Computing, Suppl. 5, 1 – 32 (1984)

The Defect Correction Approach

K. Böhmer, Marburg, **P. Hemker,** Amsterdam, and **H. J. Stetter,** Wien

Abstract

This is an introductory survey of the defect correction approach which may serve as a unifying frame of reference for the subsequent papers on special subjects.

1. Introduction

There are many ways to introduce defect corrections. In this expository article we motivate the defect correction approach from its basic idea:

For a given mathematical problem and a given approximate solution,

- define the *defect* as a quantity which indicates how well the problem has been solved,
- use this information in a *simplified version* of the problem to obtain an appropriate *correction* quantity,
- apply this correction to the approximate solution to obtain a new (better) approximate solution.

Naturally, the procedure may now be repeated.

Of course, this fundamental approach has been used in mathematics since long. We give some examples in Chapter 2. In Chapter 3 we formalize the general defect correction principle and describe several processes which implement it.

Since defect corrections are especially powerful in combination with discretizations of analytic problems, in Chapter 4 we review discretization methods and asymptotic expansions for their local and global discretization errors. In Chapter 5, we establish the general framework for the combination of defect corrections with discretization methods, and we survey a variety of algorithms of this kind. The powerful multigrid approach is interpreted as a particularly interesting application of the defect correction principle in Chapter 6.

2. Historical Examples of Defect Correction

Prototypes of defect correction are the classical procedures for the calculation of a zero of a nonlinear function in one variable: An approximation $\tilde{z}$ of the solution z^* of the problem

$$F(z)=0 \tag{2.1}$$

is substituted into F; the value of $F(\tilde{z})$ defines the defect. The simplified version of (2.1) which yields the correction of $\tilde{z}$ is some local linearization

$$\tilde{F}(z):=F(\tilde{z})+m(z-\tilde{z})=0, \tag{2.2}$$

where $m\approx F'(z^*)$; see Fig. 1. Newton's method is a more refined case where m is updated during the iteration.

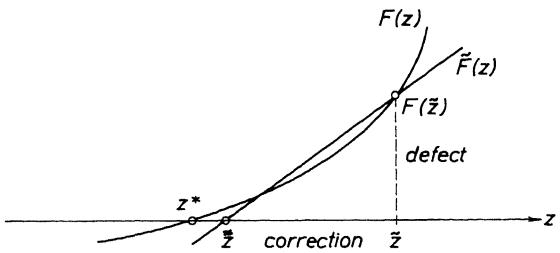

Fig. 1

Another well-known prototype is "iterative refinement" ("Nachiteration") in the numerical solution of linear algebraic equations

$$A z=b. \tag{2.3}$$

After an approximate solution $\tilde{z}$, with an unknown round-off contamination, has been obtained from a direct solution procedure, its defect $d:=A\tilde{z}-b$ is computed (with special care). Then the matrix decomposition of the previous solution process is used once more to compute a correction Δz from

$$A\,\Delta z=d. \tag{2.4}$$

Here the use of the *numerically transformed A* represents the "simplified version" of (2.3): if (2.4) could be solved exactly it would naturally yield the exact correction.

If z is a discrete approximate solution of an analytic problem, the defect formation becomes a non-trivial part of the procedure. In the late forties, L. Fox ([17], [18]) considered the discretization of a second-order boundary value problem

$$\begin{aligned} -z''(t)+p(t)\,z(t)&=q(t) \quad \text{on } (a,b),\\ z(a) \text{ and } z(b) &\text{ given}, \end{aligned} \tag{2.5}$$

by central second-order differences on an equidistant grid $\mathbb{G}$ in $[a,b]$. He suggested (though not in these terms) that a defect of an approximate solution $\tilde{z}=\{\tilde{z}(t_\nu), t_\nu\in\mathbb{G}\}$ might be defined via substitution of $\tilde{z}$ into a discretization of (2.5) which included

4-th order differences. This defect $d=\{d_\nu\}$ could be used as inhomogeneity in the problem for the correction function Δz

$$\begin{aligned} -\Delta z''(t)+p(t)\,\Delta z(t)&=d(t) \quad \text{on } (a,b),\\ \Delta z(a)=\Delta z(b)&=0; \end{aligned} \tag{2.6}$$

(2.6) could then be solved again by the basic (= "simplified") second-order discretization method. Fox considered the recursive application of this approach, with the inclusion of differences of higher and higher order into the computation of the defect d. He and others applied this method to a variety of problems, see e.g. [19], [20]. Fox's approach was later put into a more general, abstract frame-work by Pereyra ([41]–[44]) and effectively implemented; see Section 5.2.1 and Pereyra's paper in this volume.

A further generalization of the defect correction principle and an increase of the interest in the subject were initiated by the presentation of a paper "On the estimation of errors propagated in the numerical solution of ordinary differential equations" by P. E. Zadunaisky at the 1973 Dundee Conference on Numerical Analysis. Zadunaisky's heuristic technique turned out to permit an interpretation in terms of defect correction which represented a novel realization of the old idea; see Section 5.2.2. This brings us to the contemporary view of the subject.

3. General Defect Correction Principles

3.1 Basic Defect Correction Processes

We wish to "solve" the equation

$$Fz=y, \tag{3.1}$$

where $F: D\subset E\to \hat{D}\subset \hat{E}$ is a bijective continuous, generally nonlinear operator; $E, \hat{E}$ are Banach spaces. The domain D and the range $\hat{D}$ are closed subsets depending on F; $\hat{D}$ contains an appropriate neighbourhood of y. Hence, for every $\tilde{y}\in\hat{D}$ there exists, in D, exactly one solution of $Fz=\tilde{y}$; the solution of the given problem (3.1) will be called z^*.

We assume that (3.1) cannot be solved directly, but that the *defect*

$$d(\tilde{z}):=F\tilde{z}-y \tag{3.2}$$

may be evaluated for "approximate solutions" $\tilde{z}\in D$. Furthermore, we assume that we can readily solve the *approximate problem*

$$\tilde{F}z=\tilde{y} \tag{3.3}$$

for $\tilde{y}\in\hat{D}$, i.e. that we can evaluate the solution operator $\tilde{G}$ of (3.3). $\tilde{G}:\hat{D}\to D$ is an *approximate inverse* of F such that (in some appropriate sense)

$$\tilde{G}F\tilde{z}\approx\tilde{z} \quad \text{for } \tilde{z}\in D \tag{3.4}$$

and

$$F\tilde{G}\tilde{y}\approx\tilde{y} \quad \text{for } \tilde{y}\in\hat{D}. \tag{3.5}$$

Let us now assume that we know some approximation $\tilde{z} \in D$ for z^* and that we have computed its defect (3.2). In the general (nonlinear) case, there are two ways to use this information for the computation of a (hopefully better) approximation $\tilde{\tilde{z}}$ by means of solving problems of type (3.3); see Fig. 2:

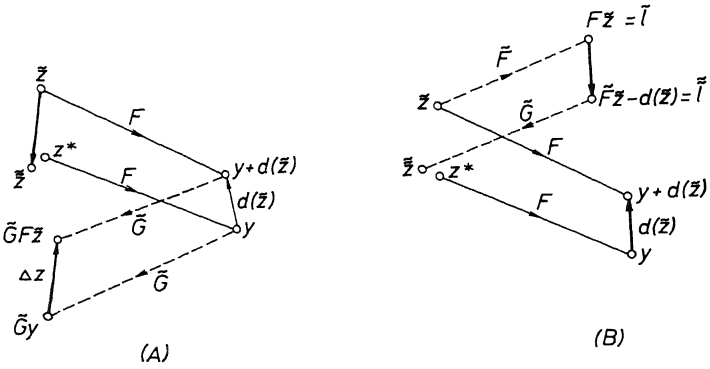

Fig. 2

(A) We compute the change Δz in the solution of (3.3) when the right hand side y is changed by $d(\tilde{z})$. We then use Δz as a correction for $\tilde{z}$, i.e. we transfer the observed change to our target problem (3.1):

$$\begin{aligned} \tilde{\tilde{z}} &:= \tilde{z} - \Delta z = \tilde{z} - [\tilde{G}(y + d(\tilde{z})) - \tilde{G}y] \\ \tilde{\tilde{z}} &:= \tilde{z} - \tilde{G}F\tilde{z} + \tilde{G}y. \end{aligned} \tag{3.6}$$

(B) We generate an equation (3.3) with solution $\tilde{z}$ and change its right-hand side $\tilde{l} = \tilde{F}\tilde{z}$ by $d(\tilde{z})$. We then take the solution of this modified equation as $\tilde{\tilde{z}}$, i.e. we again transfer the effect observed for (3.3) to our target problem (3.1):

$$\begin{aligned} \tilde{\tilde{l}} &:= \tilde{l} - d(\tilde{z}) = \tilde{l} - F\tilde{G}\tilde{l} + y, \\ \tilde{\tilde{z}} &:= \tilde{G}\tilde{\tilde{l}} = \tilde{G}[(\tilde{F} - F)\tilde{z} + y]. \end{aligned} \tag{3.7}$$

Note that it is the existence of $\tilde{G}$ and not of $\tilde{F} = \tilde{G}^{-1}$ which is essential, as is immediately clear from (3.6) and (3.7). In some respect, versions (A) and (B) appear *dual* to each other.

In both approaches, the arising problems with modified right-hand sides are often called *neighboring problems*. In some applications, the operator $\tilde{F} - F$ in (3.7) is much simpler than either $\tilde{F}$ or F so that there is an advantage in using approach (B).

The success of the basic defect correction steps (3.6) or (3.7) depends on the *contractivity* of the operations $(I - \tilde{G}F) : D \to D$ or $(I - F\tilde{G}) : \hat{D} \to \hat{D}$ resp., since (3.6) implies

$$\tilde{\tilde{z}} - z^* = (I - \tilde{G}F)\tilde{z} - (I - \tilde{G}F)z^* \tag{3.8}$$

while (3.7) implies, with $\tilde{G}l^* = z^*$,

$$\tilde{\tilde{l}} - l^* = (I - F\tilde{G})\tilde{l} - (I - F\tilde{G})l^*. \tag{3.9}$$

This contractivity is, of course, closely related to the approximate inverse property of $\tilde{G}$, cf. (3.4) and (3.5) resp.

The element $\tilde{\tilde{z}}$ which we have gained through defect correction may be used in two ways:

- We may interpret $\tilde{z}-\tilde{\tilde{z}}$ as an *estimate of the error* $\tilde{z}-z^*$ of the original approximation $\tilde{z}$,
- we may subject $\tilde{\tilde{z}}$ as our new approximation to *another defect correction step.*

The *iterative use* of the basic defect correction procedures (3.6) or (3.7) leads to the Iterative Defect Dorrection (IDeC) algorithms of Stetter [51]:

$$\text{(A)} \qquad z_{i+1} := z_i - \tilde{G}\, F\, z_i + \tilde{G}\, y, \tag{3.10}$$

$$\text{(B)} \quad l_{i+1} := l_i - F\, \tilde{G}\, l_i + y, \text{ with } z_i = \tilde{G}\, l_i; \tag{3.11}$$

for injective $\tilde{G}$, (3.11) turns into

$$z_{i+1} := \tilde{G}\,[(\tilde{F}-F)\, z_i + y]. \tag{3.11 a}$$

Usual starting values for these iterations are $z_0 = \tilde{G}\, y$ and $l_0 = y$.

The contractivity of the operators $I-\tilde{G}\,F$ or $I-F\,\tilde{G}$ resp. implies the convergence of these iterations, cf. (3.8) and (3.9): The z_i of (3.10) converge to z^* while the l_i of (3.11) converge to l^*, which implies the convergence of $z_i := \tilde{G}\, l_i$ to z^*. (Restrictions arising from an implementation in a finite computer arithmetic have been disregarded.)

