Numer. Math. 29, 425–443 (1978)

Numerische
Mathematik

The Defect Correction Principle and Discretization Methods

Hans J. Stetter*

Institut für Numerische Mathematik der Technischen Universität,
Gusshausstr. 27, A-1040 Wien, Austria

Summary. Recently, a number of closely related techniques for error estimation and iterative improvement in discretization algorithms have been proposed. In this article, we expose the common structural principle of all these techniques and exhibit the principal modes of its implementation in a discretization context.

Subject Classifications. AMS(MOS): 65J05, 65L99; CR: 5.17.

1. Introduction

During the past years, there have been numerous attempts to estimate the error of discretization methods or, equivalently, to improve the accuracy of their results. Two approaches have proved to be widely applicable: Richardson Extrapolation and Deferred Correction. Both can be used in an iterative fashion; the most notable difference between the two is the fact that Deferred Correction proceeds on the original grid of the discretization while Richardson Extrapolation needs repeated grid refinement and yet produces answers on the original grid only.

Recently a further technique of iterative improvement has been suggested. Presumably it has been used informally on various occasions; its conscious use —without iteration and for error estimations only—was promoted by Zadunaisky in [1–3] and at the 1973 Dundee conference; following this meeting Stetter [4] formalized the procedure and conceived its iterative application. Detailed analyses of particular applications were then made by Frank [6, 7], Frank and Ueberhuber [8, 9, 11] and Frank, Hertling and Ueberhuber [10, 12]. Lindberg [16], in generalizing the Deferred Correction approach, independently arrived at one version of the general technique. While Stetter had suggested to call the technique Differential Correction, Frank and Ueberhuber introduced

* Written during a sabbatical stay at Oxford University partially supported by the British Science and Research Council

0029-599X/78/0029/0425/$03.80

the term Defect Correction. In Section 5 of this paper we shall see that the wellknown Iterated Deferred Correction (or Difference Correction) technique of Fox (e.g. [13]) and Pereyra (e.g. [14, 15]) can be interpreted as a special case of the general principle of Iterative Updating Defect Correction.

In order to see the general principle and its ramifications more clearly, we shall study it at first without reference to discretizations or to differential equations. Actually, Iterative Defect Correction may be a very powerful technique in other areas of Mathematics as well. The Examples in the following are only of a demonstrative nature. Realistic implementations of Defect Correction have been discussed in many of the quoted references and numerical results have also been presented there.

2. The Basic Principle

Consider two normed linear spaces E and E^0 and a continuous (generally nonlinear) mapping F: $E \to E^0$. Let F be bijective between a domain $X \subset E$ and a domain $Y \subset E^0$, with $0 \in Y$, and assume that in all our operations we shall never leave these domains (in an application this has to be checked). Our aim is to find a good approximation to the unique solution $x^* \in X$ of

$$Fx=0 \tag{2.1}$$

with the aid of an *approximate* inverse $\tilde{G}$: $Y \to X$; $\tilde{G}$ is also assumed to be bijective and continuous. $\tilde{G}$ may be considered as the solution operator of an approximation

$$\tilde{F}x=0 \tag{2.2}$$

of (2.1).

To obtain an estimate for the error of $x_0 := \tilde{G}0$, we form the *defect* $d_0 := Fx_0$ and compute $\bar{x}_0 := \tilde{G}d_0$ (see Fig. 1). Thus $\bar{x}_0$ is the approximate solution of the equation $Fx=d_0$ which is a “neighboring problem” of (2.1) whose exact solution x_0 is known. If we assume that the error generated by our approximate solution operator $\tilde{G}$ is nearly the same for the two problems we obtain

$$x_0 - x^* \approx \bar{x}_0 - x_0. \tag{2.3}$$

This idea may either be used to estimate $x_0 - x^*$ by (2.3), or to obtain an improved approximation (see Fig. 1)

$$x_1 := x_0 - (\bar{x}_0 - x_0).$$

But then the process may be repeated ($i=0, 1, 2, \ldots$):

$$\bar{x}_i := \tilde{G}Fx_i = \tilde{G}d_i, \tag{2.4}$$

$$x_{i+1} := x_i - (\bar{x}_i - x_0) = (I - \tilde{G}F)x_i + x_0; \tag{2.5}$$

this is the principle of Iterative Defect Correction (IDeC), version A.

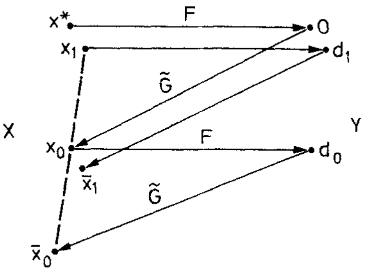

Fig. 1. Defect correction, version A

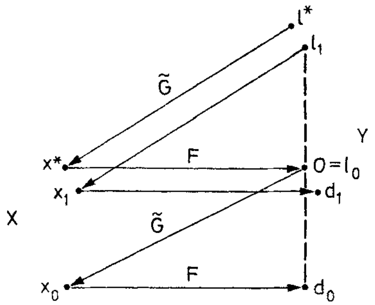

Fig. 2. Defect correction, version B

x^* of (2.1) is a fixed point of (2.5) (remember that $x_0 := \tilde{G}0$); therefore (2.5) will converge to x^* if $I - \tilde{G}F$ is a contraction in X, and the rate of convergence will depend on the (local) Lipschitz constant. Obviously, $\tilde{F}$ must reflect the local behavior of F sufficiently well.

The defect $d_0 = Fx_0$ may also be used in another way: We may attempt to obtain a better approximation to the unique element $l^* \in Y$ which satisfies

$$\tilde{G}l^* = x^*. \tag{2.6}$$

By virtue of the "neighbouring problem" idea (see Fig. 2) $l_1 := -d_0 = l_0 - d_0$ is a better approximation to l^* than the initial value $l_0 = 0$. Thus $x_1 := \tilde{G}l_1$ permits the estimate $x_0 - x_1$ for $x_0 - x^*$; also the process may be continued iteratively ($i = 0, 1, 2, \ldots$):

$$l_{i+1} := l_i - d_i = l_i - Fx_i, \tag{2.7}$$

$$x_{i+1} := \tilde{G}l_{i+1}. \tag{2.8}$$

This is version B of the IDeC principle (cf. Fig. 2); it was already suggested in Stetter [4].

When we substitute (2.8) into (2.7) we obtain

$$l_{i+1} := (I - F\tilde{G})\, l_i \tag{2.9}$$

which establishes version B as a dual of version A (cf. (2.5)); if we had started from $Fx = l_0$, $l_0 \neq 0$, we would have had to add l_0 in (2.7) and (2.9) and the duality would have been complete.

Version B may also be written as an iteration in X; (2.8) and (2.7) imply

$$x_{i+1} := \tilde{G}(l_i - d_i) = \tilde{G}(\tilde{F} - F)\, x_i; \tag{2.10}$$

this may be confronted with version A in the form (cf. (2.5))

$$x_{i+1} := x_i - (\tilde{G}Fx_i - \tilde{G}0) = (\tilde{G}\tilde{F} - \tilde{G}F)\, x_i + \tilde{G}0. \tag{2.11}$$

From (2.10) and (2.11) it is obvious that the two versions are equivalent if $\tilde{G}$ is an affine mapping

$$\tilde{G}y = \tilde{G}0 + \tilde{G}'_0 y, \qquad \tilde{G}'_0 \in \mathfrak{L}[Y, X]; \tag{2.12}$$

F need not be linear for this equivalence. With a non-linear $\tilde{G}$, on the other hand, the sequences $\{x_i\}$ obtained from versions A and B will generally differ.

