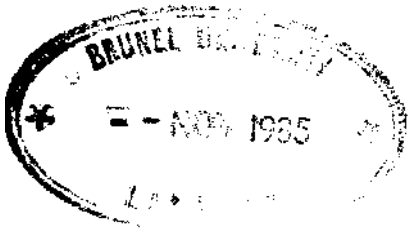
TR/21/85 October 1985

# DEFECT CORRECTION FROM A GALERKIN VIEWPOINT

By

Gerald Moore



# DEFECT CORRECTION FROM A GALERKIN VIEWPOINT

GERALD MOORE

Department of Mathematics and Statistics,

Brunel University,

Uxbridge,

Middlesex- UB8 3PH.
U.K.

## ABSTRACT

We consider the numerical solution of systems of nonlinear two point boundary value problems by Galerkin's method. An initial solution is computed with piecewise linear approximating functions and this is then improved by using higher—order piecewise polynomials to compute defect corrections. This technique, including numerical integration, is justified by typical Galerkin arguments and properties of piecewise polynomials rather than the traditional asymptotic error expansions of finite difference methods.

Subject Classifications :- AMS(MOS): 65L60; CR;G1.8

## 1. Introduction

Suppose that one uses a low accuracy (finite difference or finite element) approximation and a relatively coarse mesh to produce a numerical solution for a differential equation. If a more accurate solution is then desired one has the choice between using a finer mesh or a more accurate approximation. The former leads to larger sets of simultaneous equations to solve while the latter leads to more complexity and a larger band-width. The idea behind deferred or defect correction is to keep this complexity on the right-hand side of one's simultaneous equations and only to solve systems with the original simple matrix.

Deferred correction methods have become a very popular way of obtaining high accuracy approximations to smooth solutions of two point boundary value problems. The fundamental idea, as introduced by Fox in [5] and developed in particular by Pereyra (eg.[7,8,11]), can be seen by considering the single linear second-order problem

$$\begin{aligned} &\text{i)} \quad Ly(x) \equiv y'' = (x) + q(x)y'(x) + r(x)y(x) - f(x) \quad x \in [a.b] \\ &\text{ii)} \quad\quad y(a) = y(b) = 0. \end{aligned} \tag{1.1}$$

A basic approximation may be obtained by placing a uniform mesh over [a,b], i.e. $h = (b-a)/N$ and $x_i = a + jh$ $j = 0, \ldots, N$, and replacing the derivatives at internal mesh points by simple finite difference formulae. Thus the (N+1) - vector y is obtained by solving

$$\begin{aligned} (L^h \underset{\sim}{y}^h)_j &= (y^h_{j+1} - 2y^h_j + y^h_{j-1})/h^2 + q(x_j)(y^h_{j+1} - y^h_{j-1})/(2h) + r(x_j)y^h_j \\ &= f(x_j) \quad j=1,\ldots n-1 \\ y^h_0 &= y^h_N = 0 \end{aligned} \tag{1.2}$$

The error in this basic solution is proportional to $h$, assuming sufficient smoothness on y, but more accurate solutions may be obtained by noticing that

$$(L^h \underset{\sim}{y})_j = (Ly)(x_j) + \sum_{k=1}^{p-1} h^{2k} (T_k y)(x_j) + o(h^{2p}), \tag{1.3}$$

where $\underset{\sim}{y}$ is the (N+1) - vector of nodal values of y and the $T_k$ are higher-order differential operators. (We are now assuming more smoothness on y.) Hence if $\underset{\sim}{y}^h$ can be used to obtain an $0(h^2)$ approximation to

$T_1 y$ , i.e. if we can construct a difference operator $D^h_1$ such that

$$(T_1 y)(x_j) - (D^h_1 \underset{\sim}{y}^h)_j = O(h^2) \qquad j = 1,...,N-1, \tag{1.4}$$

then $\underset{\sim}{c}^h \equiv \underset{\sim}{y}^h + h^2 \underset{\sim}{z}^h$, where

$$L^h \underset{\sim}{z}^h = D^h_1 \underset{\sim}{y}^h, \tag{1.5}$$

will satisfy

$$\begin{aligned} L^h(\underset{\sim}{y} - \underset{\sim}{c}^h)(x_j) &= (Ly)(x_j) + h^2(T_1 y)(x_j) + 0(h^4) - f(x_j) - h^2(D^h_1 \underset{\sim}{y}^h)_j \\ &= h^2((T_1 y)(x_j) - (D^h_1 \underset{\sim}{y}^h)_j) + 0(h^4). \end{aligned} \tag{1.6}$$

The stability of L then shows that $\underset{\sim}{c}^h$ is an $0(h^4)$ approximation to $\underset{\sim}{y}$. The process may be repeated by using difference operators $D^h_k$ to approximate the differential operators T, and eventually an $0(h^{2p})$ approximation to $\underset{\sim}{y}$ is possible. This accuracy is attained while working on the same mesh and solving systems of linear equations with the same coefficient matrix, based on $L^h$, but with different right-hand sides. Of course the key theoretical problem is to show that $\underset{\sim}{y}^h$ can be used to approximate $T_1 y$ to $0(h^2)$, and similarly for the higher-order $T_k$, and this is usually achieved by showing that $\underset{\sim}{y} - \underset{\sim}{y}^h$ satisfies an asymptotic expansion: i.e.

$$\underset{\sim}{'} - \underset{\sim}{y}^h = \sum_{k=1}^{p-1} h^{2k} \underset{\sim}{e}_k + 0(h^{2p}) \tag{1.7}$$

where the e. are formed by nodal values of smooth functions $e_k(x)$. In practice the asymptotic error expansions are not needed but it is still necessary to construct the difference operators $D^h_k$.

On the other hand defect correction [3,9,12,13] relies on establishing the theoretical result that the error $\underset{\sim}{y} - \underset{\sim}{y}^h$ is smooth, i.e.., not only are the point values $0(h^2)$ but also the higher-order divided differences of point values. If then $M^h_1$ is a more accurate difference approximation to (1.1), e.g.,

$$(M^h_1 y)_j = (Ly)(x_j) + 0(h^4), \tag{1.8}$$

we construct $\underset{\sim}{v}^h \equiv \underset{\sim}{y}^h + \underset{\sim}{w}^h$ by solving

$$L^h \underset{\sim}{w}^h = \underset{\sim}{f} - M_1^h \underset{\sim}{y}^h . \tag{1.9}$$

We would then expect $\underset{\sim}{v}^h$ to be an $0(h^4)$ approximation to $\underset{\sim}{y}$ because

$$L^h (\underset{\sim}{y} - \underset{\sim}{v}^h) = (L^h - M_1^h)(\underset{\sim}{y} - \underset{\sim}{y}^h) + 0(h^4) \tag{1.10}$$

and $(L^h - M_1^h)(\underset{\sim}{y} - \underset{\sim}{y}^h)$ will be $0(h^2)$ multiplied by higher divided differences of $(\underset{\sim}{y} - \underset{\sim}{y}^h)$. As with deferred correction this idea may be repeated several times to obtain highly accurate results. Thus, in practice, the difference between deferred and defect correction is that the former requires the difference operators $D_k^h$ while the latter requires the difference operators $M_k^h$.

In this paper we present some results on using Galerkin's method (strictly a Petrov—Galerkin method) to solve nonlinear systems of first-order two point boundary value problems. Here the defect correction idea is very natural since more accurate difference operators $M_k^h$ correspond to using higher-degree piecewise polynomial spaces. (Our conception of defect correction has been particularly developed by studying the framework in [12] although there only finite difference methods are considered). The layout of the paper is as follows. In section 2 the basic Galerkin solution using piecewise linear trial functions is described and then in section 3 we compute defect corrections by using higher-degree piecewise polynomials. These methods are made practical in section 4 by analysing the effect of numerical integration and then we end with some remarks in section 5 about non-uniform meshes and higher—order differential equations.

To conclude this introduction we reproduce the following statement from p.25 of [11];-

"I wouldn't be surprised if it finally turns out that a successful implementation of high order spline methods comes about via a deferred correction type of approach, bypassing in some way the very expensive steps of high order quadrature formulae and complicated systems arising from the present approaches."

