Computing 20, 207—228 (1978)

Computing

# Iterated Defect Correction for Differential Equations Part I: Theoretical Results

R. Frank and C. W. Ueberhuber, Wien

Received May 20, 1977; revised November 3, 1977

## Abstract — Zusammenfassung

**Iterated Defect Correction for Differential Equations. Part 1: Theoretical Results.** Iterated Defect Correction (IDeC) is a technique for improving successively an approximate solution of a given problem $Fy=0$. One of the most important fields of application of this principle are differential equations. Here, IDeC can be used as a technique for increasing the order of a discretization method and thus for improving the accuracy. In this paper a metalgorithm for the class of IDeC-methods for differential equations is presented and analyzed. For every component of this metalgorithm conditions are given which guarantee a certain order of accuracy. These conditions are of particular importance for practical applications, as far as the implementation of IDeC-methods is concerned.

**Iterierte Defektkorrektur für Differentialgleichungen. Teil 1: Theoretische Resultate.** Die Iterierte Defektkorrektur (IDeC) ist ein Verfahren zur schrittweisen Verbesserung einer Näherungslösung eines gegebenen Problems $Fy=0$. Eines der wichtigsten Anwendungsgebiete dieses Prinzips sind Differentialgleichungen. Die IDeC kann dort als Methode zur Verbesserung der Ordnung eines Diskretisierungsverfahrens, und damit zur Verbesserung der Genauigkeit eingesetzt werden. In der vorliegenden Arbeit wird ein Metaalgorithmus für die Klasse der IDeC-Verfahren für Differentialgleichungen vorgestellt und analysiert. Für jeden „Baustein" dieses Metaalgorithmus werden Bedingungen angegeben, die es gewährleisten, daß eine bestimmte Ordnung erreicht wird. Diese Bedingungen sind von großer praktischer Bedeutung, wenn IDeC-Verfahren als Computer-Programme implementiert werden sollen.

## 1. Introduction

In 1966, Zadunaisky [11] introduced a new method for the estimation of global discretization errors of difference schemes. Stetter [7] took up the idea, modified it and made the proposal to employ this modified scheme in an iterative fashion for successively increasing the order of a finite difference method and thus for improving its accuracy. Zadunaisky's original scheme, used in an iterative way, was called Iterated Defect Correction (IDeC) by Frank and Ueberhuber, who first presented rigorous analyses (Frank [1], Frank, Ueberhuber [3]). The early papers on this subject considered only a few concrete realizations of a general principle. This "IDeC-principle" was pointed out and discussed by Stetter in [9], where he also proposed a large number of schemes resulting from the application of the IDeC-principle to discretization methods. In particular, Stetter made the

0010-485X/78/0020/0207/$ 04.40

proposal to use different discretization methods as components of one IDeC-method. Since a general theory for these methods is still lacking we shall present a rigorous error analysis of general IDeC-methods for the important case of (ordinary *and* partial) differential equations.

In Section 3.1 we describe a metalgorithm (in the sense of John R. Rice) for the class of IDeC-algorithms. This metalgorithm consists of blocks or components, which may be chosen in one of several ways. In Sections 3.2, ..., 3.6, we will state conditions, which must be satisfied by the respective components of an IDeC-method so that a certain order of accuracy is obtained. In Section 4 an asymptotic error analysis for the IDeC metalgorithm is given. It is a crucial question, in connection with every asymptotic error analysis, whether such a result gives an indication of the accuracy to be expected for stepsizes of practical relevance. Numerical experience gathered so far with IDeC-methods shows that, even for large stepsizes, the observed errors are in agreement with theoretical results and, moreover, that, for certain types of problems, codes based on IDeC-methods will prove to be very efficient.

Many of the considerations of this paper are not only valid for differential equations but also hold for other types of operator equations. It was our goal, however, to present *concrete conditions* which guarantee a certain improvement of the order of accuracy. For this purpose one needs a rather detailed knowledge of the structure of the variational equations whose solutions are the coefficients of the asymptotic error expansions. Consequently assumptions about the local error mapping are made in Section 2; these are satisfied for practically all discretization methods for differential equations whereas for other types of operator equations slightly different assumptions would have to be made.

## 2. Preliminaries

Throughout this paper we will deal with problems specified by:

a) a *differential equation*

$$L\,y=f(t,y) \tag{2.1}$$

with $L$ being a linear differential operator; $t:=(t_1,\ldots,t_n)\in\mathbb{R}^n$ denotes the independent variable which varies on the *interval of integration* $I=[a_1,b_1]\times\ldots\times[a_n,b_n]\subset\mathbb{R}^n$ and $y:=(y_1,\ldots,y_m)\in\mathbb{R}^m$ is the dependent variable. The order $r$ of the operator $L$ and of equation (2.1) is defined by the highest derivative occuring in $L$ (where $L$ contains partial derivatives if $n>1$). $f$ may involve (partial) derivatives of $y$ up to an order $r_f\leq r-1$,

b) *initial and/or boundary conditions*, which, in conjunction with (2.1), generate a unique solution to the problem which will be denoted by $z=(z_1(t),\ldots,z_m(t))$.

A great variety of differential equations are included within the framework of a) and b). In particular, systems of ordinary differential equations when $n=1$, and partial differential equations when $n>1$.

A more compact notation for problem a), b) is: *determine an element* $z \in E$ *such that*

$$F z = 0 \tag{2.2}$$

*where* $F: E \to E^0$ *is a nonlinear operator and* $E$ *and* $E^0$ *are normed linear spaces*[1]. $F$ has different components: one component corresponds to (2.1) and other components correspond to the initial and/or boundary conditions (an example of this notation may be found in Stetter [8], p. 2). In Chapter 1 of [8], Stetter presents a detailed analysis of discretization methods applied to operator equations (2.2). We will assume that the reader is familiar with the results and the notation of [8]. There is only a minor difference between Stetter's notation and ours: instead of his integer index $n$ (cf. Stetter [8] Def. 1.1.2, 1.1.3, 1.1.4 ...) we will use the meshwidth parameter $h$ so that the finite-difference analogue of (2.2) becomes

$$F_h \zeta_h = 0 \tag{2.3}$$

where $F_h: E_h \to E_h^0$ with $E_h$ and $E_h^0$ finite dimensional spaces. The parameter $h$ characterizes the grid and if a step-size-control is used we assume coherent gridsequences (cf. Stetter [8], p. 73).

For the subsequent analysis it is convenient to define *classes of problems*: a class of problems is characterized by $n$ and $m$, by $I$ and by the structure of the linear differential operator $L$. A *problem*, i.e. a representative of such a class of problems is characterized by its *data*, e.g. by a special right-hand side $f$, by special initial and/or boundary conditions and by special coefficients of $L$. We assume that the data of the problems are $C^\infty$-functions and that all of their derivatives are bounded and Lipschitz continuous (although the analysis can be carried out under weaker differentiability conditions). Accordingly, we will assume $E := C^\infty[I]$ and $E^0 := C^\infty[I] \times$ {components arising from initial and/or boundary conditions}. We assume that every problem of a specific class has a unique solution.

Basic tools for our later considerations are the local error mappings (Def. 1.3.1 of Stetter [8]) and their asymptotic expansions (Def. 1.3.2 of Stetter [8]). It is important to note that for all problems in a certain class and for one particular discretization method (applicable to each problem of the class) the operators $\lambda_j: E \to E^0$ (cf. (1.3.2) of Stetter [8]) have the same structure; to obtain the $\lambda_j$ for a specific problem the data of this problem must be inserted into the general "formula" of each $\lambda_j$ (which depends on the class of problems).

For most practical realizations of (2.1) and for most discretization methods $\lambda_j$ can be decomposed into

$$\lambda_j = \mu_j + \gamma_j \tag{2.4}$$

[1] In Stetter [8] $E$ and $E^0$ are assumed to be Banach-spaces. We prefer the slightly more general assumption, that $E$ and $E^0$ are normed linear spaces, since there exist definitions (of practical relevance) of norms in $E$ which do *not* imply that $E$ is complete. It is well-known that Taylor's theorem can be applied to the operator $F$, if $F$ is sufficiently often Frechet differentiable, if $E$ is a normed linear space and if $E^0$ is an $L$-normed linear space. Since the theory behind asymptotic error expansion relies on Taylor's theorem we make the further assumption that $E^0$ is an $L$-normed linear space and not only a normed linear space.

where $\mu_j$ is a linear differential operator of order $j+r$ from $E$ to $E^0$ — or more precisely: $\mu_j y$ is a linear combination of derivatives of $y$ of maximum order $j+r$, and the coefficients (or coefficient functions) of this linear combination depend on $L$ but *not* on the right-hand side $f$ of (2.1). $\gamma_j$ is a nonlinear operator from $E$ to $E^0$. To give a more precise specification of $\gamma_j$, we recall that $f(t, y) \in \mathbb{R}^m$ may contain derivatives up to an order $r_f$ of $y$ and consequently a detailed representation of the $i$-th component of $f$ reads as follows:

$$f_i(w_1, \ldots, w_v) \equiv f_i(t_1, \ldots, t_n, y_1, \ldots, y_m, \ldots, (\partial^k/\partial t_u^k)\, y_l, \ldots),$$

where $k \leq r_f$. The dimension $v$ ($v \geq m+n$) of the domain of $f$ depends not only on $n$ and $m$ but also on the number of derivatives of $y$ which actually occur in $f$. For the construction of a specific $\gamma_j$ we need terms of the following types:

$G1)$ $(\partial^k/\partial w_u^k) f_i \quad 1 \leq i \leq m, \quad 1 \leq u \leq v, \quad 0 \leq k \leq j$

$G2)$ $(\partial^k/\partial t_u^k) y_i \quad 1 \leq i \leq m, \quad 1 \leq u \leq n, \quad 0 \leq k \leq j+r_f$

$G3)$ $c_i$ ... coefficients or coefficient functions (mapping $I$ into $\mathbb{R}$), which are characterized by the class of problems and by the discretization scheme, and do *not* depend on the data of the problem.

Every component of $\gamma_j y$ is assumed to be a combination of elements of $G1$, $G2$, and $G3$ — only a finite number of multiplications and additions are allowed — into which the function $y \in E$ is inserted (different components of $\gamma_j y$ usually have different structure). The way in which $\gamma_j y$ is formed from elements of $G1$, $G2$ and $G3$ depends on the class of problems and on the discretization scheme under consideration.

In the sequel we shall assume that a decomposition like (2.4) holds.

