| \leq \|T\| \cdot \|x\|, \qquad \text{for all } x \in X. \tag{1.1.3}$$

From now on, $L(X, Y)$ denotes the set of all bounded linear operators from a Banach space X into another Banach space Y. It also follows immediately that $L(X, Y)$ is a linear space if equipped with the rules of addition and scalar multiplication introduced in Definition 1.1.1.

The proof of the following result is left as an exercise (see also [119], [125]).

Theorem 1.1.10. *The set $L(X, Y)$ is a Banach space for the norm (1.1.2).*

In a Banach space X, solving a linear equation can be stated as follows: given a bounded linear operator T mapping X into itself and some $y \in X$, find an $x \in X$ such that

$$T(x) = y. \tag{1.1.4}$$

The point x (if it exists) is called a solution of Equation (1.1.4).

Definition 1.1.11. *If T is a bounded linear operator in X and a bounded linear operator T_1 exists such that*

$$T_1 T = T T_1 = I, \tag{1.1.5}$$

where I is the identity operator in X (i.e., $I(x) = x$ for all $x \in X$), then T_1 is called the inverse of T and we write $T_1 = T^{-1}$. That is,

$$T^{-1} T = T T^{-1} = I. \tag{1.1.6}$$

If T^{-1} exists, then Equation (1.1.4) has the unique solution

$$x = T^{-1}(y). \tag{1.1.7}$$

The proof of the following result is left as an exercise (see also [130]).

Theorem 1.1.12. *(Banach Lemma on Invertible Operators) [125]. If T is a bounded linear operator in X, T^{-1} exists if and only if there is a bounded linear operator P in X such that P^{-1} exists and*

$$\|I - PT\| < 1. \tag{1.1.8}$$

If T^{-1} exists, then

$$T^{-1} = \sum_{n=0}^{\infty} (I - PT)^n P \qquad \textit{(Neumann Series)} \tag{1.1.9}$$

and

$$\left\|T^{-1}\right\| \leq \frac{\|P\|}{1 - \|I - PT\|}. \tag{1.1.10}$$

Based on Theorem 1.1.12, we can immediately introduce a computational theory for Equation (1.1.4) composed by three factors:

(A) *Existence and Uniqueness*. Under the hypotheses of Theorem 1.1.12, Equation (1.1.4) has a unique solution x^*.

(B) *Approximation*. The iteration

$$x_{n+1} = P(y) + (I - PT)(x_n) \quad (n \geq 0) \tag{1.1.11}$$

gives a sequence $\{x_n\}$ $(n \geq 0)$ of successive approximations, which converges to x^* for any initial guess $x_0 \in X$.

(C) *Error Bounds*. Clearly the speed of convergence of iteration $\{x_n\}$ $(n \geq 0)$ to x^* is governed by the estimate:

$$\|x_n - x^*\| \leq \frac{\|I - PT\|^n}{1 - \|I - PT\|} \|P(y)\| + \|I - PT\|^n \|x_0\|. \tag{1.1.12}$$

Let T be a bounded linear operator in X. One way to obtain an approximate inverse is to make use of an operator sufficiently close to T.

Theorem 1.1.13. *If T is a bounded linear operator in X, T^{-1} exists if and only if there is a bounded linear operator P_1 in X such that P_1^{-1} exists, and*

$$\|P_1 - T\| \leq \left\|P_1^{-1}\right\|^{-1}. \tag{1.1.13}$$

If T^{-1} exists, then

$$T^{-1} = \sum_{n=0}^{\infty} \left(I - P_1^{-1}T\right)^n P_1^{-1} \tag{1.1.14}$$

and

$$\left\|T^{-1}\right\| \leq \frac{\|P^{-1}\|}{1 - \left\|I - P_1^{-1}T\right\|} \leq \frac{\left\|P_1^{-1}\right\|}{1 - \left\|P_1^{-1}\right\| \|P_1 - T\|}. \tag{1.1.15}$$

Proof. Let $P = P_1^{-1}$ in Theorem 1.1.12 and note that by (1.1.13)

$$\left\| I - P_1^{-1} T \right\| = \left\| P_1^{-1} (P_1 - T) \right\| \leq \left\| P_1^{-1} \right\| \cdot \| P_1 - T \| < 1. \tag{1.1.16}$$

That is, (1.1.8) is satisfied. The bounds (1.1.15) follow from (1.1.10) and (1.1.16). That proves the sufficiency. The necessity is proved by setting $P_1 = T$, if T^{-1} exists.

The following result is equivalent to Theorem 1.1.12.

Theorem 1.1.14. *A bounded linear operator T in a Banach space X has an inverse T^{-1} if and only if linear operators P, P^{-1} exist such that the series*

$$\sum_{n=0}^{\infty} (I - PT)^n P \tag{1.1.17}$$

converges. In this case we have

$$T^{-1} = \sum_{n=0}^{\infty} (I - PT)^n P.$$

Proof. If series (1.1.17) converges, then it converges to T^{-1} (see Theorem 1.1.12). The existence of P, P^{-1} and the convergence of series (1.1.17) is again established as in Theorem 1.1.12, by taking $P = T^{-1}$, when it exists.

Definition 1.1.15. *A linear operator N in a Banach space X is said to be nilpotent if*

$$N^m = 0, \tag{1.1.18}$$

for some positive integer m.

Theorem 1.1.16. *A bounded linear operator T in a Banach space X has an inverse T^{-1} and only if there exist linear operators P, P^{-1} such that $I - PT$ is nilpotent.*

Proof. If P, P^{-1} exists and $I - PT$ is nilpotent, then series

$$\sum_{n=0}^{\infty} (I - PT)^n P = \sum_{n=0}^{m-1} (I - PT)^n P$$

converges to T^{-1} by Theorem 1.1.14. Moreover, if T^{-1} exists, then $P = T^{-1}$, $P^{-1} = T$ exists, and $I - PT = I - T^{-1}T = 0$ is nilpotent.

The computational techniques to be considered later make use of the derivative in the sense of Fréchet [125], [204].