We feel that the present paper goes a long way towards achieving this aim.

## 2. Galerkin's method for first-order systems

We consider systems of nonlinear first-order two point boundary value problems of the form

$$\text{i)} \quad v'(x) = f(x,v(x)) \qquad x \in [a,b] \tag{2.1}$$

$$\text{ii)} \quad g(v(a),v(b)) = \underset{\sim}{0},$$

where $v: [a,b] \to R^n$, $f:[a,b] \times R^n \to R^n$ and $g:R^n \times R^n \to R^n$. We assume that for $v \in \{H^1[a,b]\}^n$ the function $f(x) \equiv f(x,v(x))$ is in $\{L^2[a,b]\}^n$ and hence we can regard (2.1) as a problem of finding zeroes of the nonlinear mapping $F : \{H^1[a,b]\}^n \to \{L^2[a,b]\}^n \times R^n$ defined by

$$F(v) \equiv \begin{cases} v'(x) - f(x, v(x)) \\ g(v(a), v(b)). \end{cases} \tag{2.2}$$

We also assume that $y \{H^1[a,b]\}^n$ is a solution of $F(v) = 0$ which is isolated in the same that $F$ is (Frechet) differentiable at $y$ and its linearisation,

$$F'(y) \quad v \equiv \begin{cases} v'(x) - A_1(x)v(x) \\ B_a v(a) + B_b v(b), \end{cases} \tag{2.3}$$

has a bounded inverse from $\{L^2[a,b]\}^n \times R^n \to \{H^1[a,b]\}^n$. It is also required that the components of the nxn matrix $A_1(x)$ are in $L^\infty[a,b]$ and that the nx2n augmented matrix $B_a | B_b$ has rank n.

Now we wish to obtain an approximation to y by means of Galerkin's method. For any positive integer N we define a uniform mesh over $[a,b]$ by setting $h=(b-a)/N$ and $x_j = a+jh \qquad j=0,\ldots,N$. Then we let $S_h$ denote the $(N+1)$ - dimensional subspace of $H^1[a,b]$ consisting of continuous piecewise-linear functions with respect to the given mesh, and $T_h$ the corresponding N-dimensional subspace of $L^2[a,b]$ consisting of piecewise-constant functions. The (Petrov-) Galerkin approximation to y is obtained

by seeking a $V \in \{S_h\}^n$ such that

$$\text{i) } \langle V' - f(x,V), \sigma \rangle = 0 \qquad \forall \sigma \in \{T_h\}^n$$
$$\text{ii) } g(V(a),V(b)) = 0, \tag{2.4}$$

where $\langle .,. \rangle$ is the inner product over $\{L^2[a,b]\}^n$. Thus we have nN+n equations to determine the (N+1)n free parameters in V. The rest of the section is concerned with showing that, for sufficiently small h, (2.4) has a unique solution Y near $Y_I$, the element of $\{S_h\}^n$ which interpolates y at the mesh points $x_o....x_N$.

Let B(v,w) be the bilinear form on $\{H^1[a,b]\}^n \times \{L^2[a,b]\}^n$ defined by

$$B(v,w) \equiv \langle v' - A_x(x)v, w \rangle , \tag{2.5}$$

which is bounded

$$|B(v,w)| \leq C_1 \| v \|_{H^k} \|w\|_{L^2} \tag{2.6}$$

If $\{H^1[a,b]\}$ denotes the subspace of $\{H^1[a,b]\}^n$ satisfying the conditions $B_a v(a) + B_b v(b) - 0$, (2.3) implies that B satisfies the coercivity condition

$$\inf_{V \in \{H^1[a,b]\}^n_0} \sup_{w \in \{L^2[a,b]\}^n} |B(v,w)| \geq C_2 \|v\|_{H^1} \|w\|_{L^2} \tag{2.7}$$

for $C_2 > 0$. For sufficiently small h, this result also holds for B restricted to $\{S_h\}^n_0 \times \{T_h\}^n$, where $\{S_h\}^n_0$ denotes the subspace of $\{S_h\}^n$ satisfying $B_a V(a) + B_b V(b) = 0$. To deduce this let V be an arbitrary element of $\{S_h\}^n_0$ and $w - V' — A_1(x)V$. Since the step-functions are dense in $L^\infty[a,b]$ we may choose $T \in \{T_h\}^n$ such that $\omega w = V' - \tau$ satisfies

$$\text{i) } \| w - \omega \|_{L^2} \leq C(h) \| v\|_{H^1}$$
$$\text{ii) } \| \omega \|_{L^2} \leq 2 \| w \|_{L^2} , \tag{2..8}$$

where C(h) is independent of V and tends to zero with h.

Hence

$$\begin{aligned} |B(V,\omega)| &= |B(V,w) - B(V,w-\omega)| \\ &\geq C_2 \|V\|_{H^1} \|W\|_{L^2} - C1 \|V\|_{H^1} \|w-\omega\|_{L^2} \\ &\geq C_2 \|V\|_{H^1} \|w\|_{L^2} - C_1 C(h) \|v\|^2_{H^1} \\ &\geq (C_2 - C_1 C(h) C_2^{-1}) \|V\|_{H^1} \|w\|_{L^2} \end{aligned} \tag{2.9}$$

and for sufficiently small h we will have

$$|B(V,\omega)| \geq C_3 \|v\|_{H^1} \|w\|_{L^2} \tag{2.10}$$

with $C_3 > 0$ and independent of V and h. Thus

$$\inf_{V \in \{S_h\}^n_0} \sup_{\sigma \in \{T_h\}^n} |B(V,\sigma)| \geq c_3 \|v\|_{H^1} \|\sigma\|_{L^2} \tag{2.11}$$

and we have the coercivity result.

Now we wish to show that, for h sufficiently small, the linear equations

$$\begin{aligned} &\text{i)} \quad B(V,\sigma) = \langle z,\sigma \rangle \qquad \forall \sigma \in \{T_h\}^n \\ &\text{ii)} \quad B_a V(a) + B_b V(b) = d. \end{aligned} \tag{2.12}$$

have a unique solution $V \in \{S_h\}^n$ for any given $z \in \{L^2[a,b]\}$ and $\underset{\sim}{d} \in R^n$. Let $V_1 \ldots . V_n$ be linear functions satisfying

$$B_a V_k(a) + B_b V_k(b) = \underset{\sim}{e}_k \qquad k = 1,\ldots,n. \tag{2.13}$$

where $\underset{\sim}{e}_k$ are the unit vectors in $R^n$, and so $V_1,\ldots,V_n$ are bounded in $\{H^1[a,b]\}^n$ independently of h. By choosing $\underset{\sim}{\alpha} \in R^n$ such that $\sum_{k=1}^{n} \alpha_k V_k$ satisfies (2.12ii), (2.12) is equivalent to solving

$$B(U,\sigma) = \langle z,\sigma \rangle - B\left(\sum_{k=1}^{n} \alpha_k V_k, \sigma\right) \qquad \forall \sigma \in \{T_h\}^n \tag{2.14}$$

for $U \in \{S_h\}_0^n$ and the coercivity result (2.11) allows us to use the generalized Lax-Milgram lemma [2,10] to show that this equation has a unique solution $U_0$ satisfying

$$\|U_0\|_{H1} \leq c(\|z\|_{L2} + \|\sum_{k=1}^{n} \alpha_k V_k\|_{H1}) \tag{2.15}$$

Hence (2.12) has a unique solution $V = U_0 + \sum_{k=1}^{n} \alpha_k V_k$ satisfying

$$\|V\|_{H1} \leq C(\|z\|_{L2} + \|\underset{\sim}{d}\|_{En} \tag{2.16}$$

where $\|.\|_{En}$ denotes the Euclidean norm and C is independent of h.

