Numerische Mathematik 8, 376—391 (1966)

On Improving an Approximate Solution of a Functional Equation by Deferred Corrections*

VICTOR PEREYRA

Received October 7, 1965

Abstract. The improvement of discretization algorithms for the approximate solution of nonlinear functional equations is considered. Extensions to the method of difference corrections by FOX are discussed and some general results are proved. Applications to nonlinear boundary problems and numerical examples are given in some detail.

Introduction

Many problems of mathematical physics and applied analysis are particular instances of the problem of solving the functional equation

$$F(v) = f, \tag{1}$$

with v and f belonging to appropriate general spaces and F being an operator between these spaces.

Most of the time this equation cannot be solved in a closed form, and some approximations are required. We are interested in considering the case in which (1) is replaced by an associate, simpler problem depending on a real (small) parameter h:

$$\Phi_h(V) = g. \tag{2}$$

In this general formulation many problems can be dealt with. For example, KANTOROVICH [1948] has proved, under suitable hypothesis, several important relationships between the solutions of (1) and (2) when F is linear.

More recently, STETTER [1965] has discussed the asymptotic behavior (for $h \to 0$) of the error of discretization $e = V - \Delta_h v$, obtaining very general results and showing several applications. His main aim was to have a rigorous basis for the application of Richardson's extrapolation to the limit.

In this paper we intend to analyze in general another method for accelerating the convergence of certain approximate processes. In different contexts this method has been known as "the difference (or deferred) correction method" (FOX and GOODWIN [1949]).

In Section 1 we state the problem and give the notation and some definitions.

In Section 2 a special form of a theorem by STETTER is proved, stating the existence of an asymptotic expansion for the error of discretization.

In Section 3 a linear deferred correction procedure is presented. In Theorem 3.1 we prove that in fact this procedure will produce an improved approximate solution.

* Reproduction in Whole or in Part is Permitted for any Purpose of the United States Government. This report was supported in part by Office of Naval Research Contract Nonr-225(37) (NR-044-211) at Stanford University.

Section 4 is devoted to the analysis of an iterative deferred correction.

Sections 5 and 6 provide means of proving some of the hypotheses used in the former sections when some other hypotheses hold.

In Section 7 two applications are discussed. Finally in Section 8 some numerical results are presented.

1. Statement of the Problem and Definitions

We will consider the functional equation

$$F(v) = f. \tag{1.1}$$

The (generally non-linear) operator F will map a linear subspace of a Banach space D into a Banach space E. Problem (1.1) will always be assumed to have a unique solution $u \in D^{\dagger} \subset D$.

We are interested in accelerating the convergence of approximate methods for solving (1.1).

Let us consider now a *discretized* version of (1.1)

$$\Phi_h(V) = \Delta_h^0 f. \tag{1.2}$$

The operators Φ_h will map certain Banach spaces D_h into Banach spaces E_h, while Δ_h, Δ_h^0 will be bounded linear transformations from D into D_h and from E into E_h respectively. The possible values for the real parameter h will be: $h_0 > h_1 \ldots > 0$, $\{h_i\} = H$ being a vanishing sequence given in advance ($h_i \searrow 0$). In what follows the norms of the different spaces will appear, and to avoid cumbersome notation we will use only the symbol $\|\cdot\|$ whenever this is not confusing.

The operator in (1.2) will be assumed to have the following properties:

For each $v \in D^{\dagger}$ and $h \in H$ there exists an expansion [1]

$$\Phi_h(\Delta_h v) = \Delta_h^0 \left\{ F(v) + \sum_{j=1}^{N} h^{p_j} F_{p_j} v \right\} + O(h^{\bar{p}}), \tag{1.3}$$

where the operators F_{p_j} are given and $p_N < \bar{p}$.

The exponents appearing in (1.3) will be positive rational numbers satisfying

$$0 < p_1 < p_2 < \cdots < p_N. \tag{1.4}$$

The operators F and Φ_h will always be assumed to be at least twice Fréchet-differentiable on D and D_h respectively.

Definition 1. *If for any $e \in D_h$, a fixed $V \in D_h$ and any $h \in H$ there exists a non-negative constant K (which may depend on V) such that*

$$\|e\| \leq K \|\Phi_h'(V)\, e\|, \tag{1.5}$$

[1] The symbolic expression,

$$y(h) = O(h^k), \quad y(h) \in \text{Banach space (possibly different for each } h)$$

has the meaning,

$$\overline{\lim_{h \to 0}}\, h^{-k} \|y(h)\| \leq K, \quad K \text{ constant}.$$

then we will say that the operator $\Phi_h(V)$ is stable at V. Observe that this is equivalent to say that if $\Phi_h'(V)$ is onto, then it has an inverse and $\|[\Phi_h'(V)]^{-1}\| \leqq K$.

Definition 2. *Let q be a positive number. We will say that $U(h) \in D_h$ is an approximate solution of (1.2) if it satisfies*

$$\|\Phi_h(U(h)) - \Delta_h^0 f\| \leqq C \cdot h^q, \tag{1.6}$$

where C is a positive constant.

Whenever this is not confusing we will not mention in $U(h)$ the specific dependence on h.

Definition 3. *If $h \in H$, u is the solution of (1.1), and $U(h)$ is an approximate solution of (1.2), then the vector*

$$e(h) = U(h) - \Delta_h u \in D_h \tag{1.7}$$

will be called the global discretization error (g.d.e.) *of* (1.2).

Definition 4. *The method (1.2) having an asymptotic expansion (1.3) will be convergent of order ζ if for any $h \in H$, $\|e(h)\| \leqq C h^\zeta$, where C is a positive constant.*

In this case $U(h)$ will also be called a ζ-approximate solution of (1.1).

Definition 5. *g.d.e. admits an asymptotic expansion up to the order $p_N > 0$ if there exist $e_j \in D$, independent of $h \in H$, such that*

$$\left\| e(h) - \Delta_h \sum_{j=1}^{N} h^{p_j} e_j \right\| \leqq C_N h^{\bar{p}} \tag{1.8}$$

with $C_N > 0$ constant, and $\bar{p} > p_N$.