If $\tilde{G}$ is Frechet-differentiable, one may wish to "linearize" it, i.e. replace it by (2.12), with $\tilde{G}_0' := \tilde{G}'(0)$. This causes the two versions to coincide.

Examples. 1) Iterative Improvement in Linear Algebraic Equations: $X = Y = \mathbb{R}^n$; $Fy = Ax - b$; $\tilde{G}y$ is the result of applying the numerically obtained triangular decomposition of A to the righthand side $b + y$.

2) Simplified Newton method (in Banach spaces): Let $F\colon X \to Y$ be Frechet-differentiable and take $\tilde{F}x := Fz + F'(z)(x-z)$, $z \in X$ fixed, so that $\tilde{G}y := z - F'(z)^{-1}(Fz - y)$. Then (2.10) and (2.11) become $x_{i+1} := x_i - F'(z)^{-1} F x_i$.

3) Defect correction does not depend on local linearization: Consider the non-linear boundary value problem

$$\begin{aligned} &\text{a)}\quad x''(t) - e^{x(t)} = 0 \qquad \text{on } (-1, +1), \\ &\text{b)}\quad x(-1) = x(+1) = 0, \end{aligned} \tag{2.13}$$

to define $F\colon C^{(2)}[-1, +1] \to C[-1, +1] \times \mathbb{R}^2$. To obtain an approximate problem, take an approximation for e^x in the interval $-0.4 \leqq x \leqq 0$ (where $x^*(t)$ will lie), say $e^x \approx 0.99 + 0.81x$, and define (2.2) by

$$\begin{aligned} &\text{a)}\quad x''(t) - 0.81\,x(t) - 0.99 = 0 \qquad \text{on } (-1, +1), \\ &\text{b)}\quad x(-1) = x(+1) = 0; \end{aligned} \tag{2.14}$$

(2.14) is not a local linearization of (2.13). Since the Green's function of the differential operator in (2.14) is known, the integral representation of $\tilde{G}$ can be used to compute the x_i, $i = 0, 1, 2, \ldots$ at least numerically. The sequence converges quickly to the solution x^* of (2.13).

Note that the linearity of (2.14) is not required by the method but it is convenient for the evaluation of $\tilde{G}$.

3. Approximate Solution in a Subspace

As a step towards studying IDeC in a discretization setting, we now assume that $\tilde{G}$ is not one-to-one between Y and X but maps Y into a proper subspace of E whose intersection with X we call Ξ, and that $x^* \notin \Xi$. The image of Ξ under F we denote by H (Fig. 3). H is not a linear manifold in Y if F is nonlinear. (To visualize this situation assume that E and E^0 are spaces of continuous functions but that $\tilde{G}$ always produces a polynomial of degree $\leqq N$.) The restriction of $\tilde{G}$ to H is assumed to be bijective from H to Ξ; the restriction of F to Ξ is naturally bijective from Ξ to H.

Figure 3 shows that version A of IDeC may be carried out as previously since Ξ is a domain in a linear space. With $\xi_0 = \tilde{G}0$ (greek letters denote elements in Ξ), (2.5) becomes

$$\xi_{i+1} := (I - \tilde{G}F)\,\xi_i + \xi_0; \tag{3.1}$$

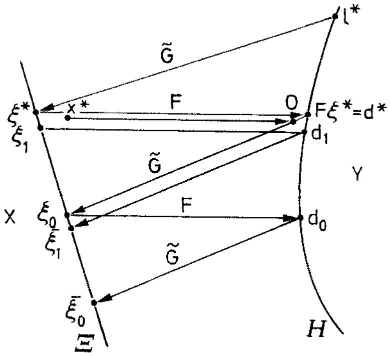

Fig. 3. Defect correction in a subspace

convergence now depends on the Lipschitz constant of the restriction of the mapping $I-\tilde{G}F$ to Ξ which we assume to be <1.

The unique fixed point $\xi^*\in\Xi$ of (3.1) is no longer characterized by (2.1). Instead ξ^* satisfies

$$\tilde{G}F\xi=\xi_0=\tilde{G}0=\tilde{G}Fx^*, \tag{3.2}$$

i.e. the defect d^* of ξ^* and 0 are mapped into the same element $\xi_0\in\Xi$ by $\tilde{G}$. The distance between ξ^* and x^* depends only on the relations between F and $\tilde{G}$ (or $\tilde{F}$) but not on the way in which ξ^* is computed (e.g. by IDeC).

Thus, if IDeC is applied in the situation of Figure 3 two independent types of errors arise:

a) the truncation error $\xi_i-\xi^*$ of the iteration (3.1),

b) the approximation error ξ^*-x^* characterized by (3.2).

The error ξ_i-x^* of the iterates will generally decrease only as long as $\|\xi_i-\xi^*\|\gg\|\xi^*-x^*\|$.

Version B of IDeC can no longer be formulated in a general situation of this type (cf. Fig. 3): The unique element $l^*\in H$ which satisfies $\tilde{G}l^*=\xi^*$ (cf. (2.6)) cannot be reached by linear combinations of elements in H.

However, it is often possible to decompose the mapping $\tilde{G}$ into a projection Δ^0 of Y into a linear subspace of E^0 and a mapping $\tilde{\Gamma}$ which is bijective from an appropriate domain H^0 in this subspace to Ξ:

$$\tilde{G}y=\tilde{\Gamma}\Delta^0 y. \tag{3.3}$$

In this case, we may form $\lambda_1:=-\Delta^0 F\xi_0\in\mathrm{H}^0$ in place of l_1 and continue (cf. (2.7) -(2.10))

$$\lambda_{i+1}:=\lambda_i-\Delta^0 F\xi_i=(I-\Delta^0 F\tilde{\Gamma})\lambda_i \tag{3.4}$$

or

$$\xi_{i+1}:=\tilde{\Gamma}\lambda_{i+1}=\tilde{\Gamma}(\tilde{\Phi}-\Delta^0 F)\xi_i, \tag{3.5}$$

where $\tilde{\Phi}: \Xi \to \mathrm{H}^0$ is the inverse of $\tilde{\Gamma}$. (3.4)/(3.5) is the appropriate form of version B of IDeC in this case.

The situation (3.3) arises whenever $\tilde{G}$ is of the form (2.12), or when its image is finite-dimensional.

Example. Consider again the BVP (2.13) and define an approximate problem (2.2) by (Π_m is the space of m-th degree polynomials):

$$\begin{aligned} &\text{a)}\quad \xi''(t) - 0.81\,\xi(t) - 0.99 = 0 \quad \text{on } \{-\tfrac{1}{2}, 0, \tfrac{1}{2}\},\\ &\text{b)}\quad \xi(-1) = \xi(+1) = 0,\\ &\text{c)}\quad \xi \in \Pi_4, \end{aligned} \tag{3.6}$$

so that $\tilde{G}$ maps Y into the finite-dimensional subspace $\Pi_4 \subset E$ while $F(\Pi_4)$ is not a subspace of E^0. The iteration (3.1) converges to the polynomial $\xi^* \in \Pi_4$ which satisfies

$$\begin{aligned} &\text{a)}\quad \xi^{*\prime\prime}(t) - e^{\xi^*(t)} = 0 \quad \text{on } \{-\tfrac{1}{2}, 0, \tfrac{1}{2}\}\\ &\text{b)}\quad \xi^*(-1) = \xi^*(+1) = 0, \end{aligned} \tag{3.7}$$

i.e. ξ^* is the "collocation solution" of (2.13) on $\{-\frac{1}{2}, 0, \frac{1}{2}\}$.