If the approximate inverse $\tilde{G}$ is an *affine* mapping, i.e. if

$$\tilde{G}\, y_1 - \tilde{G}\, y_2 = \tilde{G}'(y_1 - y_2), \qquad y_1, y_2 \in \hat{D}, \tag{3.12}$$

with a fixed linear operator $\tilde{G}'$, the two versions merge into the familiar *linear version* of the basic defect correction step

$$\tilde{\tilde{z}} = \tilde{z} - \tilde{G}'\, d(\tilde{z}) = [I - \tilde{G}'\, F]\, \tilde{z} + \tilde{G}'\, y \tag{3.13}$$

which leads to the linear IDeC algorithm

$$z_{i+1} := z_i - \tilde{G}'\, d(z_i) = [I - \tilde{G}'\, F]\, z_i + \tilde{G}'\, y. \tag{3.14}$$

Now the contractivity of $I-\tilde{G}'\,F$ (or equivalently of $I-F\,\tilde{G}'$) becomes the condition for convergence to z^*. Note that in (3.14) there is no need for F of (3.1) to be linear.

In (3.14), $\tilde{G}'\, y \in D$ is a fixed element which has to be computed only once and is usually taken as starting approximation z_0. If $y=0$, this term vanishes.

Often, the approximate inverse $\tilde{G}$ is Frechet-differentiable, i.e.

$$\tilde{G}(y + \Delta y) - \tilde{G}\, y \approx \tilde{G}'(y)\, \Delta y. \tag{3.15}$$

With this *linearization*, (3.6) yields the new approximation

$$\tilde{\tilde{z}} := \tilde{z} - \tilde{G}'(y)\, d(\tilde{z}) \tag{3.16 a}$$

while (3.7) yields $\tilde{G}\,\tilde{\tilde{l}} = \tilde{G}(\tilde{l} - d(\tilde{z}))$ and

$$\tilde{\tilde{z}} = \tilde{z} + \tilde{G}(\tilde{l} - d(\tilde{z})) - \tilde{G}\,\tilde{l} \approx \tilde{z} - \tilde{G}'(\tilde{l})\, d(\tilde{z}), \tag{3.16 b}$$

cf. also Fig. 2, (A) and (B).

Thus, from Version (A) as well as from Version (B) of nonlinear defect correction, we are led to the linear defect corrections (3.13) and (3.14), with the Frechet derivative of $\tilde{G}$ at some appropriate (generally fixed) argument as linear operator $\tilde{G}'$. (3.16) is the basis for Böhmer's ([8]–[10]) *Discrete Newton Methods*; see Section 5.2.5.

Sometimes, it is important that the neighboring problems do not deviate too far from the original problem (3.1). This may be effected through the observation that, for differentiable $\tilde{G}$,

$$\tilde{G}(y+d)-\tilde{G}y \approx \mu\left[\tilde{G}\left(y+\frac{1}{\mu}d\right)-\tilde{G}(y)\right]. \tag{3.17}$$

This transforms the basic step (3.6) of version (A) into

$$\tilde{\tilde{z}}=\tilde{z}-\mu\,\tilde{G}\left(y+\frac{1}{\mu}d(\tilde{z})\right)+\mu\,\tilde{G}(y).$$

The trick (3.17) is used by several authors.

3.2 Corrections with Varying Inverses or Defects

In this section we extend the simple idea of IDeC: We allow different approximate inverses or different defects in one iteration process. These extensions are useful mainly in connection with discretization methods; see Chapters 5 and 6. Further extensions are possible; see e.g. Hemker's paper in this volume.

For the solution of (3.1), it is not necessary to use a fixed approximate inverse in the IDeC process; as in the classical Newton method, one may use a different $\tilde{G}$ in each iteration step:

$$\text{(A)}\quad z_{i+1}:=z_i-\tilde{G}_{i+1}Fz_i+\tilde{G}_{i+1}y, \tag{3.18}$$

$$\text{(B)}\quad l_{i+1}:=l_i-F\tilde{G}_{i+1}l_i+y; \tag{3.19}$$

a similar modification for the linear IDeC is obvious.

In this way, we are able to adapt the approximate inverse during the iteration and we may try to find sequences $\{\tilde{G}_i\}$ which accelerate the convergence of the iteration. Various ways are known to design a suitable sequence $\{\tilde{G}_i\}$. We mention a few examples:

Example 1: $\tilde{G}_{i+1}=\tilde{G}(z_i)$.
The approximate inverse depends on the last iterand computed. This is the case e.g. in Newton's method for the solution of nonlinear equations, where $\tilde{G}(x)=(F'(x))^{-1}$, with $F'(x)$ the Frechet derivative of the operator F in (3.1). See also the strong Discrete Newton Methods in Section 5.2.5.

Example 2: $\tilde{G}_i=\tilde{G}(\omega_i)$.
The approximate inverse depends on a real parameter ω. This is the case e.g. in non-stationary relaxation processes for the solution of linear systems. The value ω_i may be taken from a fixed sequence of values or it may be computed adaptively during the iteration process.

Example 3: $\tilde{G}_i \in \{G_1, G_2\}$.
In each iteration step the approximate inverse is chosen from a set of two (or more) fixed approximate inverses. This is the case e.g. in Brakhage's and Atkinson's methods for the solution of Fredholm integral equations of the 2nd kind. (See Atkinson [4] and Brakhage [12].)

We now assume that F and two approximate inverses, $\tilde{G}$ and $\tilde{\tilde{G}}$, are linear operators. We consider an alternating use of these two inverses in successive IDeC steps. Then the iteration steps (cf. (3.14))

$$z_{i+\frac{1}{2}} := (I - \tilde{G}F)\, z_i + \tilde{G}\, y,$$
$$z_{i+1} := (I - \tilde{\tilde{G}}F)\, z_{i+\frac{1}{2}} + \tilde{\tilde{G}}\, y,$$

combine into a single iteration step of the form

$$z_{i+1} := (I - \tilde{\tilde{G}}F)(I - \tilde{G}F)\, z_i + (\tilde{\tilde{G}} + (I - \tilde{\tilde{G}}F)\,\tilde{G})\, y. \tag{3.20}$$

This is easily recognized as one iteration step of type (3.14) with the approximate inverse

$$G = \tilde{\tilde{G}} - \tilde{\tilde{G}}F\tilde{G} + \tilde{G} = [I - (I - \tilde{\tilde{G}}F)(I - \tilde{G}F)]\, F^{-1}. \tag{3.21}$$

The error amplification operator of the new iteration step (3.20) is obviously the product of the amplification operators of the constituent steps.

Analogously we find that a sequence of σ consecutive linear defect correction steps (3.14) may be interpreted as one combined step with an approximate inverse

$$\hat{G} = \sum_{m=0}^{\sigma-1} (I - \tilde{G}F)^m\, \tilde{G} = [I - (I - \tilde{G}F)^{\sigma}]\, F^{-1} \tag{3.22}$$

and an amplification operator $(I - \tilde{G}F)^{\sigma}$.

Another form of unsteadiness may enter into IDeC algorithms through a *varying defect definition*. This situation arises naturally when the original problem is set in an infinitely-dimensional space (differential equations, integral equations, etc.) while the numerical problem (3.1) is a *finite* (discretized) version of the original problem. Here, the exact solution z^* of (3.1) represents only an approximate solution of the original problem, and the "truncation error" of z^* constitutes a natural limit for the accuracy with which a solution of (3.1) is reasonably requested. We will study this situation in more detail in Section 5.1.

Also, there may exist a *sequence* of discretizations of the original problem

$$F_k z = y_k, \qquad k = 1, 2, \ldots, \tag{3.23}$$

with the property that the truncation error of (3.23) decreases as k increases while the evaluation of the defects

$$d_k(z) = F_k z - y_k, \qquad k = 1, 2, \ldots \tag{3.24}$$

becomes costlier. In this situation, an IDeC algorithm for the solution of (3.23), with some fixed $\bar{k} > 1$, will become more economical if defects d_k with lower values of k are used in the initial stages of the iteration and k is successively increased towards $\bar{k}$ as the accuracy of the approximation z_i increases.

The simplest such *updating* IDeC algorithms have the form ($i=0, 1, \ldots, \bar{k}-1$)

$$(A)\ z_{i+1} := z_i - \tilde{G} F_{i+1} z_i + \tilde{G} y_{i+1}, \tag{3.25}$$

$$(B)\ l_{i+1} := l_i - F_{i+1} \tilde{G} l_i + y_{i+1}, \tag{3.26}$$

or

$$(L)\ z_{i+1} := z_i - \tilde{G}' d_{i+1}(z_i) \tag{3.27}$$

cf. the basic versions (3.10)/(3.11) and their common linearization (3.14). The approximate inverse $\tilde{G}$ will generally fit all the F_k and may be used throughout. Naturally, a simultaneous updating of $\tilde{G}$ (see (3.18) and (3.19)) is possible.

Fox's difference correction procedure which was sketched in Chapter 2 is a special case of (3.26); further applications of this approach will be presented in Section 5.2.1.

3.3 Interval Defect Correction

Strict error bounds or inclusion intervals for approximate solutions have always been a goal of constructive mathematics. In order to be useful such error bounds should be realistic: they must not overestimate the actual errors by orders of magnitude. One of the tools for the generation of such error bounds is the considerate use of sets, particularly of intervals.

Let the target problem (3.1) be equivalent to the *fixed point equation*

$$z = Tz \tag{3.28}$$

where T is a contractive mapping of $D \subset E$ into itself (cf. (3.1)). Then (under suitable technical hypotheses) the iteration

$$z_{i+1} := T z_i \tag{3.29}$$

and its set counterpart

$$Z_{i+1} := T Z_i \tag{3.30}$$

both converge to the fixed point z^* of (3.28) for $z_0 \in D$ and $Z_0 \subset D$ resp. Trivially, $z^* \in Z_0$ implies $z^* \in Z_i$ for all i. In this case, the sequence $\{Z_i\}$ furnishes better and better inclusions of z^* which may be interpreted as error bounds.

In actual computation, the sets Z_i will normally be *intervals*:

$$\bar{Z} := \{z \in E : \underline{z} \leq z \leq \bar{z}\}, \tag{3.31}$$

where $\leq$ refers to some partial ordering of E. The mapping T must be expanded into a mapping $\bar{T}$ of the set of intervals in D into itself e.g. by defining $\bar{T}\bar{Z}$ as the smallest interval containing $T\bar{Z}$. Thus (3.30) becomes

$$\bar{Z}_{i+1} := \bar{T} \bar{Z}_i. \tag{3.32}$$

Under the condition

$$T\bar{Z} \subset \bar{T}\bar{Z} \quad \text{for each interval } \bar{Z} \text{ in } D, \tag{3.33}$$

which is trivial for the definition of $\bar{T}$ suggested above, the sequence $\{\bar{Z}_i\}$ also furnishes inclusions of z^* if $z^* \in \bar{Z}_0$. The convergence to the fixed point z^* is normally not preserved, however; in more than one dimensions the intervals $\bar{Z}_i$ may cease to shrink from some finite $\hat{i}$ on and no better inclusion of z^* can be obtained from (3.32).

On the other hand, it is well-known from various *fixed-point theorems* that the establishment of

$$TY \subset Y \quad \text{for some} \quad Y \subset D \tag{3.34}$$

may imply the *existence* of a fixed point z of (3.28) in Y (hence in TY) and thus the existence of a solution of (3.1). If Y is mapped into its interior proper

$$\operatorname{clos}(TY) \subset \overset{\circ}{Y} \tag{3.35}$$

then the fixed point (solution) is *unique* in Y. For more details, see e.g. the papers by Rump and Kaucher-Miranker in this volume.