Definition 1.1.17. *Let F be an operator mapping a Banach space X into a Banach space Y. If there exists a bounded linear operator L from X into Y such that*

$$\lim_{\|\Delta x\| \to 0} \frac{\|F(x_0 + \Delta x) - F(x_0) - L(\Delta x)\|}{\|\Delta x\|} = 0, \tag{1.1.19}$$

then F is said to be Fréchet-differentiable at x_0, and the bounded linear operator

$$F'(x_0) = L \tag{1.1.20}$$

is called the first Fréchet derivative of F at x_0. The limit in (1.1.19) is supposed to hold independently of the way that Δx approaches 0. *Moreover, the Fréchet differential*

$$\delta F(x_0, \Delta x) = F'(x_0)\,\Delta x \tag{1.1.21}$$

is an arbitrary close approximation to the difference $F(x_0 + \Delta x) - F(x_0)$ relative to $\|\Delta x\|$, for $\|\Delta x\|$ small.

If F_1 and F_2 are differentiable at x_0, then

$$(F_1 + F_2)'(x_0) = F_1'(x_0) + F_2'(x_0). \tag{1.1.22}$$

Moreover, if F_2 is an operator from a Banach space X into a Banach space Z, and F_1 is an operator from Z into a Banach space Y, their composition $F_1 \circ F_2$ is defined by

$$(F_1 \circ F_2)(x) = F_1(F_2(x)), \qquad \text{for all } x \in X. \tag{1.1.23}$$

It follows from Definition 1.1.17 that $F_1 \circ F_2$ is differentiable at x_0 if F_2 is differentiable at x_0 and F_1 is differentiable at $F_2(x_0)$ of Z, with (chain rule):

$$(F_1 \circ F_2)'(x_0) = F_1'(F_2(x_0))F_2'(x_0). \tag{1.1.24}$$

In order to differentiate an operator F we write:

$$F(x_0 + \Delta x) - F(x_0) = L(x_0, \Delta x)\Delta x + \eta(x_0, \Delta x), \tag{1.1.25}$$

where $L(x_0, \Delta x)$ is a bounded linear operator for given x_0, Δx with

$$\lim_{\|\Delta x\| \to 0} L(x_0, \Delta x) = L, \tag{1.1.26}$$

and

$$\lim_{\|\Delta x\| \to 0} \frac{\|\eta(x_0, \Delta x)\|}{\|\Delta x\|} = 0. \tag{1.1.27}$$

Estimates (1.1.26) and (1.1.27) give

$$\lim_{\|\Delta x\| \to 0} L(x_0, \Delta x) = F'(x_0). \tag{1.1.28}$$

If $L(x_0, \Delta x)$ is a continuous function of Δx in some ball $U(0, R)$ $(R > 0)$, then

$$L(x_0, 0) = F'(x_0). \tag{1.1.29}$$

Higher-order derivatives can be defined by induction:

Definition 1.1.18. *If F is $(m-1)$-times Fréchet-differentiable ($m \geq 2$ an integer), and an m-linear operator A from X into Y exists such that*

$$\lim_{\|\Delta x\| \to 0} \frac{\left\| F^{(m-1)}(x_0 + \Delta x) - F^{(m-1)}(x_0) - A(\Delta x) \right\|}{\|\Delta x\|} = 0, \tag{1.1.30}$$

then A is called the m-Fréchet derivative of F at x_0, and

$$A = F^{(m)}(x_0) \tag{1.1.31}$$

Higher partial derivatives in product spaces can be defined as follows: Define

$$X_{ij} = L(X_j, X_i), \tag{1.1.32}$$

where $X_1, X_2, \ldots$ are Banach spaces and $L(X_j, X_i)$ is the space of bounded linear operators from X_j into X_i. The elements of X_{ij} are denoted by L_{ij}, etc. Similarly,

$$X_{ijm} = L(X_m, X_{ij}) = L(X_m, L(X_j, X_i)) \tag{1.1.33}$$

denotes the space of bounded bilinear operators from X_k into X_{ij}. Finally, we write

$$X_{ij_1j_2\cdots j_m} = L\left(X_{jk}, X_{ij_1j_2\cdots j_{m-1}}\right), \tag{1.1.34}$$

which denotes the space of bounded linear operators from X_{jm} into $X_{ij_1j_2\cdots j_{m-1}}$. The elements $A = A_{ij_1j_2\cdots j_m}$ of $X_{ij_1j_2\cdots j_m}$ are a generalization of m-linear operators [10], [125].

Consider an operator F_i from space

$$X = \prod_{p=1}^{n} X_{j_p} \tag{1.1.35}$$

into X_i, and that F_i has partial derivatives of orders $1, 2, \ldots, m-1$ in some ball $U(x_0, R)$, where $R > 0$ and

$$x_0 = \left(x_{j_1}^{(0)}, x_{j_2}^{(0)}, \ldots, x_{j_n}^{(0)}\right) \in X. \tag{1.1.36}$$

For simplicity and without loss of generality we renumber the original spaces so that

$$j_1 = 1, j_2 = 2, \ldots, j_n = n. \tag{1.1.37}$$

Hence, we write

$$x_0 = (x_1^{(0)}, x_2^{(0)}, \ldots, x_n^{(0)}). \tag{1.1.38}$$

A partial derivative of order $(m-1)$ of F_i at x_0 is an operator

$$A_{iq_1q_2\cdots q_{m-1}} = \frac{\partial^{(m-1)} F_i(x_0)}{\partial x_{q_1} \partial x_{q_2} \cdots \partial x_{q_{m-1}}} \tag{1.1.39}$$

(in $X_{iq_1q_2\cdots q_{m-1}}$) where

$$1 \leq q_1, q_2, \dots, q_{m-1} \leq n. \tag{1.1.40}$$

Let $P(X_{q_m})$ denote the operator from X_{q_m} into $X_{iq_1q_2\cdots q_{m-1}}$ obtained from (1.1.39) by letting

$$x_j = x_j^{(0)}, \quad j \neq q_m, \tag{1.1.41}$$

for some q_m, $1 \leq q_m \leq n$. Moreover, if

$$P'(x_{q_m}^{(0)}) = \frac{\partial}{\partial x_{q_m}} \cdot \frac{\partial^{m-1} F_i(x_0)}{\partial x_{q_1} \partial x_{q_2} \cdots \partial x_{q_{m-1}}} = \frac{\partial^m F_i(x_0)}{\partial x_{q_1} \cdots \partial x_{q_m}}, \tag{1.1.42}$$

exists, it will be called the partial Fréchet derivative of order m of F_i with respect to $x_{q_1}, \dots, x_{q_m}$ at x_0.