2. Existence of an Asymptotic Expansion for g.d.e.

In STETTER [1965] it is proved that under certain conditions g.d.e. has an expansion of the form (1.8) if Φ_h has an expansion like (1.3).

We will present a simpler proof for the case in which only the first term of such an expansion is needed.

Until something else is said we will assume that the operator $\Phi_h(V)$ has the property

(α) $\quad \|\Phi_h''(V)\| < 2K_2$ for $V \in D_h$,

and also that

(β) $\quad F'(u)\, e = b$ has a unique solution $e \in D$ for any $b \in E$.

Now we can state

Theorem 2.1. *Let $U(h)$ be a p_1-approximate solution with $q \geqq p_2$, and let u be the exact solution of (1.1). Let Φ_h and Φ_h' have an asymptotic expansion (1.3) up to the order p_2, with F_{p_1} independent of h. If Φ_h is stable at $\Delta_h u$, then $e(h)$ admits an asymptotic expansion up to the order p_1, where e_1 is independent of h and satisfies*

$$F'(u)\, e_1 = -F_{p_1} u. \tag{2.1}$$

Proof. On one hand we have that, by the general Taylor-expansion and (α):

$$\|\Phi_h(U) - \Phi_h(\Delta_h u) - \Phi_h'(\Delta_h u)\, e(h)\| \leqq K_2 \cdot \|e(h)\|^2. \tag{2.2}$$

Also, by using (1.3),

$$\Phi_h(U) - \Phi_h(\Delta_h u) = \Phi_h(U) - \Delta_h^0[F(u) + h^{p_1} F_{p_1} u] + O(h^{p_2}). \tag{2.3}$$

Hence, since $\Phi_h(U) - \Delta_h^0 f = O(h^q)$ and $F(u) = f$ we obtain,

$$\Phi_h(U) - \Phi_h(\Delta_h u) = -h^{p_1} \Delta_h^0(F_{p_1} u) + O(h^{p_2}). \tag{2.4}$$

Combining (2.2) and (2.4),

$$\Phi_h'(\Delta_h u)\,\bar{e}(h) = -\Delta_h^0(F_{p_1} u) + O(h^{p^*}), \tag{2.5}$$

where $\bar{e}(h) = h^{-p_1} e(h)$ and $p^* = \min(p_1, p_2 - p_1)$.

Consider now the equation

$$F'(u)\, e = -F_{p_1} u, \tag{2.6}$$

which by (β) has a unique solution $e_1 \in D$.

We have, by our hypotheses and combination of (2.5) with (2.6),

$$\Phi_h'(\Delta_h u)[\bar{e}(h) - \Delta_h e_1] = O(h^{p^*}).$$

Using the stability, it is found that

$$\bar{e}(h) - \Delta_h e_1 = O(h^{p^*}), \tag{2.7}$$

or in other words

$$\|e(h) - \Delta_h h^{p_1} e_1\| \leqq C h^{\bar{p}} \tag{2.8}$$

with $\bar{p} = p^* + p_1 > p_1$.

3. Accelerating the Convergence

A Deferred Correction

Once the expansion (2.8) has been secured, several procedures are available in order to obtain a more accurate approximation than $U(h)$.

A well known one is Richardson's extrapolation to the limit (STETTER [1965]). This procedure requires the solution of (1.2) for two different values of h, say h_1, h_2, $(h_1 > h_2)$, and a suitable combination of these two solutions permits the term $h^{p_1} e_1$ to be eliminated giving an approximation of order $h_1^{\bar{p}}$.

We intend now to describe a different approach aimed at serving the same purpose. The underlying idea can be traced back to FOX and GOODWIN [1949] and in its present restricted form to BICKLEY, MICHAELSON and OSBORNE [1960], VOLKOV [1957], and HENRICI [1962]. It essentially consists in calculating an approximate value of e_1 by using the already computed solution (of order h^{p_1}), U. We will now assume the existence of an operator S_{p_1} such that

$$\Delta_h^0 F_{p_1} u - S_{p_1}(U) = O(h^{p^*}) \tag{3.1}$$

and range $(S_{p_1}) \subset \operatorname{range}(\Phi_h')$.

Theorem 3.1. *Under the hypotheses of Theorem 2.1, and if (3.1) is valid, then*

$$U_1 = U - h^{p_1} e^* \tag{3.2}$$

is an approximate solution of (1.1) of order $\bar{p}$, that is,

$$\|U_1 - \Delta_h u\| = O(h^{\bar{p}}).$$

Here e^* is the solution of

$$\Phi_h'(U)\, e^* = -S_{p_1}(U). \tag{3.3}$$

Proof. Calling $\eta = \bar{e}(h) - e^*$ and subtracting (3.3) from (2.5), we obtain

$$\Phi_h'(\Delta_h u)\, \eta = O(h^{p^*}), \tag{3.4}$$

since

$$\Phi_h'(\Delta_h u)\, \bar{e}(h) - \Phi_h'(U)\, e^* = \Phi_h'(\Delta_h u)\, \eta + O(h^{\bar{p}})$$

and (3.1).

From this and the stability

$$\|\eta\| = O(h^{p^*}),$$

or

$$\|U - h^{p_1} e^* - \Delta_h u\| = O(h^{\bar{p}}).$$

Remarks. i) A procedure similar to this has been empirically employed for many different problems (mainly linear) in differential and integral equations (see FOX [1962] and NOBLE [1964]).

ii) From Theorem 3.1 it is clear that this procedure will give an approximate solution with the same order of accuracy as the Richardson extrapolation technique but with much less work. In fact, both procedures consist of essentially two steps. The first step, common to both, is the computation of a p_1-approximate solution U.

For Richardson's extrapolation one has now to compute another solution with a smaller parameter, say $h/2$. That involves solving once again the non-linear problem (1.2), which in general will be of a "larger size" than the one corresponding to the original h. On the other hand, for computing the deferred correction $h^{p_1} e^*$, one only has to solve *the linear* problem (3.3) (for the *same* h). The only drawback that we can point out is the computation of $\Phi_h'(U)$ and $S_{p_1} U$, which are necessary for the deferred correction but not for Richardson's extrapolation.