For a number of practical situations $\gamma_j \equiv 0$ for each $j$. Such methods might be called *linear methods*. The analyses of the following sections, however, hold for the general case $\gamma_j \not\equiv 0$. Examples of *nonlinear* methods (in the above sense) are Runge-Kutta-methods applied to IVPs of ODEs and the "$h^2$-algorithm" applied to BVPs: $y'' = f(t, y, y'), y(a) = A, y(b) = B$.

**Note 2.1:** In some situations, whether a method is linear or non-linear depends on the choice of the discretization operator $\Delta_h^0: E^0 \to E_h^0$ (cf. Definition 1.1.2 of Stetter [8]). E.g. consider Numerov's method applied to the two-point boundary value problem $y'' = f(t, y)$, $y(a) = A$, $y(b) = B$. A *non-linear* method results if $\Delta_h^0$ is defined by $(\Delta_h^0 d)_\nu := d(t_\nu)$ and the method is *linear* for

$$(\Delta_h^0 d)_\nu := (1/12)\left(d(t_{\nu-1}) + 10\, d(t_\nu) + d(t_{\nu+1})\right),$$

where $(\Delta_h^0 d)_\nu$ is the $\nu$-th component of $\Delta_h^0 d \in E_h^0$ for an arbitrary element $d \in E^0$.

This rather artificial example shows that the choice of $\Delta_h^0$ may influence the character of the error analysis. The results of this analysis, however, are independent of the choice of $\Delta_h^0$, since the variational equations are not affected by the particular definition of $\Delta_h^0$. For the examples of nonlinear methods given above (RK-methods, ...), there exists no choice of $\Delta_h^0$ which "converts these nonlinear methods into linear ones". For $\Delta_h: E \to E_h$, we assume throughout this paper that it is given by the "natural definition":

$$[\Delta_h y]_\nu := y(t_\nu), \qquad t_\nu \in \mathbb{R}^n, \tag{2.5}$$

where $\nu$ denotes an $n$-dimensional index.

## 3. Conditions for Asymptotic Results for Iterated Defect Corrections

In this section we introduce a metalgorithm for IDeC schemes and formulate conditions on its components which guarantee the validity of Theorem 4.1 — our main result.

### *3.1 A Metalgorithm for Iterated Defect Corrections*

The IDeC principle — applied to discretization methods for differential equations — may be utilized either to estimate the global discretization error of an approximation $\zeta_h$ to $z$ (where $\zeta_h$ is the result of a discretization method $\mathbb{M}^b$ applied to $F z = 0$), or for producing iteratively approximations $\zeta_h^0, \zeta_h^1, \zeta_h^2, \ldots$ to $z$ of an increasing order. To cover both possibilities in our treatment of the IDeC-principle, we will introduce a metalgorithm, which describes one *arbitrary* IDeC-step. This step may be interpreted to be either an error-estimation procedure or one of the steps of the iterative scheme (cf. Fig. 1).

*Metalgorithm*:

Consider a given *original problem*

$$F z = 0 \tag{3.1}$$

which has been chosen from some class of problems, and a numerical approximation $\zeta_h \in E_h$ to the solution $z$ of (3.1) defined on a grid $\mathbb{G}_h$ with meshwidth $h$. According to the IDeC-principle the defect of $\zeta_h$ must be computed, which implies that $\zeta_h$ must be mapped into the domain $E$ of the operator $F$. For example, this mapping, $\nabla_h: E_h \to E$ say, can be defined via interpolation or smoothing. The defect

$$d_h := F \nabla_h \zeta_h \in E^0 \tag{3.2}$$

is used to define a new problem (*neighboring problem*)

$$F z_h - d_h = 0 \tag{3.3}$$

the exact solution of which is $z_h = \nabla_h \zeta_h \in E$. We assume that this new problem (3.3) is a member of the same class of problems as (3.1). We now choose a discretization method $\mathbb{M}$ applicable to the members of the class of problems under consideration. $\mathbb{M}$ can be applied to both (3.1) and (3.3) — on the same grid $\mathbb{G}_h$ mentioned above — yielding approximations $\xi_h \in E_h$ and $\pi_h \in E_h$, respectively. Since we know the exact solution of (3.3), the global discretization error $\pi_h$ — $\Delta_h \nabla_h \zeta_h$ of the numerical solution $\pi_h$ is known. If we now replace the unknown global discretization error $\xi_h - \Delta_h z$ in the identity

$$\Delta_h z = \xi_h - (\xi_h - \Delta_h z)$$

by $\pi_h - \Delta_h \nabla_h \zeta_h$, we obtain the "improved" approximation

$$\bar{\zeta}_h := \xi_h - (\pi_h - \Delta_h \nabla_h \zeta_h) = \Delta_h \nabla_h \zeta_h - (\pi_h - \xi_h). \tag{3.4}$$

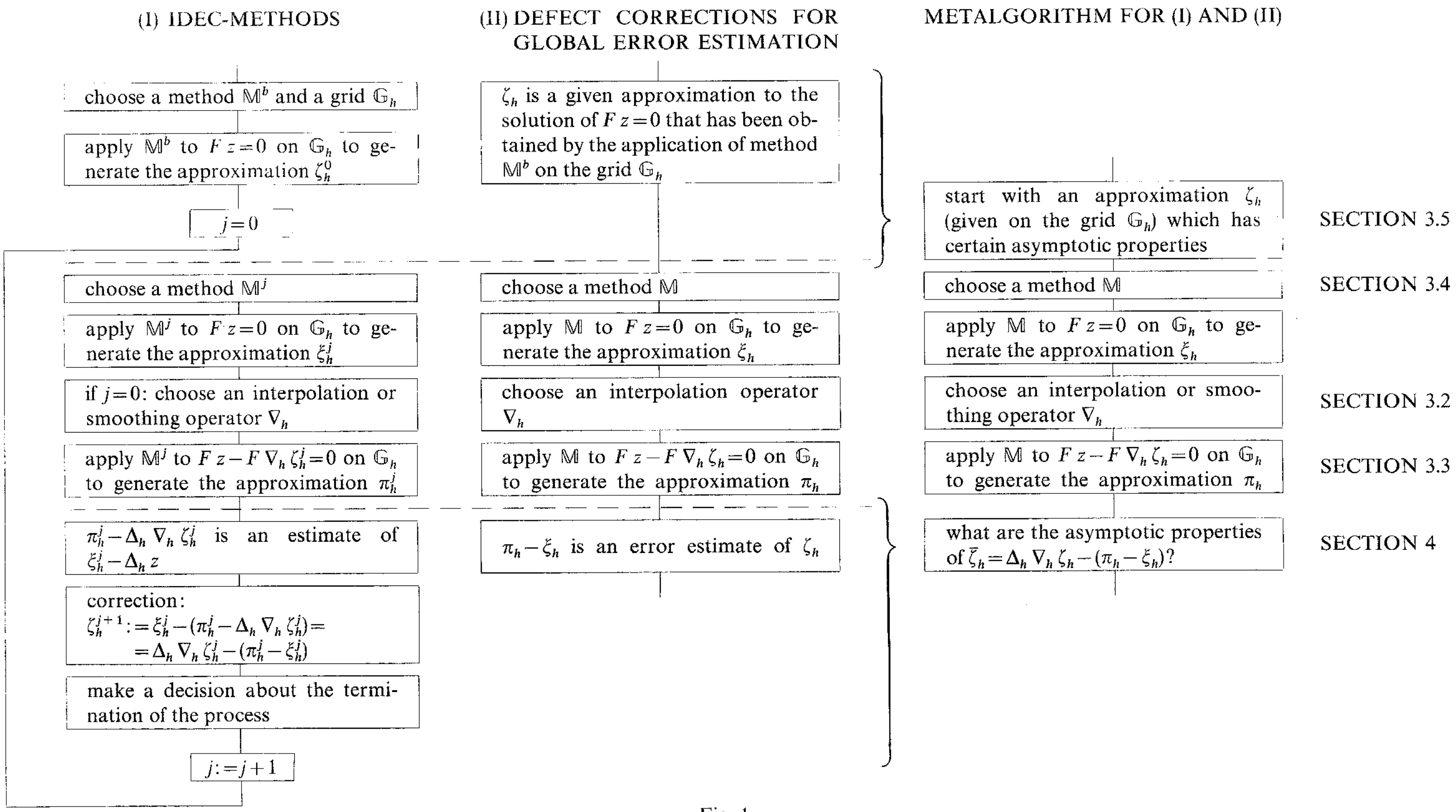
(I) IDEC-METHODS
choose a method $\mathbb{M}^b$ and a grid $\mathbb{G}_h$
apply $\mathbb{M}^b$ to $Fz=0$ on $\mathbb{G}_h$ to generate the approximation $\zeta_h^0$
$j=0$
choose a method $\mathbb{M}^j$
apply $\mathbb{M}^j$ to $Fz=0$ on $\mathbb{G}_h$ to generate the approximation $\xi_h^j$
if $j=0$: choose an interpolation or smoothing operator $\nabla_h$
apply $\mathbb{M}^j$ to $Fz-F\nabla_h\zeta_h^j=0$ on $\mathbb{G}_h$ to generate the approximation $\pi_h^j$
$\pi_h^j-\Delta_h\nabla_h\zeta_h^j$ is an estimate of $\xi_h^j-\Delta_h z$
correction:
$\zeta_h^{j+1}:=\xi_h^j-(\pi_h^j-\Delta_h\nabla_h\zeta_h^j)=$
$=\Delta_h\nabla_h\zeta_h^j-(\pi_h^j-\xi_h^j)$
make a decision about the termination of the process
$j:=j+1$
(II) DEFECT CORRECTIONS FOR GLOBAL ERROR ESTIMATION
$\zeta_h$ is a given approximation to the solution of $Fz=0$ that has been obtained by the application of method $\mathbb{M}^b$ on the grid $\mathbb{G}_h$
choose a method $\mathbb{M}$
apply $\mathbb{M}$ to $Fz=0$ on $\mathbb{G}_h$ to generate the approximation $\xi_h$
choose an interpolation operator $\nabla_h$
apply $\mathbb{M}$ to $Fz-F\nabla_h\zeta_h=0$ on $\mathbb{G}_h$ to generate the approximation $\pi_h$
$\pi_h-\xi_h$ is an error estimate of $\zeta_h$
METALGORITHM FOR (I) AND (II)
start with an approximation $\zeta_h$ (given on the grid $\mathbb{G}_h$) which has certain asymptotic properties
choose a method $\mathbb{M}$
apply $\mathbb{M}$ to $Fz=0$ on $\mathbb{G}_h$ to generate the approximation $\xi_h$
choose an interpolation or smoothing operator $\nabla_h$
apply $\mathbb{M}$ to $Fz-F\nabla_h\zeta_h=0$ on $\mathbb{G}_h$ to generate the approximation $\pi_h$
what are the asymptotic properties of $\bar{\zeta}_h=\Delta_h\nabla_h\zeta_h-(\pi_h-\xi_h)$?
SECTION 3.5
SECTION 3.4
SECTION 3.2
SECTION 3.3
SECTION 4