Furthermore, if F_i is Fréchet-differentiable m times at x_0, then

$$\frac{\partial^m F_i(x_0)}{\partial x_{q_1} \cdots \partial x_{q_m}} x_{q_1} \cdots x_{q_m} = \frac{\partial^m F_i(x_0)}{\partial x_{s_1} \partial x_{s_2} \cdots \partial x_{s_m}} x_{s_1} \cdots x_{s_m} \tag{1.1.43}$$

for any permutation $s_1, s_2, \dots, s_m$ of integers $q_1, q_2, \dots, q_m$ and any choice of points $x_{q_1}, \dots, x_{q_m}$, from $X_{q_1}, \dots, X_{q_m}$ respectively. Hence, if $F = (F_1, \dots, F_t)$ is an operator from $X = X_1 \times X_2 \times \cdots \times X_n$ into $Y = Y_1 \times Y_2 \times \cdots \times Y_t$, then

$$F^{(m)}(x_0) = \left(\frac{\partial^m F_i}{\partial x_{j_1} \cdots \partial x_{j_m}} \right)_{x = x_0} \tag{1.1.44}$$

$i = 1, 2, \dots, t$, $j_1, j_2, \dots, j_m = 1, 2, \dots, n$, is called the m-Fréchet derivative of F at $x_0 = (x_1^{(0)}, x_2^{(0)}, \dots, x_n^{(0)})$.

We now state results concerning the mean value theorem, Taylor's theorem, and Riemannian integration. The proofs are left out as exercises [125], [186].

The mean value theoremfor differentiable real functions f:

$$f(b) - f(a) = f'(c)(b - a), \tag{1.1.45}$$

where $c \in (a, b)$, does not hold in a Banach space setting. However, if F is a differentiable operator between two Banach spaces X and Y, then

$$\|F(x) - F(y)\| \leq \sup_{\bar{x} \in L(x,y)} \|F'(\bar{x})\| \cdot \|x - y\|, \tag{1.1.46}$$

where

$$L(x, y) = \{z : z = \lambda y + (1 - \lambda)x, \ 0 \leq \lambda \leq 1\}. \tag{1.1.47}$$

Set

$$z(\lambda) = \lambda y + (1 - \lambda)x, \quad 0 \leq \lambda \leq 1, \tag{1.1.48}$$

and

$$F(\lambda) = F(z(\lambda)) = F(\lambda y + (1 - \lambda)x). \tag{1.1.49}$$

Divide the interval $0 \leq \lambda \leq 1$ into n subintervals of lengths $\Delta\lambda_i$, $i = 1, 2, \dots, n$, choose points λ_i inside corresponding subintervals and as in the real Riemann integral consider sums

$$\sum_{\sigma} F(\lambda_i)\Delta\lambda_i = \sum_{i=1}^{n} F(\lambda_i)\Delta\lambda_i, \tag{1.1.50}$$

where σ is the partition of the interval, and set

$$|\sigma| = \max_{(i)} \Delta\lambda_i. \tag{1.1.51}$$

Definition 1.1.19. *If*

$$S = \lim_{|\sigma|\to 0} \sum_{\sigma} F(\lambda_i)\, \Delta\lambda_i \tag{1.1.52}$$

exists, then it is called the Riemann integral from $F(\lambda)$ from 0 and 1, denoted by

$$S = \int_0^1 F(\lambda)\, d\lambda = \int_x^y F(\lambda)\, d\lambda. \tag{1.1.53}$$

Note that a bounded operator $P(\lambda)$ on $[0, 1]$ such that the set of points of discontinuity is of measure zero is said to be integrable on $[0, 1]$.

We now state the famous Taylor theorem [103].

Theorem 1.1.20. *If F is m-times Fréchet-differentiable in $U(x_0, R)$, $R > 0$, and $F^{(m)}(x)$ is integrable from x to any $y \in U(x_0, R)$, then*

$$F(y) = F(x) + \sum_{n=1}^{m-1} \tfrac{1}{n!} F^{(n)}(x)(y-x)^n + R_m(x, y), \tag{1.1.54}$$

$$\left\| F(y) - \sum_{n=0}^{m-1} \tfrac{1}{n!} F^{(n)}(x)(y-x)^n \right\| \le \sup_{\bar{x}\in L(x,y)} \left\| F^{(m)}(\bar{x}) \right\| \frac{\|y-x\|^m}{m!}, \tag{1.1.55}$$

where

$$R_m(x, y) = \int_0^1 F^{(m)}\big(\lambda y + (1-\lambda)x\big)(y-x)^m \tfrac{(1-\lambda)^{m-1}}{(m-1)!} d\lambda. \tag{1.1.56}$$

1.2 Divided differences of operators

This section introduces the fundamentals of the theory of divided differences of a nonlinear operator. Several results are also provided using differences as well as Fréchet derivatives satisfying Lipschitz or monotone-type conditions.

Let X be a linear space. We introduce the following definition:

Definition 1.2.1. *A partially ordered topological linear space (POTL-space) is a locally convex topological linear space X which has a closed proper convex cone.*

A proper convex cone is a subset K such that $K + K \subset K$, $\alpha K \subset K$ for $\alpha > 0$, and $K \cap (-K) = \{0\}$. Thus the order relation $\leq$, defined by $x \leq y$ if and only if $y - x \in K$, gives a partial ordering that is compatible with the linear structure of the space. The cone K that defines the ordering is called the positive cone as $K = \{x \in X \mid x \geq 0\}$. The fact that K is closed implies also that intervals, $[a, b] = \{z \in X \mid a \leq z \leq b\}$, are closed sets.