Here, the projection Δ^0 in the decomposition (3.3) is to the $\mathbb{R}^5$ which may be considered a subspace of E^0 by interpreting the first 3 components as values at $-\frac{1}{2}, 0, +\frac{1}{2}$ resp. of some three parameter linear set of functions from E^0. $\tilde{\Gamma}$ computes the 5 coefficients of $\tilde{\Gamma}\eta \in \Pi_4$ from the linear system (3.6), with $\eta \in \mathbb{R}^5$ on the right hand side. (Due to symmetry and (3.6b) this is effectively only a 2×2 system.) Thus version B is possible in the form (3.5) and equivalent to (3.1).

The "approximation error" in this example is the difference between the true solution x^* of (2.13) and the collocation solution ξ^* which is independent of the way in which ξ^* is found and which limits the useful accuracy in the computation of ξ^*.

4. Approximation by Discretization

In a discretization, an infinite-dimensional original problem (2.1) is replaced by a *sequence* of approximate problems ($N \in \mathbb{N}$)

$$\tilde{\Phi}_N \xi = 0 \tag{4.1}$$

where $\tilde{\Phi}_N$ is a continuous mapping from E_N into E_N^0, both of which are N-dimensional linear spaces. Furthermore, there are sequences of mappings Δ_N and Δ_N^0 which project elements from E and E^0 into E_N and E_N^0 resp. (see, e.g., the introductory sections of Stetter [5]). Concepts like consistency and convergence refer to limits as $N \to \infty$. In standard applications, E_N and E_N^0 are spaces of functions on a grid $\mathbb{G}_N$; the grids $\mathbb{G}_N$ become arbitrarily fine as $N \to \infty$.

In the context of IDeC, we choose a particular value of N (i.e. a particular grid) and *keep it fixed* during the iterative improvement. Thus there is only *one* space E_N and E_N^0 each and (4.1) is a fixed approximate problem, with a continuous solution operator $\tilde{\Gamma}_N$ (by assumption). We restrict ourselves to domains $X_N \subset E_N$ and $Y_N \subset E_N^0$ between which Φ_N and $\tilde{\Gamma}_N$ are bijective.

In order to apply the defect correction approach in X_N and Y_N, we must be able to associate with each $\xi \in X_N$ a *defect* w.r.t. (2.1) which is an element of Y_N. For this purpose we may proceed thus (see Fig. 4):

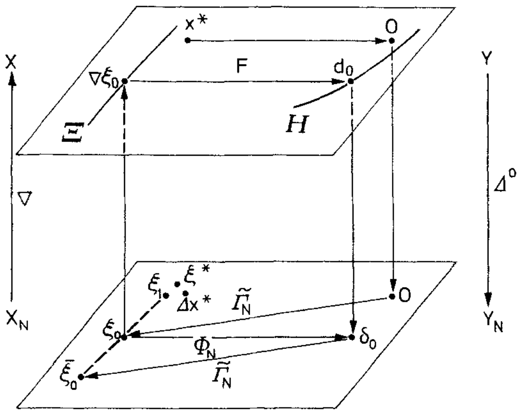

Fig. 4. Defect correction and discretization

Define a continuous mapping ∇ which maps X_N into a linear subset Ξ of X and a continuous mapping Δ^0: $Y \rightarrow Y_N$, with $\Delta^0 0 = 0$. Then define the defect δ of an element $\xi \in X_N$ by

$$\delta := \Delta^0 F \nabla \xi =: \Phi_N \xi. \tag{4.2}$$

(In a standard application, ∇ could generate a polynomial $\nabla\xi \in X$ interpolating the values of ξ at the gridpoints; the defect $d = F\nabla\xi \in Y$ can then be formed and projected into a function on the grid.)

For the purpose of IDeC, the only essential feature is that we have a continuous, bijective mapping Φ_N: $X_N \rightarrow Y_N$ (possibly after some restriction of X_N or Y_N) which *defines defects.* Then we may proceed completely within the finite-dimensional setting of X_N and Y_N. In fact, if we propose that we are looking for the unique $\xi^* \in X_N$ which solves

$$\Phi_N \xi = 0 \in E_N^0 \tag{4.3}$$

we have returned to the general situation of Section 2, with the original problem (4.3) in place of (2.1), and $\tilde{\Gamma}_N$ in place of $\tilde{G}$. Thus both versions of IDeC may be used: Version A is

$$\xi_{i+1} := \xi_i - (\bar{\xi}_i - \xi_0), \tag{4.4}$$

with

$$\bar{\xi}_i := \tilde{\Gamma}_N \delta_i = \tilde{\Gamma}_N \Phi_N \xi_i, \tag{4.5}$$

starting from $\xi_0 := \tilde{\Gamma}_N 0$, while Version B becomes (cf. also (2.10))

$$\lambda_{i+1} := \lambda_i - \delta_i = \lambda_i - \Phi_N \xi_i = [\tilde{\Phi}_N - \Phi_N]\, \xi_i, \tag{4.6}$$

$$\xi_{i+1} := \tilde{\Gamma}_N \lambda_{i+1}, \tag{4.7}$$

starting from $\lambda_0 = 0$, $\xi_0 = \tilde{\Gamma}_N 0$. (The operation $\tilde{\Phi}_N - \Phi_N$ required for the updating of λ_i may often be simpler than either $\tilde{\Phi}_N$ or Φ_N; see Example and Section 6.)

Convergence to the solution ξ^* of (4.3) or the solution λ^* of

$$\tilde{\Gamma}_N \lambda = \xi^* \tag{4.8}$$

depends on the contractive power of $(I - \tilde{\Gamma}_N \Phi_N)$: $X_N \to X_N$ or $(I - \Phi_N \tilde{\Gamma}_N)$: $Y_N \to Y_N$ resp. Again, if $\tilde{\Gamma}_N$ is linear both versions coincide.

One important distinction between versions A and B may remain in the discretization context: Consider a discretization of $Fx = d$, $d \in Y$, which requires values of the function d between gridpoints (the classical Runge-Kutta method, e.g.). Here, the mapping Δ^0 can no longer be explicitly evaluated nor can Φ_N be defined by (4.2).

In version A of the IDeC algorithm we may simply bypass the domain Y_N and define $\bar{\xi}$ by a mapping $\hat{\Gamma}_N$ from Y to X_N directly:

$$\bar{\xi} := \hat{\Gamma}_N d_i = \hat{\Gamma}_N F \nabla \xi_i. \tag{4.9}$$

Version B, however, which constructs λ_i in Y_N needs an evaluation of Δ^0 or of Φ_N.

We have yet to consider the relation between the limit ξ^* of our IDeC and the true solution x^* of the original problem (2.1). Generally, $\Delta x^* = \xi^*$ will *not* be true (at least not for a simple-minded projection Δ: $X \to X_N$). To understand the relation between ξ^* and Δx^*, we interpret (4.3) — which has been our representation of the original problem (2.1) in the finite-dimensional context — as another, more sophisticated discretization of (2.1). Then $\xi^* - \Delta x^*$ is the (global) discretization error of this new discretization for our particular value of N. Note that $\xi^* - \Delta x^*$ is determined exclusively by the relation between Φ_N and F:

$$\xi^* - \Delta x^* = (\Phi_N^{-1} \Delta^0 F - \Delta)\, x^*, \tag{4.10}$$

without any reference to the IDeC.

