OPTIMAL RECOVERY AND DEFECT CORRECTION IN THE FINITE ELEMENT METHOD

J.W. BARRETT, G. MOORE & K.W. MORTON

NUMERICAL ANALYSIS REPORT 11/83
(REVISED VERSION)

1. INTRODUCTION

A feature essential to the success of the finite element method is its creation of approximations which are optimal in some integral energy norm. This leads to the need to recover local information about the solution, e.g. point-values or derivative point-values, when additional qualitative data, e.g. smoothness, monotonicity or positivity, are assumed. (This procedure contrasts with that for finite-difference methods since there the underlying philosophy is the approximation of point-values, and hence additional information would be generated by some form of grid interpolation). The design of algorithms for this purpose could be regarded as a problem within the general field of optimal recovery (Michelli & Rivlin, 1976) but we shall refer only to the now almost classical results of Golomb & Weinberger (1959). In the finite element literature, however, there are already a number of ideas concerned with improving the accuracy of the basic Galerkin approximation, e.g. superconvergence, local averaging and defect correction, and it is also our aim to bring some of these viewpoints together in a single framework. The detailed results we shall present are preliminary in the sense that they apply only to one-dimensional problems approximated by piecewise linear or piecewise constant functions, but the framework adopted will be quite general as the eventual aim is to develop similar methods for two- and three-dimensional problems and more general approximations.

Suppose then that $a(\cdot,\cdot)$ is a symmetric, bilinear, coercive form on $H\times H$, where H is a separable Hilbert space of functions: we denote by $||\cdot||_E$ the energy norm given by

$$||v||_E^2 \equiv a(v,v) \qquad (1.1)$$

and by H_E the space H equipped with this norm. Let $u \in H_E$ be the solution we seek, to a differential problem for which $a(\cdot,\cdot)$ is the associated form, then the Galerkin approximation u^h to u from a finite-dimensional subspace $S^h \subset H_E$ is given by

$$a(u^h,v^h) = a(u,v^h) \qquad \forall v^h \in S^h. \tag{1.2}$$

From this definition we have the fundamental error projection property

$$a(u-u^h,v^h) = 0 \qquad \forall v^h \in S^h \quad , \tag{1.3}$$

and the optimal approximation result

$$||u-u^h||_E = \inf ||u-v^h||_E \qquad \forall\, v^h \in S^h. \tag{1.4}$$

It is from these properties that one deduces the standard error bounds (see Strang & Fix, 1973), the superconvergence phenomena which are so important for the practical efficacy of the finite element method, and the adaptive mesh refinement strategies which are now being developed (Babuska & Rheinboldt (1978), Reinhardt (1981) and Gartland (1984).

In the one-dimensional case, with

$$a(u,v) \equiv \int_a^b (pu'v' + quv)dx \tag{1.5}$$

and either

(i) $p > 0, \quad q \geq 0$ with $H = H^1(a,b)$ and certain boundary conditions

or

(ii) $p = 0, \; q > 0$ with $H = L^2(a,b)$

there are two well-known examples of improved accuracy for piecewise linear approximation, i.e. S^h consisting of continuous piecewise linear functions based on certain fixed nodes $\{x_j\} \in [a,b]$ with h related to the maximum distance between nodes. These are the problems $p = 1$ and $q = 0$ in (1.5), which is derived from the O.D.E. $-u'' = f$ with Dirichlet or Dirichlet/ Neumann boundary conditions, and $p = 0$ with $q = 1$ which is the best L^2 fit for u. Thus (1.3) becomes

$$\text{a)} \quad \int_a^b (u-u^h)' v^{h'} dx = 0 \qquad \forall v^h \in S^h \tag{1.6}$$

$$\text{or} \quad \text{b)} \quad \int_a^b (u-u^h) v^h dx \ 0 \ 0 \qquad \forall v^h \in S^h.$$

With (1.6a) it is well-known that u^h is exact at the nodes, i.e. $u^h(x_j) = u(x_j)$. This is a consequence of the Green's function for the underlying O.D.E. being itself piecewise linear and this property is easily destroyed by changes to the problem, for example

$$\int_a^b p(u-u^h)' v^{h'} dx = 0 \tag{1.7}$$

with p not constant. With (1.6b), on a uniform mesh, we have the widely-quoted formula

$$u(x_j) - \frac{1}{12}\{u^h(x_{j-1}) + 10u^h(x_j) + u^h(x_{j+1})\} = O(h^4) \tag{1.8}$$

in the interior if $u \in W^{4,\infty}$ The reason for this local-averaging result is that with three pieces of data, namely $u^h(x_{j-1})$, $u^h(x_j)$ and $u^h(x_{j+1})$, one can find a quadratic polynomial which fits this data and, if u is sufficiently smooth, this quadratic will be an $O(h^3)$ approximation to u locally, just as for interpolation with polynomials. However at specific points, one of them being the mid-node x_j, an extra order of convergence is obtained since the error function for the next higher degree polynomial, i.e. cubic, passes through zero there. This idea is not so unstable with respect to changes in the problem and results like (1.8) still hold for weighted L^2 fits of the form

$$\int_a^b q(u-u^h) v^h dx = 0 \qquad q > 0. \tag{1.9}$$

The importance of such formulae for improved accuracy goes far beyond the above examples. For Barrett & Morton (1980, 1981, 1984) use a symmetrizing technique on the non-self-adjoint diffusion-convection problem

$$-(au')' + (bu)' = f, \tag{1.10}$$

which aims at an approximation which is optimal in a norm

$$\int_a^b (pv'^2 + qv^2)dx , \tag{1.11}$$

where $p = \rho a^2$, $q = \rho b^2 + (\rho ab)'$ and ρ is a positive weighting factor. For a dominantly convective problem, $a << b$, we will have $p << q$ and an approximation in a norm which is "close" to that for (1.9); with this type of problem too, local recovery formulae like (1.8) exist. Furthermore, as Morton (1982a) has shown, most currently used Petrov-Galerkin methods for problems like (1.10) can be regarded as approximations to an alternative symmetrization which leads to an optimal approximation in a norm like (1.11) but with $p = a$ and $q = 0$. Thus, for model constant coefficient problems, exact values are achieved at the nodes for arbitrary f, as in Hemker (1977) with exponential test functions, but for the general variable coefficient case we do not have local recovery formulae of the quality of (1.8).