Finally, we consider the nonlinear equation (2.4) again and define the nonlinear mapping K: $\{S_h\}^n \rightarrow \{S_h\}^n$ by

$$\text{i) } B(K(V),\sigma) = \langle f(x,V) - A_1(x)V, \sigma \rangle \quad \forall \sigma \in \{T_h\}^n \tag{2.17}$$

$$\text{ii) } B_a K(V)(a) + B_b K(V)(b) = B_a V(a) + B_b V(b) - g(V(a), V(b)).$$

For sufficiently small h K has been shown to be well defined and we proceed to prove that, under certain assumptions, it is a contraction mapping in a ball about Y. The assumptions we need are that :-

a) the mapping from $\{H^1[a,b]\}^n$ to $\{L^2[a,b]\}^n$ defined by the function $f(x,v)$ is differentiate in a ball about y with each of the n components of the derivative in $L^\infty[a,b]$ and satisfying a Lipschitz condition

$$\|a_{ij}(x,v) - a_{ij}(x,w)\|_{L\infty} \leq c \|v-w\|_{H1} \tag{2.18}$$

b) the mapping $g : R^n \times R^n \rightarrow R^n$ is differentiable in a ball about $(\underset{\sim}{y}(a), \underset{\sim}{y}(b))$ and the derivative satisfies a Lipschitz condition.

Now if we write (2.17) in the form

$$\text{i)}\quad B(K(V)-Y_I,\sigma) = <f(x,v)-A_1(x)(V-Y_I) - f(x,y),\sigma> \tag{2.19}$$

$$\text{ii)}\quad B_a(K(V)-Y_I)(a)+B_b(K(V)-Y_I)(b) = B_a(V-Y_I)(a)+B_b(V-Y_I)(b)-g(V(a),V(b))$$

the result (2.16) may be used to obtain

$$\|k(V)-Y_I\|_{H1} \le C\{\|f(x,V)-f(x,Y_I)-A_1(x)(V-Y_I)\|_{L2} + \tag{2.20}$$
$$\|f(x,y)-f(x,Y_I)\|_{L2} + \|g(a),Y_I(b)) - g(V(a),V(b)) + B_a(V-Y_I)(a) + B_b(V-Y_1)(B)\|_{En}).$$

Then if we choose h so that $\|y-y_1\|_{H1}$ is sufficiently small and the Lipschitz conditions hold, K. will map a ball of radius $C\|y-Y_I\|_{L2}$ centred on $Y_I$ w.r.t. the $H^1$-norm into itself. Also if $U,V\in\{S_h\}^h$ we have

$$\text{i)}\quad B(K(U)-K(V),\sigma) - <f(x,U)-f(x,V)-A_1(x)(U-V),\sigma> \tag{2.21}$$

$$\text{ii)}\quad B_a(K(U)-K(V))(a) + B_b(K(U)-KCV))(b) = g(V(a),V(b))-g(U(a),U(b)) + B_a(U-V)(a) + B_b(U-V)(b)$$

and applying (2.16) again leads to

$$\|K(U)-K(V)\|_{H1} \le C(\|U-Y_I\|_{H1} + \|V-Y_I\|_{H1} + \|y-Y_I\|_{H1}\ \|U-V\|_{H1}$$

for sufficiently small h, and hence to K being a contraction in the above ball. Thus K will have a unique fixed point $Y_0$ in this ball which will also be a locally unique solution of (2,4).

It is now simple to produce some error estimates of $y-Y_0$ because we immediately have the "superconvergence" result

$$\|Y_I-y_0\|_{H1} \le C\|y-Y_I\|_{L2} \tag{2.22}$$

Thus

$$\|y-Y_0\|_{H1} \le \|y-Y_I\|_{H1} + \|Y_I-Y_0\|_{H1} \tag{2.23}$$

which is 0(h) if $y\in\{H^2[a,b]\}^n$, and

$$\|y-Y_0\|_{L2} \le \|y-Y_I\|_{L2} + \|Y_1-Y_0\|_{L2} \tag{2.24}$$

which is $0(h)$ if $y \in \{H^1[a,b]\}^n$ and $0(h^2)$ if $y \in \{H^2[a,b]\}^n$.

Similar results follow in $L^\infty$ by making us of relationship

$$C\| z \|_{L2} \leq \| z \|_{L\infty} \leq C \| z \|_{H1} \tag{2.25}$$

for $z \in \{H^1[a,b]\}^n$ i. e.

$$\begin{aligned} &\text{i)} \quad \|Y_I - Y_0\|_{L\infty} \leq C \| y - Y_I \|_{L2} \\ &\text{ii)} \quad \| y - Y_0 \|_{L\infty} \leq Ch^2 \| y \|_{W2,\infty} \text{ etc.} \end{aligned} \tag{2.26}$$

In the next section we shall require some results about the linearization of (2.2) at $Y_0$. For sufficiently small h this will exist and we write it as

$$\begin{aligned} &\text{i)} \quad v'(x) - \acute{A}(x)\, v(x) \\ &\text{ii)} \quad \overset{\circ}{B}_a v(a) + \overset{\circ}{B}_b v(b) , \end{aligned} \tag{2.27}$$

where the components of $A(x)$ are in $L^\infty[a,b]$, cf. (2.3). If we define the bounded bilinear form B on $\{H^1[a,b]\}^n \times \{L^2[a,b]\}^n$ by

$$\overset{\circ}{B}(v,w) \equiv \langle v' - \acute{A}(x)\, v, \; w \rangle \tag{2.28}$$

then we wish to show that the linear equations

$$\begin{aligned} &\text{i)} \quad \overset{\circ}{B}_a (v\sigma) = \langle z, \sigma \rangle \qquad \forall \sigma \in \{T_h\}^n \\ &\text{ii)} \quad \overset{\circ}{B}_a v(a) + \overset{\circ}{B}_b v(b) = \underset{\sim}{d} \end{aligned} \tag{2.29}$$

have a unique solution $V \in (S_h\}^n$ for any given $z \in \{L^2[a,b]\}^n$ and $\underset{\sim}{d} \in R^n$, cf. (2.12), and that (2.16) holds. This is achieved by showing that B satisfies a coercivity condition on $\{S_h\}^n{}_0 \times \{T_h\}^n$, which follows from

$$\begin{aligned} |\overset{\circ}{B}(v,\sigma)| &\geq |B_a(v),\sigma)| - |B(v,\sigma) - \overset{\circ}{B}(v,\sigma)| \\ &\geq (C_3 - c\|y - Y_0\|_{H^1}) \|V\|_{H1} \|\sigma\|_{L2} \end{aligned} \tag{2.30}$$

cf. (2.11).

## 3. Higher accuracy through defect correction.

The essence of defect correction for Galerkin methods is to introduce a mapping P from $(S_h\}^n$ to another sub space of $\{H^1[a,b]\}^n$ with superior approximating power, e.g. a cubic spline subspace. Then

$$P(Y_0)^1 - f(x,PY_0) \tag{3.1}$$

is regarded as an estimate of the error in $Y_0$ and we can calculate a sequence $\{Y_k\} \subset \{S_h\}^n$ of (hopefully) better approximations to $Y_I$ through the iteration

$$\text{i)} \quad \overset{\circ}{B}(Y_k - Y_{k-1},\sigma) = - \langle (PY_{k-1})' - f(x,PY_{k-1}),\sigma\rangle \quad \forall\sigma\in\{T_h\}^n \tag{3.2}$$

$$\text{ii)} \quad \overset{\circ}{B}_a(Y_k - Y_{k-1},\sigma)\,(a) + \overset{\circ}{B}_b(Y_k - Y_{k-1})\,(b) = - g((PY_{k-1})\,(a), (PY_{k-1})\,(b)),$$

which is well-defined for sufficiently small h. In this section we shall deduce that $\{Y_k\}$ approaches Y. in the $\infty$ - norm by making smoothness assumptions on f . We shall use A. to denote the divided forward difference

$$\Delta v(x) \equiv (v(x+h) - v(x)/h \tag{3.3}$$

with

$$\Delta^j v(x) = \Delta(\Delta^{j-1}v(x)), \tag{3.4}$$

and for $v\in L^\infty[a,b]$ we define the semi-norms

$$|v|_{DJ} = \|\Delta^j v\|_{L\infty} \tag{3.5}$$

where the $\infty$- norm is taken over the interval $[a,x_{N-j}]$. These definitions extend naturally to vectors and we note that $|\cdot|_{D0} = \|\,.\,\|_{L\infty}$ .