Fig. 1

The two most common applications of the above metalgorithm are (compare Fig. 1):

(i) *the order of accuracy of a given discretization method* $\mathbb{M}^b$ (solution: $\zeta_h^0$) *is increased by the iterative use of* (3.4), which yields successively $\zeta_h^1, \zeta_h^2, \ldots$. $\zeta_h$ and $\bar{\zeta}_h$ of the metalgorithm may be identified with any of the approximations $\zeta_h^j$ and $\zeta_h^{j+1}$, respectively, $j=0, 1, 2, \ldots$. In our earlier papers we only considered the special case $\mathbb{M} \equiv \mathbb{M}^b \equiv \mathbb{M}^0 \equiv \mathbb{M}^1 \equiv \mathbb{M}^2 \equiv \ldots$. In this case the analysis of Section 4 shows that $\zeta_h^j$ is of order $(j+1)\,q$ if $\mathbb{M}$ is of order $q$ (and if certain assumptions about $\nabla_h, \ldots$ are satisfied), and so this scheme provides higher order approximations by using only a low order discretization method (which is a desirable feature for some types of problems). Moreover, these IDeC-methods offer the advantage that $\mathbb{M}$ has to be applied to (3.1) only once (since $\zeta_h^0 = \xi_h^0 = \xi_h^1 = \ldots$).

(ii) to *produce estimates of the global discretization error associated with* $\zeta_h$. Let us assume that $\nabla_h$ is defined via an interpolation process, i.e. $\Delta_h \nabla_h \zeta_h = \zeta_h$. The correction (3.4) then yields

$$\bar{\zeta}_h = \xi_h - (\pi_h - \zeta_h) = \zeta_h - (\pi_h - \xi_h). \tag{3.5}$$

It will turn out in Section 4 that $\bar{\zeta}_h$ is an approximation of order $p+q$, if $\zeta_h$ is of order $p$ and $\mathbb{M}$ is a method of order $q$ and some further assumptions are satisfied. As a consequence $\pi_h - \xi_h$ is an excellent estimate of the global discretization error $\zeta_h - \Delta_h z$:

$$\pi_h - \xi_h = \zeta_h - \bar{\zeta}_h = \zeta_h - \Delta_h z + 0\,(h^{p+q}).$$

If $\nabla_h$ is *not* an interpolation operator, i.e. $\Delta_h \nabla_h \zeta_h \neq \zeta_h$, then

$$\pi_h - \xi_h = \zeta_h - \Delta_h z + 0\,(h^{p+q})$$

is true only under the additional assumption that

$$\Delta_h \nabla_h \zeta_h - \zeta_h = 0\,(h^{p+q}).$$

Consequently, this last condition must be checked in such situations.

**Note 3.1.1:** When constructing $\pi_h$ and $\xi_h$, it is assumed that the grid $\mathbb{G}_h$ is used. For the purpose of global error estimation only, Stetter [9] suggested the use of a coarser grid $\mathbb{G}_{\bar{h}}$, with a meshwidth $\bar{h} > h$, for the computation of $\pi_{\bar{h}}$ and $\xi_{\bar{h}}$. This reduces the computational effort. (3.4) then yields

$$\bar{\zeta}_h := \Delta_h \bar{\nabla}_{\bar{h}}\, \xi_{\bar{h}} - (\Delta_h \bar{\nabla}_h\, \pi_{\bar{h}} - \Delta_h \nabla_h \zeta_h)$$

where $\bar{\nabla}_h : E_{\bar{h}} \to E$ is an interpolation or smoothing operator. The analysis of Section 4 extends easily to the situations when $\bar{h} \neq h$ (provided $\bar{\nabla}_{\bar{h}}$ is defined appropriately).

### *3.2 Choice of the Interpolation or Smoothing Operator* $\nabla_h$

For IDeC-methods (alternative (i) of Section 3.1), it is essentially the choice of $\nabla_h$ which determines the maximum attainable order $J$ (which cannot be exceeded, even by an arbitrarily large number of IDeC-steps). For error estimation schemes (alternative (ii) of Section 3.1) $\nabla_h$ (together with the order of $\mathbb{M}$) determines the "quality" of the estimate $\pi_h - \xi_h$ (as it is reflected in the order of accuracy of $\bar{\zeta}_h$). Thus, for both alternatives the choice of $\nabla_h$ will be influenced by the desired maximum attainable order $J$ (in the second alternative only *one* IDeC-step is considered).

The question of how to choose $\nabla_h$ will be answered in two steps. Initially, we point out the necessity of taking piecewise functions into consideration and discuss the implications of such a choice. Secondly, if a certain maximum attainable order $J$ is sought, we give a criterium for choosing functions which are potentially suitable for defect corrections. For instance, piecewise polynomial functions form an appropriate class, but the degrees of the polynomials have to be chosen according to the desired value of $J$.

Usually, the argument $\zeta_h$ of $\nabla_h$ (the numerical solution of the original problem) has a large number of components, and so it is appropriate in many situations to define $\nabla_h$ in a piecewise manner: for instance $\nabla_h \eta_h$ (where $\eta_h$ denotes an arbitrary element of $E_h$), may be a piecewise continuous polynomial, a spline-function, .... For a piecewise function $\nabla_h \eta_h$, which is defined on a partition $\{I_s^h, s=1\,(1)\,S(h)\}$ of $I$, its differentiability is usually determined by its smoothness at the boundaries of the $n$-dimensional subintervals $I_s^h$. It will turn out to be convenient to classify the operators $\nabla_h$ according to the smoothness of $\nabla_h \eta_h$:

$D1)$ $\nabla_h \eta_h \in C^{(i)}[I]$, $\quad 0 \leq i < r-1$

$D2)$ $\nabla_h \eta_h \in C^{(i)}[I]$, $\quad r-1 \leq i$, where $r$ denotes the order of the operator $L$.

Note that $\nabla_h \zeta_h \in C^{(i)}[I]$ does not agree with our former assumption $\nabla_h \zeta_h \in E = C^{\infty}[I]$, which was made in Section 3.1 for explanatory purposes only.

The space, which contains $\nabla_h \eta_h$, and which is characterized by the partition $\{I_s^h, s=1\,(1)\,S(h)\}$ and by the differentiability properties of its elements, will be denoted by $E^h$ (this function space should not be confused with the finite dimensional space $E_h$). Corresponding to the definition of $E$, we will assume that the restriction of an element of $E^h$ to $I_s^h$ belongs to $C^{\infty}[I_s^h]$. $E^h$ depends on the meshwidth $h$ of the partition $\{I_s^h\}$. To give this dependence a precise meaning, let us assume for explanatory purposes that $I=[0,1]^n$ and let $\mathbb{G}_h$ be an equidistant grid $\mathbb{G}_h := \{v\,h,\ v=0\,(1)\,w\,K;\ h=\frac{1}{wK}\}^n$. We assume that the sequence of meshwidths is characterized by $K \to \infty$ and by a fixed $w \in \mathbb{N}$. For every stepsize $h$ we consider a subgrid $\bar{\mathbb{G}}_h$ of $\mathbb{G}_h$ which is defined as follows: $\bar{\mathbb{G}}_h := \{\mu\,w\,h,\ \mu=0\,(1)\,K\}^n$. We assume that the partition $\{I_s^h\}$ is induced by the set

$$B_h := \bigcup_{i=1}^{n} \bigcup_{\mu=1}^{K-1} b_h^{i\mu},$$

with

$$b_h^{i\mu} := [a_1, b_1] \times \ldots \times [a_{i-1}, b_{i-1}] \times \{\mu\,w\,h\} \times [a_{i+1}, b_{i+1}] \times \ldots \times [a_n, b_n]$$

where $[a_l, b_l] = [0, 1]$, $l = 0\,(1)\,n$. The assumption of a fixed $w$ for $h \to 0$ $(K \to \infty)$ implies

(i) for $h \to 0$, $\mathbb{G}_h$ and $\bar{\mathbb{G}}_h$ are refined in exactly the same way

(ii) if $t \in B_{h'}$ for some $h'$, then there exist infinitely many $h'' < h'$ (with $h'$, $h'' \in \{1/w\,K;\ K \in \mathbb{N}\}$), such that $t \in B_{h''}$.

According to its definition, $w$ characterizes the number of points of $\mathbb{G}_h$ that lie inside the (closed) subintervals $I_s^h$: this number is $(w+1)^n$. Therefore, $w$ is strongly related to the definition of $\nabla_h$; e.g., for piecewise polynomial functions, $w$ is equal to the degree of the polynomials and for spline functions $w=1$. These considerations have to be slightly generalized for non-equidistant grids, in particular coherent grid sequences (as defined on p. 73 of Stetter [8]) have to be assumed for the one-dimensional grids along the coordinate axes.

Property (ii) permits the proper definition of expressions like "$\| \varphi_h(t) \| = 0(h^{\text{power}})$, for every $t \in B_h$" (where $\varphi_h(t) \in \mathbb{R}^m$).

The introductory part of this section was devoted to a precise definition of $E^h$. Obviously $E \subset E^h$ and in the following we assume $\Delta_h: E^h \to E_h$ and $\nabla_h: E_h \to E^h$ (replacing $\Delta_h: E \to E_h$, $\nabla_h: E_h \to E$, as introduced in Section 3.1).