With respect to $\Phi_h'(U)$ we can say that, if Newton's method is being used for solving problem (1.2), then it will already be available from the first part of the computation.

The extra computation (and derivation) of $S_{p_1}(U)$ is rapidly compensated for in problems wherein decreasing h by a factor α increases the size of the problem by a power $n \geqq 2$ of $1/\alpha$ (for instance, in elliptic partial differential equations).

iii) The purpose of introducing the notion of an approximate solution of (1.2) (Definition 2) will appear clearly in the applications. At this point we can say that since the exact solution of (1.2) will only be an approximation to the solution of (1.1), it is of no use to solve (1.2) exactly. Moreover, the condition $q \geqq p_1$ says how incomplete this solution can be. This idea is present in the works of DOUGLAS [1961] and HENRICI [1962], in connection with the solution of mildly non-linear elliptic equations and two-point boundary value problems respectively, and it has a very important practical value since only "incomplete" solutions are computationally available.

The expansion in Theorem (2.1) will generally be needed for the construction of the operator S_{p_1}.

4. Iterated Deferred Corrections

In Section 3 we developed a way of eliminating the first term in the expansion of g.d.e. It had the advantages of involving only the solution of a linear problem and preserving the "size" of the main equation. If a longer expansion is available one can ask whether it is possible to eliminate further terms in a similar fashion. This seems to be unlikely, since in eliminating the first term we have thrown information away which cannot be regained within a linear process.

However, if we are willing to loose the linearity, then it is possible to describe an iterative procedure which will give an asymptotically more accurate solution for each (non-linear) step.

This procedure will have an advantage over the successive Richardson's extrapolation in that the parameter h will not have to be changed.

It will also have a disadvantage in that for each special problem some expressions will have to be worked out. In order to carry out the proof of the next theorem we have to introduce a new definition.

Definition 6. *A differentiable operator $\Phi_h(V)$ is said to have the mean value property* (m.v.p.) *if for each $V_1, V_2 \in D_h$, there exists a linear operator $M(V_1, V_2)$ such that*

$$\Phi_h(V_1) - \Phi_h(V_2) = M(V_1, V_2)(V_1 - V_2),$$

and

$$\|M(V_1, V_2) - \Phi'(V)\| = o(1)$$

for $V_1, V_2 \to V$.

Theorem 4.1. *Let Φ_h have m.v.p. with $M(V_1, V_2)$ non-singular. Let us assume also that for $V_1, V_2 \in D_h$, $\|[M(V_1, V_2)]^{-1}\| < K$, where K does not depend on V_1, V_2. If an expansion* (1.3) *is valid, $U^{(k)}$ $(k<N)$ is a p_k-approximate solution of* (1.1), *and there exist operators $S_{p_j}^{(k)}$ such that*

$$h^{p_j} \Delta_h^0 F_{p_j} u - S_{p_j}^{(k)}(U^{(k)}) = O(h^{p_{k+1}}) \qquad (j = 1, \ldots, k), \tag{4.1}$$

then the solution of the equation

$$\Phi_h(U^{(k+1)}) = \Delta_h^0 f + \sum_{j=1}^{k} S_{p_j}^{(k)}(U^{(k)}) \tag{4.2}$$

is a p_{k+1}-approximate solution of (1.1).

Proof. By (1.3) we have that

$$\Phi_h(\Delta_h u) = \Delta_h^0 \left[f + \sum_{j=1}^{k} h^{p_j} F_{p_j} u\right] + O(h^{p_{k+1}}). \tag{4.3}$$

Hence, by subtracting (4.2) from (4.3), we obtain

$$M e^{(k+1)} = O(h^{p_{k+1}}), \tag{4.4}$$

and the hypothesis on M implies

$$e^{(k+1)} = \Delta_h u - U^{(k+1)} = O(h^{p_{k+1}}).$$

With this result it is simple now to describe *the iterated deferred corrections procedure* (i.d.c.p.). In order to obtain approximate solutions of increasing accuracy we will proceed as follows.

i) Obtain $U^{(1)}$ by solving

$$\|\Phi_h(U^{(1)}) - \Delta_h^0 f\| \leqq K_1 h^{q_1} \tag{4.5}$$

with $q_1 \geqq p_1$, $K_1 \geqq 0$.

ii) For $k = 1, 2, \ldots, N-1$ compute $U^{(k+1)}$ satisfying

$$\left\|\Phi_h(U^{(k+1)}) - \left[\Delta_h^0 f + \sum_{j=1}^{k} S_{p_j}^{(k)}(U^{(k)})\right]\right\| \leqq K_{k+1} h^{q_{k+1}}, \tag{4.6}$$

where $q_{k+1} \geqq p_{k+1}$, $K_{k+1} \geqq 0$. Then, and from Theorem 4.1, $U^{(k+1)}$ will satisfy

$$\|\Delta_h u - U^{(k+1)}\| \leqq C \cdot h^{p_{k+1}}. \tag{4.7}$$

On writing (4.5) and (4.6) we have shown explicitly how "incomplete" the approximate solutions $U^{(k)}$ can be.

5. Operators of Monotonic Type

We will now assume that the spaces E_h, E_h^0 introduced in Section 1 have also a compatible structure of Archimedean lattices.

Thus it makes sense to consider operators Φ_h whose Fréchet derivatives are of monotonic type (SCHRÖDER [1961]).

In this case Φ_h' will be non-singular and $[\Phi_h']^{-1}$ will be isotone.

In the usual way, for any $V \in E_h$ (or E_h^0) we have a notion of absolute value $|V| \in E_h$ (or E_h^0): $0 \leqq |V| = \sup(V, -V)$.

In many practical problems of monotonic type one has information not on $\Phi_h'(V)$ but on a certain (generally simpler, and independent of V) linear operator, say B_h.

The following Theorem gives a sufficient condition for the operator Φ_h to be stable, in terms of assumed bounds for $\|B_h^{-1}\|$.