Example 1.2.2. Some simple examples of POTL-spaces are:

(1) $X = E^n$, n-dimensional Euclidean space, with

$$K = \left\{(x_1, x_2, ..., x_n) \in E^n \mid x_i \geq 0,\ i = 1, 2, ..., n\right\};$$

(2) $X = E^n$ with $K = \{(x_1, x_2, ..., x_n) \in E^n \mid x_i \geq 0,\ i = 1, 2, ..., n-1, x_n = 0\}$;
(3) $X = C^n[0, 1]$, continuous functions, maximum norm topology, pointwise ordering;
(4) $X = C^n[0, 1]$, n-times continuously differentiable functions with

$$\|f\| = \sum_{k=0}^{n} \max \left| f^{(K)}(t) \right|, \text{ and pointwise ordering;}$$

(5) $C = L^p[0, 1]$, $0 \leq p \leq \infty$ usual topology,

$$K = \left\{f \in L^p[0, 1] \mid f(t) \leq 0 \text{ a.e.}\right\}.$$

Remark 1.2.3. Using the above examples, it is easy to see that the closedness of the positive cone is not, in general, a strong enough connection between the ordering and the topology. Consider, for example, the following properties of sequences of real numbers:

(1) $x_1 \leq x_2 \leq \cdots \leq x^*$, and $\sup\{x_n\} x^*$ implies $\lim_{n\to\infty} x_n = x^*$;
(2) $\lim_{n\to\infty} x_n = 0$ implies that there exists a sequence $\{y_n\}$ with $y_1 \geq y_2 \geq \cdots \geq 0$, $\inf\{y_n\} = 0$ and $-y_n \leq x_n \leq y_n$;
(3) $0 \leq x_n \leq y_n$, and $\lim_{n\to\infty} y_n = 0$ imply $\lim_{n\to\infty} x_n = 0$.

Unfortunately, these statements are not true for all POTL-spaces:

(a) In $X = C[0, 1]$ let $x_n(t) = -t^n$. Then $x_1 \leq x_2 \leq \cdots \leq 0$, and $\sup\{x_n\} = 0$, but $\|x_n\| = 1$ for all n, so $\lim_{n\to\infty} x_n$ does not exist. Hence (1) does not hold.
(b) In $X = L^1[0, 1]$ let $x_n(t) = n$ for $\frac{1}{n+1} \leq t \leq \frac{1}{n}$ and zero elsewhere. Then $\lim_{n\to\infty} \|x_n\| = 0$ but clearly property (2) does not hold.
(c) In $X = C^1[0, 1]$ let $x_n(t) = \frac{t^n}{n}$, $y_n(t) = \frac{1}{n}$. Then $0 \check{S} x_n \leq y_n$, and $\lim_{n\to\infty} y_n = 0$, but $\|x_n\| = \max\left|\frac{t^n}{n}\right| + \max\left|t^{n-1}\right| = \frac{1}{n} + 1 > 1$; hence x_n does not converge to zero.

We will now devote a brief discussion of certain types of POTL-spaces in which some of the above statements are true.

Definition 1.2.4. *A POTL-space is called regular if every order-bounded increasing sequence has a limit.*

Remark 1.2.5. Examples of regular POTL-spaces are E^n and L^p, $0 \leq p \leq \infty$, whereas $C[0, 1]$, $C^n[0, 1]$ and $L^\infty[0, 1]$ are not regular, as was shown in (a) of the above remark. If $\{x_n\}$ $n \geq 0$ is a monotone increasing sequence and $\lim_{n\to\infty} x_n = x^*$ exists, then for any k_0, $n \geq k_0$ implies $x_n \geq x_{k_0}$. Hence $x^* = \lim_{n\to\infty} x_n \geq x_{k_0}$, i.e., x^* is an upper bound on $\{x_n\}$ $n = 0$. Moreover, if y is any other upper bound, then $x_n \leq y$, and hence $x^* = \lim_{n\to\infty} x_n \leq y$, i.e., $x^* = \sup\{x_n\}$. This shows that in any POTL-space, the closedness of the positive cone guarantees that, if a monotone increasing sequence has a limit, then it is also a supremum. In a regular space, the converse of this is true; i.e., if a monotone increasing sequence has a supremum, then it also has a limit. It is important to note that the definition of regularity involves both an order concept (monotone boundedness) and a topological concept (limit).

Definition 1.2.6. *A POTL-space is called normal if, given a local base U for the topology, there exists a positive number* η *so that if* $0 \leq x \in V \in U$ *then* $[0, x] \subseteq \eta^U$.

Remark 1.2.7. If the topology of a POTL-space is given by a norm then this space is called a partially ordered normed space (PON)-space. If a PON-space is complete with respect to its topology then it is called a partially ordered Banach space (POB)-space. According to Definition 1.2.6. A PON-space is normal if and only if there exists a positive number α such that

$$\|x\| \leq \alpha \|y\| \quad \text{for all} \quad x, y \in X \quad \text{with} \quad 0 \leq x \leq y.$$

Let us note that any regular POB-space is normal. The converse is not true. For example, the space $C[0, 1]$, ordered by the cone of nonnegative functions, is normal but is not regular. All finite-dimensional POTL-spaces are both normal and regular.

Remark 1.2.8. Let us now define some special types of operators acting between two POTL-spaces. First we introduce some notation if X and Y are two linear spaces then we denote by (X, Y) the set of all operators from X into Y and by $L(X, Y)$ the set of all linear operators from X into Y. If X and Y are topological linear spaces, then we denote by $LB(X, Y)$ the set of all continuous linear operators from X into Y. For simplicity, the spaces $L(X, X)$ and $LB(X, X)$ will be denoted by $L(X)$ and $LB(X)$. Now let X and Y be two POTL-spaces and consider an operator $G \in (X, Y)$. G is called isotone (resp. antitone) if $x \geq y$ implies $G(x) \leq G(y)$ (resp. $G(x) \leq G(y)$). G is called nonnegative if $x \geq 0$ implies $G(x) \geq 0$. For linear operators, the nonnegativity is clearly equivalent with the isotony. Also, a linear operator is inverse nonnegative if and only if it is invertible and its inverse is nonnegative. If G is a nonnegative operator, then we write $G \geq 0$. If G and H are two operators from X into Y such that $H - G$ is nonnegative, then we write $G \leq H$. If Z is a linear space, then we denote by $I = I_z$ the identity operator in Z (i.e., $I(x) = x$

for all $x \in Z$). If Z is a POTL-space, then we have obviously $I \geq 0$. Suppose that X and Y are two POTL-spaces and consider the operators $T \in L(X, Y)$ and $S \in L(Y, X)$. If $ST \leq I_x$ (resp. $ST \geq I_x$), then S is called a left subinverse (resp. superinverse) of T and T is called a right subinverse (resp. superinverse) of S. We say that S is a subinverse of T if S is a left as well as a right subinverse of T.