Thus we have the same situation as in Section 3: The error $\xi_i - \Delta x^*$ is composed of two independent parts:

a) the truncation error $\xi_i - \xi^*$ of the iteration,

b) the discretization error $\xi^* - \Delta x^*$ of the defect defining discretization (4.3).

Since

$$|\, \|\xi^* - \Delta x^*\| - \|\xi_i - \xi^*\| \,| \leqq \|\xi_i - \Delta x^*\| \leqq \|\xi^* - \Delta x^*\| + \|\xi_i - \xi^*\|, \tag{4.11}$$

some knowledge about $\xi^* - \Delta x^*$ is necessary for a considerate termination of the iteration.

Example. Consider the standard discretization of the BVP (2.13)

$$\begin{aligned} &\text{a)}\quad \frac{1}{h^2}[\xi(t_{n-1})-2\xi(t_n)+\xi(t_{n+1})]-e^{\xi(t_n)}=0, \qquad n=1(1)N-1, \\ &\text{b)}\quad \xi(t_0)=\xi(t_N)=0, \end{aligned} \tag{4.12}$$

and take (for simplicity) $N=4$, $h=\frac{1}{2}$ (fixed) so that ξ: $\mathbb{G}_4:=\{-1, -\frac{1}{2}, 0, \frac{1}{2}, 1\} \to \mathbb{R}$. Let V define the interpolating polynomial $\in \Pi_4$ and let Δ^0 restrict the function part of an element $y \in E^0$ to $\{-\frac{1}{2}, 0, \frac{1}{2}\}$.

Then $\Delta^0 F$ is the operator of (3.7) and the defect defining function $\Phi_4 = \Delta^0 F V$: $X_4 \to Y_4$ can easily be found. In our simple situation, with $\xi(-\frac{1}{2})=\xi(+\frac{1}{2})$ and $\xi(\pm 1)=0$, we obtain for the second derivative of the interpolating polynomial $V\xi$

$$\begin{aligned} (V\xi)''(0) &= -10\xi(0)+\tfrac{32}{3}\xi(\tfrac{1}{2}) \\ (V\xi)''(\tfrac{1}{2}) &= \quad 2\xi(0)-\tfrac{16}{3}\xi(\tfrac{1}{2}) \end{aligned}$$

and

$$\Phi_4 \xi = \begin{cases} -10\xi(0)+\frac{32}{3}\xi(\frac{1}{2})-e^{\xi(0)} & \text{at } t=0, \\ \quad 2\xi(0)-\frac{16}{3}\xi(\frac{1}{2})-e^{\xi(\frac{1}{2})} & \text{at } t=\pm\frac{1}{2}. \end{cases} \tag{4.13}$$

In the expression for $\tilde{\Phi}_4 - \Phi_4$ occurring in version B of IDeC (cf. (4.6)) the exponentials in (4.12) and (4.13) cancel. In version A which differs from version B since (4.12) is non-linear, a similar economy does not occur.

Our "defect" will vanish precisely when $V\xi$ is the collocation solution of the Example in Section 3; thus our present limit ξ^* is the restriction to $\mathbb{G}_4$ of that polynomial and the "discretization error" $\xi^* - \Delta x^*$ is the error of the collocation on $\mathbb{G}_4$.

5. Updating the Defect Function

With our reinterpretation of (4.3), IDeC has acquired a new aspect: It has now become an iterative technique for the solution of the more sophisticated discrete problem (4.3) which presumably could not have been solved directly.

Furthermore, this interpretation suggests a possibility for extension: Since it is of dubious value to approach the solution ξ^* of (4.3) too closely (see (4.11)), we may rather wish to take only a few steps (or only *one*) towards solving (4.3) and then set up a *new* discrete problem (4.3) with an improved function Φ_N. We can now embark on a new IDeC, starting with our latest approximation to the old ξ^*. This updating procedure may be repeated and become part of an overall iterative scheme.

If we agree to update Φ_N, which is our means for computing the defect of an approximation ξ, after each iteration, we must have a sequence of functions Φ_N^i: $X_N \to Y_N$, $i=1,2,\ldots$. Iterative Updating Defect Correction (IUDeC) now takes the form

$$\xi_{i+1} := \xi_i - (\tilde{\Gamma}_N \Phi_N^{i+1} \xi_i - \xi_0) \qquad \text{(Version A)} \tag{5.1}$$

or (cf. (4.6))

$$\lambda_{i+1} := \lambda_i - \Phi_N^{i+1} \tilde{\Gamma}_N \lambda_i = [\tilde{\Phi}_N - \Phi_N^{i+1}] \xi_i \qquad \text{(Version B).} \tag{5.2}$$

Both (5.1) and (5.2) are started with $i=0$ and $\xi_0 = \tilde{\Gamma}_N 0$ which is the result of our original discrete problem (4.1). Note that only problems of the type $\tilde{\Phi}_N \xi = \delta$ are solved throughout so that no solution operator other than $\tilde{\Gamma}_N$ appears.

We have now to distinguish the solutions $\xi_i^* \in X_N$ of the various problems

$$\Phi_N^i \xi = 0. \tag{5.3}$$

In step (5.1), we have (cf. Section 2)

$$\|\xi_{i+1} - \xi_{i+1}^*\| \leqq L_{i+1}^A \|\xi_i - \xi_{i+1}^*\| \tag{5.4}$$

and similarly in step (5.2)

$$\|\lambda_{i+1} - \lambda_{i+1}^*\| \leqq L_{i+1}^B \|\lambda_i - \lambda_{i+1}^*\| \tag{5.5}$$

where L_{i+1}^A and L_{i+1}^B are Lipschitz constants for $I - \tilde{\Gamma}_N \Phi_N^{i+1}$ and $I - \Phi_N^{i+1} \tilde{\Gamma}_N$ resp. Alternatively, (see (5.2)) we may consider version B in the form

$$\xi_{i+1} - \xi_{i+1}^* = \tilde{\Gamma}_N [\tilde{\Phi}_N - \Phi_N^{i+1}] \xi_i - \tilde{\Gamma}_N [\tilde{\Phi}_N - \Phi_N^{i+1}] \xi_{i+1}^* \tag{5.6}$$

and regard the contractivity of $\tilde{\Gamma}_N [\tilde{\Phi}_N - \Phi_N^{i+1}]$, which may be advantageous because of the simpler structure of $\tilde{\Phi}_N - \Phi_N^{i+1}$.

If these Lipschitz constants are <1 we progress in each step, but in each step *towards a new goal* ξ_{i+1}^*. Thus, IUDeC is reasonable only if the ξ_i^* have a consistent behavior, i.e. if they approach Δx^*. Since (cf. (4.11))

$$|\, \|\xi_i^* - \Delta x^*\| - \|\xi_i - \xi_i^*\| \,| \leqq \|\xi_i - \Delta x^*\| \leqq \|\xi_i^* - \Delta x^*\| + \|\xi_i - \xi_i^*\| \tag{5.7}$$

a decrease in $\|\xi_i - \xi_i^*\|$ should be supported by a decrease in $\|\xi_i^* - \Delta x^*\|$. Therefore the discrete problems (5.3) should have smaller and smaller global discretization errors (for our fixed N!) as i increases.