We are now in the position to formulate conditions on $\nabla_h$ which guarantee the desired maximum attainable order $J$:

*For an arbitrary sequence of functions* $\{y_h\}_{h\to 0}$, $y_h \in E^h$, *which satisfies*

$$\| D^k y_h \| = 0(1), \qquad h \to 0, \qquad k = 0, 1, 2, \ldots \tag{3.6}$$

*the following relation must hold*

$$\| D^k \nabla_h \Delta_h y_h - D^k y_h \| = 0(h^{\max(0,\, J-(k-r))}), \qquad k = 0, 1, 2, \ldots. \tag{3.7}$$

For $n=1$, $D^k := d^k/dt^k$; for $n>1$, $D^k := \partial^k/(\partial t_1^{k_1} \ldots \partial t_n^{k_n})$ with $\Sigma\, k_\nu = k$, $0 \le k_\nu \le k$ and we assume that (3.6) holds for all the different derivatives $D^k$ (corresponding to the possible choices of $k_1, \ldots, k_n$) and implies (3.7) for all these derivatives $D^k$. If necessary onesided derivatives have to be formed at the boundaries of the subintervals $I_s^h$. If $y \in C^{(i)}[I]$ is an element of $E^h$, then $D^k y$ is normally *not* an element of $E^h$, since the differentiability of $D^k y$ at $t \in B_h$ is generally lower than that of $y$. For $k > i$, $D^k$ usually defines at those points *two* different values of the one-sided derivatives. We assume that $\| D^k y \|$ for $y \in E^h$ is given by

$$\| D^k y \| := \max_{s;\, t \in I_s^h} \| D^k y(t) \|$$

where the norm used on the right hand side of this expression is the $\mathbb{R}^m$-norm.

**Note 3.2.1:** For most of the realistic definitions of $\nabla_h$, the quality of the approximation $D^k \nabla_h \Delta_h y_h$ for $D^k y_h$ is increasing $k$; this fact is taken into account in formula (3.7); thus interpolation and smoothing functions of practical relevance (piecewise polynomial functions, spline functions, etc.) are included in our analysis.

**Example 3.2.1:** *Lagrange interpolation with piecewise continuous polynomials of fixed degree M*:

$\alpha$) $n=1$: For these functions (3.6) implies

$$\| D^k \nabla_h \Delta_h y_h - D^k y_h \| = 0(h^{\max(0,\, M+1-k)}), \qquad k = 0, 1, 2, \ldots.$$

This relation may be derived e.g. from formula (3.3.15) of Hildebrand [6] for $k \leq M$ and is trivially satisfied for $k > M$ since $D^k \nabla_h \Delta_h y_h \equiv 0$. A comparison of the above formula with (3.7) shows that piecewise polynomials with $M \geq J + r - 1$ satisfy "(3.6)$\Rightarrow$(3.7)" for a given $J$. Therefore, these functions are a suitable choice for IDeC-methods, if one has a maximum attainable order $J$ in mind; the particular choice $M = J + r - 1$, however, will be prefered for economical reasons.

$\beta$) $n > 1$: If in problem (2.1) no mixed derivatives occur (but only derivatives $D^k = \partial^k / \partial t_i^k$) then for any discretization scheme of practical interest no mixed derivatives are involved in $\mu_j$ and $\gamma_j$. It is therefore not necessary in these cases to include mixed derivatives in our theoretical investigations. In particular, in the requirement "(3.6)$\Rightarrow$(3.7)", the symbol $D^k$ denotes only derivatives $\partial^k / \partial t_i^k$. Then a simple generalization of (3.3.15) of Hildebrand [6] shows that "(3.6)$\Rightarrow$(3.7)" holds, if $M \geq J + r - 1$. If, in problem (2.1), mixed derivatives occur it is still possible to show

$$\| D^k \nabla_h \Delta_h y_h - D^k y_h \| = 0 \, (h^{M+1-k});$$

but, for $k_i \leq M$ and $\sum_i k_i > M$, the mixed derivatives $D^k \nabla_h \Delta_h y_h$ do not in general vanish and negative powers of $h$ can therefore occur in the above relation. Consequently, for such problems, "(3.6)$\Rightarrow$(3.7)" cannot be guaranteed for piecewise Lagrange interpolation.

**Example 3.2.2:** *Spline interpolation with natural spline functions of degree* $M = J + r - 1$. These functions also satisfy the assumption "(3.6)$\Rightarrow$(3.7)" as for instance Lemma 7.1 of Swartz and Varga [10] shows for equidistant grids.

With respect to computational complexity considerations, it turns out that, for increasing degree, the computation needed for spline functions grows faster than the computation needed for piecewise polynomial functions (which involve the use of "weightmatrices"; cf. for instance Frank, Ueberhuber [5]). For IVPs, where forward step methods are highly desirable, only piecewise polynomial interpolation (which is a *local* interpolation scheme) seems to be suitable, whereas spline-functions would require the storage of the approximations $\zeta_h^0, \zeta_h^1, \ldots$ over the whole integration interval. For problems of order $r \geq 2$, piecewise polynomials have the disadvantage of being members of the lower differentiability class $D\,1$, which demands the choice of a method $\mathbb{M}$ which is applicable to piecewise problems (cf. Sections 3.3 and 3.4).

For the lower differentiability class $D\,1$, we have in addition to "(3.6)$\Rightarrow$(3.7)" the requirement:

*If* $t \in b_h^{i\mu}$ *for* $1 \leq i \leq n$ *and* $1 \leq \mu \leq K - 1$ *then*

$$\| D_+^k \, (\nabla_h \zeta_h)\,(t) - D_-^k \, (\nabla_h \zeta_h)\,(t) \| = 0 \, (h). \qquad k = J + r, J + r + 1, \ldots \tag{3.8}$$

where $D_+^k$ and $D_-^k$ denote $\lim\limits_{t_l \to \mu w h + 0} \partial^k / \partial t_i^k$ and $\lim\limits_{t_l \to \mu w h - 0} \partial^k / \partial t_i^k$ respectively.

If $\nabla_h \zeta_h$ consists of polynomials of degree $M = J + r - 1$ (as it does e.g. in Example 3.2.1 and 3.2.2) the $(M+1)$-st and all further derivatives vanish and (3.8) is trivially satisfied. Otherwise the validity of (3.8) has to be derived from the properties of $\nabla_h$ and $\zeta_h$.

### 3.3 *Piecewise Problems*

The solution of the neighboring problem (3.3) (namely, the piecewise function $\nabla_h \zeta_h$) is normally not an element of $E$, but of the more general space $E^h$ (cf. Section 3.2). Consequently, the defect $d_h = F \nabla_h \zeta_h$ belongs generally not to $E^0$ but to the space $E^{0h} := C^{(i-r)}[I] \times \{$components belonging to initial and/or boundary conditions$\}$. For $i-r<0$, we assume that the elements of $E^{oh}$ have jump discontinuities along $B_h$ (this will be the most usual case; only for $D2$ with $i \geq r$ will this not happen). In accordance with the definition of $E^h$, we assume that the restriction of the "functional part" of an arbitrary element $d \in E^{0h}$ ot $I_s^h$ belongs to $C^\infty[I_s^h]$.

As a consequence of having defined $\nabla_h$ to be piecewise, the neighboring problems (3.3) are members of a class of *piecewise problems* $\{F z = 0;\ F: E^h \to E^{0h}\}$. Because $E \subset E^h$ and $E^0 \subset E^{0h}$, all *smooth problems* $\{F z = 0;\ F: E \to E^0\}$ are included in the more general class of piecewise problems. Thus (3.1) and (3.3) are both members of the same class of problems, a property, which is essential for the applicability of the IDeC-principle.

The question of existence and uniqueness of a solution of a smooth problem is normally covered by theorems of analysis whereas this question has to be discussed for piecewise problems (note that the existence of a solution of problem (3.3) is trivial: $\nabla_h \zeta_h$ is a solution; however, the existence of a solution of an *arbitrary* problem of our class of piecewise problems is not obvious). The greatest difficulty arises for $D1$, where the uniqueness of the solution is normally lost: additional degrees of freedom are introduced through the introduction of jump discontinuities in derivatives of order less than $r$, at the points $t \in B_h$. As a consequence, additional conditions have to be included in the operator $F$, to guarantee the existence and uniqueness of a solution. Consequently, *we require, that under the assumptions*

$U$ 1) *for $D$ 1, the differences between $D_+^k$, $D_-^k$, $k = i+1\,(1)\,r-1$, are prescribed at the points $t \in B_h$* (concerning the definition of $D_+^k$, $D_-^k$ cf. (3.8)),

$U$ 2) *for $D$ 2, no additional conditions are applied to $F$,*

*the existence and uniqueness of the solution of piecewise problems* (from the specific class under consideration) *is guaranteed.*

Note that in the case $U1$) the prescribed differences between the derivatives $D_+^k$, $D_-^k$ (the "jump conditions"), contribute to the data of the operator $F$, and accordingly components have to be added to the elements of $E^{0h}$. These additional components are functions which map

$$b_h^{l,\mu} := [a_1, b_1] \times \ldots \times [a_{l-1}, b_{l-1}] \times \{\mu\, w\, h\} \times [a_{l+1}, b_{l+1}] \times \ldots \times [a_n, b_n] \subset B_h$$

into the space $\mathbb{R}^m$. For each of the given "jump-functions" $\varphi_{k,l,\mu}$ the jump condition can be written as

$$D_+^k(z)(t) - D_-^k(z)(t) = \varphi_{k,l,\mu}(t),\ t \in b_h^{l,\mu},\ k = i+1\,(1)\,r-1,\ \mu = 1\,(1)\,K-1,\ l = 1\,(1)\,n$$

where $z$ denotes the solution of the problem under consideration. For $n=1$ the functions $\varphi_{k,l,\mu}$ reduce to $m$-vectors.

Recall that $\nabla_h \zeta_h$ is used for the construction of the data of (3.3). Therefore the jump functions of (3.3) are defined by

$$\varphi_{k,l,\mu} := D_+^k (\nabla_h \zeta_h)(t) - D_-^k (\nabla_h \zeta_h)(t)$$

and thus all data of (3.3) are well defined. For the problem (3.1), these jumps are zero. Particular examples of the technique for prescribing the jump conditions have been discussed by Frank [1] and Frank and Hertling [2].

**Note 3.3.1:** For the norm of the spaces $E^h$ and $E_h$ we will assume the maximum norm in agreement with Section 2.2.1 of Stetter [8]. The norm of an element $y \in E^{0h}$ is formed by summing up the norm of the functional part of $y$ which belongs to $C^{(i-r)}(I)$, the norm of the components which belong to initial and/or boundary conditions and the norm of the components belonging to the jump conditions. The number of this third type of components grows like $1/nh \sim 1/h$ for $h \to 0$; that means: if the norm of each of the jump components behaves like $0(h^{p+1})$ and if the norms of the other components behave like $0(h^p)$ then $\| y_h \| = 0(h^p)$. In our later applications this will be the typical situation.