Theorem 5.1. *If $\Phi_h'(V)$, B_h and $(\Phi_h'(V) - B_h)^{-1}$ are operators of monotonic type, and*

$$\|B_h^{-1}\| \leqq K \tag{5.1}$$

then Φ_h is stable.

Proof. It is enough to prove that for any $w \geqq 0$,

$$[\Phi_h']^{-1} w \leqq B_h^{-1} w. \tag{5.2}$$

In fact,

$$(B_h^{-1} - [\Phi_h']^{-1})^{-1} = \Phi_h'(\Phi_h' - B_h)^{-1} B_h. \tag{5.3}$$

Since the product of inverse-monotonic operators is inverse-monotonic, it follows that $B_h^{-1} - [\Phi_h']^{-1}$ is isotone and from the definition of isotonicity we obtain that, $0 \leqq (B_h^{-1} - [\Phi_h']^{-1})\, w$ and (5.3) follows. We can write now,

$$e = [\Phi_h']^{-1}(\Phi_h' e),$$

and from this

$$\|e\| \leqq \|B_h^{-1} |\Phi_h' e|\,\| \leqq K \|\Phi_h' e\|;$$

and, according to Definition 1, Φ_h is stable.

6. A Sufficient Condition for Convergence

In many cases convergence can be proved from the properties of Section 2 if some additional hypothesis is valid. For instance,

Theorem 6.1. *Let u be the solution of* (1.1) *and $U(h)$ an approximate solution of* (1.2) *with exponent $q \geqq p_2$. If an expansion* (1.3) *with $N=2$ is valid, Φ_h has the mean value property and the $M(U, \Delta_h u)$ of Definition 6 has an inverse bounded in norm, then the method* (1.2) *is convergent of order p_1.*

Proof. If instead of using the development (2.2) we use the mean value property, then we obtain instead of (2.4) the equations

$$M\,e(h) = O(h^{p_1}), \tag{6.1}$$

and if $K \geqq \|M^{-1}\|$, then it follows that

$$\|e(h)\| \leqq K h^{p_1}. \tag{6.2}$$

Observation. The point in using m.v.p. is that it eliminates the term in $\|e(h)\|^2$ from the discussion. Observe that m.v.p. is not an automatic property for arbitrary, (let us say) twice Fréchet differentiable non-linear operators.

7. Applications

We will consider now several applications of the methods described in the preceding Sections.

7.1. Two-Point Boundary Value Problem

We want to solve the problem

$$\begin{aligned} F^1(y) &= -y''(x) + f\big(x, y(x), y'(x)\big) = 0 \quad \text{for} \quad x \in [a, b], \\ F^0(y) &= y(a) - \alpha = 0, \\ F^2(y) &= y(b) - \beta = 0. \end{aligned} \tag{7.1}$$

In Pereyra [1965] the author has considered a simpler instance of (7.1) for which $y'(x)$ was not present in the differential equation. We would like to discuss this problem at length, since we feel that all the important features of d.c.p. can be displayed here in an environment of median technical difficulty, not so complicated as to obscure the issues and not too simple as to be trivial. Furthermore, besides the paper mentioned above, we do not know of any publication in which this application has been discussed rigorously and in detail.

In order to insure existence and uniqueness of a solution of (7.1) we will assume that

$$f_y(x, y, z) \geqq 0, \qquad |f_z(x, y, z)| < K \tag{7.2}$$

in a certain bounded region $\Omega = [a, b] \times B \times B'$. Let us call that solution $y(x)$.

The different spaces are $D = C[a, b]$, the space of continuous functions on $[a, b]$, and $E = C[a, b] \times R^2 \cdot F(y)$ will be defined on $C^2[a, b]$.

In order to define a discretization of (7.1), we introduce the vanishing sequence $H = \{h_i\}$, where $h_i = \dfrac{b-a}{q_i}$ (q_i positive integers, $q_i \to \infty$ for $i \to \infty$), and $h_0 < 2/K$.

Then $D_{h_i}=E_{h_i}=R^{q_i+1}$ the q_i+1-dimensional real space. The link between the spaces corresponding to the continuous and discrete cases will be provided by the operators

$$\Delta_{h_i}:\ y(x)\to\{y(x_j)\}\ \text{ with }\ x_j=a+jh_i, \qquad j=0,1,\ldots,q_i,$$
$$\Delta^{01}_{h_i}:\ y(x)\to\{y(x_j)\}_{j=1,\ldots,q_i-1},$$
$$\Delta^{00}_{h_i}:\ y(x)\to y(a), \qquad \Delta^{02}_{h_i}:\ y(x)\to y(b).$$

In what follows we will use p to symbolize a fixed, but otherwise arbitrary q_i.

The norms involved will be the L_∞ norms for vectors and matrices.

We can now define a discrete version of (7.1):

$$\begin{aligned}
&[\Phi_h(V)]_j = h^{-2}(-V_{j-1}+2V_j-V_{j+1})+f(x_j,V_j,(V_{j+1}-V_{j-1})/2h)=0,\\
&\qquad\qquad j=1,\ldots,p-1,\\
&[\Phi_h(V)]_0 = V_0-\alpha=0,\\
&[\Phi_h(V)]_p = V_p-\beta=0,
\end{aligned} \tag{7.3}$$

where $H\ni h=(b-a)/p$.