A special IUDeC algorithm has been proposed a long time ago and used extensively: The *difference correction* procedure of Fox (e.g. [13]) and its refined versions due to Pereyra and others (e.g. [17], [18]) are immediately recognizable as procedures of the form (5.2). Here, the operations Φ_N^{i+1} are formed by adding higher and higher order differences to the basic discretization (4.1). Thus the ξ_i^* become solutions of higher and higher order discretizations of (2.1) and $\|\xi_i^* - \Delta x\|$ decreases accordingly, at least for a sufficiently fine grid $\mathbb{G}_N$. This more general nature of Pereyra's Deferred Correction approach has been discovered independently by Lindberg [16] who has also applied it to new classes of problems. Our exposition shows that this use of an increasingly accurate defect defining function Φ_N^i can be naturally extended to all kinds of IDeC procedures in relation with discretization methods. There is no prerequisite on the form of the Φ_N^i; they may just as well be derived by interpolation and subsequent substitution into (2.1) (as in (4.2)) or yet by some other method.

At first, one may believe that IUDeC permits a (theoretically) arbitrarily close approximation of Δx^*, with the aid of the solution operator $\tilde{\Gamma}_N$. However, we remain in the same spaces X_N and Y_N (i.e. on the same grid $\mathbb{G}_N$) throughout; hence there must be a limit to the accuracy achievable without knowledge of x^*. E.g., if the formation of the Φ_N^i involves polynomial interpolation, we cannot go beyond a certain degree on a grid with $N+1$ gridpoints.

Thus there is a discrete problem

$$\Phi_N^I \xi = 0 \tag{5.8}$$

beyond which we cannot or do not wish to go. Then ξ_I^* takes the place of ξ^* in Section 4, and we may interpret IUDeC as an iterative technique for computing ξ_I^*. Now the error $\|\xi_I^* - \Delta x^*\|$ of (5.8) remains unaffected by all our manipulations and establishes the background against which the progress in the iterations has to be judged.

On the other hand, we may also identify (5.8) with (4.3); then we find that we are able to use *simpler* defect defining functions Φ_N^i in early stages of the iterations of Section 4 and possibly save computational effort. The efficiency of this approach will depend on the balancing of the terms in (5.7).

6. Ways of Choosing the Defect Defining Functions

For a given problem (2.1) and a given basic discretization (4.1) with solution operator $\tilde{\Gamma}_N$, the various IDeC or IUDeC algorithms are defined solely by their associated defect defining functions Φ_N or $\Phi_N^1, \ldots, \Phi_N^I$. Let us consider a few standard ways of choosing Φ_N when (2.1) is a system of ordinary differential equations on a fixed interval $[0, T]$, with either initial or boundary conditions.

I. Global Interpolation

Here we use the approach (4.2) and define a continuous and piecewise differentiable counterpart $V\xi \in X$ to our approximation $\xi \in X_N$; for $V\xi$ we may form the defect by substitution into the differential equation and into the boundary conditions if they are not automatically satisfied. Polynomial interpolation (which was originally suggested by Zadunaisky) is a natural and computationally simple way; however, one need not use one polynomial over the whole interval $[0, T]$. Instead one may define subintervals $[\bar{t}_{m-1}, \bar{t}_m]$, with a suitable number of gridpoints in each, and interpolate on each subinterval with a polynomial of an appropriate degree.

Of course, continuity has to be enforced. Continuity of first derivatives across $\bar{t}_m$ is not necessary in first order systems if the piecewise continuity of the defect is taken into account in the numerical solution procedure $\tilde{\Gamma}_N$. In higher order differential equations, one must either have the appropriate continuity of derivatives, or one may apply a technique devised by Frank [7]: He extends the definition of the original problem (2.1) by permitting solutions with jumps in derivatives at the fixed points $\bar{t}_m$. To retain uniqueness, he specifies the jump sizes (to be zero) in additional "boundary" conditions. Thus, non-zero jumps contribute to the defect and influence the evaluation of $\tilde{l}_N$. For details, see Frank [7].

In conjunction with an IUDeC algorithm, the degrees of the piecewise polynomials (and the lengths of the subintervals) may be increased as the iteration proceeds.

Other interpolation functions may be used where it is appropriate and computationally feasible.

In the case of global interpolation, the solution ξ^* of (4.3) or ξ_i^* of (5.3) is characterized by the fact that its interpolation $P\xi^*_{(i)}$ generates a defect in the differential equations (and possibly in the boundary conditions) which vanishes under the projection Δ^0. Normally this will mean that the defect vanishes at the gridpoints of $\mathfrak{G}_N$. In this case, the discrete problems (4.3) or (5.3) are classical *collocation problems* and IDeC is a computational technique for the approximate solution of a collocation method for the given ODE-problem (cf. the Examples in Sections 3 and 4). This connection between IDeC and collocation was first displayed by Frank and Ueberhuber [11]; it has played a prominent role in the use of IDeC for initial value problems with stiff ODEs (see Frank and Ueberhuber [9]).

II. Local Interpolation

To compute the value of the defect δ at $t_n \in \mathfrak{G}_N$, we may consider values $\xi(t_{n_\mu})$, $\mu = 0(1)m$, with n_μ ranging over a neighbourhood of n, interpolate the $\xi(t_{n_\mu})$ by a polynomial $P_n(t)$ of degree m and form the defect of P_n in the differential equation *at* t_n. This may be done for each $t_n \in \mathfrak{G}_N$, with – if necessary – special provisions for points near the boundary of $[0, T]$.

In this approach a different polynomial P_n is generally used for each $t_n \in \mathfrak{G}_N$, therefore we have no element $P\xi \in X$ to be used in (4.2). But we can still define the defect by means of the original problem, because *with ODEs* the evaluation of $[Fx](t)$ requires information about x only at t. If the function F of the original problem (2.1) contains a global operation $\left(\text{e.g. } \int\limits_0^T k(t,\tau)\,x(\tau)\,d\tau\right)$ then the local interpolation approach is not feasible.

It may be desirable to have the n_μ situated symmetrically w.r.t. n even near the boundary. This can often be achieved by an extension of ξ to arguments outside $[0, T]$ as suggested by Fox [13].

III. Finite Differences

In many situations it is possible to formulate very accurate discrete analoga (4.3) of (2.1); but their direct solution would require an enormous computational effort, particularly for nonlinear ODEs. We can now use these accurate finite difference approximations to define the defect of iterates in IDeC. Often, there is also a natural sequence of more and more accurate discretizations (5.3) which can be used in an IUDeC algorithm. Note that $\tilde{\Phi}_N$ and Φ_N^i in (5.2) may differ only by a combination of higher order differences; hence the term "difference correction" introduced by Fox is very suggestive.

Actually, Fox [13] suggested and discussed this version of IUDeC as early as 1957 in its application to two-point boundary value problems. His use of central difference corrections corresponds to the local interpolation approach (see II. above) while further refinements by Pereyra (e.g. [18]) are more in the spirit of the global interpolation approach, though formulated in terms of differences throughout. It is clear that difference correction and polynomial interpolation are very similar and may actually lead to identical expressions for the defect in some cases.

The advantage of the interpolation approach is its ease of application even in complicated situations. Lindberg [16] who has described the IUDeC principle in remarkable generality is quite limited in his applications by considering defect defining functions only in the form of finite difference operators. See also Frank, Hertling, Ueberhuber [12] for a discussion of this aspect.

7. Order Considerations

In the analysis of discretization procedures, one considers the rate at which certain quantities tend to zero if the grids $\mathfrak{G}_N$ become arbitrarily fine (cf. beginning of Section 4). Ordinarily, the refinement of $\mathfrak{G}_N$ is characterized by a meshwidth parameter h_N and the rate of decrease is expressed in terms of powers of h_N.