### *3.4 Choice of the Method* $\mathbb{M}$

Of the method $\mathbb{M}$, we primarily require that it be applicable to the piecewise problems under consideration. Because discretization methods are usually designed for smooth problems they meet difficulties when confronted with piecewise problems having jump discontinuities at the points $t \in B_h$. Also, for $i \geq r$, if $i$ is not substantially larger than $r$, the low differentiability of the data (along $B_h$) may cause difficulties. Discretization methods, however, which satisfy the following conditions, are applicable to our piecewise problems:

*A*1) $\mathbb{M}$, applied to a certain concrete problem, requires the solution of a system of equations, and we assume that for every equation which results from discretizing the differential equation (but *not* from discretizing jump conditions, ...) only gridpoints from *one* of the (closed) intervals $I_s^h$ are involved.

*A*2) For $D1$ the method $\mathbb{M}$ must comprise the discretized jump conditions (cf. Section 3.3). We suppose that the jumps (occuring along $B_h$) are discretized in a linear way, i.e. components of $\gamma_j$ (cf. Section 2) corresponding to jump conditions vanish. The respective components of $\mu_j$ are assumed to be linear combinations of terms $D_+^k z(t) - D_-^k z(t)$, $t \in B_h$, $k \leq j+r-1$.

Examples for methods satisfying $A1$ *and* $A2$ may be found for instance in Frank [1] and Frank and Hertling [2].

**Note 3.4.1:** Multistep methods for IVPs of ODEs do not satisfy $A1$, whereas Runge-Kutta methods are applicable (cf. Frank and Ueberhuber [3]).

*The essential assumption about* $\mathbb{M}$ *is that Theorem* 1.3.1 *of Stetter* [8] *must be applicable*, i.e. $\mathbb{M}$ admits asymptotic expansions. This assumption is sufficient for an asymptotic analysis of defect corrections, but one might think of situations where Theorem 1.3.1 of [8] is not applicable and iterated defect corrections work nevertheless well. For instance, for stiff ODEs the authors carried out numerical experiments with IDeC-methods based on the implicit Euler scheme and obtained promising results, despite the wellknown fact that the error expansion of Theorem 1.3.1 of [8] does not reflect the true behavior of the global error for large values of $h \cdot L$ (where $L$ denotes the Lipschitz constant of the problem under consideration). Another example is given by BVPs for elliptic PDEs which often have a solution that is not sufficiently smooth at the vertices of the integration interval, and so Theorem 1.3.1 of [8] is *not* applicable. Preliminary numerical experiments, however, have shown that also in this case iterated defect corrections can be applied successfully (cf. Frank and Hertling [2]).

An example where the assumptions of Theorem 1.3.1 of [8] are not satisfied and iterated defect corrections do *not* work, are linear multistep methods for IVPs of ODEs (refer also to Note 3.4.1).

### *3.5 Assumptions About $\zeta_h$*

According to our metalgorithm (cf. Section 3.1), $\zeta_h$ may be either the result of a discretization method $\mathbb{M}^b$ applied to the original problem (3.1), or the result of preceding IDeC-steps. To cover both cases *we require that $\zeta_h$ satisfies*

$$\begin{gathered}\zeta_h = \Delta_h (z + h^p e_{h,p} + h^{p+1} e_{h,p+1} + \ldots + h^{J+r-1} e_{h,J+r-1}) + R_h \\ \Delta_h : E^h \to E_h,\ e_{h,j} \in E^h,\ R_h \in E_h,\ \| R_h \| = 0 (h^{J+r})\end{gathered} \tag{3.9 a}$$

*with*

$$B1) \qquad \| D^k e_{h,j} \| = \begin{cases} 0 (h^{\max(-p, \min(0, J-j))} & k \le r \\ 0 (h^{\max(-p, \min(0, J-j-(k-r)))}) & k > r \end{cases} \tag{3.9 b}$$

$$j = p (1) J + r - 1,\ k = 0, 1, 2, \ldots$$

*$B$2) additional assumption for $D$1: for every $t \in B_h$*

$$\| D^k_+ e_{h,j} (t) - D^k_- e_{h,j} (t) \| = \begin{cases} 0 (h^{\max(-p+1, \min(1, J-j))}) & i < k \le r \\ 0 (h^{\max(-p+1, \min(1, J-j-(k-r)))}) & k > r \end{cases} \tag{3.9 c}$$

$$j = p (1) J + r - 1, \qquad k = 0, 1, 2, \ldots .$$

In Section 4 we will conclude inductively from (3.9) that $\bar{\zeta}_h$ satisfies

$$\bar{\zeta}_h = \Delta_h (z + h^{\bar{p}} \bar{e}_{h,\bar{p}} + \ldots + h^{J+r-1} \bar{e}_{h,J+r-1}) + \bar{R}_h \tag{3.10 a}$$

with $\| \bar{R}_h \| = 0 (h^{J+r})$ and $\bar{p} := \min (p + q, J)$, where $q$ denotes the order of the method $\mathbb{M}$. In addition we will show that (3.10 b) and (3.10 c) hold. They are obtained from (3.9 b) and (3.9 c), respectively, by replacing $p$ by $\bar{p}$.

In Section 3.2 we introduced $J$ as the maximum attainable order. This meaning of $J$ follows immediately from (3.10): (3. 10 b) for $k = 0$ implies $\bar{e}_{h,\bar{p}} = 0 (1)$, i.e. $\bar{\zeta}_h$ is an approximation of order $\bar{p} = \min (p + q, J)$; thus the order of the defect corrections is limited by $J$. Consequently we make the obvious requirements:

$$J > p \ \text{ and } \ J > q.$$

If $\zeta_h$ is the result of a discretization method $\mathbb{M}^b$ which admits an asymptotic error expansion, then the coefficients $e_{h,j}$ of (3.9) are functions which are independent of $h$ $(e_{h,j} \equiv e_j)$; $B1)$ is trivially satisfied and the jumps of the derivatives, occuring in $B2)$, are zero (since (3.1) is a smooth problem, this follows from $A2$ of Section 3.4 and from Theorem 1.3.1 of Stetter [8]). According to the smoothness assumptions applying for problem (3.1), $\zeta_h$ posseses an asymptotic expansion up to an arbitrary order; in particular, up to an order $J+r-1$ (remainder term of order $J+r$). The order $J+r-1$, instead of $J$ (the maximum attainable order), was chosen because of technical reasons (mainly because $\lambda_j$ contains derivatives up to order $j+r$).

The general formula (3.9) covers not only those situations where $\zeta_h$ posseses an asymptotic expansion, but allows $\zeta_h$ to be the result of a discretization scheme, where the coefficient functions $e_{h,j}$ *do* depend on $h$, i.e. where (3.9) is *not* an asymptotic error expansion in the usual sense. In this situation, it is necessary to make additional assumptions about the smoothness of the functions $e_{h,j}$ for $h \to 0$; in particular, the assumption (3.9 b) for $k=0$. The rather complicated relations (3.9 b) for $k>0$ and (3.9 c) had to be introduced to cover the case when $\zeta_h$ is the result of a preceding IDeC-step.

**Note 3.5.1:** If $\zeta_h$ is the result of a *strongly stable linear multistep method* applied to IVPs of ODEs, then the functions $e_{h,j}$ are given by:

$$e_{h,j}(t) = e_j(t) + \sum_i x_i^{t/h} w_{ji}(t/h) \tag{3.11}$$

(cf. (4.4.21) of Stetter [8]), where the $x_i$ are the extraneous zeros of the characteristic polynomial $\rho$ with $|x_i|<1$ for all values of $i$. The $k$-th derivative of $x_i^{t/h}$ tends to zero faster than any given power of $h$ with $h \to 0$ so that $\| D^k e_{h,j}(t) \| = 0(1)$ which shows that $B1)$ is satisfied. $B2)$ is satisfied trivially. As a consequence, the IDeC-principle is applicable if $\mathbb{M}^b$ is a strongly stable linear multistep method, whereas IDeC-methods with $\mathbb{M}$ or $\mathbb{M}^j$ being a linear multistep method are *not* possible (cf. Section 3.4).

**Note 3.5.2:** If in Examples 3.2.1 and 3.2.2 the degree $M$ of the polynomials is chosen to be $J+r-1$, the maximum attainable order is $J$. In practical implementarions of IDeC-methods (alternative (i) of Section 3.1), however, the iteration will often be stopped before $J$ is reached (e.g. if already an "earlier" iterate satisfies an accuracy requirement). On the other hand, the question arises as to whether successive increases in the degrees of the polynomials used during the IDeC-iteration (say by introducing $\mathbb{V}_h^j$ with degrees of the polynomials $M^0 \leq M^1 \leq \ldots$) implies a corresponding increase in the maximum attainable order. From our analysis this question must be answered in the negative: If for the IDeC-step "$\zeta_h \to \bar{\zeta}_h$", as described in our metalgorithm, piecewise interpolation polynomials of degree $M = J+r-1$ are chosen, then the induction of Section 4 shows that the expansion (3.10) can be written up to order $J+r-1$, i.e. the expansion (3.9) and (3.10) are "of the same length". In a subsequent IDeC-step, proceeding from $\bar{\zeta}_h$, (3.10) takes the place of (3.9). Thus the expansion for the result of this new IDeC-step is also of length $J+r-1$, irrespective of the degree $\bar{M} > M$, used in this step.

**Note 3.5.3:** In $B1$) and $B2$) we did not assume an upper bound for the index $k$, due to our assumption about the smoothness of the data of (3.1). If these data are not functions in $C^\infty$, the analyses of Section 4 are also possible, but an upper bound for $k$ (which depends on the definition of $\mathbb{V}_h$, and which will be at least as large as $J+r$) must be introduced, and the smoothness of the data limits the maximum attainable order.

**Note 3.5.4:** (3.10) shows that an IDeC-step which goes from $\zeta_h$ to $\bar{\zeta}_h$ improves the order of approximation from $p$ to $\bar{p}$. Lemma 4.1 shows that the functions $\mathbb{V}_h \zeta_h$, $\mathbb{V}_h \bar{\zeta}_h$, ... have the same order of accuracy as the "discrete solutions" $\zeta_h$ and $\bar{\zeta}_h$ namely $p$ and $\bar{p}$. This result may be of some importance for practical implementations, because it shows how to obtain easily an accurate solution $(\mathbb{V}_h \bar{\zeta}_h(t))$ at (output-) points $t$ not belonging to $\mathbb{G}_h$.