Due to (3.33), the satisfaction of (3.34) or (3.35) for $\bar{T}$ also implies the above conclusions about the fixed point of T. In finite dimensions, with the definition (3.31) of intervals as closed sets, $\bar{T}\bar{Z}$ is automatically closed. Furthermore, the inclusion of intervals with *machine elements* as bounds ("machine intervals") may simply be checked on a computer. Thus, an implementation $\bar{T}$ of T which maps the set of machine intervals in D into itself and which satisfies (3.33) is a suitable tool for a fully computational and mathematically rigorous proof of the existence and uniqueness of a solution of (3.1) in some computed interval about an approximate solution $\tilde{z}$ of (3.1).

Since the evaluation of interval mappings is computationally expensive, one will not attempt to iterate (3.32) and test $\bar{Z}_{i+1} \subset \bar{Z}_i$ after each iteration. Instead one will generate a good approximation $\tilde{z}$ for z^* in a conventional way (e.g. by defect correction); then one will consider intervals

$$\bar{Z} = \tilde{z} + \Delta Z, \tag{3.36}$$

as candidates for the establishing of

$$\bar{T}\bar{Z} \subset \bar{Z}. \tag{3.37}$$

Normally, ΔZ will have (machine element) bounds $-\Delta z$, Δz with $\| \Delta z \|$ small. (3.37) then implies the inclusion

$$-\Delta z \leq \tilde{z} - z^* \leq \Delta z. \tag{3.38}$$

Naturally, it will largely depend on the design of $\bar{T}$ whether we will be able to verify (3.37) for an interval (3.36) with a small ΔZ. The *defect correction principle* serves as a good basis for the construction of suitable mappings $\bar{T}$.

For simplicity, we start from a defect correction setting with a linear (or linearized) approximate inverse $\tilde{G}'$ (see (3.14)):

$$Tz := z - \tilde{G}'(Fz - y). \tag{3.39}$$

Since we plan to evaluate T – or its implementation $\bar{T}$ – for intervals $\bar{Z} = \tilde{z} + \Delta Z$ where $\tilde{z}$ has a small defect, we look for a *linearization* of the defect computation in a

neighborhood of $\tilde{z}$ which is *extendible to intervals*. The appropriate tool is an interval matrix L which satisfies, for some $\rho > 0$,

$$F z \in F\tilde{z} + L(z - \tilde{z}) \quad \text{for } \|z - \tilde{z}\| \leq \rho, \tag{3.40}$$

(The right-hand side of (3.40) is a set because the elements of L are intervals.) If F is differentiable, then any $L \supset \bigcup_{\|z-\tilde{z}\| \leq \rho} \{F'(z)\}$ is adequate.

Letting z range over $Z = \tilde{z} + \Delta z$ in (3.40) we obtain

$$FZ \subset F\tilde{z} + L\Delta Z \quad \text{for "small" } \Delta Z \tag{3.41}$$

and (see (3.39))

$$\begin{aligned} TZ &\subset \tilde{z} + \Delta Z - \tilde{G}'(F\tilde{z} - y) - \tilde{G}' L \Delta Z \\ &=: \tilde{z} + \hat{T}(\tilde{z})\Delta Z. \end{aligned} \tag{3.42}$$

This mapping $\hat{T}$ maps small sets ΔZ (i.e. sets contained in a neighborhood of 0) into the sets

$$\Delta \hat{Z} = \hat{T}(\tilde{z})\Delta Z := -\tilde{G}' d(\tilde{z}) + (I - \tilde{G}' L)\Delta Z. \tag{3.43}$$

Obviously (see (3.42)) $\Delta \hat{Z} \subset \Delta Z$ implies $T(\tilde{z} + \Delta Z) \subset \tilde{z} + \Delta Z$ and thus the desired inclusion $z^* \in \tilde{z} + \Delta Z$ (cf. (3.38)).

If $\widehat{T}$ is an interval implementation of $\hat{T}$ which satisfies

$$\hat{T}\Delta Z \subset \widehat{T}\Delta Z \tag{3.44}$$

then the computational verification of ($^\circ$ means "open interior")

$$\widehat{\Delta Z} := \widehat{T}(\tilde{z})\Delta Z \subset \overset{\circ}{\Delta Z} \tag{3.45}$$

implies

$$z^* \in \tilde{z} + \widehat{\Delta Z} \tag{3.46}$$

as well as the uniqueness of z^* in that interval.

(3.43) represents an *interval version of defect correction* which is computationally feasible and efficient at the same time. For the establishment of (3.45), it is essential that the defect $d(\tilde{z}) = F\tilde{z} - y$ of $\tilde{z}$ is small and that $I - \tilde{G}' L$ is sufficiently contractive. Applications of this approach for the algorithmic generation of guaranteed inclusion intervals for the solutions of various problems may be found in the papers by Rump and Kaucher-Miranker in this volume.

4. Discretization of Operator Equations

Defect correction in the general sense of Chapter 3 does not refer to special properties of the spaces E and $\hat{E}$ in which the equation (3.1) is posed. They may be infinitely dimensional function spaces and the mappings $F, \tilde{F}, \tilde{G}$ etc. may be analytic operations. For numerical computation, however, we must resort to *finite dimensions*: Elements of spaces must be specified by N-tuples of real numbers. If our original problem (3.1) is in an infinitely dimensional setting, we must model it in finite dimensions.

This may either be achieved by a "projection" of the problem into finite-dimensional subspaces of E and $\hat{E}$, or by an explicit design called *discretization* since it involves the use of discrete analogs of "continuous" elements and operations. Galerkin's method is a well-known example of the first approach while numerical quadrature formulas for integrals and difference quotients for derivatives exemplify the discretization approach. Here, functions are generally represented by their values at a finite set of arguments, the "grid"; they become *grid functions* from the grid points into some $\mathbb{R}^m$.

The discretization approach derives its flexibility and power from the simultaneous consideration of a parametrized infinite set or sequence of grids. This set contains increasingly "fine" grids, with increasingly many gridpoints. The corresponding spaces E^h of grid functions have higher and higher dimensions tending to infinity as the grid parameter $h \in \mathcal{H} \subset \mathbb{R}_+$ tends to zero.

It is this *asymptotic aspect* of the modelling which dominates much of the theory of discretization methods. Within a given computation, one can utilize only one grid or a few different grids; here the goal is to choose the grid(s) sufficiently fine so that the finite dimensional model becomes a sufficiently close replica of the original problem while the computational effort (which naturally increases with the refinement of the grid) remains sufficiently low. The knowledge of the asymptotic behavior helps in the choice of the right balance.

Although we assume that the reader is familiar with discretization methods, we give a synopsis of some fundamental concepts so that their use in the defect correction setting may be well understood. In more complicated situations, the formal definitions may not be fully applicable; they should then be taken as a guide for the analysis.

4.1 Fundamental Concepts in Discretization

The structural pattern underlying the discretization of an operator equation (cf. (3.1))

$$F z = y, \tag{4.1}$$

with $F : D \subset E \to \hat{D} \subset \hat{E}$, is the following:

$$h \in \mathcal{H}: \quad \begin{array}{ccc} E & \xrightarrow{F} & \hat{E} \\ {\scriptstyle \Delta^h}\downarrow & & \downarrow{\scriptstyle \hat{\Delta}^h} \\ E^h & \xrightarrow{F^h} & \hat{E}^h \end{array} \tag{4.2}$$

Thus (4.1) is modelled by an inifinite set ($h \in \mathcal{H}$) of problems

$$F^h z^h = y^h, \qquad h \in \mathcal{H}, \tag{4.3}$$

where each F^h maps a suitable domain in E^h into $\hat{E}^h$. Elements of the original spaces E and $\hat{E}$ are mapped into their "discrete images" in the finite dimensional Banach spaces E^h and $\hat{E}^h$ resp. by the surjective linear discretization mappings Δ^h and $\hat{\Delta}^h$

resp. A *discretization method* has to specify these spaces and mappings; in particular it must specify how the discrete operations F^h are obtained from the "continuous" operation F.

For some more sophisticated discretization methods, the element y^h in (4.3) is not $\hat{\Delta}^h y$ but it depends on F^h and z^h as well. In order that we may neglect this fine point, we write (4.1) equivalently as

$$(F-y)\,z := F z - y = 0 \tag{4.4}$$

and its set of discrete models as

$$(F-y)^h z^h = 0, \qquad h \in \mathscr{H}. \tag{4.5}$$

Note that $(F-y)$: $D \to \hat{E}$ maps $z \in D$ into its defect $d(z)$ with regard to the given problem (4.1).

The parameter set $\mathscr{H} \subset \mathbb{R}_+$ must have 0 as an accumulation point so that the limit $\{h \to 0, h \in \mathscr{H}\}$ is well-defined. Often $\mathscr{H}$ will be an infinite sequence, e.g. $\left\{h = \frac{1}{N},\ N \in \mathbb{N}\right\}$. Generally, the value of h is a measure of the refinement of the grid; for equidistant grids in $\mathbb{R}^1$, h may simply be the grid spacing or *stepsize*.

Distances in $E(\hat{E})$ and $E^h(\hat{E}^h)$ must correspond asymptotically, hence we require

$$\begin{aligned} &\lim_{h\to 0} \| \Delta^h z \|_{E^h} = \| z \|_E \quad \text{for } z \in E, \\ &\lim_{h\to 0} \| \hat{\Delta}^h y \|_{\hat{E}^h} = \| y \|_{\hat{E}} \quad \text{for } y \in \hat{E}. \end{aligned} \tag{4.6}$$

Note that $\lim_{h\to 0}$ always means $\lim_{h\to 0,\, h\in\mathscr{H}}$ and that the dimensions of E^h and $\hat{E}^h$ increase beyond limit as $h \to 0$.

In the setting of (4.2), the well-known asymptotic concepts of the theory of discretization methods are easily formulated. Concepts often carry the labels "global" or "local" if they are related to quantities in E^h or $\hat{E}^h$ resp. E.g., the mapping $E \to \hat{E}^h$ defined by

$$\Lambda^h(z) := (F-y)^h \Delta^h z - \hat{\Delta}^h (F-y) z \in \hat{E}^h, \ h \in \mathscr{H}, \ z \in D, \tag{4.7}$$

is called the *local discretization error* of the discretization (4.5) of (4.4).

The discretization is called *consistent* if

$$\lim_{h\to 0} \| \Lambda^h(z) \|_{\hat{E}^h} = 0, \qquad z \in D; \tag{4.8}$$

it is called *consistent of order p* if

$$\| \Lambda^h(z^*) \|_{\hat{E}^h} = \| (F-y)^h \Delta^h z^* \|_{\hat{E}^h} = 0(h^p), \ h \to 0, \tag{4.9}$$

where z^* is the solution of (4.4). Note that $\Lambda^h(z^*)$ is the *defect of* $\Delta^h z^*$ *in the discrete problem* (4.5).

On the other hand, with $(z^h)^*$ the solution of (4.5), the *global discretization error* is

$$e^h := (z^h)^* - \Delta^h z^* \in E^h, \quad h \in \mathscr{H}, \tag{4.10}$$

and *convergence* (*of order p*) means

$$\lim_{h \to 0} \| e^h \|_{E^h} = 0 \quad \text{and} \quad \| e^h \|_{E^h} = 0(h^p). \tag{4.11}$$

From the practical, nonasymptotic point of view, one would prefer estimates of the sort ($h \in \mathscr{H}$)

$$\begin{aligned} \| \Lambda^h(z^*) \|_{\hat{E}^h} &\leq \hat{C} h^p, \\ \| e^h \|_{E^h} &\leq C h^p, \end{aligned} \tag{4.12}$$

with realistic constants $\hat{C}$ and C. We will return to this in Section 4.2.