If we apply this analysis in relation to IDeC, we are considering the following situation: We have a fully defined IDeC-(or IUDeC-)algorithm on a given grid $\mathfrak{G}_N$, with mappings $\tilde{\Gamma}_N$ and Φ_N or Φ_N^i between domains X_N and Y_N associated with $\mathfrak{G}_N$. We now consider $\mathfrak{G}_N$ and these mappings as elements of infinite sequences $\{\mathfrak{G}_N\}$, $\{\tilde{\Gamma}_N\}$, $\{\Phi_N^i\}$ of such quantities where the refinement h_N of $\mathfrak{G}_N$ tends to zero as $N \to \infty$. Each quantity which arises in the course of our IDeC-algorithm has now become a function of N, and quantities or differences of quantities that tend to zero as $N \to \infty$ may be related to powers of h_N. Only the quantities of the original problem (2.1) remain invariant in this limit process.

In our previous analysis we have not denoted the reference to N with many quantities and we will retain this notation. But it should be kept in mind that an assertion like $\|\xi_i - \xi_i^*\| = O(h^{(i+1)p})$ is meaningless for one IDeC-algorithm on a fixed grid; it is only an assertion about the limit behavior of the sequence of values $\|\xi_i - \xi_i^*\|$ arising from the sequence of IDeC-algorithms defined above. Also $e_1 = O(h^{p_1})$ and $e_2 = O(h^{p_2})$, with $p_1 > p_2$, implies only the existence of an $\bar{h}$ such that $|e_1| < |e_2|$ holds for all $h_N \in (0, \bar{h})$; but $\bar{h}$ may be far too small to be meaningful in any computation. Thus, results of an order analysis should be used only as guide-lines regarding the relative sizes of quantities. Finally, remember that in a statement like $\|\xi_i - \Delta x^*\| = O(h^r)$ the choice of the norm is not irrelevant and that different norms may lead to different values of r. For the elements of the sequence $\{\xi_i - \Delta x^*\}_{N \in \mathbb{N}}$ come from different spaces E_N, with $\dim E_N \to \infty$ as $N \to \infty$. (Cf. e.g. Section 2.2 of Stetter [5].)

The essential difficulty arising in an order analysis of IDeC-algorithms is the following: Since $\tilde{\Phi}_N$ and Φ_N are both approximations to the same differential equation, their difference will normally amount to a high order difference operation (or a differentiation process after interpolation); consider, e.g., the original difference correction procedure of Fox [13]. In a naive analysis, the maximum norm of an r-th order difference quotient of some ξ_i in an IDeC-algorithm would increase like h^{-r}; but actually it remains bounded if the original problem is sufficiently smooth and the algorithm has been set up correctly. To establish this fact one has to ascertain that ξ_i possesses an asymptotic expansion in powers of h, with coefficients which are projections into X_N of smooth functions $\bar{e}_{ij}$ in X independent of h:

$$\xi_i = \Delta x^* + \sum_{j=\bar{p}_i}^{J_i} h^j \Delta \bar{e}_{ij} + \bar{\rho}_i(h), \qquad \text{with } \bar{\rho}_i \in X_N,\ \|\bar{\rho}_i\| = O(h^{J_i+1}). \tag{7.1}$$

The existence of such expansions for the solutions of many classes of standard discretization algorithms can be derived from general theorems (e.g. in Stetter [5]). If the discrete problems (4.1) and (4.3) resp. (5.3) belong to these classes then asymptotic expansions of type (7.1) exist for the starting element ξ_0 of the IDeC algorithm and for the limit element(s) ξ^* resp. ξ_i^*. By subtraction, we arrive at an asymptotic expansion of the form

$$\xi_0 = \xi^* + \sum_{j=\tilde{p}}^{J_0} h^j \Delta e_{0j} + \rho_0(h), \qquad \|\rho_0\| = O(h^{J_0+1}), \tag{7.2}$$

where we have assumed that the discretizations (4.1) and (4.3) are of orders $\tilde{p}$ and p resp., with $p \geqq \tilde{p}$.

This order assumption also implies (cf. e.g. Stetter [5], Section 1.1.2) that

$$\begin{aligned} &\|(\tilde{\Phi}_N \Delta - \Delta^0 F) e\| = O(h^{\tilde{p}}) \\ &\|(\Phi_N \Delta - \Delta^0 F) e\| = O(h^p) \end{aligned} \tag{7.3}$$

for any sufficiently smooth $e \in X$. Some analysis will normally reveal the existence of mappings D_N: $X \to Y$ such that

$$(\tilde{\Phi}_N - \Phi_N) \Delta e = h^{\tilde{p}} \Delta^0 D_N e \tag{7.4}$$

where D_N may be expanded

$$D_N = D^0 + h D^1 + h^2 D^2 + \cdots + h^J D_N^J \tag{7.5}$$

with D^j, $j = O(1) J-1$, independent of h and $\|D_N^J\| = O(1)$. (Consider, e.g. the expansion of a $\tilde{p}$-th order difference in terms of derivatives.)

Let us now consider the generation of $\xi_1 \in X_N$ in Version B of IDeC, starting with the expansion (7.2). We will assume that all operations are sufficiently Frechet-differentiable. (4.6) becomes

$$\begin{aligned} \lambda_1 = [\tilde{\Phi}_N - \Phi_N]\, \xi_0 = [\tilde{\Phi}_N - \Phi_N]\, \xi^* + [\tilde{\Phi}_N - \Phi_N]'(\xi^*) \sum_{j=\tilde{p}}^{J_0} h^j \Delta e_{0j} + \cdots \\ + [\tilde{\Phi}_N - \Phi_N]'(\xi^*)\, \rho_0(h) + \cdots; \end{aligned} \tag{7.6}$$

the Frechet-Taylor expansion is terminated at an appropriate order by a remainder term. Substitution of (7.4)/(7.5) into (7.6) and collection of terms by powers of h yields an asymptotic expansion for λ_1 whose coefficients we denote by $\Delta^0 d_{1j}$.

By (4.3) and (4.8), $[\tilde{\Phi}_N - \Phi_N]\,\xi^* = \lambda^*$ and, by (7.4), the lowest exponent of h in the expansion is now $2\tilde{p}$!

Thus (7.6) leads to

$$\lambda_1 = \lambda^* + \sum_{j=2\tilde{p}}^{J_1} h^j \Delta^0 d_{1j} + \sigma_1(h), \qquad \|\sigma_1\| = O(h^{J_1+1}). \tag{7.7}$$

We have in any case to assume that the discretization (4.1) is stable, which implies $\|\tilde{\Gamma}_N'(\lambda)\| = O(1)$ for all $\lambda \in Y_N$ sufficiently close to 0 (cf. e.g. Stetter [5], Section 1.1.4). Therefore we obtain from (7.7), with (4.7) and (4.8),

$$\xi_1 = \xi^* + \tilde{\Gamma}_N'(\lambda^*) \sum_{j=2\tilde{p}}^{J_1} h^j \Delta^0 d_{1j} + \cdots + \tilde{\Gamma}_N'(\lambda^*)\,\sigma_1(h) + \cdots. \tag{7.8}$$

(7.8) suffices to establish $\|\xi_1 - \xi^*\| = O(h^{2\tilde{p}})$; but to obtain

$$\xi_1 = \xi^* + \sum_{j=2\tilde{p}}^{J_1} h^j \Delta e_{1j} + \rho_1(h), \qquad \|\rho_1\| = O(h^{J_1+1}), \tag{7.9}$$

we have to interpret the elements $\varepsilon_{1j} = \tilde{\Gamma}_N'(\lambda^*)\,\Delta^0 d_{1j} \in X_N$ as projections of elements which arise as solutions of variational problems associated with (2.1).