We finally end this section by noting that for the theory of partially ordered linear spaces, the reader may consult M.A. Krasnosel'skii [128], [129], Vandergraft [199], or Argyros and Szidarovszky [43].

The concept of a divided difference of a nonlinear operator generalizes the usual notion of a divided difference of a scalar function in the same way in which the Fréchet derivative generalizes the notion of a derivative of a function.

Definition 1.2.9. *Let F be a nonlinear operator defined on a subset D of a linear space X with values in a linear space Y, i.e., $F \in (D, Y)$ and let x, y be two points of D. A linear operator from X into Y, denoted $[x, y]$, which satisfies the condition*

$$[x, y](x - y) = F(x) - F(y) \tag{1.2.1}$$

is called a divided difference of F at the points x and y.

Remark 1.2.10. If X and Y are topological linear spaces, then we shall always assume the continuity of the linear operator $[x, y]$. (Generally, $[x, y] \in L(X, Y)$ if X, Y are POTL-spaces then $[x, y] \in LB(X, Y)$).

Obviously, condition (1.2.1) does not uniquely determine the divided difference, with the exception of the case when X is one-dimensional. An operator $[\cdot, \cdot]: D \times D \to L(X, Y)$ satisfying (1.2.1) is called a divided difference of F on D. If we fix the first variable, we get an operator

$$\left[x^0, \cdot\right]: D \to L(X, Y). \tag{1.2.2}$$

Let x^1, x^2 be two points of D. A divided difference of the operator (1.2.2) at the points x^1, x^2 will be called a divided difference of the second order of F at the points x^0, x^1, x^2 and will be denoted by $\left[x^0, x^1, x^2\right]$. We have by definition

$$\left[x^0, x^1, x^2\right]\left(x^1 - x^2\right) = \left[x^0, x^1\right] - \left[x^0, x^2\right]. \tag{1.2.3}$$

Obviously, $\left[x^0, x^1, x^2\right] \in L(X, L(X, Y))$.

Let us now state a well-known result due to Kantorovich concerning the location of fixed points, which will be used extensively later [125].

Theorem 1.2.11. *Let X be a regular POTL-space and let x, y be two points of X such that $x \leq y$. If $H: [x, y] \to X$ is a continuous isotone operator having the property that $x \leq H(x)$ and $y \geq H(y)$, then there exists a point $z \in [x, y]$ such that $H(z) = z$.*

We now assume that X and Y are Banach spaces. Accordingly we shall have $[x, y] \in LB(X, Y)$, $[x, y, z] \in LB(X, LB(X, Y))$. As we will see in later chapters, most convergence theorems in a Banach space require that the divided differences of F satisfy Lipschitz conditions of the form:

$$\|[x, y] - [x, z]\| \le c_0 \|y - z\| \tag{1.2.4}$$

$$\|[y, x] - [z, x]\| \le c_1 \|y - z\| \tag{1.2.5}$$

$$\|[x, y, z] - [u, y, z]\| \le c_2 \|x - y\| \quad \text{for all} \quad x, y, z, u \in D. \tag{1.2.6}$$

It is a simple exercise to show that if $[\cdot, \cdot]$ is a divided difference of F satisfying (1.2.4) or (1.2.5), then F is Fréchet-differentiable on D and we have

$$F'(x) = [x, x] \quad \text{for all} \quad x \in D. \tag{1.2.7}$$

Moreover, if (1.2.4) and (1.2.5) are both satisfied, then the Fréchet derivative F' is Lipschitz continuous on D with Lipschitz constant $I = c_0 + c_1$.

We shall also give an example of divided differences of the first and of the second order in the finite-dimensional case. We shall consider the space $|R^q$ equipped with the Chebysheff norm, which is given by

$$\|x\| = \max\{|x_i| \in \mathbf{R} : 1 \le I \le q\} \quad \text{for} \quad x = (x_1, x_2, ..., x_q) \in \mathbf{R}^q. \tag{1.2.8}$$

It follows that the norm of a linear operator $L \in LB(\mathbf{R}^q)$ represented by the matrix with entries I_{ij} is given by

$$\|L\| = \max\left\{\sum_{j=1}^q |I_{ij}| \mid |1 \le i \le q\right\}. \tag{1.2.9}$$

We cannot give a formula for the norm of a bilinear operator. However, if B is a bilinear operator with entries b_{ijk}, then we have the estimate

$$\|B\| \le \max\left\{\sum_{j=1}^q \sum_{k=1}^q |b_{ijk}| \ |1 \le i \le q\right\}. \tag{1.2.10}$$

Let U be an open ball of $|R^q$ and let F be an operator defined on U with values in $\mathbf{R}^q$. We denote by $f_1, ..., f_q$ the components of F. For each $x \in U$ we have

$$F(x) = (f_1(x), ..., f_q(x))^T. \tag{1.2.11}$$

Moreover, we introduce the notation

$$D_j f_i(x) = \frac{\partial f(x)}{\partial x_j}, \quad D_{kj} f_i(x) = \frac{\partial^2 f_i(x)}{\partial x_j \partial x_k}. \tag{1.2.12}$$

Let x, y be two points of U and let us denote by $[x, y]$ the matrix with entries

$$[x, y]_{ij} = \frac{1}{x_j - y_j}\left(f_i(x_1, ..., x_j, y_{j+1}, ..., y_q) - f_i(x_1, ..., x_{j-1}, y_j, ..., y_q)\right). \tag{1.2.13}$$