Naturally, a small defect $\Lambda^h(z^*)$ of $\Delta^h z^*$ in (4.5) should imply the closeness of $\Delta^h z^*$ to the solution $(z^h)^*$ of (4.5). For this, we need a Lipschitz bound on the inverse of $(F-y)^h$:

$$\| z^h - \bar{z}^h \|_{E^h} \leq S \, \| (F-y)^h z^h - (F-y)^h \bar{z}^h \|_{\hat{E}^h}, \tag{4.13}$$

for $z^h, \bar{z}^h \in D^h$. If S is a *uniform* constant for $h \in \mathscr{H}$ (i.e. for $h \to 0$), consistency of order p implies convergence of the same order, cf. (4.9)–(4.11). This uniformity in h of the *condition of* (4.5) with respect to perturbations of its right-hand side is called *stability*. In nonlinear problems, the size of the perturbation (i.e. the distance of the images under $(F-y)^h$, see (4.13)) may have to be restricted for (4.13) to hold. For more details on discretization methods, the relevant literature should be consulted, e.g. Stetter [49]. There one may also find examples of well-known discretizations described in the context of the above framework.

4.2 Asymptotic Expansions

Consider a discrete approximation $z^h \in E^h$ of the solution $z^* \in E$ of an analytic target problem (4.4) which has been obtained by a well-defined discretization algorithm. It is natural to expect that – as a function of the discretization parameter h – the global discretization error $e^h := z^h - \Delta^h z^*$ (see (4.10)) will have more structure than the uniform bound $C h^p$ (see (4.12)). In particular, the quantity $e^h/h^p \in E^h$ may be a discrete approximation of a fixed element $e_p \in E$:

$$\frac{1}{h^p} e^h = \frac{1}{h^p} [z^h - \Delta^h z^*] = \Delta^h e_p + r_p^h \text{ with } \lim_{h \to 0} \| r_p^h \| = 0. \tag{4.14}$$

Such an asymptotic statement makes sense only when we consider a parametrized situation as described in Section 4.1. Nevertheless, if $e_p \in E$ in (4.14) is known and if the remainder term r_p^h is sufficiently small for the value of h used in a given computation, (4.14) implies

$$z^h - \Delta^h z^* \approx h^p \Delta^h e_p \tag{4.15}$$

or

$$z_\nu^h - z^*(t_\nu^h) \approx h^p e_p(t_\nu^h), \tag{4.16}$$

which conveys considerably more qualitative and quantitative information than (4.12). The asymptotic statement (4.14), which may also be written as

$$z^h = \Delta^h [z^* + h^p e_p] + o(h^p), \tag{4.17}$$

is called an *asymptotic expansion of the global discretization error* of the set $\{z^h\}_{h \in \mathcal{H}}$ defined by the discretization algorithm under consideration. A more general asymptotic expansion (to order $J > p$) of the global discretization error is given by

$$z^h = \Delta^h [z^* + h^p e_p + h^{p+1} e_{p+1} + \ldots + h^J e_J] + o(h^J), \tag{4.18}$$

where the e_j, $j = p(1)J$, are fixed (i.e. h-independent) elements from E. $p \geq 1$ is the order of convergence of the discretization, cf. (4.11).

In special cases, (4.18) contains only *even powers* of h. A discretization algorithm which generates z^h with such an expansion is often called *symmetric*.

The deliberate consideration and use of asymptotic expansions in discretization algorithms began with Gragg's thesis [26]; the first general discussion was presented in Stetter [48], see also Stetter [49]. At first, the main objective was the justification of *Richardson extrapolation* where the mere existence of an asymptotic expansion (4.18) is used for the construction of a higher order approximation from several approximations z^{h_i}, $h_i \in \mathcal{H}$, $i = 1(1)r$, obtained on grids $\mathbb{G}^{h_i}$ with a non-empty intersection. In Romberg's quadrature method of this type, the asymptotic expansion for the underlying trapezoidal rule was trivially supplied by the Euler-McLaurin sum formula; now Gragg's results permitted the design of the Gragg-Bulirsch-Stoer extrapolation algorithm for initial value problems in ordinary differential equations (Gragg [27], Burlisch-Stoer [15]). This is also a nontrivial example of a symmetric discretization algorithm; the progression of the expansion (4.18) in even powers of h makes the extrapolation particularly effective.

An asymptotic expansion for a set $\{z^h\}_{h \in \mathcal{H}}$ of discrete approximations of $z \in E$ implies that the z^h have inherited a certain "*smoothness*" from z, uniformly in h. This means that *difference quotients* of such grid functions up to some order *remain bounded* for $h \to 0$ in spite of their $O(h^r)$ denominators and that they approximate the respective derivatives of z. This fact plays a central role in the design and analysis of more sophisticated discretization algorithms, e.g. those combining defect correction with discretization.

For illustration, assume that $E = C[0, 1]$ and the E^h employ equidistant grids with stepsize h. Consider the k-th forward difference quotients $\frac{1}{h^k} \delta^k z^h \in E^h$ defined by

$$(\delta^k z^h)_\nu := \sum_{\kappa=0}^{k} (-1)^{k-\kappa} \binom{k}{\kappa} z^h_{\nu+\kappa} \tag{4.19}$$

(with some appropriate modification for the last k gridpoints). Assume that $\{z^h\}_{h \in \mathcal{H}}$ represents a discrete approximation of order $p \geq 1$ of a fixed element $z^* \in C^{(p+1)}[0, 1] \subset E$ so that (4.11) holds. This implies, for $k \leq p$,

$$\frac{1}{h^k} \delta^k z^h = \frac{1}{h^k} \delta^k (\Delta^h z^*) + O(h^{p-k}) = \Delta^h z^{*(k)} + O(h) + O(h^{p-k}) \tag{4.20}$$

and difference quotients of an order $k > p$ may no longer be assumed to be bounded as $h \to 0$.

For $k = p$, the validity of (4.17), instead of (4.11), changes (4.20) into $\frac{1}{h^p} \delta^p z^h = \Delta^h z^{*(p)} + o(1)$. More generally, if $z^* \in C^{(J)}[0,1]$ and $e_j \in C^{(J-j)}[0,1]$, $j = p(1)J$, in the asymptotic expansion (4.18), then the difference quotients of the z^h are bounded uniformly in h up to order J and converge to the respective derivatives of z^* as $h \to 0$.

The existence of an asymptotic expansion of the type (4.18) of the global discretization error of a given discretization (4.5) of a target problem (4.4) depends on the smoothness of the analytic problem and its solution z^* and on the structure of the discretization. One prerequisite is the existence of an *asymptotic expansion of the local discretization error* (to order J)

$$(F - y)^h \Delta^h z^* = \hat{\Delta}^h [h^p d_p + \ldots + h^J d_J] + o(h^J), \tag{4.21}$$

with h-independent elements $d_j \in \hat{E}$; $p \geq 1$ is the order of consistency of the discretization, cf. (4.9).

In simple situations, e.g. one-step methods for initial value problems of ordinary differential equations, (4.21) is essentially sufficient for the validity of (4.18), and the e_j may be established as solution of variational problems

$$F'(z^*) e_j = \bar{d}_j, \tag{4.22}$$

whose right-hand sides $\bar{d}_j \in \hat{E}$ depend on the d_i, $i \leq j$, in (4.21), the e_i, $i < j$, and higher derivatives of the original problem operator (4.4).

In more complex situations, notably for discretizations of partial differential equations on non-trivial domains, even the establishment of the existence of a *principal term* $e_p \in E$ of the global discretization error (cf. (4.14) – (4.17)) is not always possible. In other cases, the concept of (4.18) has to be modified by a relaxation of the complete h-independence of the e_j. For more details, the relevant literature should be consulted, e.g. Pereyra-Proskurowski-Widlund [46], Böhmer [11], Munz [38], Marchuk-Shaidurov [37], Lin-Zhu [34a].

Asymptotic expansions of the local and global discretization errors also play a crucial role in the design of *error estimation procedures* for discretization algorithms. Such a procedure is called *asymptotically correct* if it generates estimates δ^h for the local or ε^h for the global discretization error which satisfy, at least locally (i.e. at a given gridpoint t_ν^h)

$$\delta^h(t_\nu^h) = [\Lambda^h(z^*)](t_\nu^h)\,(1 + o(1)), \tag{4.23}$$

resp. $h \to 0$.

$$\varepsilon^h(t_\nu^h) = e^h(t_\nu^h)\,(1 + o(1)), \tag{4.24}$$

Estimates of the local discretization error are used for the control and design of the grid in adaptive discretization algorithms while estimates of the global discretization error e^h assess the accuracy of the computed approximation.

5. Defect Correction and Discretization

5.1 The Fundamental Algorithmic Pattern

The combination of defect correction and discretization creates a situation with many facets. This permits the design of a variety of algorithms which superficially look quite different. We will therefore at first establish the common structural basis of these algorithms.

From the point of view of defect correction, we have again a problem (3.1) whose solution we strive to approximate by solving simpler problems (3.3). But from the point of view of discretization, both (3.1) and (3.3) are only discretizations of the true target problem (4.1). The discretizations and the generated approximations may be considered in their dependence upon the discretization parameter h, which brings a new – asymptotic – aspect into defect correction.

We begin with two discrete problems

$$F^h z^h = y^h \tag{5.1}$$

and

$$\tilde{F}^h z^h = \tilde{y}^h, \tag{5.2}$$

both of which are discretizations of the same target problem

$$F z = y. \tag{5.3}$$

The discretization parameter h in (5.1) and (5.2) is kept fixed at first.

For simplicity, we assume that both discretizations employ the same grids and grid function spaces E^h and $\hat{E}^h$. (5.1) is supposed to produce the better approximation of the image $\Delta^h z^*$ in E^h of the desired solution z^* of (5.3); but the computational solution of (5.1) is assumed to be considerably more costly than that of (5.2) – or even impossible in an immediate way. Thus it is natural to use some variant of defect correction in the spaces E^h and $\hat{E}^h$ to compute a good approximation of $(z^h)^*$, the solution of (5.1), by means of (5.2) and defects $d^h(\tilde{z}^h) := F^h \tilde{z}^h - y^h$ with respect to (5.1).

As indicated in Section 3.1, we may use an approximation $\tilde{\tilde{z}}^h$ obtained from a given approximation $\tilde{z}^h$ via a defect correction step (3.6) or (3.7) in one of two ways:

a) *Error estimation*: We may interpret $\tilde{z}^h - \tilde{\tilde{z}}^h$ as an estimate of the global discretization error $\tilde{z}^h - \Delta^h z^*$ of our initial approximation $\tilde{z}^h$. Yet the legitimacy of this approach is no longer so clear: Whereas the second right-hand term in

$$\tilde{z}^h - \Delta^h z^* = [\tilde{z}^h - \tilde{\tilde{z}}^h] + [\tilde{\tilde{z}}^h - (z^h)^*] + [(z^h)^* - \Delta^h z^*] \tag{5.4}$$

is made small relative to the first term by a sufficiently contractive defect correction step, the third term cannot be affected at all by defect correction in the spaces E^h and $\hat{E}^h$. It represents the global discretization error of the better discretization (5.1) and it must be small relative to the left hand side if the first term is to dominate.

b) *Iterative improvement*: We regard $\tilde{\tilde{z}}^h$ as our "given" approximation for another defect correction step; this may be iterated. But again there is a new, limiting aspect: We really don't want to approximate $(z^h)^*$ but $\Delta^h z^*$; hence it is not sensible to carry

the approximation of $(z^h)^*$ beyond the level of its global discretization error $(z^h)^* - \Delta^h z^*$:

$$\tilde{\tilde{z}}^h - \Delta^h z^* = [\tilde{\tilde{z}}^h - (z^h)^*] + [(z^h)^* - \Delta^h z^*]. \tag{5.5}$$

Again, the last term is not affected at all by an iterative use of defect correction to diminish the preceeding term.