A considerable amount of technical reasoning may be necessary to get from (7.6) to (7.7) and from (7.8) to (7.9) for a particular problem and particular discretizations (4.1) and (4.3). If these transitions are possible in the form stated above, we may continue the process (compare (7.9) with (7.2)) and arrive at the following order assertions:

Let $\tilde{p}$ and p be the order of (4.1) and (4.3) resp. Then in Version B of an IDeC-algorithm based upon (4.1) and (4.3)

$$\|\xi_i - \xi^*\| = O(h^{(i+1)\tilde{p}}), \qquad i = 1(1)I, \tag{7.10}$$

where I depends on the order J_0 of the remainder term in (7.2) and on the decrease in the order of the remainder term encountered in each iteration. Furthermore (cf. (4.11))

$$\|\xi_i - \Delta x^*\| = O(h^{\min((i+1)\tilde{p},\,p)}), \qquad i = 1(1)I. \tag{7.11}$$

By a similar line of arguments, one arrives at the same assertions for the ξ_i arising in Version A of an IDeC-algorithm (compare Frank [7], Frank and Ueberhuber [8]).

In an IUDeC-algorithm, we assume that (5.3) is of order p_i and that each ξ_i^* possesses an asymptotic expansion of type (7.1); this leads to expansions

$$\xi_{i+1}^* - \xi_i^* = \sum_{j=\min(p_i,\,p_{i+1})}^{J} h^j \cdot \Delta \hat{e}_{ij} + \hat{\rho}_i(h), \qquad i = 1, 2, \ldots. \tag{7.12}$$

Replace ξ^* by ξ_1^* and λ^* by λ_1^* in (7.2)–(7.9). If $\min(p_1, p_2) \geqq 2\tilde{p}$, we may use (7.12) to turn (7.9) into an expansion

$$\xi_1 = \xi_2^* + \sum_{j=2\tilde{p}}^{J_1} h^j \cdot \Delta e_{1j} + \rho_1(h), \tag{7.13}$$

with new coefficients e_{1j} and a new remainder term ρ_1; the identification of (7.13) and (7.2) permits again a recursive argument and the following assertion holds:

Let $\tilde{p}$ and p_i, $i=1,2,\ldots$, be the orders of the discretizations (4.1) and (5.3) resp. and assume that $\min(p_i, p_{i+1}) \geqq (i+1)\tilde{p}$. Then in an IUDeC-algorithm based upon (4.1) and (5.3)

$$\|\xi_i - \xi_{i+1}^*\| = O(h^{(i+1)\tilde{p}}), \qquad i = 1(1)I, \tag{7.14}$$

with I determined similarly as in (7.10). Furthermore (cf. (5.7))

$$\|\xi_i - \Delta x^*\| = O(h^{(i+1)\tilde{p}}), \qquad i = 1(1)I. \tag{7.15}$$

The result (7.15) agrees with Pereyra's asymptotic results (e.g. [19]) for his Iterated Deferred Correction Algorithm for two-point boundary value problems. (7.15) has also been derived by Lindberg [16] using arguments and assumptions similar to ours.

8. Simplifications for Error Estimation

For an appraisal of the global discretization error $\xi_0 - \Delta x^*$ of the approximate numerical solution ξ_0 of (2.1), it would often be sufficient to obtain the correct order of magnitude and the sign of the error. In this case, defect correction should be applied in as "cheap" a form as possible and it may appear ineconomic to spend another application of $\tilde{\Gamma}_N$, even if some effort can be saved in the second application, e.g. by reusing Jacobians.

Instead let us interpret $\xi_0 = \tilde{\Gamma}_N 0$ as an *intermediate* iterate $\tilde{\tilde{\xi}}_i$ of an IDeC-iteration based on a simpler discretization

$$\tilde{\tilde{\Phi}}_N \xi = 0 \tag{8.1}$$

of (2.1) than (4.1). We may then use the simpler solution operator $\tilde{\tilde{\Gamma}}_N$ of (8.1) to compute $\tilde{\tilde{\xi}}_{i+1}$ and to estimate the error of $\xi_0 = \tilde{\tilde{\xi}}_i$ by

$$\xi_0 - \Delta x^* = \tilde{\tilde{\xi}}_i - \Delta x^* \approx \tilde{\tilde{\xi}}_i - \xi^* \approx \tilde{\tilde{\xi}}_i - \tilde{\tilde{\xi}}_{i+1} = \xi_0 - \tilde{\tilde{\xi}}_{i+1}. \tag{8.2}$$

Obviously, the validity of (8.2) rests on the assumption that the solution ξ^* of (4.3) is sufficiently close to Δx^* and that the IDeC-operator associated with (8.1) is sufficiently contracting:

$$\|\xi_0 - \Delta x^*\| \leqq \|\xi_0 - \tilde{\tilde{\xi}}_{i+1}\| + \|\xi^* - \Delta x^*\| + \|\tilde{\tilde{\xi}}_{i+1} - \xi^*\|.$$

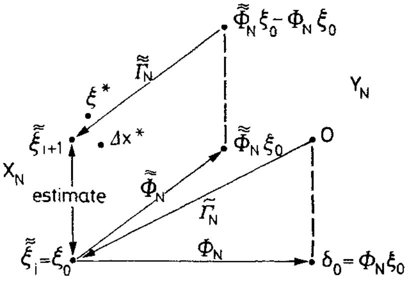
$\tilde{\tilde{\Phi}}_N\xi_0 - \Phi_N\xi_0$
$\tilde{\tilde{\Gamma}}_N$
ξ^*
Y_N
$\tilde{\tilde{\xi}}_{i+1}$
Δx^*
$\tilde{\tilde{\Phi}}_N\xi_0$
0
X_N
estimate
$\tilde{\tilde{\Phi}}_N$
$\tilde{\Gamma}_N$
$\tilde{\tilde{\xi}}_i = \xi_0$
Φ_N
$\delta_0 = \Phi_N\xi_0$

Fig. 5. Error estimation by a different operator

Still we have to be careful. In Version A, we now have (cf. (4.4))

$$\tilde{\tilde{\xi}}_{i+1} := \tilde{\tilde{\xi}}_i - (\tilde{\tilde{\Gamma}}_N \Phi_N \tilde{\tilde{\xi}}_i - \tilde{\tilde{\xi}}_0) = \xi_0 - (\tilde{\tilde{\Gamma}}_N \phi_N \xi_0 - \tilde{\tilde{\Gamma}}_N 0) \tag{8.3}$$

where $\tilde{\tilde{\Gamma}}_N 0$ is normally not available. Thus, in place of one application of $\tilde{\Gamma}_N$ we are faced with two applications of $\tilde{\tilde{\Gamma}}_N$.

In Version B, we have (cf. (4.6))

$$\tilde{\tilde{\xi}}_{i+1} := \tilde{\tilde{\Gamma}}_N [\tilde{\tilde{\Phi}}_N - \Phi_N] \tilde{\tilde{\xi}}_i = \tilde{\tilde{\Gamma}}_N [\tilde{\tilde{\Phi}}_N \xi_0 - \Phi_N \xi_0]. \tag{8.4}$$

Again $\tilde{\tilde{\Phi}}_N \xi_0$ is not available; but for a sufficiently simple $\tilde{\tilde{\Phi}}_N$ its evaluation should be very cheap. Thus for a judicious choice of (8.1), (8.4)/(8.2) should provide an economical way of estimating the global discretization error of ξ_0 with moderate accuracy; see also Figure 5.