The linear operator $[x, y] \in LB(\mathbf{R}^q)$ defined in this way obviously satisfies condition (1.2.1). If the partial derivatives $D_j f_i$ satisfy some Lipschitz conditions of the form

$$\left|D_j f_i\left(x_1, ..., x_k + t, ..., x_q\right) - D_j f_i\left(x_1, ..., x_k, ..., x_q\right)\right| \leq p^i_{jk}\,|t| \tag{1.2.14}$$

then condition (1.2.4) and (1.2.5) will be satisfied with

$$c_0 = \max\left\{\frac{1}{2}\sum\nolimits_{j=1}^{q}\left(p^i_{jj} + \sum\nolimits_{k=j+1}^{q} p^i_{jk}\right)|1 \leq i \leq q\right\} \tag{1.2.15}$$

and

$$c_1 = \max\left\{\frac{1}{2}\sum\nolimits_{j=1}^{q}\left(p^i_{jj} + \sum\nolimits_{k=1}^{j-1} p^i_{jk}\right)|1 \leq i \leq q\right\}. \tag{1.2.16}$$

We shall prove (1.2.4) only as (1.2.5) can be proved similarly.

Let x, y, z be three points of U. We shall have in turn

$$\begin{aligned}[x, y]_{ij} - [x, z] = \sum_{k=1}^{q}\Big\{&\left[x, \left(y_1, ..., y_k, z_{k+1}, ..., z_q\right)\right]_{ij} \\ &- \left[x\left(y_1, ..., y_{k-1}, z_k, ..., z_q\right)\right]_{ij}\Big\} \quad \text{by (1.2.13).}\end{aligned} \tag{1.2.17}$$

If $k \leq j$ then we have

$$\begin{aligned}&\left[x, \left(y_1, ..., y_k, z_{k+1}, ..., z_q\right)\right]_{ij} - \left[x, \left(y_1, ..., y_{k-1}, z_k, ..., z_q\right)\right]_{ij} \\ &\quad= \frac{1}{x_j - z_j}\left\{f_i\left(x_1, ..., x_j, z_{j+1}, ..., z_q\right) - f_i\left(x_1, ..., x_{j-1}, z_j, ..., z_q\right)\right\} \\ &\qquad - \frac{1}{x_j - z_j}\left\{f_i\left(x_1, ..., x_j, z_{j+1}, ..., z_q\right) - f_i\left(x_1, ..., x_{j-1}, z_j, ..., z_q\right)\right\} = 0.\end{aligned}$$

For $k = j$ we have

$$\begin{aligned}&\left|\left[x, \left(y_1, ..., y_j, z_{j+1}, ..., z_q\right)\right]_{ij} - \left[x, \left(y_1, ..., y_{j-1}, z_j, ..., z_q\right)_{ij}\right]\right| \\ &\quad= \Bigg|\frac{1}{x_j - y_j}\left\{f_i\left(x_1, ..., x_j, z_{j+1}, ..., z_q\right) - f_i\left(x_1, ..., x_{j-1}, y_i, z_{j+1}, ..., z_q\right)\right\} \\ &\qquad - \frac{1}{x_j - y_j}\left\{f_i\left(x_1, ..., x_j, z_{j+1}, ..., z_q\right) - f_i\left(x_1, ..., x_{j-1}, z_j, ..., z_q\right)\right\}\Bigg| \\ &\quad= \Bigg|\int_0^1\left\{D_j f_i\left(x_1, ..., x_j, y_j + t\left(x_j - y_j\right), z_{j+1}, ..., z_q\right)\right. \\ &\qquad\left. - D_j f_i\left(x_1, ..., x_j, z_j + t\left(x_j - z_j\right), z_{j+1}, ..., z_q\right)\right\}dt\Bigg| \\ &\quad\leq \left|y_j - z_j\right| p^i_{jj}\int_0^1 t dt = \frac{1}{2}\left|x_j - z_j\right| p^i_{jj}\end{aligned}$$

(by (1.2.14)).

Finally for $k > j$ we have using (1.2.13) and (1.2.17) again

$$\left|\left[x, \left(y_1, ..., y_k, z_{k+1}, ..., z_q\right)\right]_{ij} - \left[x, \left(y_1, ..., y_{k-1}, z_k, ..., z_q\right)\right]_{ij}\right|$$

$$= \left|\frac{1}{x_j - y_j}\left\{f_i\left(x_1, ..., x_j, y_{j+1}, ..., y_k, z_{k+1}, ..., z_q\right)\right.\right.$$

$$- f_i\left(x_1, ..., x_{j-1}, y_j, ..., y_k, z_{k+1}, ..., z_q\right)$$

$$- f_i\left(x_1, ..., x_j, y_{j+1}, ..., y_{k-1}, z_k, ..., z_q\right)$$

$$\left.\left. + f_i\left(x_1, ..., x_{j-1}, y_j, ..., y_{k-1}, z_k, ..., z_q\right)\right\}\right|$$

$$= \left|\int_0^1 \left\{f_i\left(x_1, ..., x_{j-1}, y_j + t\left(x_j - y_j\right), y_{j+1}, ..., y_k, z_{k+1}, ..., z_q\right)\right.\right.$$

$$\left.\left. - f_i\left(x_1, ..., x_{j-1}, y_j + t\left(x_j - y_j\right), y_{j+1}, ..., y_{k-1}, z_k, ..., z_q\right)\right\} dt\right|$$

$$\leq |y_k - z_k|\, p^i_{jk}.$$

By adding all the above, we get

$$\left|[x, y]_{ij} - [x, z]_{ij}\right| \leq \frac{1}{2}\left|y_j - z_j\right| p^i_{jj} + \textstyle\sum_{k=j+1}^q |y_k - z_k|\, p^i_{jk}$$

$$\leq \|y - z\| \left\{\frac{1}{2}\textstyle\sum_{j=1}^q \left(p^i_{jj} + \sum_{k=j+1}^q p^i_{jk}\right)\right\}.$$

Consequently, condition (1.2.4) is satisfied with c_0 given by (1.2.15). If each f_j has continuous second-order partial derivatives that are bounded on U, we have

$$p^i_{jk} = \sup\left\{\left|D_{jk} f_i\,(x)\right|\ |x \in U\right\}.$$

In this case $p^i_{jk} = p^i_{kj}$ so that $c_0 = c_1$.