The Fréchet derivative of Φ_h is

$$\begin{aligned}
&\{\Phi_h'(V)\,e\}_j = h^{-2}\Big\{-\Big(1+\frac{h}{2}f_z^j(V)\Big)e_{j-1}+\big(2+h^2f_y^j(V)\big)e_j-\\
&\qquad -\Big(1-\frac{h}{2}f_z^j(V)\Big)e_{j+1}\Big\} \qquad j=1,\ldots,p-1,\\
&\{\Phi_h'(V)\,e\}_0 = e_0,\\
&\{\Phi_h'(V)\,e\}_p = e_p.
\end{aligned} \tag{7.4}$$

The notation for the partial derivatives of f in (7.4) is, $f_z^j(V)=f_z(x_j,V_j,(V_{j+1}-V_{j-1})/2h)$, and so on. For $v\in C^{2N+3}[a,b]=D^\dagger$ we have the expansions

$$\begin{aligned}
&\Phi_h^1(\Delta_h v)=\Delta_h^{01}\Big\{F^1(v)+\sum_{j=1}^{N}h^{2j}\Big[\frac{(-2)}{(2j+2)!}v^{(2j+2)}+\\
&\qquad +g_{2j}(v,v''',\ldots,f_z^{(\nu)})\Big]\Big\}+O(h^{2N+1}),\\
&\Phi_h^0(\Delta_h v)=\Delta_h^{00}F^0(v), \qquad \Phi_h^2(\Delta_h v)=\Delta_h^{02}F^2(v).
\end{aligned} \tag{7.5}$$

The functions g_{2j} can be computed by reordering

$$\sum_{\nu=1}^{N}\frac{f_z^{(\nu)}(v)}{\nu!}\Big[\sum_{j=1}^{N}\frac{h^{2j}}{(2j+1)!}v^{(2j+1)}\Big]^{\nu}=h^2\Big(f_z'\frac{v^{(3)}}{3!}\Big)+h^4\Big(f_z'\frac{v^{(5)}}{5!}+f_z^{(2)}\frac{v^{(3)}}{2\cdot 3!}\Big)+h^6\ldots.$$

Since $\Phi_h(V)$ clearly has the mean value property (with $[M(V_1,V_2)]_j=[\Phi_h'(\widetilde{V}_j)]_j$, the subindex meaning the j-th rows of the corresponding matrices, where $\widetilde{V}_j$ are different intermediary points for each row), if we are able to show that it is also stable then, by Theorem 6.1, we will have that it is convergent of order 2.

We will next show that $\Phi_h'(V)$ is of monotonic type. The order considered is componentwise in R^p.

Lemma 7.1. *The operator $\Phi_h'(V)$ of (7.4) is of monotonic type for any $V \in D_h$.*

Proof. Let $e \in E_h$, and assume that

$$\Phi_h'(V)\, e \geqq 0 .$$

We want to prove that $e \geqq 0$. Suppose that for some j, $0<j<p$, $e_j \leqq e_i$ for all $0<i<p$ and $e_j<0$.

But

$$\{\Phi_h'(V)\, e\}_j = -\lambda_1 e_{j-1} + \tfrac{1}{2}[2 + h^2 f_y^j(V)]\, e_j - \lambda_2 e_{j+1} \geqq 0 ,$$

or

$$0 > e_j \geqq 2 \cdot (\lambda_1 e_{j-1} + \lambda_2 e_{j+1})/[2 + h^2 f_y^j(V)] ,$$

with

$$\lambda_1 + \lambda_2 = 1 , \qquad \lambda_1, \lambda_2 > 0 .$$

Hence, we obtain that

$$e_j \geqq \min(e_{j-1}, e_{j+1}) ,$$

and from this follows that $e_j = e_{j-1} = e_{j+1}$.

Repeating this argument we will finally obtain that for all $0 \leqq i \leqq p$, $e_i = e_j < 0$. But this is a contradiction.

From Lemma 7.1 we have in particular that $\Phi_h'(V)$ is invertible and that $[\Phi_h'(V)]^{-1}$ has positive elements.

Corollary 1. *If e is the solution of*

$$[\Phi_h'(V)\, e]_j = 0, \qquad j = 1, \ldots, p-1 ,$$

$$e_0 = \alpha , \qquad e_p = \beta$$

then

$$\min(0, \min(\alpha, \beta)) \leqq e_j \leqq \max(0, \max(\alpha, \beta)) . \tag{7.6}$$

Proof. It is enough to observe that, for $1 \leqq j \leqq p-1$,

$$e_j \leqq \max(e_{j-1}, e_{j+1}) \quad \text{if} \quad e_j \geqq 0$$

or

$$e_j \geqq \min(e_{j-1}, e_{j+1}) \quad \text{if} \quad e_j \leqq 0 ,$$

and a reasoning similar to the one used in Lemma 7.1 shows that e_j cannot have either a positive maximum or a negative minimum at an interior point (maximum principle). Observe that both Lemma 7.1 and its corollary are also valid for the operator $M(V_1, V_2)$, since the arguments of $f_y^j(v)$ did not play any role in the proof, the only relevant property being the positivity of $f_y(V)$.

Theorem 7.1. *$\Phi_h(V)$ is stable on D (uniformly in h and V). Moreover, the operators $M(V_1, V_2)$ are non-singular and their inverses are uniformly bounded in norm.*

Proof. If in Theorem 5.1 we take $M(V_1, V_2)$ as $\Phi_h'(V)$ and $\operatorname{diag}(f_y^j(\widetilde{V}_j))$ as B_h then it is clear that these operators and $\Theta(\widetilde{V}) = M(V_1, V_2) - \operatorname{diag}(f_y^j(\widetilde{V}))$ satisfy the hypothesis of the theorem and, consequently $M(V_1, V_2)$ will have an inverse bounded in norm if Θ has this property. Since this will be valid for arbitrary arguments $\widetilde{V}_j$ in M, let $\widetilde{V}_j = V$ for $j = 0, \ldots, p$. In this case $M \equiv \Phi_h'(V)$ and we have from the above result that Φ_h is stable. That these properties are uniform will stem from the fact that the bounds on $\|\Theta^{-1}\|$ will be independent of $\tilde{V}_j$ and h.