In initial value problems, one may, e.g., use the plain explicit Euler method for (8.1) (except when dealing with stiff problems). The values of $f(t_n, \xi_0(t_n))$ will normally have been computed anyway for $t_n \in \mathfrak{G}_N$, so the evaluation of $\tilde{\tilde{\Phi}}_N \xi_0$ is trivial. If the evaluation of the defect $\delta_0 = \Phi_N \xi_0$ at t_{n-1} does not require information about t_n, the evaluation of (8.4)/(8.2) may even be done "in parallel" with the computation of ξ_0 so that an estimate of $\xi_0(t_n) - x^*(t_n)$ is available together with $\xi_0(t_n)$. Such a procedure has been successfully implemented and tested by the author ([20]).

Another possibility which might be particularly useful in boundary value problems consists in the use of a $\tilde{\tilde{\Phi}}_N$ which works on a coarser grid than $\mathfrak{G}_N$ (naturally a subgrid of $\mathfrak{G}_N$). Thus the dimension of the system of simultaneous equations could be reduced, at the expense of obtaining an error estimate only on the coarser grid. The similarity with Richardson Extrapolation comes to one's mind; actually there is only a small step from here to a procedure which is a "Version B" of Richardson Extrapolation as we will explain in a separate paper.

In Stetter [4], the suggestion had been made (although differently phrased) that $\tilde{\tilde{\Phi}}_N$ be some discrete approximation of a linearization of (2.1), which would make $\tilde{\tilde{\Gamma}}_N$ of the form (cf. (2.12))

$$\tilde{\tilde{\Gamma}}_N \delta = \tilde{\tilde{\Gamma}}_N 0 + \tilde{\tilde{\Gamma}}'_N \delta, \qquad \tilde{\tilde{\Gamma}}'_N \in \mathfrak{L}[Y_N, X_N].$$

In this case, (8.3) or (8.4) may be combined with (8.2) into

$$\xi_0 - \Delta x^* \approx \tilde{\tilde{\xi}}_i - \tilde{\tilde{\xi}}_{i+1} = \tilde{\tilde{\Gamma}}'_N \Phi_N \xi_0. \tag{8.5}$$

However, the application of $\tilde{\tilde{\Gamma}}'_N$ will normally involve Jacobians of the original problem and thus it may be expensive even with a primitive discretization (8.1).

Example. Consider the Example in Section 4. For $\mathbb{N}=4$, $h=\frac{1}{2}$, the value of the solution ξ_0 of (4.12) at $t=0$ is -0.363428 (to 6 decimal places), and from (4.13) we obtain $(\Phi_4 \xi_0)(0) \approx -0.010523$.

Now choose (4.12) with $N=2$, $h=1$, for $\tilde{\tilde{\Phi}}_4$. Then $(\xi_0(\pm 1)=0!)$

$$(\tilde{\tilde{\Phi}}_4 \xi_0)(0) = -2\xi_0(0) - e^{\xi_0(0)}. \approx 0.031567.$$

To find $\tilde{\tilde{\xi}}_1(0)$ we have thus to solve (cf. (8.4)):

$$-2\tilde{\tilde{\xi}}_1(0) - e^{\tilde{\tilde{\xi}}_1(0)} = -0.042090$$

which yields $\tilde{\tilde{\xi}}_1(0) \approx -0.367334$ and -0.003906 as our estimate for the error of $\xi_0(0)$ by (8.2). The true error is -0.004628.

9. Conclusion

In this paper we have tried to exhibit the general structure of Defect Correction which is a basic principle of Numerical Analysis. Discretization is a form of approximation which does not rely on linearization; therefore it is an excellent field for the application of the Defect Correction Principle which also does not depend on linearization. It is hoped that this exposition may help to understand the present implementations of Defect Correction more fully and that it may also point the way towards new applications.

References

1. 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. New York: Academic Press 1966
2. Zadunaisky, P.E.: On the accuracy in the numerical computation of orbits. In: Periodic orbits, stability and resonances (G.E.O. Giacaglia, ed.), pp. 216–227. Dordrecht: Reidel 1972
3. Zadunaisky, P.E.: On the estimation of errors propagated in the numerical integration of ordinary differential equations. Numer. Math. **27**, 21–39 (1976)
4. Stetter, H.J.: Economical global error estimation. In: Stiff differential systems (R.A. Willoughby, ed.), pp. 245–258. New York-London: Plenum Press 1974
5. Stetter, H.J.: Analysis of discretization methods for ordinary differential equations. Berlin-Heidelberg-New York: Springer 1973
6. Frank, R.: Schätzungen des globalen Diskretisierungsfehlers bei Runge-Kutta-Methoden, ISNM Vol. 27, pp. 45–70. Basel-Stuttgart: Birkhäuser 1975
7. Frank, R.: The method of iterated defect-correction and its application to two-point boundary value problems. Part I: Numer. Math. **25**, 409–419 (1976); Part II: Numer. Math. **27**, 407–420 (1977)

8. Frank, R., Ueberhuber, C.W.: Iterated defect correction for Runge-Kutta methods. Report No. 14/75, Inst. f. Numer. Math., Technical University of Vienna, 22 pp., 1975
9. Frank, R., Ueberhuber, C.W.: Iterated defect correction for the efficient solution of stiff systems of ordinary differential equations. Nordisk Tidskr. Informationsbehandling (BIT) **17**, 146–159 (1977)
10. Frank, R., Hertling, J., Ueberhuber, C.W.: Iterated defect correction based on estimates of the local discretization error. Report No. 18/76, Inst. f. Numer. Math., Technical University of Vienna, 21 pp., 1976
11. Frank, R., Ueberhuber, C.W.: Collocation and iterated defect correction. In: Numerical treatment of differential equations (R. Bulirsch, R.D. Grigorieff, J. Schröder, eds.), pp. 19–34. Lecture notes in Mathematics, Vol. 631. Berlin-Heidelberg-New York: Springer 1978
12. Frank, R., Hertling, J., Ueberhuber, C.W.: An extension of the applicability of iterated deferred corrections. Report No. 23/76, Inst. f. Numer. Math., Technical University of Vienna, 20 pp., 1976 (to appear in Math. Comput.).
13. Fox, L.: The numerical solution of two-point boundary value problems in ordinary differential equations. Oxford: University Press 1957
14. Pereyra, V.L.: On improving an approximate solution of a functional equation by deferred corrections. Numer. Math. **8**, 376–391 (1966)
15. Pereyra, V.L.: Iterated deferred corrections for nonlinear operator equations. Numer. Math. **10**, 316–323 (1967)
16. Lindberg, B.: Error estimation and iterative improvement for the numerical solution of operator equations. Report UIUCDCS-R-76-820, Dept. of Computer Science, Univ. of Ill., Urbana, 1976
17. Daniel, J.W., Martin, A.J.: Implementing deferred corrections for Numerov's difference method for second-order two-point boundary-value problems. Report CNA-107, Center for Numerical Analysis, Univ. of Texas, Austin, 1975
18. Lentini, M., Pereyra, V.L.: Boundary problem solvers for first order systems based on deferred corrections. In: Numerical solutions of boundary value problems for ordinary differential equations (A.K. Aziz, ed.). New York: Academic Press 1975
19. Pereyra, V.L.: Iterated deferred corrections for nonlinear boundary value problems. Numer. Math. **11**, 111–125 (1968)
20. Stetter, H.J.: Global error estimation in Adams PC-codes. TOMS (to appear)

Received May 3, 1977