### *3.6 Continuity Assumptions for the Variational Equations*

In Section 4 we will proof asymptotic results for IDeC-methods essentially by comparing corresponding terms of asymptotic error expansions for the problems (3.1) and (3.3), both solved with the same method $\mathbb{M}$. The coefficients of these asymptotic expansions satisfy *linear* differential equations (variational equations) belonging to the same class of problems as (3.1) and (3.3). We therefore require a continuous dependence of the solution (and its derivatives) of a linear problem on its data. Since the data of the neighboring problem (3.3) depend on $h$ the same is true for the data of the corresponding variational equations. Thus we consider two sequences (depending on the parameter $h$) of linear problems:

$$F_h^1\, z_h^1 \equiv G_h\,(y_h^1)\, z_h^1 + g_h^1 = 0 \tag{3.12 a}$$

$$F_h^2\, z_h^2 \equiv G_h\,(y_h^2)\, z_h^2 + g_h^2 = 0 \tag{3.12 b}$$

$$F_h^1, F_h^2 : E^h \to E^{0h};\ y_h^1, y_h^2 \in E^h;\ g_h^1, g_h^2 \in E^{0h};$$

$$G_h : E^h \to \mathrm{Lin}\,[E^h \to E^{0h}].$$

Now we introduce the projector $P$, which is defined on $E^{0h}$, such that those components of an arbitrary element $g \in E^{0h}$, which belong to initial and/or boundary conditions and jump-conditions are omitted in $P\,g$.

*We require that, for arbitrary sequences* (3.12), *the relations*

$$\| g_h^1 - g_h^2 \| = 0\,(h^{\max(0,\,\min(p,\,s_1))}) \tag{3.13 a}$$

$$\| D^k P\, g_h^1 - D^k P\, g_h^2 \| = 0\,(h^{\max(0,\,\min(p,\,s_1-k))}) \quad k=1,2,3,\ldots \tag{3.13 b}$$

$$\| D^k y_h^1 - D^k y_h^2 \| = 0\,(h^{\max(0,\,\min(p,\,s_2-k))}) \quad k=0,1,2,\ldots \tag{3.13 c}$$

*imply*

$$\| D^k z_h^1 - D^k z_h^2 \| = 0\,(h^{\max(0,\,\min(p,\,s_3))}), \quad k=0\,(1)\,r-1 \tag{3.14 a}$$

$$\| D^k z_h^1 - D^k z_h^2 \| = 0\,(h^{\max(0,\,\min(p,\,s_3-(k-r)))}), \quad k \geq r \tag{3.14 b}$$

*where* $s_1, s_2 \in \mathbb{N}$, $s_3 = \min\,(s_1, s_2)$.

For $D1$ the jump conditions inherent in $g_h^1 - g_h^2$ of (3.13) are of order $\max\,(0, \min\,(p, s_1)) + 1$ (cf. Note 3.3.1).

**Note 3.6.1:** The question arises how it may be verified in given situations, that (3.13) implies (3.14). In our earlier papers we discussed this question for special situations:

$\alpha$) For IVPs of first order ODEs — situation $D2$ — (cf. Frank, Ueberhuber [3]) the assertion of Lemma 2 [3] is identical with (3.14 a).

$\beta$) For BVPs of second order ODEs — situation $D1$ — (cf. Frank [1]) the assertion of Lemma 4.1 [1] is identical with (3.14 a). The considerations of [1] show that it is indeed important to require that the order of the jump conditions is $\max(0, \min(p, s_1)) + 1$ (cf. Note 3.3.1).

For $\alpha$) and $\beta$), (3.14 b) follows after the total differentiation of (3.12) with respect to the independent variable $t$.

For PDEs there might exist situations where the proposition "(3.13)$\Rightarrow$(3.14)" can be proved by reducing the PDEs to ODEs (via the method of lines).

*For $D1$ we further require, for a given sequence of linear problems*

$$F_h z_h \equiv G_h(y_h) z_h + g_h = 0, \tag{3.15}$$

*that for every $t \in B_h$ the relations*

$$\| D_+^k z_h(t) - D_-^k z_h(t) \| = 0(h^{\max(1, \min(p+1, s_4))}), \qquad k = i+1\,(1)\,r-1 \tag{3.16 a}$$

$$\| D_+^k P g_h(t) - D_-^k P g_h(t) \| = 0(h^{\max(1, \min(p+1, s_5-k))}), \quad k = 0, 1, 2 \ldots \tag{3.16 b}$$

$$\| D_+^k y_h(t) - D_-^k y_h(t) \| = 0(h^{\max(1, \min(p+1, s_6-k))}), \qquad k = i+1, i+2, \ldots \tag{3.16 c}$$

*imply*

$$\| D_+^k z_h(t) - D_-^k z_h(t) \| = 0(h^{\max(1, \min(p+1, s_7-(k-r)))}), \qquad k \geq r \tag{3.17}$$

*where $s_4, s_5, s_6, s_7 \in \mathbb{N}$, $s_7 = \min(s_4, s_5, s_6)$.*

**Note 3.6.2:** The validity of "(3.16)$\Rightarrow$(3.17)" may be derived, for most classes of problems, by the total differentiation of (3.15) at $t \in B_h$.

**Note 3.6.3:** For $D1$ we assumed that jump conditions for the jumps of the $k$-th derivatives, $k = i+1\,(1)\,r-1$, are included in $F$. Therefore the validity of (3.16 a) depends on the particular sequence of functions $\{\varphi_{h,k,l,\mu}\}_{h\to 0}$ for the considered sequence of problems (3.15).

## 4. Asymptotic Analysis

In this section an asymptotic analysis of our IDeC-metalgorithm will be presented. The main result is the following theorem:

**Theorem 4.1:** *Under the assumptions made in Section* 3, *in particular, if* (3.9) *is valid for $\zeta_h$, if $\mathbb{M}$ is a method of order $q$, and if* (3.6) *implies* (3.7) *for $\nabla_h$, the result $\bar{\zeta}_h$ of an IDeC-step satisfies*

$$\bar\zeta_h - \Delta_h z = 0\,(h^{\min(p+q,\,J)}) = 0\,(h^{\bar p}) \quad \textit{for} \quad h \to 0 \tag{4.1}$$

*provided that the maximum attainable order $J$ is larger than $p$ and $q$. Moreover, $\bar\zeta_h$ satisfies* (3.10), *i.e. further IDeC-steps are possible if $J > p+q$.*

### *4.1 Main Idea of the Proof*

Consider $\zeta_h$ satisfying (3.9), and the corresponding neighboring problem (3.3). If the method $\mathbb{M}$ is applied to the original problem (3.1) and to the neighboring problem (3.3), we have the asymptotic error expansions

$$\xi_{\bar h} = \Delta_{\bar h}\,(z + \bar h^q\, e_q + \bar h^{q+1}\, e_{q+1} + \ldots + \bar h^{J+r-1}\, e_{J+r-1}) + R \tag{4.2}$$

$$\pi_{\bar h} = \Delta_{\bar h}\,(\nabla_h\, \zeta_h + \bar h^q\, \hat e_{h,q} + \bar h^{q+1}\, \hat e_{h,q+1} + \ldots + \bar h^{J+r-1}\, \hat e_{h,J+r-1}) + \hat R_h \tag{4.3}$$

where $e_j$, $\hat e_{h,j} \in E^h$, $\bar h$ is the meshwidth parameter and $\| R \| = 0\,(\bar h^{J+r})$, $\| \hat R_h \| = 0\,(\bar h^{J+r})$, $R$, $\hat R_h \in E_{\bar h}$. (The notation $\bar h$ for the meshwidth parameter is used, to avoid confusion with the meshwidth parameter $h$, which determines $\zeta_h$ and therefore problem (3.3)). Note that for $D\,1$, where the elements of $E^{0h}$ contain "jump-components", normally the jump conditions for $e_j$ vanish, whereas, the jump conditions for $\hat e_{h,j}$ do *not* vanish. An example of the expansions (4.2) and (4.3) in the case $D1$ may be found in Frank [1]. The dependence of the data of the neighboring problem (3.3) on $h$ reflected by the subscript $h$ of $\hat e_{h,j}$ and $\hat R_h$ in (4.3). For the following asymptotic analyses, we will not consider only a fixed value of $h$ but a sequence of meshwidth parameters with $h \to 0$. This assumption requires the more precise notation

$$\| \hat R_h \| \le \mathrm{const}\,(h) \cdot \bar h^{J+r} \tag{4.4}$$

instead of $\| \hat R_h \| = 0\,(\bar h^{J+r})$.

As we assumed the same grid $\mathbb{G}_h$ for the computation of $\zeta_h$, $\pi_h$ and $\xi_h$ (cf. Section 3.1), we will choose $\bar h = h$. The subtraction of (4.3) from (4.2) after this substitution gives

$$\begin{aligned}(\xi_h - \Delta_h\, z) - (\pi_h - \Delta_h\, \nabla_h\, \zeta_h) = h^q\, \Delta_h\,(e_q - \hat e_{h,q}) + h^{q+1}\, \Delta_h\,(e_{q+1} - \hat e_{h,q+1}) + \ldots \\ \ldots + h^{J+r-1}\, \Delta_h\,(e_{J+r-1} - \hat e_{h,J+r-1}) + R - \hat R_h.\end{aligned} \tag{4.5}$$

If

$S\,1)$ $\quad \| e_j - \hat e_{h,j} \| = 0\,(h^{\max(0,\,\min(p,\,J-j))}), \quad j = q\,(1)\,J+r-1 \qquad (4.6)$

$S\,2)$ for const $(h)$ of (4.4) there exists a constant $C$ independent of $h$ such that const $(h) \le C$,

we can conclude from (4.5):

$$\begin{aligned}\bar\zeta_h - \Delta_h\, z = \xi_h - (\pi_h - \Delta_h\, \nabla_h\, \zeta_h) - \Delta_h\, z = (\xi_h - \Delta_h\, z) - (\pi_h - \Delta_h\, \nabla_h\, \zeta_h) = \\ = \Delta_h\,(h^{\bar p}\, \bar e_{h,\bar p} + \ldots + h^{J+r-1}\, \bar e_{h,J+r-1}) + \bar R_h\end{aligned} \tag{4.7}$$

with $\bar R_h = 0\,(h^{J+r})$, which is (3.10 a) of the desired result. In (4.7), the terms $\bar e_{h,j}$ are defined by the relations

$$h^j\, \bar e_{h,j} = h^{j-p}\,(e_{j-p} - \hat e_{h,j-p}), \quad j = p+q\,(1)\,J-1 \tag{4.8 a}$$

$$h^J \bar{e}_{h,j}=h^{J-p}(e_{J-p}-\hat{e}_{h,J-p})+\ldots+h^J(e_J-\hat{e}_{h,J})\ \text{[2]} \qquad (4.8\,\mathrm{b})$$

$$h^j \bar{e}_{h,j}=h^j(e_j-\hat{e}_{h,j}), \qquad j=J+1(1)\,J+r-1. \qquad (4.8\,\mathrm{c})$$