Hence we will show that $\|\Theta^{-1}\|$ is bounded. Let $\varphi(x)\in C^2[a,b]$ be a negative function for which

$$\Theta(\tilde{V})(\Delta_h\varphi)<0, \quad \text{and} \quad \frac{\max(|\varphi(a)|,|\varphi(b)|)}{\min_{x\in[a,b]}|\Theta(\tilde{V})(\Delta_h\varphi(x))|}\leq C(K,(b-a)). \tag{7.7}$$

We have that

$$0\leq[\Delta_h\varphi]_i=\sum_{s=1}^{p-1}[\Theta^{-1}]_{is}[\Theta(\tilde{V})(\Delta_h\varphi)]_s+[\Theta^{-1}]_{i0}|\varphi(a)|+[\Theta^{-1}]_{ip}|\varphi(b)|,$$

and from (7.7) we obtain the inequality

$$\sum_{s=1}^{p-1}[\Theta^{-1}]_{is}|[\Theta(\Delta_h\varphi)]_s|\leq\max(|\varphi(a)|,|\varphi(b)|), \tag{7.8}$$

since $\max([\Theta^{-1}]_{i0},[\Theta^{-1}]_{ip})\leq 1$. Finally we obtain from (7.7) and (7.8) that

$$\max_i\sum_{s=1}^{p}|[\Theta^{-1}]_{is}|\leq C(K,b-a)$$

or, in other words, $\|\Theta^{-1}\|_\infty\leq C$, where C only depends on the bound K of the partial derivative f_z and on the length of the interval $[a,b]$. That a function $\varphi(x)$ with properties (7.7) exists is shown in detail in BERS [1953], Section 3. Theorem 7.1 has been proved.

In conclusion, Φ_h is stable and convergent of order 2, and we can apply any of the deferred correction algorithms of Sections 3 and 4. We will now assume that $f(x,y,y')$ is sufficiently differentiable as a function of its three arguments, which in particular will imply that the solution $y(x)$ of (7.1) has continuous derivatives up to the order necessary in the following discussion.

For the linear, one-step correction, which will give a fourth order approximate solution, we will develop some special formulas in order to approximate $y'''(x)$ and $y^{\mathrm{IV}}(x)$ at the interior points. Since $p^*=\min(p_1,p_2-p_1)=2$ we need these approximations to be of order h^2 (see (3.1)).

Lemma 7.2. *Let U be an approximate solution of (7.3) with $q\geq 2$. The expressions*

$$\begin{aligned}\delta f_j=\delta f(x_j,U_j,(U_{j+1}-U_{j-1})/2h)&=f(x_{j+1},U_{j+1},(U_{j+2}-U_j)/2h)-\\&-f(x_{j-1},U_{j-1},(U_j-U_{j-2})/2h)\end{aligned}$$

and the similarly defined $\delta^2 f_j$ satisfy

$$\begin{aligned}y'''(x_j)-(\delta f_j)/2h&=O(h^2),\\ y^{\mathrm{IV}}(x_j)-(\delta^2 f_j)/h^2&=O(h^2)\end{aligned} \tag{7.9}$$

for $j=2,\ldots,p-2$. (A non symmetric formula has to be used at $j=1$, $n-1$.)

Proof. We will prove (7.9) for δf_j and exactly the same argument can be used in order to obtain the second formula. First of all we recall that

i) $y''(x)=f(x,y(x),y'(x))$,

ii) $e_j=U_j-y(x_j)=O(h^2)$,

iii) $e''(x)=f_y e(x)+f_{y'}e'(x)+[\frac{1}{12}y^{\mathrm{IV}}(x)-\frac{1}{6}f_{y'}y'''(x)]$,

and finally

iv) $e_j = h^2 e(x_j) + O(h^4)$.

Also from (i) we have that

v) $\delta_x y''(x) = \delta_x f(x, y(x), y'(x)) = 2h y'''(x) + O(h^3)$.

Hence, it is enough to show that δf_j approaches $\delta_x f(x)$ as $O(h^3)$. In fact,

$$\delta[f(x, y(x), y'(x)) - f(x, U, \delta U/2h)] = h^2 \delta[f_y e(x) + f_{y'}(\delta e(x)/2h + y'''(x))] + O(h^4)$$

and

$$\delta[f(x) - f_j]/2h = h^2 \frac{d}{dx}[f_y \cdot e(x) + f_{y'}(e'(x) + y'''(x))] + O(h^4),$$

and from the differentiability properties of all the involved functions we have that

$$\delta f(x) - \delta f_j = O(h^3),$$

and Lemma 7.2 is proved.

From this lemma we can now define

$$[S_2(U)]_j = \frac{h^{-2}}{12} \delta^2 f_j - \frac{h^{-1}}{12} f^j_{y'}(U)\, \delta f_j, \tag{7.10}$$

which satisfies condition (3.1) since also $f^j_{y'}(U) - f_{y'}(y(x_j)) = O(h^2)$. Hence, by Theorem 3.1 we can obtain by solving (3.3) and using (3.2) an approximate solution U_1 of order 4.

For the iterated deferred corrections, besides the increased differentiability requirements, it is necessary to define the operators $S^{(k)}_{p_i}$ of (4.1).

As before, the approximations to the different derivatives will be in terms of differences, either of the successive approximate solutions $U^{(k)}$ or of the values of the right hand side $f(x, y, y')$ at this $U^{(k)}$. Formulas (7.5) and (4.1) show that at the k-th step we need to approximate the quantities $h^{2j} u^{(2j+1)}(x)$ and $h^{2j} u^{(2j+2)}(x)$ $(j \leq k)$ up to the order $2k+2$ in h.

In Volkov [1957], § 2 we find a general set up of formulae for numerical differentiation. They allow us to approximate derivatives of any order with all the accuracy we may need and with quite a general distribution of tabular points. The detailed discussion of the practical application of these formulae to nonlinear boundary value problems will be carried out elsewhere. Let us only remark that it is clear from Volkov's formulae and our requirements that the use of central differences will soon require points outside of the interval $[a, b]$. Even if unsymmetric differences are used, care would have to be taken in order to have enough points, especially if several iteration are planned. In fact, the maximum number of iterations desired and the set of formulae chosen for approximating the derivatives at the different points will impose a new restriction on the largest step h_0 (minimum number of points) which can be allowed.

7.2. Mildly Nonlinear Partial Differential Equations of Elliptic Type

Here the continuous problem is

$$\begin{aligned} F^1(z) &= \Delta z(x, y) - f(x, y, z, z_x, z_y) = 0 \quad \text{for} \quad (x, y) \in D, \\ F^2(z) &= z(x, y) - g(x, y) = 0 \quad \text{for} \quad (x, y) \in \partial D, \end{aligned} \tag{7.11}$$

where $g(x, y)$ is a given function. Let $\{V_j\}$ be the nodal points of a square mesh of width h which covers $\bar{D}$.