In the previous section we showed that

$$\|Y_1 - Y_0\|_{D0} \le Ch^2 \|y\|_{H2} .$$

Nov we wish to show that extra smoothness conditions on f and y imply that higher-order differences of $Y_1 - Y_0$ are also $0(h^2)$. For some integer $p \geq 2$, assume that f(x,v(x)) has the following expansion in a ball about $y \in \{H^1[a,b]\}^n$.

$$F(x,v(x)) = f(x,y(x)) + \sum_{\ell=1}^{P} A_\ell(x)\,(v-y)^\ell(x) + R(x,v(x)), \tag{3.7}$$

where each of the $n^{\ell+1}$ components of the multi-linear operators $A_\ell(x)$ is in $W^{p-2,\infty}[a,b]$ and the remainder term satisfies

$$\| R(x,v) \|_{L\infty} \leq C \| (v-y)^P \|_{L\infty}. \tag{3.8}$$

Then if we write the $a^{th}$ difference, $l \leq s \leq p$ of the $i^{th}$ component of $Y_1 - Y_0$ as

$$\Delta^s (Y_I - Y_0)_i (x_j) = h^{-1} \Delta^{s-1} (\langle f(x,y) - f(x,Y_0), \sigma_i^{j\leftarrow} \rangle), \tag{3.9}$$

where a. is the element of $\{T_h\}^n$ whose every component is zero apart from the $i^{th}$ and this is unity from $x_j$ to $x_{j+1}$ and zero elsewhere, we may use the expansion to obtain

$$|\Delta^s (Y_I - Y_0)_i (x_j)| \leq C \Big( \sum_{\ell=1}^{p-1} \sum_{\alpha=0}^{s-1} hm(\alpha,\ell) \left\| y - Y_0 \right\|_{D\alpha}^{\ell} + h^{1-s} \left\| y - Y_0 \right\|_{L\infty}^{P} \Big) \tag{3.10}$$

where

$$m(\alpha, \ell) = \begin{cases} 0 & \alpha \leq p - \ell \\ p - \ell - \alpha & \alpha > p - \ell \end{cases} \tag{3.11}$$

Hence, if we assume that $y \in \{W^{p+1,\infty}[a,b]\}$ so that the interpolation error $|y-Y_I|_{D\alpha}$ is $0(h^2)$ $0 \leq \alpha \leq p-l$, we may replace $y-Y_0$ by $(y-Y_I) + (Y_1 - Y_0)$ in (3.10) and show by induction that

$$| Y_1 - Y_0 |_{Ds} \leq C h^2 \qquad s=0, \ldots, P. \tag{3.12}$$

The rest of this section is devoted to shoving that the iteration (3.2) leads to

$$|Y_I - Y_k|_{Ds} = 0(h^{(k+1)}) \tag{3.13}$$

for s=0, ..... ,p-k: and k=0, .....,P-1. In fact we shall generalise (3.2) by allowing P to change at each iteration so that

$$\text{i)} \quad \overset{\circ}{B}(Y_k - Y_{k-1}, \sigma) = - \langle (P_k Y_{k-1})' - f(x, P_k Y_{k-1}), \sigma \rangle \qquad \forall \sigma \in \{T_h\}^n$$

$$\text{ii)} \quad \overset{\circ}{B}_a(Y_k - Y_{k-1}, \sigma)\,(a) + \overset{\circ}{B}_b(Y_k - Y_{k-1})\,(b) = - g((P_k Y_{k-1})\,(a), (P_k Y_{k-1})\,(b)). \tag{3.14}$$

Conditions on the $P_k$ will be developed as we proceed hut it will always be assumed that each is a nodal interpolatory mapping, thus $(P_k\ V)\,(x.) = V(x_j)$ $j = 0, ..., N$, whose range is a space of continuous piecewise polynomials over [a,b]. (However note that $P_k$ is not necessarily a linear mapping) .

Also we assume that, if $y \in \{W^2(k+1),^{\infty}[a,b]\}^n$ and $V \in \{S_h\}^n$ is sufficiently close to $Y_I$, the following stability and approximation properties hold:-

$$\text{a)} \quad \| P_k Y_I - P_k V \|_{L\infty} \leq C \| Y_I - V \|_{L\infty} ,$$

$$\text{b)} \quad \max_{j=1,\ldots,N} \{ \| (P_k Y_I - P_k V)'' \|_{L\infty(x_{j-1}, x_j)} \} \leq C \max_{\alpha=0,1,2} \{ | Y_I - V |_{D\alpha} \}, \tag{3.15}$$

$$\text{c)} \quad \| y - P_k Y_I \|_{L\infty} \; Ch^{2(k+1)} \; \| y \|_{W^2_{(k+1),}} ,$$

where C is independent of h and V or y. Some natural choices for the $\{P_k\}$ are listed below.

1) Globally C° piecewise $(2k+l)^{th.}$-degree polynomials, where the polynomial on the subinterval $x_j., x_{j+1}$ is obtained by interpolation at the points $x_{j-k}, .... , x_{j+k+1}$ .

2) Globally $C^2$ piecewise $(2k+1)^{th.}$-degree polynomial splines.

3) Globally $C^1$ piecewise $(2k+1)^{th.}$- degree polynomials, where the polynomial on the subinterval $[xj,x_{j\cdot+1}]$ is obtained by interpolation at the points $x_{j-m}$, $x_{j-m}$ ,...,$x_{j+m+1}$ and has derivative values f(x-V) at the points $x_{j-m}$, ..., $x_{j+m+1}$ if k is odd and at the points $x_{i-m+1}$,...,$x_{i+m}$. if k is even. (here m - (k-l)/2 if k is odd and m = k/2 if k is even).

4) Globally $C^{2k-1}$ piecewise $(2k+1)^{th.}$-degree polynomial splines interpolating function values and taking derivative values f (x,V) at the nodes.

With each choice there is the question of what to do near the ends of the mesh and this problem is considered at the end of the section.

First we obtain a bound on the $L^\infty$-norm of $Y_I$-$Y_K$, in terms of higher differences of $Y_I$ -$Y_{k-1}$ by rewriting (3.14) as

$$\text{i)} \quad \overset{\circ}{B}(Y_k - Y_k,\sigma) = - \langle \acute{A}(x)\,(Y_I - Y_{k-1} - (P_kY_I - P_kY_{k-1})), \sigma\rangle + \langle f(x,y) - f(x, P_kY_{k-1}) - \acute{A}(x)\,(P_kY_I - P_kY_{k-1})), \sigma\rangle \tag{3.16}$$

$$\text{ii)} \quad \overset{\circ}{B}_a(Y_I - Y_k)\,(a) + \overset{\circ}{B}_b(Y_I - Y_k)\,(b) = g\,(Y_{k-1}(a)\,,\ Y_{k-1}\,(b)) - g\,(Y_I\,(a)\,, Y_I\,(b)) + \overset{\circ}{B}_a(Y_I - Y_{k-1})\,(a) + \overset{\circ}{B}_b(Y_I - Y_{k-1})\,(b)$$

and then applying (2.29). We need a stronger Lipschitz condition on the derivative of f with the $H^1$ -norm replaced by the $L^\infty$-norm in (2.18). This together with the conditions (3.15) means that for $y \in \{W^{2(k+1),\infty}\,[a,b]\}^n$

$$\| Y_I - Y_k \|_{H1} \le C\{h^2 \max_{\alpha=0,1,2} (|\,Y_I - Y_{k-1}\,|_{D\alpha}),\ + |\,Y_I - Y_{k-1}\,|^2_{D0} + h^{2(k+1)}\,\|\,y\,\|_{W^{2(k+1),\infty}}\} \tag{3.17}$$

and so if $|Y_I-Y_{k-1}\,|_{D0}$ is $0(h^2)$ then

$$|\,Y_I - Y_k\,|_{D0} \le C(h^2 \max_{\alpha=0,1,2} \{|\,Y_I - Y_{k-1}\,|_{D\alpha}\} + h^{2(k+1)}\,\|\,y\,\|_{W^{2(k+1),\infty}}\}\,. \tag{3.18}$$