To derive (3.10 b) and (3.10 c), we have to verify the stronger assertion $S1'$ instead of $S1$:

$$S1')\qquad \| D^k e_j-D^k \hat{e}_{h,j} \| = \begin{cases} 0\,(h^{\max(0,\min(p,J-j))}), & k<r \\ 0\,(h^{\max(0,\min(p,J-j-(k-r))}), & k\geq r \end{cases} \qquad (4.9)$$

$$j=q\,(1)\,J+r-1, \qquad k=0,1,2,\ldots$$

$$\| D^k_+ \hat{e}_{h,j}(t)-D^k_- \hat{e}_{h,j}(t) \| =0\,(h^{\max(1,\min(p+1,J-j))}), \qquad i<k<r \quad (4.10\,\mathrm{a})$$

$$\| D^k_+ \hat{e}_{h,j}(t)-D^k_- \hat{e}_{h,j}(t) \| =0\,(h^{\max(1,\min(p+1,J-j-(k-r)))}) \qquad k\geq r \quad (4.10\,\mathrm{b})$$

where (4.10) for every $t\in B_h$ is required only for $D1$. It is easily proved that (4.9) and (4.10) imply (3.10).

It will be the main task of Section 4.2 to show the validity of $S\,1'$ and of $S\,2$ under the assumptions of Section 3.

### *4.2 Technical Details of the Proof*

*Proof of* $S1'$:

For technical reasons we need an auxiliary operator $\bar{\nabla}_h: E_h \to E^h$ which is defined by:

$C1)$ $\Delta_h \bar{\nabla}_h \eta=\eta$, $\eta\in E_h$, i.e., $\bar{\nabla}_h$ is an interpolation operator,

$$C2)\qquad \| D^k \bar{\nabla}_h \eta_h \| =0\,(h^{\max(1,J-(k-r))}), \qquad (4.11)$$

where $\{\eta_h\}$ denotes a sequence of elements of $E_h$ with $\| \eta_h \| =0\,(h^{J+r})$, $h\to 0$.

For instance $\bar{\nabla}_h$ may be chosen to be piecewise continuous Lagrange interpolation with polynomials of degree $J+r-1$; for $n>1$ no mixed derivatives $D^k$ are admitted according to $\beta$) of Example 3.2.1. Now we prove several lemmas:

**Lemma 4.1:** *Under the assumptions of Section* 3

$$\| D^k z-D^k \nabla_h \zeta_h \| =0\,(h^{\max(0,\min(p,J-(k-r)))}), \qquad k=0,1,2,\ldots. \qquad (4.12)$$

*Proof*: An auxiliary function $\Psi_h\in E^h$ is defined for every meshwidth parameter:

$$\Psi_h:=z+h^p e_{h,p}+h^{p+1} e_{h,p+1}+\ldots+h^{J+r-1} e_{h,J+r-1}+\bar{\nabla}_h R_h \qquad (4.13)$$

where $e_{h,j}\in E^h$ are the coefficients of the expansion (3.9). Note that, from (3.9 b) and from (4.11), it follows that $\{\Psi_h\}_{h\to 0}$ satisfies (3.6). The following estimation is possible, since $\zeta_h=\Delta_h \Psi_h$:

$$\| D^k z-D^k \nabla_h \zeta_h \| \leq \| D^k z-D^k \Psi_h \| + \| D^k \Psi_h-D^k \nabla_h \zeta_h \| \leq$$

$$\leq h^p \| D^k e_{h,p} \| +h^{p+1} \| D^k e_{h,p+1} \| +\ldots+h^{J+r-1} \| D^k e_{h,J+r-1} \| +$$

$$+\| D^k \bar{\nabla}_h R_h \| + \| D^k \Psi_h-D^k \nabla_h \Delta_h \Psi_h \|$$

which yields, together with (3.9), (4.11) and (3.7) the required result. □

[2] (4.8 b) reads for $J-p<q$: $h^J \bar{e}_{h,J}=h^q(e_q-\hat{e}_{h,q})+\ldots$.

For $D1$ we need the additional lemma:

**Lemma 4.2:** *Under the assumptions of Section* 3 *we have for every* $t \in B_h$

$$\| D_+^k (\nabla_h \zeta_h)(t) - D_-^k (\nabla_h \zeta_h)(t) \| = 0\,(h^{\max(1, \min(p+1, J-(k-r)))}), \quad k = i+1, i+2, \ldots \quad (4.14)$$

*Proof*: To derive (4.14), we will first investigate the order of the jumps of the auxiliary function $\Psi_h$ at the points $t \in B_h$:

$$\| D_+^k \Psi_h(t) - D_-^k \Psi_h(t) \| \leq \| D_+^k z(t) - D_-^k z(t) \| + h^p \| D_+^k e_{h,p}(t) - D_-^k e_{h,p}(t) \| + \ldots +$$
$$+ \| D_+^k (\bar{\nabla}_h R_h)(t) - D_-^k (\bar{\nabla}_h R_h)(t) \| = 0\,(h^{\max(1, \min(p+1, J-(k-r)))}) \qquad (4.15)$$

which follows from (3.9 c) and (4.11). Assume for the moment that $k < J + r$: From the identity $\nabla_h \zeta_h \equiv \nabla_h \Delta_h \Psi_h$ and from (3.7) it follows for $t \in B_h$:

$$\| D_+^k (\nabla_h \zeta_h)(t) - D_-^k (\nabla_h \zeta_h)(t) \| = \| D_+^k (\nabla_h \zeta_h)(t) - D_+^k \Psi_h(t) + D_+^k \Psi_h(t) -$$
$$- D_-^k \Psi_h(t) + D_-^k \Psi_h(t) - D_-^k (\nabla_h \zeta_h)(t) \| \leq$$
$$\leq \| D_+^k \nabla_h \Delta_h \Psi_h(t) - D_+^k \Psi_h(t) \| + \| D_+^k \Psi_h(t) - D_-^k \Psi_h(t) \| +$$
$$+ \| D_-^k \Psi_h(t) - D_-^k \nabla_h \Delta_h \Psi_h(t) \| =$$
$$= 0\,(h^{J-(k-r)}) + 0\,(h^{\max(1, \min(p+1, J-(k-r)))}) + 0\,(h^{J-(k-r)}).$$

Together with (3.8), which covers the case $k \geq J + r$, the assertion follows. □

With the help of the above two Lemmas (which follow essentially from the induction hypothesis (3.9)) we are now in a position to prove $S1'$. For this purpose we consider the variational equations (cf. Theorem 1.3.1 [8]) for $e_j$ and $\hat{e}_{h,j}$ respectively:

$$F'(z)\, e_j = d_{h,j}^1 = -\lambda_j z - \sum_{l=1}^{j-q} \lambda_l'(z)\, e_{j-l} + g_j(e_q, \ldots, e_{j-q}) \qquad (4.16\text{ a})$$

$$F'(\nabla_h \zeta_h)\, \hat{e}_{h,j} = d_{h,j}^2 = -\hat{\lambda}_{h,j}(\nabla_h \zeta_h) - \sum_{l=1}^{j-q} \hat{\lambda}_{h,l}'(\nabla_h \zeta_h)\, \hat{e}_{h,j-l} + \hat{g}_{h,j}(\hat{e}_{h,q}, \ldots, \hat{e}_{h,j-q}) \quad (4.16\text{ b})$$

where $\lambda_j$, $\hat{\lambda}_{h,j}: E^h \rightarrow E^{0h}$ are the coefficient operators of the local error mapping of problem (3.1) and (3.3), respectively. In both formulas, (4.16 a) and (4.16 b), the same operator $F'$ occurs, because the defect (inherent in the neighboring problem) does *not* depend on the dependent variable $y$, so that its Frechet derivative with respect to $y$ vanishes.

To proof $S1'$, we proceed in an inductive fashion, i.e. we assume that the validity of (4.9) (and the validity of (4.10) for $D1$) has already been verified for all $j'$ with $q \leq j' < j$ and derive (4.9) (and (4.10)) for $j$. For $j = q$ it will turn out that the validity of (4.9) and (4.10) follows immediately from Lemma 4.3, Lemma 4.4 and (4.22) of Lemma 4.6, because in this case only $e_q$ and $\hat{e}_{h,q}$ appear in the variational equations (4.16).

In Section 3.6 we required that for linear problems "(3.13)⇒(3.14)" and "(3.16)⇒(3.17)". To apply this requirement we will identify the problems (4.16 a) and (4.16 b) with (3.12 a) and (3.12 b), respectively; i.e. the following substitutions are made:

| Notation of Section 3.6 | $G_h$ | $y_h^1$ | $y_h^2$ | $g_h^1$ | $g_h^2$ | $z_h^1$ | $z_h^2$ |
|---|---|---|---|---|---|---|---|
| Notation of Section 4 | $F'$ | $z$ | $\nabla_h \zeta_h$ | $d_{h,j}^1$ | $d_{h,j}^2$ | $e_j$ | $\hat{e}_{h,j}$ |

. (4.17)

Thus, the essence consists in verifying that, after these substitutions, the relations (3.13) and (3.16) hold, and that the parameters $s_1, \ldots$ in those relations are such that (3.14) and (3.17) are identical with (4.9) and (4.10).