As the limiting quantity $(z^h)^* - \Delta^h z^*$ is unknown, it seems impossible to design a sound algorithm on this basis. At this point, the asymptotic aspect of discretizations comes into play. It is *assumed* that the same defect correction step is carried out for smaller and smaller values of h and – relative to this hypothetical limit process – the asymptotic orders in powers of h of the terms in (5.4) or (5.5) resp. are determined. These *asymptotic relative sizes* are then employed as guidance for the algorithmic use of the above procedures for a *fixed* value of h. This reasoning is fully in line with the traditional analysis and design of discretization algorithms: There the relative importance of terms has always been judged by their asymptotic orders as $h \to 0$, and "higher order terms" have been neglected in favor of "lower order terms".

The main difficulty in an *asymptotic analysis* of a defect correction discretization algorithm is the rigorous assessment of the asymptotic contractivity of the defect correction procedure. Assume that (5.1) and (5.2) are consistent discretizations of (5.3) of orders p and $\tilde{p}$ resp., $p > \tilde{p}$. Then one may rather easily establish a Lipschitz bound $O(h^p)$ for the contraction mapping $I - \tilde{G}^h F^h$ or its dual $I - F^h \tilde{G}^h$ (cf. (3.8) and (3.9)) if one only considers sets $\{z^h\}_{h \in \mathcal{H}}$ of elements $z^h = \Delta^h z \in E^h$ arising from the discretization of *fixed* elements $z \in D$ with *sufficient smoothness*. On the other hand, in the norms of the spaces E^h and $\hat{E}^h$ the contraction is generally $O(1)$ only. Thus a serious analysis requires the establishment of suitable asymptotic expansions for the elements involved in the algorithm (cf. (4.18)) or some other more specific investigations.

If – for suitably chosen norms – the defect correction procedure is contractive of $O(h^{\tilde{p}})$, i.e. if

$$\| \tilde{\tilde{z}}^h - (z^h)^* \| = O(h^{\tilde{p}}) \, \| \tilde{z}^h - (z^h)^* \|, \tag{5.6}$$

one has the following asymptotic sizes for the terms in (5.4), with $\tilde{z}^h - (z^h)^* = O(h^q)$:

$$\tilde{\tilde{z}}^h - (z^h)^* = O(h^{q+\tilde{p}}), \quad \tilde{z}^h - \tilde{\tilde{z}}^h = O(h^q), \tag{5.7}$$

while $(z^h)^* - \Delta^h z^* = O(h^p)$ independently of the defect correction. Thus, $\tilde{z}^h - \tilde{\tilde{z}}^h$ is a sensible *asymptotic error estimate* for the discretization error of $\tilde{z}^h$ if $q < p$.

For IDeC, we have from (5.5) and (5.6), with $z_0^h - (z^h)^* = O(h^q)$,

$$\begin{aligned} z_i^h - \Delta^h z^* &= [z_i^h - (z^h)^*] + [(z^h)^* - \Delta^h z^*] \\ &= O(h^{q+i\tilde{p}}) + O(h^p). \end{aligned}$$

Thus iterative defect correction makes sense only until $q + i\tilde{p} \geq p$. If the iteration is started with $z_0^h = (\tilde{z}^h)^*$, the solution of (5.2), so that $q = \tilde{p}$, we have the well-known result

$$z_i^h - \Delta^h z^* = O(h^{\min((i+1)\tilde{p}, p)}); \tag{5.8}$$

i.e. the "orders" of the approximations generated byIDeC increase in steps of $\tilde{p}$ until the limiting order p is reached.

This limit can only be extended if the defect defining equation (5.1) is updated during the iteration procedure in a way which leads to larger and larger orders p. Obviously, $p_i=(i+1)\tilde{p}$ is a reasonable strategy. This leads to the following optimal convergence result for *updating* IDeC:

$$z_i^h - A^h z^* = O(h^{(i+1)\tilde{p}}) = O(h^{p_i}). \tag{5.9}$$

An essential aspect of the algorithmic pattern just explained is the fact that the "better" discretization (5.1) enters only through *defects* formed relatively to it:

$$d^h(z^h) := F^h z^h - y^h. \tag{5.10}$$

Actually it is this *defect defining function* $d^h : D^h \to E^h$ which is explicitly designed while the associated target discretization (5.1) is only implicitly defined through $d^h(z^h)=0$. This fundamental pattern is somehow present in all combinations of defect correction and discretization.

5.2 Some Defect Correction Discretization Algorithms

In the following, some historically important defect correction discretization algorithms and some major subsequent developments will be sketched. We will take the liberty to describe the algorithmic approaches not always in the terms of their proposers but rather within our own framework of concepts and notations. This permits a more concise presentation and easier cross references. Due to space limitations, we have to restrict ourselves to a few typical developments. For a more detailed and well annotated survey of the historical development, the reader is referred to Skeel [47].

5.2.1 Pereyra's Iterated Deferred Corrections

In Chapter 2, we have sketched Fox's suggestion for improving the finite-difference solution of (2.5) by the use of higher order differences. In his view, the standard second order discretization originated as the lowest order term in an expansion of the second derivative in (2.5) into an infinite series of differences. Approximations of higher order terms of this expansion could be introduced in "reruns": The higher order differences were formed for the previously obtained approximate solution, in analogy to the relaxation approach.
In our terminology, the second order discretization corresponds to (5.2) while the successively longer partial sums of the difference expansion generate a sequence of defect defining functions (3.24) (cf. (5.10)), and the whole process becomes a (linear) updating IDeC algorithm (3.27).

In 1965, this approach was considerably generalized into the principle of Iterative Deferred Corrections by V. Pereyra [41]: He identified Fox's difference expansion as a particularly simple case of an asymptotic expansion (4.21) of the local

discretization error for the employed basic discretization (5.2). Therefore he aimed at the construction of difference operators $S_{i+1}^h: E^h \to \hat{E}^h$ such that (cf. (4.21))

$$\hat{\Delta}^h \sum_{j=p}^{i+1} h^j d_j - S_{i+1}^h(z_i^h) = O(h^{i+2}) \tag{5.11}$$

would follow from $z_i^h - \Delta^h z^* = O(h^{i+1})$. A new approximation $z_{i+1}^h \in E^h$ could then be obtained from

$$(\tilde{F} - \tilde{y})^h z_{i+1}^h = S_{i+1}^h(z_i^h), \tag{5.12}$$

with the original discretization operation $(\tilde{F} - \tilde{y})^h$ (cf. (4.5)); it was shown to satisfy

$$z_{i+1}^h - \Delta^h z^* = O(h^{i+2}) \tag{5.13}$$

under rather weak technical assumptions. In a subsequent paper [43], the construction of S_{i+1}^h was essentially reduced to a problem of interpolation which was solved very efficiently in Björck-Pereyra [7].

Actually, Pereyra had assumed that the asymptotic expansion (4.21) proceeded in *multiples of* $\tilde{p}$ so that the order of the error z_{i+1}^h was increased by $\tilde{p}$ over that of z_i^h. In symmetric discretizations, $\tilde{p}$ would be 2, cf. the remark below (4.18). This approach led to powerful codes for the numerical solution of nonlinear two-point boundary value problems; see also Pereyra's paper in this volume.

In our framework, this is an example of version (*B*) of the updating IDeC algorithm (3.26) in the context of Section 5.1. The basic discretization in (5.12) corresponds to (5.2) while the S_{i+1}^h-operators define defects for successively improved discretizations (5.1): Obviously, at each level i of the Deferred Correction procedure, the fictitious target discretization is

$$(F_{i+1} - y_{i+1})^h z^h := (\tilde{F} - \tilde{y})^h z^h - S_{i+1}^h(z^h) = 0 \tag{5.14}$$

since an iteration of (5.12), with S_{i+1}^h *kept fixed*, would generate the solution of (5.14). Due to (5.11), with an expansion containing powers $h^{j\tilde{p}}$ only, the local discretization error of (5.14) is $O(h^{(i+1)\tilde{p}})$. This corresponds to the optimal strategy leading to the convergence result (5.9).

5.2.2 Zadunaisky's Global Error Estimation

Zadunaisky who was concerned with the finite-difference solution of orbit differential equations had presented his ideas at various astronomers' meetings since 1964 (e.g. [54]); but only his presentation at the 1973 Dundee Numerical Analysis Conference caught the attention of numerical analysts.

In Zadunaisky's approach (see [50], [55]), it is assumed that an approximate solution $\tilde{z}^h$ of an initial value problem for the system

$$z'(t) = f(t, z(t)) \tag{5.15}$$

has been obtained by some standard discretization method; an estimate for the *global error* (4.10) of $\tilde{z}^h$ is requested. For this purpose, a *defect* of $\tilde{z}^h$ is formed and used in the following way:

(i) The values of $\bar{z}^h$ on the grid are interpolated by piecewise polynomials (in each component).

(ii) For these polynomials, the defect in (5.15) may trivially be formed; this generates a defect function $\bar{d}^h(t)$ on the entire interval of integration.

(iii) The "neighboring" initial value problem

$$\bar{z}'(t) = f(t, \bar{z}(t)) + \bar{d}^h(t), \tag{5.16}$$

whose exact solution is the piecewise polynomial function of (i) and (ii), is solved with the discretization method used previously for (5.15); this produces a grid function $\tilde{z}^h$.

(iv) For (5.16), from the *exact values* $\bar{z}^h$ and the *approximate values* $\tilde{z}^h$, the values $\Delta z^h := \tilde{z}^h - \bar{z}^h$ of the global discretization error of the discretization are computed. These values are used as estimates of the global error of the same discretization method for (5.15).

In our framework of concepts, the original problem (5.15) naturally corresponds to (5.3) and its discretization to (5.2). The target discretization (5.1) is found by the consideration of $d^h(z^h) = 0$: It is a *polynomial collocation method*; its solution would consist of piecewise polynomials which satisfy (5.15) at the gridpoints.

The algorithmic procedure (i) – (iv) is a prototype of defect correction version (*A*), see (3.6); the numerical solution $\tilde{z}^h$ of (5.16) corresponds to $\tilde{\tilde{z}}$. The usage is then precisely as discussed in Section 5.1. Under assumption (5.6), the order p of the (implied) polynomial collocation method must be greater than the order $\tilde{p}$ of the discretization method for asymptotically correct error estimation.

The experimental results of Zadunaisky's heuristically conceived procedure were so remarkable that they prompted further analysis and generalization. In the end, this led to the conceptual schemes of Sections 3.1 and 5.1.

5.2.3 IDeC with Polynomial Collocation

In presenting and interpreting Zadunaisky's approach, Stetter [50] proposed the iterative use of the technique ((3.10) instead of (3.6), Section 5.1); he also conjectured the fundamental order result (5.8). At Vienna, a group under R. Frank began to clarify the theoretical foundations and to investigate the practical applicability.

It is clear that there are numerous ways of arriving at a defect function $\bar{d}^h$ for the *neighboring problem* (5.16) through polynomial interpolation. There is not only the choice of the degree to be made but there are also different ways of "joining" the polynomial pieces: The interpolation intervals may be disjoint except for their boundary gridpoints or they may overlap, with each polynomial piece used only for the middle part of its interval.

In applications to second order equations (boundary value problems), the discontinuity of the first derivatives at junctions must be accounted for, a problem which was cleverly solved by Frank [22]. In applications to initial value problems, on the other hand, one has to decide whether one solves each of the successive neighboring

problems in an IDeC algorithm over the whole interval of integration ("global connecting strategy") or whether one iterates on a shorter interval first and then continues with the best value at the endpoint to the next partial interval ("local connecting strategy").

The effects of these and other algorithmic decisions were studied by the group in Vienna. More important, the analytic foundations of the whole approach were revealed. In order to establish the fundamental contractivity assumption (5.6) for defect correction with a given type of polynomial interpolation, Frank and his colleagues had to study the asymptotic expansions of quantities like the numerical solution $\bar{z}^h$ of the neighboring problem (3.16). This quantity depends upon h in a *twofold* way: through the function $\bar{d}^h$ in (3.16) which has been obtained from $\tilde{z}^h$ and through the discretization applied to (3.16). With respect to these problems, there are now a good number of rigorous results (e.g. [22]–[25]), other situations may be dealt with analogously. For a completely algebraic discussion, see Hairer [32].