Finally we bound higher differences of $Y_I$-$Y_k$ in terms of lower differences plus differences of the previous iterate $Y_I$-$Y_{k-1}$. This is achieved by considering, for $1\le s\le p-k$.

$$\begin{aligned}\Delta^s(Y_I-Y_k)_i(x_j) = h^{-1}\Delta^{s-1}\{&<A_I(x)(Y_I-Y_k),\sigma_i^{j+}> + <(\overset{\circ}{A}-A_1)(x)(Y_I-Y_k),\sigma_i^{j+}>\\ &-<A_1(x)(Y_I-Y_{k-1}-(P_kY_I-P_kY_{k-1})),\sigma_i^{j+}> + <(A_1-\overset{\circ}{A})(x)(Y_I-Y_{k-1}),\sigma_i^{j+}>\\ &+<f(x,y)-f(x,P_kY_{k-1})-A_1(x)(P_kY_I-P_kY_{k-1}),\sigma_i^{j+}>\}. \end{aligned} \tag{3.19}$$

We also need $P_k$ to satisfy conditions like (3.15) but now involving higher differences; i.e. if $y\in\{W^{2(k+1)+t,\infty}[a,b]\}^n$ and $V$ is close to $Y_I$, then for $0\le t\le p-k-1$ and $1\le k\le p-1$

$$\begin{aligned} &\text{a)}\ |P_kY_I-P_kV|_{Dt} \le C \max_{0\le\alpha\le t}\{|Y_I-V|_{D\alpha}\},\\ &\text{d)}\ \max_{j=1,..,N-t}\{\|\Delta^t(P_kY_I-P_kV)'\|_{L^\infty(x_{j-1},x_i)}\} \le C\max_{0\le\alpha\le t+2}\{|Y_I-V|_{D\alpha}\},\\ &\text{c)}\ |y-P_kY_I|_{Dt} \le Ch^{2(k+1)}\|y\|_{W^{2(k+1)+t,\infty}}\}, \end{aligned} \tag{3.20}$$

where $C$ is independent of $V$ or $y$. (Examples of $\{P_k\}$ satisfying these conditions will be given at the end of the section.) We shall also shortly wish to expand $A(x)$ about $A_1(x)$ in powers of $y$-$Y_0$ and in order to linearise through the expansion (3.7) we assume that the (Frechet) derivative of the remainder term satisfies

$$\|R'(x,v)\|_{L^\infty} \le C\|v-y\|_{L^\infty}^{p-1} \tag{3.21}$$

Now applying (3.20) and (3.21) to (3.19), and using the fact that $y$-$Y_0$ is $0(h^2)$ in the various difference norms, gives

$$\begin{aligned}\left|\Delta^s(Y_I-Y_k)(x_j)\right| \le C\{&\sum_{\alpha=0}^{s-1}|Y_I-Y_k|_{D\alpha} + h^2\sum_{\alpha=0}^{s-1}|Y_I-Y_k|_{D\alpha}\\ &+\sum_{\alpha=0}^{s-1}|y-P_kY_I|_{D\alpha} + \sum_{\ell=2}^{p-1}\sum_{\alpha=0}^{s-1}h^{m(\alpha,\ell)}|(y-P_kY_{k-1})^\ell|_{D\alpha}\\ &+h^{1-s}\|y-P_kY_{k-1}\|_{L^\infty}^{p-1}\}. \end{aligned} \tag{3.22}$$

Hence by splitting $y-P_k y_{k-1}$ into $(y-P_k y_I) + (P_k Y_I - P_k Y_{k-1})$ and using (3.12) and (3.18) we can show by induction that (3.13) holds.
Thus after p-1 iterations we will have, for $y \in \{W^{2p,\infty}[a,b]\}^n$,

$$\| Y_I - Y_{p-1} \|_{W^{1,\infty}} = 0(h^{2p}) \tag{3.23}$$

and high accuracy is attained at the mesh points.

To conclude this section we consider the problem of defining the mappings $\{P_k\}$ on the subintervals near a and b so that conditions (3.20) will hold. There are a number of possibilities.

1) If globally $C^\circ$ piecewise $(2k+1)^{th.}$ -degree polynomials are used then values at $x_{\_k},\ldots,x_{-1}$ and $x_{n+1},\ldots, x_{n+k}$ can be obtained by extrapolation. Thus we set

$$\Delta^{P+k+1} V(x_j) = 0 \quad j = -k,\ldots,-1$$

$$\nabla^{P+k+1} V(x_j) = 0 \quad j = N+1,\ldots,N+k$$

where $\nabla$ denotes the backward (divided) difference.

2) If globally $C^{2k}$ piecewise $(2k+l)^{th.}$ -degree polynomial splines are used then our end conditions are

$$\Delta^{p+k-1} s''V(x_j) = 0 \quad j = 0,\ldots,k-l$$

$$\square^{p+k-1} s,''(x_j) = 0 \quad j = N-k+l,\ldots,N$$

where s (x) is the spline.

3) If globally $c^1$ piecewise $(2k+1)^{th.}$ -degree polynomials are used then function values are required at $x_{-m}, \ldots,x_{-1}$ and $x_{N+1},\ldots\ldots_{N+m}$ while derivative values are needed at $x_{-m,}\ldots\ldots.N+m$ While derivates values are required at $x_{-m}, \ldots x_{-1}$ and $x_{N+1},\ldots,x_{N+m}$ if k is odd and at $x_{l-m}, \ldots,x_{-l}$ and $x_{N+I},\ldots,x_{n+m-1}$ if k is even. (Again $m = (k-1)/2$ if k is odd and $m = k/2$ if k is even).

Thus we set

$$\Delta^{P+k+1} V(x_j) = 0 \quad j= -m,...,-1$$

$$\nabla^{p+k+1} V(x.) = 0 \quad j= N+1,...,N+m$$

together with

$$\Delta^{p+k} f(x_j,V(x_j)) = 0 \quad j = -m,\ldots,-1$$

$$\nabla^{P+k} f(x.,V(x.)) = 0 \quad j = N+1,...,N+m$$

if k is odd or

$$\Delta^{p+k} f(x_j,V(x_j)) = 0 \quad j= 1-m,...,-1$$

$$\nabla^{p+k} f(x.,V(x.)) = 0 \quad j= N+1,...,N+m-1$$

if k is even.

4) If globally C piecewise $(2k+1)^{th.}$ degree polynomial splines are used then our end conditions are

$$\Delta^{p+k\ 1} s''(x_j) = 0 \quad j=0,...,k-2$$

$$\Delta^{p+k-1} s''(x) = 0 \quad j=N+2-k,...,N$$

An alternative idea, which can be used instead of 1) and 3) above, is to adapt a method which has been used successfully is deferred correction with finite differences [5,6,7,8] . Thus we assume that our differential equation is valid and smooth over a slightly larger interval $[a-\varepsilon,b+\varepsilon]$ . An approximation over this larger interval may then be obtained from (2.4), with $Y_0$ outside [a,b] computed as for initial value problems, and hence we shall already have the extra values required to compute an appropriate $P_1 Y_0$. This technique can be used repeatedly with (3.14) to ensure that we always have sufficient values of Y. exterior to [a,b] in order to compute $P_{k+1}\ Y_k$ . On account of its simplicity and avoidance of high-order extrapolation it is this technique that we would recommend in practice.

## 4. Including numerical quadrature

In practice the discrete equations developed in the last two sections cannot be used, because the exact calculation of various integrals is either impossible or too time-consuming. Thus in this section we approximate these integrals by chosen quadrature formulae and show the accuracy required therein for the defect correction results to be retained.