Moreover, consider again three points x, y, z of U. Similarly with (1.2.17), the second divided difference of F at x, y, z is the bilinear operators defined by

$$[x, y, z]_{ijk} = \tfrac{1}{y_k - z_k}\left\{\left[x, \left(y_1, ..., y_k, z_{k+1}, ..., z_q\right)\right]_{ij}\right.$$
$$\left. - \left[x, \left(y_1, ..., y_{k-1}, z_k, ..., z_q\right)\right]_{ij}\right\}. \tag{1.2.18}$$

It is easy to see as before that $[x, y, z]_{ijk} = 0$ for $k < j$. For $k = j$ we have

$$[x, y, z]_{ijj} = \left[x_j, y_j, z_j\right]_t f_i\left(x_1, ..., x_{i-1}, t, z_{j+1}, ..., z_q\right) \tag{1.2.19}$$

where the right-hand side of (1.2.19) represents the divided difference of $f_i\big(x_1, ..., x_{j-1}, t, z_{j+1}, ..., z_q\big)$ as a function of t, at the points x_j, y_j, z_j. Using Genocchi's integral representation of divided differences of scalar functions [154], we get

$$[x, y, z]_{ijj} = \int_0^1 \int_0^1 t D_{jj} f_i \left(x_1, ..., x_{j-1}, x_j \right.$$
$$\left. + t\left(y_j - x_j\right) + ts\left(z_j - y_j\right), z_{j+1}, ..., z_q\right) dsdt. \quad (1.2.20)$$

Hence, for $k > j$ we obtain

$$[x, y, z]_{ijk} = \frac{1}{(y_k - z_k)\left(x_j - y_j\right)} \left\{ f_i \left(x_1, ..., x_j, y_{j+1}, ..., y_k, z_{k+1}, ..., z_q\right) \right.$$
$$- f_i \left(x_1, ..., x_j, x_{j+1}, ..., y_{k-1}, z_k, ..., z_q\right)$$
$$- f_i \left(x_1, ..., x_{j-1}, y_j, ..., y_k, z_{k+1}, ..., z_q\right)$$
$$\left. + f_i \left(x_1, ..., x_{j-1}, y_j, ..., y_{k-1}, z_k, ..., z_q\right)\right\}$$
$$\times \frac{1}{x_j - y_j} \int_0^1 \left\{ D_k f_i \left(x_1, ..., x_j, y_{j+1}, ..., y_{k-1}, z_k + t\left(y_k - z_k\right), z_{k+1}, ..., z_q\right) \right.$$
$$\left. - D_k f_i \left(x_1, ..., x_{j-1}, y_j, ..., y_{k-1}, z_k + t\left(y_k - z_k\right), z_{k+1}, ..., z_q\right)\right\} dt$$
$$= \int_0^1 \int_0^1 D_{kj} f_i \left(x_1, ..., x_{j-1}, y_j \right.$$
$$\left. + s\left(x_j - y_i\right), y_{j+1}, ..., y_{k-1}, z_k + t\left(y_k - z_k\right), z_{k+1}, ..., z_q\right) dsdt. \quad (1.2.21)$$

We now want to show that if

$$\left| D_{kj} f_i \left(v_1, ..., v_m + t, ..., v_q\right) - D_{kj} f_i \left(v_1, ..., v_m, ...v_q\right) \right| \le q_{km}^{ij} |t|$$
$$\text{for all} \quad v = \left(v_1, ..., v_q\right) \in U, \quad 1 \le i, j, k, m \le q, \quad (1.2.22)$$

then the divided difference of F of the second order defined by (1.2.18) satisfies condition (1.2.6) with the constant

$$c_2 = \max_{1 \le i \le q} \sum_{j=1}^{q} \left\{ \frac{1}{6} q_{jj}^{ij} + \frac{1}{2} \sum_{m=1}^{j-1} q_{jm}^{ij} + \frac{1}{2} \sum_{k=j+1}^{q} q_{kj}^{ij} + \sum_{k=j+1}^{q} \sum_{m=1}^{j-1} q_{km}^{ij} \right\}. \quad (1.2.23)$$

Let u, x, y, z be four points of U. Then using (1.2.18), we can easily have

$$[x, y, z]_{ijk} - [u, y, z]_{ijk} = \sum_{m=1}^{q} \left\{ \left[\left(x_1, ..., x_m, u_{m+1}, ..., u_q\right), y, z\right]_{ijk} \right.$$
$$\left. \left[\left(x_1, ..., x_{m-1}, u_m, ..., u_q\right), y, z\right]_{ijk} \right\}. \quad (1.2.24)$$

If $m = j$, the terms in (1.2.24) vanish so that using (1.2.21) and (1.2.22), we deduce that for $k > j$

$$
\begin{aligned}
&\left|[x,y,z]_{ijk}-[u,y,z]_{ijk}\right| \\
&\quad=\Bigg|\sum_{m=1}^{j-1}\int_0^1\int_0^1\Big\{D_{kj}f_i\left(x_1,...,x_m,u_{m+1},...,u_{j-1},y_j+s\left(x_j-y_j\right),\right.\\
&\qquad\qquad\left.y_{j+1},...,y_{k-1},z_k+t\left(y_k-z_k\right),z_{k+1},...,z_q\right)\\
&\qquad-D_{kj}f_i\left(x_1,...,x_{m-1},u_m,...,u_{j-1},y_j+s\left(x_j-y_j\right),\right.\\
&\qquad\qquad\left.y_{j+1},...,y_{k-1}z_k+t\left(y_k-z_k\right),z_{k+1},...,z_q\right)\Big\}\,dsdt\\
&\qquad+\int_0^1\int_0^1\Big\{D_{kj}f_i\left(x_1,...,x_{j-1},y_j+s\left(x_j-y_j\right),\right.\\
&\qquad\qquad\left.y_{j+1},...,y_{k-1},z_k+t\left(y_k-z_k\right),z_{k+1},...,z_q\right)\\
&\qquad-D_{kj}f_i\left(x_1,...,x_{j-1},y_j+s\left(u_j-y_j\right),\right.\\
&\qquad\qquad\left.y_{j+1},...,y_{k-1},z_k+t\left(y_k-z_k\right),z_{k+1},...,z_q\right)\Big\}\,dsdt\Bigg|\\
&\quad\leq\frac{1}{2}\left|x_j-u_j\right|q_{kj}^{ij}+\sum_{m=1}^{j-1}\left|x_m-u_m\right|q_{km}^{ij}.
\end{aligned}
$$