A further major effort has been devoted to the application of defect correction techniques based upon polynomial collocation to *stiff* initial value problems. For such problems, the asymptotic discretization theory sketched in Chapter 4 does not describe the observed behavior except for unreasonably small steps. Instead, the stability of the computation at a fixed (large) stepsize becomes the dominating issue. Since this situation can only be dealt with by implicit discretizations (with a large computational effort per step), it is particularly challenging to improve the efficiency by the use of IDeC. A number of important results have been gained in this respect (e.g. [25], [52]) and software based on IDeC is under development. One particular aspect has been studied in more detail in Frank's paper in this volume.

All the investigations mentioned in this section concern version (*A*) of IDeC. They show that this version is equally flexible as the version (*B*) approach used by Pereyra.

5.2.4 The Approaches of Lindberg and Skeel

In 1976, Lindberg [35], [36] proposed and analyzed the general idea of defect correction, version (*B*), in the discretization context independently of the investigations on defect correction techniques in Vienna. He had realized that it was not at all necessary to start from the local discretization error of (5.2) for the generation of suitable right-hand sides in the deferred corrections (5.12), cf. (4.21) and (5.11). Instead he suggested the use of an "arbitrary" discretization (5.1), with an order p greater than the order $\tilde{p}$ of (5.2), for an estimation of the local discretization error.

His basic algorithmic pattern

$$\tilde{F}^h \underline{\xi}^h = -F^h \underline{z}^h \tag{5.17}$$

is identical with (3.7) applied to (5.1)/(5.2) under the assumptions that $\tilde{z}^h$ solves (5.2) and $\tilde{y}^h = y^h$ (or both vanish). The quantity $\underline{z}^h - \underline{\tilde{z}}^h$ is recommended for global error estimation, cf. Section 5.1.

Furthermore, with a sequence of discretization operators F_i^h of order $(i+1)\tilde{p}$, Lindberg proposed the updating IDeC, version (*B*),

$$\tilde{F}^h z_{i+1}^h = \tilde{F}^h z_i^h - F_{i+1}^h z_i^h, \quad i=0,1,\ldots, \tag{5.18}$$

as a generalization of Pereyra's deferred corrections (5.12); this corresponds to the application of (3.26) in the discretization context (with $y_{i+1}^h = 0$ for a easier notation). For the construction of the defect defining operators F^h and F_i^h, he considered various finite-difference techniques and the use of local interpolants.

Lindberg [36] analyzes his algorithmic patterns (5.17) and (5.18) by means of asymptotic expansions. Under appropriate assumptions (even too strong in some respects) he proves the order results (5.7), with $q=\tilde{p}$, and (5.9). He also applies his results in a number of interesting situations.

Skeel [47] must be credited for his successful effort to relate the various historical developments which have formed the basis of defect correction discretization algorithms. He points out many interesting details far beyond what we have been able to sketch on these few pages.

Skeel [47] then introduces a theoretical framework which makes it easier to deal with the assumptions needed in an analysis of defect correction steps. The main objective of the use of asymptotic expansions by Frank and his colleagues and by Lindberg was the establishment of sufficient smoothness for various quantities, cf. our respective remarks in Section 4.2 (e.g. (4.19), (4.20)) and Section 5.1. Therefore Skeel introduces discrete Sobolev norms which include the values of differences up to a specified order. This leads to a more natural formulation of the essential assumptions underlying the results of Frank et al. and of Lindberg.

Skeel gives a general analysis of one step of an IDeC discretization procedure. Although it is clear that this analysis is applicable to each stage of the iteration, the rigorous recursive verification of the necessary assumptions for the results (5.8) and (5.9) may be quite difficult in realistic situations where asymptotic expansion results are not available.

A similar analysis, based on discrete Sobolev norms and avoiding asymptotic expansions, has been given independently by Hackbusch [29] for the case of linear operators.

5.2.5 Böhmer's Discrete Newton Methods

The Discrete Newton Methods (*DNM*) of Böhmer ([8]–[10]) have been based on the linear version (3.14) of IDeC, which avoids some algorithmic difficulties arising in the nonlinear case; they have a remote relation to Ortega-Rheinboldt's discretized Newton methods ([40]). The *DNM* have the form (cf. (3.14))

$$\widetilde{F^*(z_0)}^h (z_{i+1}^h - z_i^h) = -d^h(z_i^h), \quad i=0,1,\ldots; \tag{5.19}$$

here the linear discretization operator $\widetilde{F^*(z_0)}^h$ originates from the application of the basic discretization method (corresponding to (5.2)) to an approximation $F^*(z_0)$ of the Frechet derivative $F'(z_0)$ of the operator F in the analytic problem (5.3).

The defect $d^h(z^h)$ is formed by a local "prolongation" of $z^h \in E^h$ into a function $z \in E$ and a subsequent substitution of z into (5.3): this analytic defect is discretized back into the space E^h by an operator Ω^h which is related to F^h of (5.2) by an "additivity condition"

$$\widetilde{(F-y)}^h z^h = \tilde{F}^h z^h + \Omega^h y. \tag{5.20}$$

If this defect defining operation has an order p and if z_0^h is the solution of (5.2) and $z_0 \in E$ its prolongation, then the order result (5.8) may again be proved for the sequence z_i^h defined by (5.19), see [9], [10]. The same result is also shown to hold for a nonlinear version under weaker conditions than those imposed by Lindberg.

"Strong" *DNM*, with updated exact Frechet derivatives, have the quadratic convergence of the classical Newton process: but the increased effort outweighs this advantage.

An application of *DNM* to an important physical problem has been discussed in the paper by Böhmer et al. in this volume. The error control in Schönauer's software for partial differential equations (see his paper in this volume) has some relation to *DNM* but has been conceived independently.

5.2.6 Brakhage's Defect Correction for Integral Equations

Already in 1960, Brakhage [12] had devised an iterative method for linear Fredholm integral equations which is based on defect evaluation and also represents a step towards multigrid methods of the second kind as they will be explained in Section 6.4.

Brakhage discretized the integrals in a Fredholm equation

$$z(s) - \int_a^b k(s,t)\, z(t)\, dt = y(s) \qquad s \in [a,b], \tag{5.21}$$

by quadrature rules on a fine and on a coarse grid, with grid parameters h and H respectively. By a restriction of s to the gridpoints, he obtained two discretizations (linear systems of equations)

$$F^h z^h = (I - K^h) z^h = y^h \tag{5.22}$$

and

$$F^H z^H = (I - K^H) z^H = y^H \tag{5.23}$$

corresponding to (5.1) and (5.2). His idea was – as in IDeC – to use defects from (5.22) for the correction of solutions of (5.23).

In the present situation, the computation of the defect $d^h(z^H)$ in (5.22) of solution z^H on the H-grid is facilitated by the fact that z^H may be interpolated in $[a,b]$ by the use of (5.21), with only the integral replaced by the quadrature rule. The defect thus obtained is then subject to a smoothing by one more application of the quadrature operator K^h on the fine grid. With this smoothed defect $\bar{d}^h(z^H)$, Brakhage's iteration

$$F^H(z_{i+1}^H - z_i^H) = -\bar{d}^h(z_i^H) \tag{5.24}$$

is an immediate example of a linear IDeC, cf. (3.14).

Brakhage gives a rigorous proof for the convergence of the sequence z_i^H to the solution z^h of (5.22) using the Banach fixed point theorem. He also gives computable

estimates for $\| z_i^H - z^h \|$ based on accuracy results for the quadrature rules K^h and K^H. His method was further developed by Atkinson [4] and, in connection with multi-grid methods, by Hackbusch [30], Hemker-Schippers [33], and Mandel; see Section 6.4 and Mandel's paper in this volume.

6. Multi-level and Multi-grid Methods

6.1 The Use of Different Levels of Discretization

As explained in Chapter 4, the discretizations (4.3) for different values of $h \in \mathcal{H}$ are closely related to the original problem. In many cases it is possible to use this relation to design efficient numerical methods for the solution of (4.3) with a specified $\bar{h} \in \mathcal{H}$ and for the efficient computation of a sufficiently accurate approximation z^h of z^*.

Over the years many authors rediscovered the possibility to use the information obtained with the solution of $F^H z^H = 0$, where the meshwidth $H > h$, $H \in \mathcal{H}$, is relatively large, for the solution of the problem $F^h z^h = 0$ with small h. This idea is used for differential equations [1, 3] as well as for integral equations [12]. In its simplest form, a solution z^H can be used as a starting approximation in an iterative process for the solution of $F^h z^h = 0$. This is particularly advantageous for the solution of nonlinear problems because of the *mesh independence principle* (*MIP*).

It states that, under suitable technical assumptions, the number of Newton iterations

$$(F^h)'(z_k^h)(z_{k+1}^h - z_k^h) = -F^h z_k^h$$

needed to attain a specified accuracy in the solution of $F^h z^h = 0$ is *independent of* h for all sufficiently fine grids, if the iteration is started from $z_0^h := \Delta^h z_0$, $z_0 \in E$ fixed. The *MIP* in its basic form for special second order boundary value problems is due to Allgower-McCormick [3], it was extended to general operators by Allgower-Böhmer [1]. For a combination with IDeC see Allgower-Böhmer-McCormick [2].

Another application of discretizations on grids with different sizes is *multigrid iteration*. Here the problem to be solved is a discrete equation on a (very) fine grid. To accelerate the iterative solution of the discrete system of equations, one makes use of the discretization on coarser grids. The multigrid iteration is most simply explained by first considering only two grids: the two-grid method. Then the principle is easily extended to more grids. This approach is used in Sections 6.2 – 6.3.

A further recursive extension of the use of different levels of discretization is possible: When we have solved, by multigrid iteration, the discrete problem on some grid of our grid sequence, we may use the interpolated result as initial approximation for the solution of the discrete problem on the next finer grid. This process is called a *Full Multigrid Method*. In the following section we will restrict ourselves to multigrid iteration.

The multigrid approach evolved in the 1970's from the efforts of various researchers. It owes particularly much to the impetus of *A.* Brandt who has promoted it by his papers, his lectures, and his program developments since 1972, see e.g. [13], [14].

6.2 *The Two-grid Method*

The two-grid method is a non-stationary defect correction iteration in which only two different approximate inverses are used; see Section 3.2. These two different iteration steps are:

(i) a relaxation step (e.g. Jacobi, Gauss-Seidel, the incomplete LU-decomposition iteration, etc.) on the fine grid, and

(ii) a coarse grid correction.