When quadrature is used our basic Galerkin approximation to y, labelled $Y^{Q}_{o}$, is obtained by solving the equations

$$\begin{aligned} &\text{i) } \langle V',\sigma\rangle - Q_0[\langle f(x,V),\sigma\rangle] = 0 \qquad \forall\sigma\in\{T_h\}^n \\ &\text{ii) } g(V(a),V(b)) = 0 , \end{aligned} \tag{4.1}$$

cf. (2.4). Here $Q_0[\langle f(x,V),\sigma\rangle]$ is an approximation to $\langle f(x,V),\sigma\rangle$ derived from a linear quadrature rule for approximating real-valued functions of the form

$$\int_a^b w(x)\, dx . \tag{4.2}$$

In order to apply the quadrature successfully we shall henceforth assume that f (x,V) is a vector of functions in C[a,b] for $V\in\{S_h\}^n$ near y. If $E_0[\langle v,\sigma\rangle]$ denotes the quadrature error $\langle v,\sigma\rangle - Q_0[\langle v,\sigma\rangle]$ we only assume for the moment the following boundedness and approximation results:-

a) if v is vector of continuous functions

$$|Q_0[\langle v,\sigma\rangle]| \le C\Big(\sum_{j=1}^{N} h \|v\|^2_{L^\infty[x_{j-1},x_j]}\Big)^{\frac{1}{2}} \|\sigma\|_{L^2}$$

b) if $v\in\{H^2[a,b]\}^n$

$$\|E_0[\langle v,\sigma\rangle]\| \le Ch^2 \|v\|_{H2}\|\sigma\|_{L2} \tag{4.3}$$

c) if $A(x)$ is an nxn matrix with components in $W^{2,\infty}[a,b]$ and $V \in \{S_h\}^n$

$$|E_Q[\langle A(x)V,\sigma\rangle]| \leq Ch^2 \|v\|_{HI} \|\sigma\|_{L2}$$

with C independent of v,V or h. It is easily checked that the two most natural choices of quadrature method, based on the trapezoidal or midpoint rules, both satisfy the above conditions.

We now introduce conditions in order to show that, for h sufficiently small, (4.1) will have a locally unique solution in $\{S_h\}^n$ near $Y_0$ and $Y_1$. This is achieved, cf. (2.17), by considering the fixed points of the nonlinear mapping $K^Q: \{S_h\}_n \rightarrow \{S_h\}_n$ defined by

$$\begin{aligned} &\text{i) } B(K^Q(V),\sigma) = Q_0[\langle f(x,V),\sigma\rangle] - \langle A_1(x)V,\sigma\rangle \quad \forall \sigma \in \{T_h\}^n \\ &\text{ii) } B_a K^Q(V)(a) + B_b k^Q(v)(b) = B_a V(a) + B_b V(b) - g(V(a),V(b)). \end{aligned} \tag{4..4}$$

As with (2.17) we proceed to show that $K^Q$ is a contraction mapping in a suitable ball about $Y_I$ in the $H^1$-norm. By rewriting (4.4) in the form

$$\begin{aligned} \text{i) } B(k^Q(v)-Y_I,\sigma) &= Q_0[\langle f(x,V) - f(x,Y_I) - A_1(x)(V-Y_I),\sigma\rangle] \\ &- E_0[\langle A_1(x)(V-Y_I),\sigma\rangle] - E_0[\langle f(x,y),\sigma\rangle] \\ &- Q_0[\langle f(x,y) - f(x,Y_I)\rangle] \end{aligned} \tag{4.5}$$

$$\text{ii) } B_a(K^Q(V)-Y_I)(a) + B_b(K^Q(V)-Y_I)(b) = B_a(V-Y_I)(a) + B_b(V-Y_I)(b) - g(V(a),V(b))$$

we see that, provided $y \in \{H^3[a,b]\}^n$ and the components of $A(x)$ are in $W^{2,\infty}[a,b]$, we may use (2.16),(4.3) and the Lipschitz continuity of the derivatives of f and g at (2.18) to obtain

$$\|K^Q(V)-Y_I\|_{H^1} \leq C(\|V-Y_I\|_{H^1}^2 + \|y-Y_I\|_{H^1}\|V-Y_I\|_{H^1} + h^2\|V-Y_I\|_{H^1} + h^2\|y\|_{H^3}) \tag{4.6}$$

for h sufficiently small. Consequently $K^Q$ will map a ball of radius $0(h^2)$ centred on $Y_1$ in the $H^1$-norm into itself. Also if $U,V \in \{S_h\}^n$ we have

$$\text{i) } B(K^Q(U) - K^Q(V),\sigma) = Q_0[<f(x,U) - f(x,V) - A_1(x)(U-V),\sigma>] - E_0[<A_1(x)(U-V),\sigma>] \quad \forall\sigma\in\{T_h\}^n \tag{4.7}$$

$$\text{ii) } B_a(K^Q(U)-K^Q(V))(a) + B_b(K^Q(U)-K^Q(V))(b) - g(V(a),V(b)) - g(U(a),U(b)) + B_a(U-V)(a) + B_b(U-V)(b)$$

and so, for sufficiently small h,

$$\|K^Q(U)-K^Q(V)\|_{H^1} \le C(\|u-Y_1\|_{H^1} + \|V-Y_1\|_{H^1} + h^2)\,\|U-V\|_{H^1} \tag{4.8}$$

and $K^Q$ will be a contraction in the above ball. Hence (4.1) will have a locally unique solution near $Y_I$ and this is our $Y_0^Q$ which satisfies

$$\|Y_I - Y_0^Q\|_{H^1} = Ch^2\,\|y\|_{H^3}\,. \tag{4.9}$$

We now define our defect correction iterates $\{Y_k^Q\}$ by

$$\begin{aligned} &\text{i) } \overset{\circ}{B}{}_k^Q(Y_k^Q - Y_{k-1}^Q,\sigma) = Q_k[<f(x,P_kY_{k-1}^Q),\sigma>] - <(P_kY_{k-1}^Q)',\sigma> \quad \forall\sigma\in\{T_h\}^n \\ &\text{ii) } \overset{\circ}{B}{}_a^Q(Y_k^Q - Y_{k-1}^Q)(a) + \overset{\circ}{B}{}_a^Q(Y_k^Q - Y_{k-1}^Q)(b) = -g((P_kY_{k-1}^Q)(a),(P_kY_{k-1}^Q)(b)), \end{aligned} \tag{4.10}$$

where $\{Q_1\}$ are a sequence of quadrature rules upon which we shall shortly place conditions. The bounded bilinear form $\overset{\circ}{B}{}^Q(V,\sigma)$ over $\{S_h\}^n$ x $\{T_h\}^n$ and the linear mappings $\overset{\circ}{B}{}_a^Q, \overset{\circ}{B}{}_b^Q$ are derived from the linearisation of (4.1) at $Y_0^Q$ i.e.,

$$\overset{\circ}{B}{}^Q(V,\sigma) \equiv <V', \sigma - Q_0[<\overset{\circ}{A}{}^Q(x)V,\sigma>] \tag{4.11}$$

where $\overset{\circ}{A}{}^Q(x)$ is the linearisation of f(x.v) at $Y_0^Q$ of course if $z\in\{L^2[a,b]\}^n$ and deR are given we require that the linear equations

$$\begin{aligned} &\text{i) } \overset{\circ}{B}{}^Q(V,\sigma) = <z,\sigma> \quad \forall\sigma\in\{T_h\}^n \\ &\text{ii) } \overset{\circ}{B}{}_a^Q)V(a) + \overset{\circ}{B}{}_b^Q)V(b) = \underset{\sim}{d} \end{aligned} \tag{4.12}$$

are uniquely solvable with

$$\|V\|_{H^1} \le C(\|Z\|_{L^2} + \|\underset{\sim}{d}\|_E n), \tag{4.13}$$

for sufficiently small h. However since

$$\text{i)}\ |B(V,\sigma) - \overset{\circ}{B}{}^{Q}(V,\sigma)| = |E_0[<A_1(x)V,\sigma>] - Q_0[<A_1 - \overset{\circ}{A}{}^{Q})(x)V,\sigma>]|$$

$$\leq ch\ \|V\|_{H^1}\ \|\sigma\|_{L^2} \qquad (4.14)$$

$$\text{ii)}\ |(B_a - \overset{\circ}{B}{}^{Q}_a)V(a) + (B_b - \overset{\circ}{B}{}^{Q}_b)V(b)| \leq Ch^2\|V\|_{H^1}$$

we may use simple pertubation arguments to obtain (4.13), cf. (2.29).