The assertion of Lemma 4.1 is identical with (3.13 c) for $s_2 = J + r$. For the derivation of (3.13 a) and (3.13 b), we conclude from Lemma 4.1 and Lemma 4.2 the following lemmas about the components of the decomposition (cf. Section 2):

$$\lambda_j = \mu_j + \gamma_j, \qquad \hat{\lambda}_{h,j} = \hat{\mu}_{h,j} + \hat{\gamma}_{h,j}.$$

**Lemma 4.3:** *Under the assumptions of Section* 3:

$$\| \mu_j z - \hat{\mu}_{h,j} (\nabla_h \zeta_h) \| = 0 \, (h^{\max(0, \min(p, J-j))}), \qquad j = q \, (1) \, J + r - 1 \tag{4.18}$$

*Proof*:

α) The functional part of the elements $\mu_j z, \hat{\mu}_{h,j} (\nabla_h \zeta_h) \in E^{0h}$ are linear combinations of $D^k z$ and $D^k (\nabla_h \zeta_h)$, respectively, where $k \leq j + r$. The coefficients or coefficient functions occuring in this linear combinations depend only on the operator $L$, which is the same for both problems (3.1) and (3.3). For $k \leq j + r$ we have (cf. (4.12)):

$$\| D^k z - D^k \nabla_h \zeta_h \| = 0 \, (h^{\max(0, \min(p, J-(k-r)))}) = 0 \, (h^{\max(0, \min(p, J-j))}).$$

β) In the situation $D1$ we must take the jump conditions into consideration. Since (3.1) is a smooth problem we have

$$\| D_+^k z(t) - D_-^k z(t) \| = 0, \qquad t \in B_h$$

whereas $\nabla_h \zeta_h$ has jumps whose behaviour is described in Lemma 4.2 (cf. (4.14)). For $k \leq j + r - 1$ (cf. $A2$ of Section 3.4) we can conclude:

$$\| (D_+^k z(t) - D_-^k z(t)) - (D_+^k (\nabla_h \zeta_h)(t) - D_-^k (\nabla_h \zeta_h)(t)) \| =$$
$$= 0 \, (h^{\max(1, \min(p+1, J-(k-r)))}) = 0 \, (h^{\max(0, \min(p, J-j)) + 1})$$

which together with Note 3.3.1 gives the required result. □

**Lemma 4.4:** *Under the assumptions of Section* 3

$$\| \gamma_j z - \hat{\gamma}_{h,j} (\nabla_h \zeta_h) \| = 0 \, (h^{\max(0, \min(p, J-j))}), \qquad j = q \, (1) \, J + r - 1. \tag{4.19}$$

*Proof* (Sketch): According to our assumption (cf. $A2$ of Section 3.4) the jump conditions of $\gamma_j$ and $\hat{\gamma}_{h,j}$ vanish. Thus it remains to investigate the functional part of $\gamma_j z, \hat{\gamma}_{h,j} (\nabla_h \zeta_h) \in E^{0h}$, whose structure was discussed in Section 2. If e.g. $\gamma_j z$ contains a term $f_z(t, z) D^2 z$ and the corresponding term of $\hat{\gamma}_{h,j} (\nabla_h \zeta_h)$ is $(\partial / \partial z) [f(t, \nabla_h \zeta_h) + d_h(t)] D^2 (\nabla_h \zeta_h) = f_z(t, \nabla_h \zeta_h) D^2 (\nabla_h \zeta_h)$ then the following estimation is possible:

$$\| f_z(t, z) D^2 z - f_z(t, \nabla_h \zeta_h) D^2 (\nabla_h \zeta_h) \| \leq \| f_z(t, z) D^2 z - f_z(t, z) D^2 (\nabla_h \zeta_h) \| +$$
$$+ \| f_z(t, z) D^2 (\nabla_h \zeta_h) - f_z(t, \nabla_h \zeta_h) D^2 (\nabla_h \zeta_h) \|$$

and from Lemma 4.1 and the Lipschitz continuity of $f_z$ the desired order can be derived. With exactly the same technique situations can be covered, where $\gamma_j$ contains more complicated terms, and the required result follows from the assumptions about the structure of $\gamma_j$ (cf. Section 2). □

Lemma 4.3 gives together with Lemma 4.4

$$\| \lambda_j z - \lambda_{h,j}(\nabla_h \zeta_h) \| = 0\,(h^{\max(0,\min(p,J-j))}), \qquad j = q\,(1)\,J + r - 1. \tag{4.20}$$

For $j > q$ the inhomogeneities of (4.16), $d^1_{h,j}$ and $d^2_{h,j}$, contain terms of the form $\lambda'_l(z)\, e_{j-2}$ and $\lambda'_{h,l}(\nabla_h \zeta_h)\, \hat{e}_{h,j-2}$, respectively, beside $\lambda_j z$ and $\hat{\lambda}_{h,j}(\nabla_h \zeta_h)$. For these terms the following estimation is possible.

**Lemma 4.5:** *From the assumptions of Section* 3 *and from the validity of* $S1'$ *for* $j' < j$, *it follows*

$$\| \lambda'_l(z)\, e_{j-l} - \hat{\lambda}'_{h,l}(\nabla_h \zeta_h)\, \hat{e}_{h,j-l} \| = 0\,(h^{\max(0,\min(p,J-j))}). \tag{4.21}$$

*Proof* (Sketch): If in the following relation

$$\lambda'_l(z)\, e_{j-l} - \hat{\lambda}_{h,l}(\nabla_h \zeta_h)\, \hat{e}_{h,j-l} = (\mu'_l(z)\, e_{j-l} - \hat{\mu}'_{h,l}(\nabla_h \zeta_h)\, \hat{e}_{h,j-l}) +$$
$$+ (\gamma'_l(z)\, e_{j-l} - \hat{\gamma}'_{h,l}(\nabla_h \zeta_h)\, \hat{e}_{h,j-l})$$

the first term (in parantheses) of the right-hand side is replaced by $(\mu_l(e_{j-l}) - \hat{\mu}_{h,l}(\hat{e}_{h,j-l}))$ [$\mu_l$ is a *linear* differential operator!], then the desired result (4.21) can be easily derived inductively from the assumed validity of (4.9) for $j' < j$ (and from (4.10) for $j' < j$ in the case $D1$) and from the assumptions about $\mu_l$ and $\gamma_l$ made in Section 2. □

Similar considerations for the terms $g_j(e_q, \ldots, e_{j-q})$ and $\hat{g}_{h,j}(\hat{e}_{h,q}, \ldots, \hat{e}_{h,j-q})$ lead to the desired relation (3.13 a) with $s_1 = J - j$ (cf. (4.17)). To establish the validity of (3.13 b) with $s_1 = J - j$ we need the following lemma:

**Lemma 4.6:** *Under the assumptions of Section* 3

$$\| D^k P \lambda_j z - D^k P \hat{\lambda}_{h,j}(\nabla_h \zeta_h) \| = 0\,(h^{\max(0,\min(p,J-j-k))}), \qquad k = 1, 2, \ldots \tag{4.22}$$

$$\| D^k P \lambda'_l(z)\, e_{j-l} - D^k P \hat{\lambda}'_{h,l}(\nabla_h \zeta_h)\, \hat{e}_{h,j-l} \| = 0\,(h^{\max(0,\min(p,J-j-k))}),$$
$$l < j, \qquad k = 1, 2, \ldots. \tag{4.23}$$

This lemma can be proved in an analogous way to that of Lemmas 4.3, 4.4 and 4.5.

Thus we have proved that (3.13) (with the substitutions (4.17)) holds, and consequently (3.14) is valid, which is identical with the assertion (4.9) for $j$. Similar considerations lead to relations (3.16) with $s_7 = J - j$, if (3.15) is identified with (4.16 b), i. e. if the following substitutions are made:

| Notation of (3.15) | $G_h$ | $y_h$ | $g_h$ | $z_h$ |
|---|---|---|---|---|
| Notation of (4.16 b) | $F'$ | $\nabla_h \zeta_h$ | $d^2_{h,j}$ | $\hat{e}_{h,j}$ |

*Sketch of Proof of S2:*

To proof $S2$, all terms forming the remainder of (1.3.12) of Stetter [8] have to be discussed separately. The desired result then follows from the smoothness assumptions on $f(t, y)$ and from the above lemmas. □

## 5. Conclusion

In this paper, we have provided a basis for the practical implementations of the IDeC-principle applied to discretizations for differential equations. The components of an IDeC-method have been investigated step by step w.r.t. the conditions which they must satisfy in order to produce an IDeC-method with a certain desired order of accuracy. In Part 2 of this paper we will demonstrate practical implications of the results presented in Part 1.

### References

[1] Frank, R.: The Method of Iterated Defect Correction and its Application to Two-Point Boundary Value Problems Part I: Num. Math. *25*, 409—419 (1976); Part II: Num. Math. *27*, 407—420 (1977).

[2] Frank, R., Hertling, J.: Die Anwendung der Iterierten Defektkorrektur auf das Dirichletproblem, Report No. 20/76, Inst. F. Num. Math., Technical University of Vienna, 1976 (to appear).

[3] Frank, R., Ueberhuber, C. W.: Iterated Defect Correction for Runge Kutta Methods, Report No. 14/75, Inst. f. Num. Math., Technical University of Vienna, 1975.

[4] Frank, R., Ueberhuber, C. W.: Collocation and Iterated Defect Correction, in: Lecture Notes in Mathematics, Vol. 631, pp. 19—34. Berlin-Heidelberg-New York: Springer 1978.

[5] Frank, R., Ueberhuber, C. W.: Iterated Defect Correction for the Efficient Solution of Stiff Systems of Ordinary Differential Equations. BIT *17*, 146—159 (1977).

[6] Hildebrand, F. B.: Introduction to Numerical Analysis, 2nd ed. New York: McGraw-Hill 1974.

[7] Stetter, H. J.: Economical Global Error Estimation, in: Stiff Differential Systems (Willoughby, R. A., ed.), pp. 245—258. New York-London: Plenum Press 1974.

[8] Stetter, H. J.: Analysis of Discretization Methods for Ordinary Differential Equations. Berlin-Heidelberg-New York: Springer 1973.

[9] Stetter, H. J.: The Defect Correction Principle and Discretization Methods, Report No. 26/77, Inst. f. Num. Math., Technical University of Vienna, 1977 (to appear).

[10] Swartz, B. K., Varga, R. S.: Error bounds for Spline and *L*-Spline Interpolation. J. Approximation Theory *6*, 6—49 (1972).

[11] Zadunaisky, P. E.: A Method for the Estimation of Errors Propagated in the Numerical Solution of a System of Ordinary Differential Equations, in: Proc. Astron. Union, Symposium No. 25. Academic Press 1966.

[12] Zadunaisky, P. E.: On the Estimation of Errors Propagated in the Numerical Integration of Ordinary Differential Equations. Num. Math. *27*, 21—39 (1976).

Dr. R. Frank, Dr. C. W. Ueberhuber
Institut für Numerische Mathematik
Technische Universität Wien
Gußhausstraße 27—29
A-1040 Wien
Austria