The approximate inverse in the coarse grid correction for the solution of $F^h z^h = y^h$ is given by

$$\tilde{G}^h = \Delta_H^h (F^H)^{-1} \hat{\Delta}_h^H. \tag{6.1}$$

Here Δ_H^h denotes the prolongation (interpolation) of a solution from a coarse grid to a fine grid; $\hat{\Delta}_h^H$ denotes the restriction (averaging, weighting) of the residual from a fine grid to a coarse grid. Thus, a coarse grid correction step in the two-grid method (for a linear problem) reads

$$z_{i+1}^h = z_i^h - \Delta_H^h (F^H)^{-1} \hat{\Delta}_h^H (F^h z_i^h - y^h). \tag{6.2}$$

One step in the two-grid method (TGM) iteration consists of p relaxation sweeps, followed by a coarse grid correction and, again, followed by q relaxation sweeps. This step of the linear two-grid algorithm for the solution of $F^h z^h = y^h$, is given in the following quasi-ALGOL program.

```
proc TGM (vector z^h, y^h)                    z_i^h → z_{i+1}^h
begin
    for j to p do relax (F^h, z^h, y^h);      p relaxation steps
    d^h := F^h * z^h − y^h;                   defect computation
    d^H := Δ̂_h^H d^h;                         transfer to coarser grid
    solve (F^H, c^H, d^H);                    solution of F^H c^H = d^H
    c^h := Δ_H^h c^H;                         transfer to finer grid
    z^h := z^h − c^h;                         correction
    for j to q do relax (F^h, z^h, y^h);      q relaxation sweeps
end
```

In this procedure the right-hand-side y^h and an approximate solution z^h are given; by the procedure the given z^h (i.e. z_i^h) is updated and changed into the new iterate z_{i+1}^h. The error amplification operator of one step of this linear two-grid method is given by (cf. (3.20))

$$M_{TGM,p,q}^h = (I^h - B^h F^h)^q (I^h - \Delta_H^h (F^H)^{-1} \hat{\Delta}_h^H F^h)(I^h - B^h F^h)^p, \tag{6.3}$$

where B^h is the approximate inverse associated with the relaxation process. We also may write

$$M_{TGM,p,q}^h = (M_{REL}^h)^q (((F^h)^{-1} - \Delta_H^h (F^H)^{-1} \hat{\Delta}_h^H) F^h) (M_{REL}^h)^p, \tag{6.4}$$

where M_{REL}^h is the error amplification operator of the relaxation sweep. In the latter expression the operator

$$(F^h)^{-1} - \Delta_H^h (F^H)^{-1} \hat{\Delta}_h^H,$$

determines the relative convergence between the operators F^h and F^H.

Convergence proofs of the multigrid method are based on the analysis of (6.4). E.g. for an elliptic differential equation, Hackbusch [28] proves the following theorem:

If

(i) the operators F^h and F^H are relatively convergent of order α, i.e.

$$\| (F^h)^{-1} - \Delta_H^h (F^H)^{-1} \hat{\Delta}_h^H \| \leq C \cdot H^\alpha,$$

(ii) the relaxation process for F^h satisfies a proper smoothing property of order α, i.e.

$$\| F^h (M_{REL}^h)^\nu \| < C_0(\nu) \cdot h^{-\alpha},$$

with $C_0(\nu)$ independent of h, $C_0(\nu) \to 0$ as $\nu \to \infty$,

(iii) the discretizations F^h and F^H satisfy the regular relative mesh property, i.e.

$$H/h < C,$$

then the error amplification operator satisfies

$$\| M_{TGM,p,q}^h \|_{E \to E} < C \cdot C_0(p), \tag{6.5}$$

where C, $C_0(p)$ are independent of h and $C_0(p) \to 0$ as $p \to \infty$. □

The most difficult part in the application of this theorem in practical situations is the verification of condition (ii).

6.3 The Multigrid Method

In the two-grid method we have to solve exactly one coarse grid problem $F^H c^H = d^H$ in each iteration step. In the multigrid method (MGM) we solve this problem only approximately by applying a few iteration steps of the same MGM on the coarser level. Then we have to solve directly a discrete problem only on the very coarsest grid in the sequence. This may be a relatively simple task because of the small number of gridpoints on that grid. With σ iteration steps of the MGM used to approximate $(F^H)^{-1}$, the multigrid method is given in the following quasi-ALGOL program:

proc MGM (**integer** level, **vector** z^h, y^h)
if level $= 0$
then solve (F^h, z^h, y^h);
else
  for j **to** p **do** relax (F^h, z^h, y^h);
  $v^H := 0$;
  $d^H := \hat{\Delta}_h^H (F^h * z^h - y^h)$;
  for m **to** σ **do** MGM (level $-1, v^H, d^H$);
  $z^h := z^h - \Delta_H^h v^H$;
  for j **to** q **do** relax (F^h, z^h, y^h);
end if

Fig. 3 shows how the computation proceeds between the various grid levels in one step of MGM for level $= 3$, i.e. in the computation of z_{i+1}^h from z_i^h with the aid of 3 increasingly coarser grids:

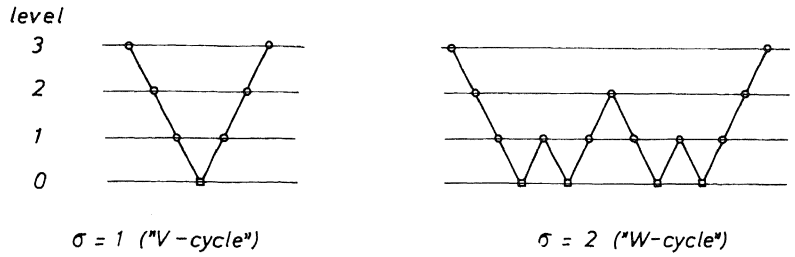

Fig. 3

We denote the error amplification operator of a MGM-step on the h-level of discretization by $M^h_{MGM,p,q,\sigma}$, on the next coarser level we denote it by M^H_{MGM}. In the MGM-cycle the approximate inverse of the coarse grid correction is not given by (6.1), because in the algorithm $(F^H)^{-1}$ is only approximated by σ steps of a defect correction process (viz. MGM on the H-level). The amplification operator of this process is M^H_{MGM}, and hence, as was shown in Section 3.2 (3.22), the approximate inverse obtained by σ iteration steps of MGM is

$$(I^H-(M^H_{MGM})^\sigma)\,(F^H)^{-1}.$$

Replacing $(F^H)^{-1}$ by this expression in (6.1) and (6.2) we find for the amplification operator of the coarse grid correction in MGM

$$I^h-\Delta^h_H(I^H-(M^H_{MGM})^\sigma)\,(F^H)^{-1}\,\hat{\Delta}^H_h F^h. \tag{6.6}$$

Using equation (6.4), we infer

$$M^h_{MGM,p,q,\sigma}=M^h_{TGM,p,q}+(M^h_{REL})^q\,\Delta^h_H(M^H_{MGM})^\sigma(F^H)^{-1}\,\hat{\Delta}^H_h(M^h_{REL})^p. \tag{6.7}$$

Under conditions (i)–(iii) in Section 6.2 and a few other technical conditions, we derive from (6.7)

$$\|M^h_{MGM,p,q,\sigma}\|\le\|M^h_{TGM,p,q}\|+C\,\|M^H_{MGM}\|^\sigma. \tag{6.8}$$

Here we have a recursive expression, where the rate of convergence of the MGM on level h is expressed by the rate of convergence of the TGM on level h and of the MGM on level H. Further, we know that on the second coarsest level h_1 we have $M^{h_1}_{MGM}=M^{h_1}_{TGM}$.

By (6.5) we have $\|M^h_{TGM,p,q}\|\le C<1$, if p is large enough. Hence we can find a σ such that $\|M^H_{MGM}\|<1$. Often a small value of σ (e.g. $\sigma=2$) can be shown to be sufficient to have

$$\|M^h_{MGM,p,q,\sigma}\|\le C<1$$

on all levels, C independent of h.

The essential difference between multigrid iteration and a simple relaxation iteration (e.g. Gauss-Seidel) is the fact that $\|M^h_{MGM}\|$ is strictly less than one, whereas $\|M^h_{REL}\|\to 1$ for $h\to 0$. Further, when $\sigma<(H/h)^d$, where d is the dimension of the problem, the overall amount of computational work in a MGM-cycle is simply proportional to the amount of work in a relaxation cycle. Therefore, for small h, MGM iteration is essentially more efficient than straightforward relaxation.

In the previous *MGM* algorithm, the fact was used that all F^h are linear: Only corrections had to be computed on the lower grid levels. If the F^h are *nonlinear*, we have to compute *corrected solutions* on the lower grid levels, too. In a *TGM*, we will simply pass the smoothed approximation z^h and its defect d^h to the coarser grid and solve

$$F^H v^H = F^H(\Delta_h^H z^h) + \hat{\Delta}_h^H d^h.$$

Then the observed change $c^H = v^H - \Delta_h^H z^h$ is passed back and applied as correction: $z^h \to z^h - \Delta_H^h c^H$. Thus the nonlinear *MGM* takes the form:

proc *FAS* (**integer** level, **vector** z^h, y^h)
if level $=0$
then solve (F^h, z^h, y^h);
else
  for j **to** p **do** relax (F^h, z^h, y^h);
  $v^H := z^H := \Delta_h^H z^h$;
  $d^H := -F^H * z^H + \hat{\Delta}_h^H (F^h * z^h - y^h)$;
  for m **to** σ **do** *FAS* (level $-1, v^H, d^H$);
  $z^h := z^h - \Delta_H^h (v^H - z^H)$;
  for j **to** q **do** relax (F^h, z^h, y^h)
end if

This algorithm is called *FAS* (full approximation scheme) by Brandt (e.g. [14]) because discrete approximations are available on all levels. If the *FAS*-algorithm is imbedded in a Full Multigrid Method (Section 6.1), then coarse level approximations are available before *FAS* iteration is started and, hence, forming $z^H := \Delta_h^H z^h$ can sometimes be omitted. Also the device (3.17) may be employed: The defect d^h may be divided by some $\mu > 1$ before it is used while the resulting change c^H is multiplied by the same μ.

6.4 A Multigrid Method of the Second Kind

We consider the Fredholm integral equation of the 2nd kind, with a compact kernel,

$$z(s) - \int_a^b k(s,t)\, z(t)\, dt = y(s),$$

or, in operator notation,

$$Fz := z - Kz = y.$$

Further, we consider a sequence of discretizations

$$F^p z := z - K^p z = y^p, \qquad p = 0, 1, 2, \ldots, \tag{6.9}$$

where K^p and y^p denote discretizations on a mesh h_p with $h_p \to 0$ as $p \to \infty$. A simple method to solve (6.9) is by means of successive substitution

$$z_{i+1} = K^p z_i + y^p.$$

It converges if $\| K^p \| < 1$ and, for a compact operator, K^p has a smoothing property.

or $p>0$, a coarse grid correction is possible by the use of a coarser grid solution operator as approximate inverse

$$\tilde{G}^p := (F^{p-1})^{-1} = (I - K^{p-1})^{-1}.$$

ombination of one smoothing step and one coarse grid correction yields

$$\begin{aligned} M^p_{TGM} &= (I - (F^{p-1})^{-1} F^p) K^p \\ &= (I - K^{p-1})^{-1} (K^p - K^{p-1}) K^p. \end{aligned}$$

nder suitable conditions it can be shown [33] that – with the trapezoidal rule ed for discretization – we have

$$\| M^p_{TGM} \| \leq \| (I - K^{p-1})^{-1} \| \, \| (K^p - K^{p-1}) K^p \| < C h_p^2, \text{ for } p \to \infty. \quad (6.10)$$

he *TGM* still needs the solution of eq. (6.9) on level $p-1$. This solution can, again, e approximated by σ *MGM* iterations on level $p-1$. Using equation (3.22) we btain

$$\begin{aligned} M^p_{MGM} &= (I - (I - (M^{p-1}_{MGM})^\sigma) (F^{p-1})^{-1} F^p) K^p \\ &= M^p_{TGM} + (M^{p-1}_{MGM})^\sigma (K^p - M^p_{TGM}). \end{aligned}$$

his leads to a recursion relation similar to (6.8)

$$\| M^p_{MGM} \| \leq \| M^p_{TGM} \| + \| M^{p-1}_{MGM} \|^\sigma (\| K^p \| + \| M^p_{TGM} \|). \quad (6.11)$$

nder suitable conditions, it can be derived from (6.10) and (6.11) that, for $\sigma = 2$ and $M^0_{TGM} \|$ small enough,

$$\| M^p_{MGM} \| = O(h_p^2) \quad \text{as} \quad p \to \infty.$$

his is the typical behavior of the multigrid iteration of the second kind: the finer the iscretization of the analytical problem the faster the convergence of the iterative rocess for the solution of the discrete system. The essential difference between the *GMs* of the 1st and of the 2nd kind is in the type of the operator to which the ultigrid principle is applied. A regular differential operator $L : D \Rightarrow D$ maps a space ith a stronger into a space with a weaker topology: whereas the compact integral perator K maps a space with a weaker into one with a stronger topology. onsequently, for differential equations we can get $\| M^h_{MGM} \|$ bounded and strictly ss than one uniformly in h, while for integral equations we can have $\| M^h_{MGM} \|$ ounded by $O(h^m)$ for some $m>0$. These multigrid convergence factors have to be mpared with the convergence factors for straightforward relaxation. There we nd $\| M^h_{REL} \| \approx 1 - O(h^m)$ for problems of the 1st kind, and $\| M^h_{REL} \| \approx C < 1$ for roblems of the 2nd kind.