To prove error results for $\{Y_I - Y^Q_k\}$ we must first show that the higher differences of $Y_I - Y^Q_0$ are $0(h^2)$. (That $\|Y_I - Y^Q_0\|_{L^\infty}$ is $0(h^2)$ already follows from (4.9).) Thus the basic quadrature rule $Q_0$ must be chosen so that it varies smoothly from one subinterval to the next and so does its error; i-e. for $0\leq s\leq p-1$

a) if v is a vector of continuous functions

$$\max_{i=1,\ldots,n}\ \max_{j=1,\ldots,N-s}\ \{|\Delta^s(Q_0[<v,\sigma_i^{j-}>])|\} \leq Ch\ \max_{0\leq\alpha\leq s}\ \{|V|_{D\alpha}\},$$

b) if $v\in\{W^{S+2,\infty}[a,b]\}^n$ (4.15)

$$\max_{i=1,\ldots,n}\ \max_{j=1,\ldots,N-s}\ \{|\Delta^s(E_0[<v,\sigma_i^{j-}>])|\} \leq Ch^3\|v\|_{W^{s+2,\infty}}\ .$$

Note that in the usual case of $Q_0$ being based on a composite rule, e.g. composite trapezoidal or composite midpoint, (4.15) will follow from (4.3) because we can take the differences inside $Q_0[.]$ and $E_0[.]$. Now our equation for the higher-order differences is

$$\Delta^s(Y_I - Y^Q_0)_i\ (x_j)\ h^{-1}\Delta^{s-1}\ (Q_0[<f(x,y) - f(x,Y^Q_0),\sigma_i^{j+}>] + E_0[<f(x,y),\sigma_i^{j+}>]) \qquad (4.16)$$

and, with (4.15), we may follow through the argument after (3.9) to obtain

$$Y_I - Y^Q_0|_{D^s} \leq Ch^2 \qquad s=0,\ldots,p \qquad (4.17)$$

if $y\in\{W^{p+1,\infty}[a,b]\}^n$.

Our next task is to bound the $L^\infty$-norm of $Y_I - Y^Q_k$ in terms of higher differences of $Y_I - Y^Q_{k-1}$ and so we rewrite (4.10) in the form

$$
\begin{aligned}
\text{i) } \overset{\circ}{B}{}^{Q}(Y_I - Y_k^Q, \sigma) \;=\; & Q_0[<A_1 - \overset{\circ}{A}{}^{Q})(x)(Y_I - Y_{k-1}^Q), \sigma>] + E_0[<A_1(x)(Y_I - Y_{k-1}^Q), \sigma>] \\
& - <A_1(x)(Y_I - Y_{k-1}^Q - (P_k Y_I - P_k Y_{k-1}^Q)), \sigma> \qquad (4.18) \\
& + Q_k[<f(x,y) - f(x, P_k Y_{k-1}^Q) - A_1(x)(P_k Y_I - P_k Y_{k-1}^Q), \sigma>] \\
& - E_k[<A_1(x)(P_k Y_I - P_k Y_{k-1}^Q), \sigma>] + E_k[<f(x,y), \sigma>]
\end{aligned}
$$

$$
\begin{aligned}
\text{ii) } \overset{\circ}{B}a^{Q}(Y_I - Y_k^Q)(a) + \overset{\circ}{B}b^{Q}(Y_I - Y_k^Q)(b) \;=\; & g(Y_{k-1}^Q(a), Y_{k-1}^Q(b)) - g(Y_I(a), Y_I(b)) \\
& + \overset{\circ}{B}a^{Q}(Y_I - Y_{k-1}^Q)(a) + \overset{\circ}{B}b^{Q}(Y_I - Y_{k-1}^Q)(b),
\end{aligned}
$$

cf. (3.16). We shall need the stronger Lipschitz condition on the derivative of f mentioned after (3.16) and the following stability and error results for $Q_k$ k=0,...,p-l :-

a) for v a vector of continuous functions

$$Q_k[<v,\sigma>] \le C\,\|v\|_{L\infty}\,\|\sigma\|_{L2} \quad ,$$

b) if $v \in \{W^{2(k+1),\infty}[a,b]\}^n$ (4.19)

$$E_k[<v,\sigma>] \le Ch^{2(k+1)}\,\|v\|_{W^{2(k+1),\infty}}\,\|\sigma\|_{L^2}$$

c) if A(x) is an nxn matrix with components in $W^{2,\infty}[a,b]$ and $V \in \{S_h\}^n$ sufficiently close to $Y_I$

$$E_k[<A(x)(P_k Y_I - P_k V), \sigma>] \le Ch^2 \max_{\sigma=0,1,2}\{|Y_I - Y|_{D\alpha}\}.$$

Thus provided $\|Y_I - Y_0^Q\|_{L^\infty}$ is $0(h^2)$ and $y \in \{W^{2k+3,\infty}[a,b]\}^n$. we may repeat the argument in (3-17) and (3.18) to obtain

$$\|Y_I - Y_k^Q\|_{D0} \le C(h^2 \max_{\sigma=0,1,2}\{|Y_I - Y_{k-1}|_{D\alpha}\} + h^{2(k+1)}\,\|y\|_{W^{2k+3,\infty}}). \qquad (4.20)$$

Finally a bound is obtained for the higher differences of $Y_I$-$Y^{Q}{}_k$ and the equation analogous to (3.19) is

$$
\begin{aligned}
\Delta^s(Y_I - Y_k^Q)_i(x_j) = h^{-1}\Delta^{s-1}\{ & Q_0[<A_1(x)(Y_I - Y_k^Q), \sigma_i^{j+}>] - Q_0[<(A_1 - \mathring{A}^Q)(x)(Y_I - Y_k^Q), \\
& \sigma_i^{j+}>] \qquad (4.21) \\
& + Q_0[<(A_1 - \mathring{A}^Q)(x)(Y_I - Y_k^Q), \sigma_i^{j+}>] + E_0[<A_1(x)(Y_I - Y_{k-1}^Q), \sigma_i^{j+}>] \\
& - <A_1(x)(Y_I - Y_{k-1}^Q - (P_k Y_I - P_k Y_{k-1}^Q)), \sigma_i^{j+}> \\
& + Q_k[<f(x,y) - f(x, P_k Y_{k-1}^Q) - A_1(x)(P_k - Y_I - P_k Y_{k-1}^Q), \sigma_i^{j+}>] \\
& - E_k[<A_1(x)(P_k - Y_I - P_k Y_{k-1}^Q), \sigma_i^{j+}>] + E_k[<f(x,y), \sigma_i^{j+}>]
\end{aligned}
$$

Our quadrature rules $\{Q_k\}$ must vary smoothly enough from one subinterval to the next so that divided differences do not introduce powers of h (cf.(4.15)) and so we need the conditions for $O \le s \le p-k-1$, $l \le k \le p-1$:—

a) if v is a vector of continuous functions

$$\max_{i=1,\ldots,n} \max_{j=1,\ldots,N-s} \{ | \Delta^s (Q_k [< v, \sigma_i^{j-} >] | \} \le Ch \max_{\alpha=0,\ldots,s} \{ | v |_{D\alpha} \} ,$$

b) if $v \in \{W^{2(k+l)+s}[a,b]\}^n$

$$\max_{i=1,\ldots,n} \max_{j=1,\ldots,N-s} \{ | \Delta^s (E_k [< v, \sigma_i^{j-} >] | \} \le Ch^{2k+3} \| v \|_{W^{2(k+1)+s}} \qquad (4.22$$

c) if A(x) is an nxn matrix with components in $W^{s+2,\infty}[a,b]$ and $V \in \{S_h\}^n$ sufficiently close to $Y_I$

$$\max_{i=1,\ldots,n} \max_{j=1,\ldots,N-s} \{ | \Delta^s (E_k [< A(x)(P_k Y_I - P_k V), \sigma_i^{j-} > I) | \} \le Ch^3 \max_{\alpha=0,\ldots,s+2} \{ | Y_I - V |_{D\alpha} \} ,$$