Similarly for $k=j$, we obtain in turn

$$
\begin{aligned}
&\left|[x,y,z]_{ijj}-[u,y,z]_{ijj}\right|\\
&=\Bigg|\int_0^1\int_0^1 t\Big\{D_{jj}f_i\left(x_1,...,x_{j-1},x_j+t\left(y_j-x_j\right)+ts\left(z_j-y_j\right),z_{j+1},...,z_q\right)\\
&\quad-D_{jj}f_i\left(x_1,...,x_{j-1},u_j+t\left(y_j-u_j\right)+ts\left(z_j-y_j\right),z_{j+1},...,z_q\right)\Big\}\,dsdt\\
&\quad+\sum_{m=1}^{j-1}\int_0^1\int_0^1 t\Big\{D_{jj}f_i\left(x_1,...,x_m,u_{m+1},...,u_{j-1},\right.\\
&\qquad\qquad\left.x_j+t\left(x_j-y_j\right)+ts\left(z_j-y\right)j,z_{j+1},...,z_q\right)\\
&\qquad-D_{jj}f_i\left(x_1,...,x_{m-1},u_m,...,u_{j-1},\right.\\
&\qquad\qquad\left.x_j+t\left(y_j-x_j\right)+ts\left(z_j-y_j\right),z_{j+1},...,z_q\right)\Big\}\,dsdt\Bigg|\\
&\leq\frac{1}{6}\left|x_j-u_j\right|q_{jj}^{ij}+\frac{1}{2}\sum_{m=1}^{j-1}\left|x_m-u_m\right|q_{jm}^{ij}.
\end{aligned}
$$

Finally using the estimate (1.2.10) of the norm of a bilinear operator, we deduce that condition (1.2.6) holds with c_2 given by (1.2.23).

We make an introduction to the problem of approximating a locally unique solution x^* of the nonlinear operator equation $F(x)=0$, in a POTL-space X. In particular, consider an operator $F\colon D\subseteq X\to Y$ where X is a POTL-space with values in a POTL-space Y. Let x_0, y_0, y_{-1} be three points of D such that

$$x_0 \le y_0 \le y_{-1}, \quad \left[x_0, y_{-1}\right],$$

and denote

$$\begin{aligned}
D_1 &= \left\{(x, y) \in X^2 \mid x_0 \le x \le y \le y_0\right\}, \\
D_2 &= \left\{(y, y_{-1}) \in X^2 \mid x_0 \le u \le y_0\right\}, \\
D_3 &= D_1 \cup D_2. \qquad (1.2.25)
\end{aligned}$$

Assume there exist operators $A_0 \colon D_3 \to LB\,(X, Y)$, $A \colon D_1 \to L\,(X, Y)$ such that:

(a)

$$F\,(y) - F\,(x) \le A_0\,(w, z)\,(y - x) \quad \text{for all} \quad (x, y), (y, w) \in D_1, \ (w, z) \in D_3; \qquad (1.2.26)$$

(b) the linear operator $A_0\,(u, v)$ has a continuous nonsingular nonnegative left subinverse;

(c)

$$F\,(y) - F\,(x) \ge A\,(x, y)\,(y - x) \quad \text{for all} \quad (x, y) \in D_1; \qquad (1.2.27)$$

(d) the linear operator $A\,(x, y)$ has a nonnegative left superinverse for each $(x, y) \in D_1$

$$F\,(y) - F\,(x) \le A_0\,(y, z)\,(y - x) \quad \text{for all} \quad x, y \in D_1, \ (y, z) \in D_3. \qquad (1.2.28)$$

Moreover, let us define approximations

$$F\,(y_n) + A_0\,(y_n, y_{n-1})\,(y_{n+1} - y_n) = 0 \qquad (1.2.29)$$

$$F\,(x_n) + A_0\,(y_n, y_{n-1})\,(x_{n+1} - x_n) = 0 \qquad (1.2.30)$$

$$y_{n+1} = y_n - B_n F\,(y_n)\, n \ge 0 \qquad (1.2.31)$$

$$x_{n+1} = x_n - B_n^1 F\,(x_n) \quad n \ge 0, \qquad (1.2.32)$$

where B_n and B_n^1 are nonnegative subinverses of $A_0\,(y_n, y_{n-1})$ $n \ge 0$.

Under very natural conditions, hypotheses of the form (1.2.26) or (1.2.27) or (1.2.28) have been used extensively to show that the approximations (1.2.26) and (1.2.30) or (1.2.31) and (1.2.32) generate two sequences $\{x_n\}$ $n \ge 1$, $\{y_n\}$ $n \ge 1$ such that

$$x_0 \le x_1 \le \cdots \le x_n \le x_{n+1} \le y_{n+1} \le y_n \le \cdots \le y_1 \le y_0 \qquad (1.2.33)$$

$$\lim_{n\to\infty} x_n = x^* = y^* = \lim_{n\to\infty} y_n \quad \text{and} \quad F\left(x^*\right) = 0. \qquad (1.2.34)$$

For a complete survey on these results, we refer to the works of Potra [164] and Argyros and Szidarovszky [42]–[44].

Here we will use similar conditions (i.e., like (1.2.26), (1.2.27), (1.2.28)) for two-point approximations of the form (1.2.29) and (1.2.30) or (1.2.31) and (1.2.32).

Consequently, a discussion must follow on the possible choices of the linear operators A_0 and A.