The discretization will be given as usual (FORSYTHE and WASOW [1960]) by the system of difference equations

$$[\Phi_h(V)]_j = h^{-2}(4V_j - V_j^N - V_j^S - V_j^E - V_j^W) + \\ + f(x, y, V_j, (V_j^E - V_j^W)/2h, \quad (V^N - V_j^S)/2h) = 0 \tag{7.12}$$

for each V_j which with its four closest neighbors is contained in $\bar{D}$. This problem has been discussed in BERS [1953], and all the necessary properties and conditions can be obtained from there. In VOLKOV [1957] an iterated deferred corrections algorithm for $\Delta u = f(x, y)$ is discussed. The treatment of a general boundary, needing interpolation, can be taken from there. In order to make our description simpler we will assume that the boundary ∂D is such that it does not require interpolation, for any $h \in H$, or in other words that all the interior grid points V_j are regular. We will also assume that the given functions $g(x, y)$ and $f(x, y, z, z_x, z_y)$ have enough regularity properties and that the solution $z(x, y)$ of (7.11) is sufficiently differentiable. In this case the boundary equations will simply be

$$[\Phi_h(V)]_j = V_j - g(x, y). \tag{7.12'}$$

Both in formulas (7.12) and (7.12') the (x, y) represents the node in D corresponding to V_j.

With these hypotheses the treatment parallels the one of Section 7.1 and we will not repeat it here. $F_2 = \frac{1}{12}\left(\frac{\partial^4 z}{\partial x^4} + \frac{\partial^4 z}{\partial y^4}\right) + \frac{1}{6}\left(f_{z_x}\frac{\partial^3 z}{\partial x^3} + f_{z_y}\frac{\partial^3 z}{\partial y^3}\right)$ needs to be approximated in order to compute the linear deferred correction.

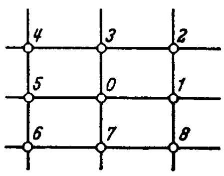

Fig. 1

If, instead of using the approximate solution V, one uses the values of the right hand side at V, then it is possible to approximate F_2 using only second differences in every given direction. This will reduce the complexity of the problem at points close to the boundary.

In fact, for any sufficiently differentiable function $z(x, y)$

$$12F_2 = \frac{\partial^4 z}{\partial x^4} + \frac{\partial^4 z}{\partial y^4} = \Delta^2 z(x, y) - 2\frac{\partial^4 z}{\partial x^2 \partial y^2}. \tag{7.13}$$

In turn, if $V_j - z(x, y) = O(h^2)$, then

$$12F_2 = \Delta_h f(x, y, z, \delta_x z/2h, \delta_y z/2h) - 2S(z) + O(h^2), \tag{7.14}$$

where $S(z)$ is defined by

$$S(z_0) = h^{-4}(4z_0 - 2(z_1 + z_3 + z_5 + z_7) + z_2 + z_4 + z_6 + z_8),$$

with z_i defined in Fig. 1.

Formula (7.14) would be used every time that at least one z_i $(i = 1, \ldots, 8)$ belongs to the boundary ∂D.

When solving the system of nonlinear Eqs. (7.12) (or any similar discretization) some iterative technique will be needed. Let us suppose that Newton's method can be applied successfully. At each stage of this outer iteration it will be necessary to solve a large system of linear equations.

In order to do so, generally, an iterative technique will also be used. To minimize the amount of work in this inner iteration it is important that the matrix which represents the Fréchet derivative Φ_h' be as simple as possible.

If a fourth order approximation in h is desired then, broadly speaking, we have available three different kinds of techniques.

i) We may use a fourth order discretization instead of the one of second order in (7.12). For instance, one may use the 9-point approximation given in BRAMBLE and HUBBARD [1962].

ii) We may solve (7.12) twice with different steps and then use Richardson's extrapolation to the limit.

iii) We may solve (7.12) once and then perform a linear deferred correction as described above.

Let us suppose, as an example, that D is a square with sides of length one.

The most significant figure here is the number of nodal points, which in turn gives the number of equations involved. Suppose h is such that we have 100 equations. It is clear that (i) will have a more complicated matrix (less zero entries) than either (ii) or (iii), which in turn will imply more arithmetic operations at each inner iteration step. Moreover, the outer iteration will need to be more "complete" than in the latter cases. (ii) and (iii) have already been compared in Section 3. This is one instance in which method (iii) noticeably involves less computation than (ii).

The conclusion is that for multidimensional problems (iii) may be a valuable technique.

For the iterative deferred correction we will have an increasing task in setting up the necessary approximations to the successive error terms, but the possibility of obtaining more and more accurate results with a fixed h must also be born in mind.

To end this Section we can say that many other applications are possible. A few more are listed in STETTER [1965], and those examples show that most of the time many of the necessary properties will have already been proved. One need only follow the general guidelines in either Theorem 3.1 or Theorem 4.1 in order to generate approximations with the proper accuracy.

8. Numerical Examples

Let us consider the two-point boundary value problem

$$-y'' = x y' e^{-2y},$$
$$y(1) = 0,$$
$$y(2) = \ln 2$$

whose solution is $y(x) = \ln x$.

An Extended ALGOL program which implements the discretization described in Section 7.1 has been written for the Burroughs B 5500 at the Stanford University Computation Center.