6.5 *Software for Multigrid Methods*

With respect to the multigrid methods several software developments are going on. The first known software in this area is contained on MUG-tape. This is a tape with arious software related to multigrid, by A. Brandt and coworkers. From 1978 it has irculated among interested parties in several updated versions.

Another piece of MG-software that is continuously being updated is the subroutine PLTMG by R. E. Bank and A. H. Sherman [5]. It solves an elliptic boundary value problem in a two-dimensional domain. It uses a finite element procedure for the discretization and it has an automatic adaptive refinement of a user-provided crude triangulation.

Other MG-software emphasizes the efficient solution of the discrete systems that are obtained from various kinds of discretization of a more or less general elliptic partial differential equation in two dimensions. Here we mention the program MGOO by the Bonn-group [21], the program BOXMG by J. E. Dendy [16] and MGD1 by P. Wesseling [53]. For the last program several variants for different situations have been constructed by Z. Novak [39] and P. de Zeeuw [34].

Also some MG-software specially tuned for vector-machines is available. Vectorized versions for 7-point discretizations in a rectangle are available in portable FORTRAN, viz. the subroutines MGD1V and MGD5V by P. de Zeeuw. Another program specially designed for the solution of the Poisson equation on a Cyber 205 is mentioned by Barkai and Brandt [6].

References

[1] Allgower, E. L., Böhmer, K.: A mesh independence principle for operator equations and their discretizations. GMD Report, Bonn, 1984.
[2] Allgower, E. L., Böhmer, K., McCormick, S. F.: Discrete defect corrections: the basic ideas. ZAMM *62*, 371–377 (1982).
[3] Allgower, E. L., McCormick, S. F.: Newton's method with mesh refinements for numerical solution of nonlinear two-point boundary value problems. Numer. Math. *29*, 237–260 (1978).
[4] Atkinson, K. E.: Iterative variants of the Nyström method for the numerical solution of integral equations. Numer. Math. *22*, 17–33 (1973).
[5] Bank, R. E., Sherman, A. H.: An adaptive multi-level method for elliptic boundary value problems. Computing *26*, 91–105 (1982).
[6] Barkai, D., Brandt, A.: Vectorized multigrid Poisson solver for the CDC CYBER 205. Appl. Math. and Computation *48*, 215–227 (1983).
[7] Björck, A., Pereyra, V.: Solution of Vandermonde systems of equations. Math. Comp. *24*, 893–903 (1970).
[8] Böhmer, K.: Defect corrections via neighbouring problems. I. General theory. MRC Report, University of Wisconsin-Madison No. 1750 (1977).
[9] Böhmer, K.: Discrete Newton methods and iterated defect corrections, I. General theory, II. Initial and boundary value problems in ordinary differential equations. Berichte Nr. 10, 11, Universität Karlsruhe, Fakultät für Mathematik (1978).
[10] Böhmer, K.: Discrete Newton methods and iterated defect corrections. Numer. Math. *37*, 167–192 (1981).
[11] Böhmer, K.: Asymptotic expansion for the discretization error in linear elliptic boundary value problems on general regions. Math. Z. *177*, 235–255 (1981).
[12] Brakhage, H.: Über die numerische Behandlung von Integralgleichungen nach der Quadraturformelmethode. Numer. Math. *2*, 183–196 (1960).
[13] Brandt, A.: Multi-level adaptive solutions to boundary-value problems. Math. Comp. *31*, 333–390 (1977).
[14] Brandt, A., Dinar, N: Multigrid solutions to elliptic flow problems. In: Numerical Methods for Partial Differential Equations (Parter, S., ed.), pp. 53–147. Academic Press 1979.
[15] Bulirsch, R., Stoer, J.: Numerical treatment of ordinary differential equations by extrapolation methods. Numer. Math. *8*, 1–13 (1966).

[16] Dendy, J. E., jr.: Black box multigrid. J. Comp. Phys. *48*, 366 – 386 (1982).
[17] Fox, L.: Some improvements in the use of relaxation methods for the solution of ordinary and partial differential equations. Proc. Roy. Soc. London *A 190*, 31 – 59 (1947).
[18] Fox, L.: The solution by relaxation methods of ordinary differential equations. Proc. Cambridge Phil. Soc. *45*, 50 – 68 (1949).
[19] Fox, L., ed.: Numerical Solution of Ordinary and Partial Differential Equations. Oxford: Pergamon Press 1962.
[20] Fox, L.: The Numerical Solution of Two-point Boundary Value Problems in Ordinary Differential Equations. Oxford: University Press 1957.
[21] Foerster, H., Witsch, K.: Multigrid software for the solution of elliptic problems on rectangular domains: MGOO. In: Multigrid Methods (Hackbusch, W., Trottenberg, U., eds.), pp. 427 – 460. (Lecture Notes in Mathematics, Vol. 960.) Berlin-Heidelberg-New York: Springer 1982.
[22] Frank, R.: The method of iterated defect-correction and its application to two-point boundary value problems. Numer. Math. *25*, 409 – 419 (1976).
[23] Frank, R., Hertling, J., Ueberhuber, C. W.: An extension of the applicability of iterated deferred corrections. Math. Comp. *31*, 907 – 915 (1977).
[24] Frank, R., Ueberhuber, C. W.: Iterated defect correction for differential equations, part I: Theoretical results. Computing *20*, 207 – 228 (1978).
[25] Frank, R., Ueberhuber, C. W.: Iterated defect correction for the efficient solution of stiff systems of ordinary differential equations. BIT *17*, 46 – 159 (1977).
[26] Gragg, W. B.: Repeated extrapolation to the limit in the numerical solution of ordinary differential equations. Thesis, UCLA, 1963.
[27] Gragg, W. B.: On extrapolation algorithms for ordinary initial value problems. SIAM J. Num. Anal. *2*, 384 – 403 (1965).
[28] Hackbusch, W.: Error analysis of the nonlinear multigrid method of the second kind. Apl. Mat. *26*, 18 – 29 (1981).
[29] Hackbusch, W.: Bemerkungen zur iterierten Defektkorrektur und zu ihrer Kombination mit Mehrgitterverfahren. Report 79-13, Math. Institut Universität Köln (1979); Rev. Roum. Math. Pures Appl. *26*, 1319 – 1329 (1981).
[30] Hackbusch, W.: Die schnelle Auflösung der Fredholmschen Integralgleichung zweiter Art. Beiträge Numer. Math. *9*, 47 – 62 (1981).
[31] Hackbusch, W.: On the regularity of difference schemes – Part II: Regularity estimates for linear and nonlinear problems. Ark. Mat. *21*, 3 – 28 (1982).
[32] Hairer, E.: On the order of iterated defect corrections. Numer. Math. *29*, 409 – 424 (1978).
[33] Hemker, P. W., Schippers, H.: Multiple grid methods for the solution of Fredholm equations of the second kind. Math. Comp. *36*, 215 – 232 (1981).
[34] Hemker, P. W., Wesseling, P., De Zeeuw, P. M.: A portable vector code for autonomous multigrid modules. In: PDE Software: Modules, Interfaces and Systems (Engquist, B., ed.), pp. 29 – 40. Procs. IFIP WG 2.5 Working Conference, North-Holland, 1984.
[34a] Lin Qun, Zhu Qiding: Asymptotic expansions for the derivative of finite elements, J. Comp. Math. *2* (1984, to appear).
[35] Lindberg, B.: Error estimation and iterative improvement for the numerical solution of operator equations. Report UIUCDS-R-76-820 (1976).
[36] Lindberg, B.: Error estimation and iterative improvement for discretization algorithms. BIT *20*, 486 – 500 (1980).
[37] Marchuk, G. I., Shaidurov, V. V.: Difference Methods and Their Extrapolations. Berlin-Heidelberg-New York: Springer 1983.
[38] Munz, H.: Uniform expansions for a class of finite difference schemes for elliptic boundary value problems. Math. Comp. *36*, 155 – 170 (1981).
[39] Novak, Z., Wesseling, P.: Multigrid acceleration of an iterative method with application to transonic potential flow. (To appear.)
[40] Ortega, J. M., Rheinboldt, C. W.: On discretization and differentiation of operators with applications to Newton's method. SIAM J. Numer. Anal. *3*, 143 – 156 (1966).
[41] Pereyra, V.: The difference correction method for nonlinear two-point boundary value problems. – Techn. Rep. CS 18, Comp. Sc. Dept., Stanford Univ., California, 1965.
[42] Pereyra, V.: Accelerating the convergence of discretization algorithms. SIAM J. Numer. Anal. *4*, 508 – 533 (1967).

[43] Pereyra, V.: Iterated deferred corrections for nonlinear operator equations. Numer. Math. *10*, 316–323 (1967).
[44] Pereyra, V.: Iterated deferred corrections for nonlinear boundary value problems. Numer. Math. *11*, 111–125 (1968).
[45] Pereyra, V.: Highly accurate numerical solution of quasilinear elliptic boundary-value problems in n dimensions. Math. Comp. *24*, 771–783 (1970).
[46] Pereyra, V., Proskurowski, W., Widlund, O.: High order fast Laplace solvers for the Dirichlet problem on general regions. Math. Comp. *31*, 1–16 (1977).
[47] Skeel, R. D.: A theoretical framework for proving accuracy results for deferred corrections. SIAM J. Numer. Anal. *19*, 171–196 (1981).
[48] Stetter, H. J.: Asymptotic expansions for the error of discretization algorithms for nonlinear functional equations. Numer. Math. *7*, 18–31 (1965).
[49] Stetter, H. J.: Analysis of Discretization Methods for Ordinary Differential Equations. Berlin-Heidelberg-New York: Springer 1973.
[50] Stetter, H. J.: Economical global error estimation. In: Stiff Differential Systems (Willoughby, R. A., ed.), pp. 245–258. New York-London: Plenum Press 1974.
[51] Stetter, H. J.: The defect correction principle and discretization methods. Numer. Math. *29*, 425–443 (1978).
[52] Ueberhuber, C. W.: Implementation of defect correction methods for stiff differential equations. Computing *23*, 205–232 (1979).
[53] Wesseling, P.: A robust and efficient multigrid method. In: Multigrid Methods (Hackbusch, W., Trottenberg, U., eds.), pp. 613–630. (Lecture Notes in Mathematics, Vol. 960.) Berlin-Heidelberg-New York: Springer 1982.
[54] Zadunaisky, P. E.: A method for the estimation of errors propagated in the numerical solution of a system of ordinary differential equations. In: The Theory of Orbits in the Solar System and in Stellar Systems. Proc. of Intern. Astronomical Union, Symp. 25, Thessaloniki (Contopoulos, G., ed.). 1964.
[55] Zadunaisky, P. E.: On the estimation of errors propagated in the numerical integration of ordinary differential equations. Numer. Math. *27*, 21–40 (1976).

Prof Dr. K. Böhmer
Fachbereich Mathematik
Philipps-Universität
D-3550 Marburg
Federal Republic of Germany

Dr. P. W. Hemker
Centre for Mathematics and Computer Science
Department of Numerical Mathematics
Kruislaan 413
NL-1098 SJ Amsterdam
The Netherlands

Prof. Dr. H. J. Stetter
Institut für Angewandte und
Numerische Mathematik
Technische Universität Wien
Wiedner Hauptstrasse 6–10
A-1040 Wien
Austria