Then provided that $y \in W^{2k+3+s,\infty}[a,a]\}^n$, at the components of $A_1(x)$

are in $W^{P,\infty}[a,b]$, are the $\| Y_I - Y_{k-l}^Q \|_{L^\infty}$ is $0(h^2)$, we will have

$$| \Delta^s (Y_I - Y_k^Q)_i (x_j) | \le C \left( \sum_{\alpha=0}^{s-1} | Y_I - Y_k^Q |_{D^\alpha} + h^2 \sum_{\alpha=0}^{s+1} | Y_I - Y_k^Q |_{D^\alpha} + h^{2(k+1)} \| y \|_{W^{2k+3+s,\infty}} \right) \qquad (4.23$$

(cf.(3.22)) and can deduce by induction that

$$| Y_I - Y_k^Q |_{D^s} \le C h^{2(k+1)} \qquad (4.24$$

for $0 \le S \le p-k$ and $0 \le k \le p-1$. Consequently if $y \in \{w2^{p+1,\infty}[a,b]\}^n$ we will have

$$\| Y_I - Y_{P-1}^Q \|_{W^{1,\infty}} = O(h^{2P}) \qquad (4.25)$$

after p-1 iterations and the analogue with the exact integration case is maintained.

To include this section we make some remarks about the choice of quadrature rules Q, to satisfy (4.19) and (4.22). The most obvious idea is to base Q, on a composite Gauss-Legendre rule with (k+1) points in each subinterval. This may not, however, be the most efficient choice with regard to the number of function evaluations. Another possibility is to let s(x) be a (2k+l) degree spline interpolating the integrand at the knots and then integrating the spline exactly over each subinterval by the formula

$$\int_{x_j}^{x_{j+1}} s(x)dx \equiv h(s(x_j) + S(x_{j+1}))/2 - \sum_{m-1}^{k} B_{2m} h^{2m} (s^{(2m-1)}(xj+1) - s^{(2m-1)}(x_j)) / (2m)! \quad , \tag{4.26}$$

where the B's are the Bernouilli numbers. However special end conditions, as introduced in previous section, must be used, A third idea is to let p(x) be the (2k+l) degree polynomial which interpolates the integrand at the nodes $x_{j-k}$ , ,,..,$x_{j+k+1}$ . and to integrate this polynomial exactly over $[x_j,x_{j+1}]$. This integral may be computed easily from interpolation formulae, eg. pk $\int_{x_j}^{x_{j+1}}$ p (x) dx then

$$P_k = p_k\_1 + h\beta_k \delta^{2k}((p(x_i)+p(x_{i+1})/2) \tag{4.27}$$

wher $\delta$ is the usual central difference, $p_0 = h\,(p(x_j) + p\,(x_{j+1}))/2$ and

| $\beta_1$ | $\beta_2$ | $\beta_3$ | $\beta_4$ |
|---|---|---|---|
| −1/12 | 11/720 | −191/6040 | 2497/3628800 |

etc. In fact this technique, combined with the extension idea at the end of the previous section for obtaining nodal values outside [a,b] , would seem to be the most useful practical procedure since then (4.10) reduces to simple finite difference equations. Finally we make the point that if the quadrature rules $\{Q_k\}$ use only mesh point values then the resulting equations are independent of the choice of $\{P_k\}$ since the latter are nodal interpolatory mappings.

## 5. Generalizations

In this final section we make some remarks about how the previous defect correction results may be extended to a non-uniform mesh and to higher-order differential equations.

If the mesh is non-uniform with h denoting the length of the largest subinterval, it is easily checked that the existence and approximation results of section 2 still hold. The main difference with defect correction results of section 3 is that the non-uniformity of the mesh interferes with the smoothness of the divided differences of $Y_I$-$Y_k$ and in general we can only expect an improvement of 0(h) per iteration. This can easily be proved without any of the expansions or extra conditions developed in section 3 since we may work in the $H^1$ -norm. Thus using the defect iteration (3.14), and re-writing it as in (3.16), we only need the conditions

$$\begin{aligned} &\text{a)} \quad \|P_k V\|_{H^1} \leq C\|V\|_{H^1} \\ &\text{b)} \quad \|y - P_k Y_I\| \leq Ch^{k+2}\|y\|_{H^{k+2}} \end{aligned} \tag{5.1}$$

in order to deduce that

$$\|Y_I - Y_k\|_{H^1} \leq C(h\|Y_I - Y_{k-1}\|_{H^1} + \|Y_I - Y_{k-1}\|^2_{H^1} + h^{k+2}\|y\|_{H^{k+2}}). \tag{5.2}$$

Consequently if $y \in \{H^{p+1}[a,b]\}$ then after (p-1) iterations we shall have

$$\left\|Y_I - Y_{p-1}\right\|_{H^1} \leq 0(h^{p+1}). \tag{5.3}$$

Furthermore it is not difficult to develop conditions for the quadrature rules so that these results are retained when numerical integration is used. (Of course in practice one would only use a non-uniform mesh if y lacked smoothness or had large derivative values in certain sub-domains of [a,b]. The interplay of this phenomenon and defect correction is a difficult topic and not considered in this paper,)

With regard to higher-order differential equations, work is currently being carried out and we only make some observations here. Results for a single linear second-order problem were given in [1], using continuous

piecewise linear trial and test functions, and these can be generalized. One possibility for higher-order equations is to use $(2k—1)^{th}$ degree splines as trial and test functions for even order (2k) problems and $(2k-1)^{th}$ degree splines as trial functions with $(2k-2)^{th}$ degree splines as test functions for odd order (2k-l) problems. It may be considered however, this leads to matrices with a larger than necessary band—width. Alternatively one may try to generalize the ideas behind the $H^1$- and $H^{-1}$ - Galerkin methods [4].

References

1. Barrett, J.W., Moore, G., Morton, K.W.: Optimal recovery and defect correction in the finite element method. University of Reading Num. Anal. Report 11/83. Submitted to I.M.A. J. of Num.Anal.

2. Babuska, I., Aziz, A.K.: Survey lectures on the mathematical foundations of the finite element method. Proc. Symp. on The Mathematical Foundations of the Finite Element Method with Applications to Partial Differential Equations, Maryland (A.K. Aziz, ed.), 1972.

3. Christiansen, J., Russell, R.D.: Deferred corrections using uncentred end formulae. Numerische Math. 35,21-33 (1980).

4. Dupont, T.: A unified theory of superconvergence for Galerkin methods for two-point boundary problems. SIAM J. numer. Analysis 13, No.3, 362-368 (1976).

5. Fox, L.: Some improvements in the use of relaxation methods for the solution of ordinary and partial differential equations. Proc. Roy. Soc. A 190,31-59(1947).

6. Fox, L.: Numerical methods for boundary-value problems. Proc. Conf. on Computational Techniques for Ordinary Differential Equations, Manchester (I. Gladwell, D.K. Sayers, eds.), 1978.

7. Keller, H.B.: Numerical solution of two-point boundary-value problems. CBMS Regional Conference Series 24, SIAM, Philadelphia (1976).

8. Keller, H.B., Pereyra, V.: Difference methods and deferred corrections for ordinary boundary-value problems. SIAM J. numer. Analysis 16, No. 2, 241-259 (1979).

9. Lees, M.: Discrete methods for nonlinear two-point boundary value problems. Proc. Symp. on Numerical Solution of Partial Differential Equations, Maryland (J.H. Bramble, ed.), 1966.

10. Morton, K.W. : Finite element methods for non-self adjoint problems. Proc. SERC Summer School, Lancaster (P.R. Turner, ed.), Springer Lecture Notes in Maths 965,1982.

11. Pereyra, V.: High-order finite—difference solution of differential equations. Stanford University Comp. Sci. Dept. Report STAN-CA-73-348(1973).

12. Skeel, R.D.: A theoretical framework for proving accuracy results for deferred corrections. SIAM J. numer. Analysis 19, No. 1, 171-196 (1981).

13. Stetter, H.J.: The defect correction principle and discretization methods. Numerische Math. 29, 425-443(1978).