The system of non-linear equations was solved by Newton's method (this solution is called h^2-APP. on Table 1), and then a linear deferred correction

Table 1

Init. val.	2-APP.	4-APP.	Ex. sol.
$8.66433\,97571_{10}-02$	$1.17733\,79749_{10}-01$	$1.17770\,79795_{10}-01$	$1.17783\,03566_{10}-01$
$1.73286\,79514_{10}-01$	$2.23073\,40768_{10}-01$	$2.23134\,10329_{10}-01$	$2.23143\,55132_{10}-01$
$2.59930\,19272_{10}-01$	$3.18379\,33305_{10}-01$	$3.18446\,36827_{10}-01$	$3.18453\,73112_{10}-01$
$3.46573\,59029_{10}-01$	$4.05396\,60218_{10}-01$	$4.05459\,37547_{10}-01$	$4.05465\,10811_{10}-01$
$4.33216\,98786_{10}-01$	$4.85451\,56983_{10}-01$	$4.85503\,40779_{10}-01$	$4.85507\,81578_{10}-01$
$5.19860\,38543_{10}-01$	$5.59575\,89913_{10}-01$	$5.59612\,49691_{10}-01$	$5.59615\,78794_{10}-01$
$6.06503\,78300_{10}-01$	$6.28587\,81812_{10}-01$	$6.28606\,34127_{10}-01$	$6.28608\,65943_{10}-01$

(Section 3) was applied (h^4-APP.). The results obtained are displayed in Table 1. A linear interpolation between the two ends was taken as the initial approximation.

The step used for the results in Table 1 was $h=\frac{1}{8}$.

Table 2

Init. val.	2-APP.	4-APP.
$8.66433\,97571_{10}-02$	$1.17733\,79749_{10}-01$	$1.17779\,00778_{10}-01$
$1.73286\,79514_{10}-01$	$2.23073\,40768_{10}-01$	$2.23140\,88343_{10}-01$
$2.59930\,19272_{10}-01$	$3.18379\,33305_{10}-01$	$3.18451\,89955_{10}-01$
$3.46573\,59029_{10}-01$	$4.05396\,60218_{10}-01$	$4.05463\,80533_{10}-01$
$4.33216\,98786_{10}-01$	$4.85451\,56983_{10}-01$	$4.85506\,86029_{10}-01$
$5.19860\,38543_{10}-01$	$5.59575\,89913_{10}-01$	$5.59615\,06788_{10}-01$
$6.06503\,78300_{10}-01$	$6.28587\,81812_{10}-01$	$6.28608\,10170_{10}-01$

6-APP.	Ex. sol.
$1.17783\,00047_{10}-01$	$1.17783\,03566_{10}-01$
$2.23143\,49185_{10}-01$	$2.23143\,55132_{10}-01$
$3.18453\,65526_{10}-01$	$3.18453\,73112_{10}-01$
$4.05465\,03623_{10}-01$	$4.05465\,10811_{10}-01$
$4.85507\,75453_{10}-01$	$4.85507\,81578_{10}-01$
$5.59615\,74283_{10}-01$	$5.59615\,78794_{10}-01$
$6.28608\,63474_{10}-01$	$6.28608\,65943_{10}-01$

The same problem was solved using i.d.c.p. and the same step size. The results of this experiment are shown in Table 2. The norms of the corresponding g.d.e. were,

$$\|U^{(1)}-\Delta_h u\| = 7.44\times 10^{-5}, \qquad \|U^{(2)}-\Delta_h u\| = 4.03\times 10^{-6},$$
$$\|U^{(3)}-\Delta_h u\| = 7.59\times 10^{-8}.$$

The number of inner iterations were respectively 3, 3 and 5 in order to reduce the norms of the residuals below 2×10^{-5}, 4×10^{-7} and 6×10^{-9}, those being the bounds indicated by the theory.

Acknowledgement. I would like to express my most sincere thanks to all people who helped me in preparing this paper and in general throughout my stay at Stanford University.

To Professor GENE GOLUB who called my attention to the problem and gave me continuous encouragement and generous advice.

To Professor G. STRANG of M.I.T., whose suggestions helped to simplify the presentation.

To JOHN WELSCH and JAMES VARAH for illuminating discussions, and to RICHARD BARTELS who proofread the manuscript.

Last, but not least, to Mrs. LYNNE LEONTOVICH who typed this manuscript.

References

BERS, L.: On mildly nonlinear difference equations of elliptic type. Jour. Res. Nat. Bur. of Stand. **51**, 229–236 (1953).

BICKLEY, W. G., S. MICHAELSON, and M. R. OSBORNE: On finite-difference methods for the numerical solution of boundary-value problems. Proc. Roy. Soc. London A **262**, 219–236 (1961).

BRAMBLE, J. H., and B. E. HUBBARD: On the formulation of finite difference analogues of the Dirichlet problem for Poisson's equation. Numer. Math. **4**, 313–327 (1962).

DOUGLAS, J.: Alternating direction iteration for mildly nonlinear elliptic difference equations. Numer. Math. **3**, 92–98 (1961).

FORSYTHE, G. E., and W. R. WASOW: Finite-difference methods for partial differential equations. New York: Wiley 1960.

FOX, L.: Numerical solution of ordinary and partial differential equations. Oxford: Pergamon Press 1962.

—, and E. T. GOODWIN: Some new methods for the numerical integration of ordinary differential equations. Proc. Camb. Phil. Soc. **45**, 373–388 (1949).

HENRICI, P.: Discrete variable methods in ordinary differential equations. New York: Wiley 1962.

KANTOROVICH, L. V.: Functional analysis and applied mathematics. Uspehi. Mat. Nauk. (N.S.) **3**, 89–185 (1948) [Russian]. Translated by C. D. BENSTER, Nat'l. Bur. Stand., 1948.

NOBLE, B.: The numerical solution of nonlinear integral equations. In: Publication No. 11 of the Math. Res. Center, U.S. Army. The University of Wisconsin Press. Edited by P. M. ANSELONE (1964).

PEREYRA, V.: The difference correction method for non-linear two-point boundary value problems. Comp. Sc. Dept. Tech. Report CS 18, Stanford University (1965).

SCHRÖDER, J.: Linear operators with positive inverse. MRC Tech. Report No. 230. The University of Wisconsin (1961).

STETTER, H. J.: Asymptotic expansions for the error of discretization algorithms for non-linear functional equations. Numer. Math. **7**, 18–31 (1965).

VOLKOV, E. A.: An analysis of one algorithm of heightened precision of the method of nets for the solution of Poisson's equation. Vych. Mat. **1**, 62–80 (1957) [Russian]. Translated by R. BARTELS — Comp. Sc. Dept. Tech. Report CS 27, Stanford University (1965).

Mathematics Research Center
The University of Wisconsin
Madison, Wisconsin 53706 (USA)