These least-squares type problems are also of great significance for evolutionary problems and several authors have exploited the fact that, for such problems, Galerkin methods lead naturally to least-squares approximations to the solution - see Cullen & Morton (1980). At the very least, this fact needs to be taken account of in the graphical presentation of final results for problems where non-linear effects are pronounced. Moreover, it is increasingly being recognised that some sort of recovery or post-processing at each time step can give significantly improved accuracy - see van Leer (1979), Morton (1982b, 1984).

The formula (1.8) assumes that the underlying true solution u is smooth, but for many problems this assumption will not hold over at least part of the domain. In diffusion convection problems there may be sharp boundary layers and evolutionary problems may well involve shocks and other discontinuities. To deal with these different cases Barrett & Morton (1980) adopted the following general technique for local recovery. Any a priori data on u was incorporated into a local approximation u_R with a number of free parameters. Then the error projection property (1.3) was exploited by requiring that

$$a(u_R - u^h, \phi_j) = 0 \tag{1.12}$$

for a sufficient number of local basis functions ϕ_j of S^h to determine these parameters. Using (1.3), this approach can be regarded as determining u_R so that it fits known data functionals $a(u, \phi_j)$. In a part of the domain where u is smooth a polynomial form for u_R could be chosen, however, in, for example, a boundary layer, an exponential form would be more appropriate.

If we move on to the case where p and q are of comparable magnitude in (1.5), the above approach can still be followed but local recovery results are disappointing and it is better to recover globally, i.e. u_R satisfies

$$a(u_R - u^h, v^h) = 0 \qquad \forall v^h \in S^h \ . \tag{1.13}$$

Here u_R would belong to a space of the same dimension as S^h which, for smooth problems, would generally be chosen to be piecewise-polynomials of higher-order e.g. cubic splines with suitable end conditions. Clearly this formulation is equivalent to a Petrov-Galerkin approximation of u with piecewise linear test functions and higher-order trial functions,

$$a(u_R - u, v_h) = 0 \qquad \forall v^h \in S^h. \tag{1.14}$$

As such it is subsumed by standard Petrov-Galerkin error analysis which shows that u_R attains the appropriate higher-order of accuracy, e.g. fourth order for cubic splines, if u is sufficiently smooth. However the direct solution of (1.14) would hardly be efficient, especially if u^h

had already been computed, and so we propose that u_R should be calculated iteratively by using (1.14) in a defect correction mode. This involves only the solution of systems of equations with the symmetric, positive-definite, Galerkin stiffness matrix for S^h and different right-hand sides.

The ideas in this paper can often be considered within the framework of Golomb & Weinberger (1959), and this connection is noted at various points. Their basic problem is the approximation of an unknown element u of a Hilbert space H when certain data on u is given. More specifically it is assumed that

(i) the values of certain bounded linear functional $F_j(u)$ $j=1 \to n$ are given,

(ii) a bound for $b(u,u)$ is known, where $b(\cdot,\cdot)$ is a symmetric, bilinear, coercive form on H.

Their main result is that the "optimal" approximation to u is given by

$$\bar{u} = \sum_{j=1}^{n} \alpha_j \theta_j \quad , \tag{1.15}$$

where the θ_j are the representers of the F_j in H equipped with inner-product $b(\cdot,\cdot)$, and the α_j are determined by the conditions

$$F_j(\bar{u}) = F_j(u) \qquad j=1 \to n \quad . \tag{1.16}$$

If an approximation to any bounded linear functional of u is required, say $F(u)$, i.e. any further information about u is needed, then the best available estimate is $F(\bar{u})$, with the sharp error bound

$$|F(u) - F(\bar{u})|^2 \leqq \{b(\theta,\theta) - b(\bar{\theta},\bar{\theta})\}\{b(u,u) - b(\bar{u},\bar{u})\} \quad , \tag{1.17}$$

where θ is the representer of F and $\bar{\theta}$ its orthogonal projection with respect to $b(\cdot,\cdot)$ onto span$\{\theta_j\}$. From our viewpoint the F_j would be either $u^h(x_j)$ or $a(u,\phi_j)$ and $b(\cdot,\cdot)$ some integral energy norm which implied smoothness properties by including higher derivatives in its definition: thus, for instance, the nodal superconvergence of Galerkin approximations is a direct application of (1.17). Also although this framework seems inapplicable to the case where the qualitative assumptions on u are other than those of smoothness, our method of dealing with examples of such problems in §2.4 has underlying similarities.

The present paper is divided into two sections. In §2 we investigate recovery from weighted L^2 fits like (1.9), with the given data consisting of functionals of the form $\int qu\phi_j$ or $u^h(x_j)$. Local recovery formulae for smooth solutions u, generalising (1.8), are deduced in §2.2 and §2.3 and we specifically mention the connections with previous work. In §2.4 two examples of local recovery for non-smooth u are considered, but we note that the study of this type of problem is still at an early stage. In §3 recovery from O.D.E. formulations like (1.5) is investigated again with functionals $a(u,\phi_j)$ or $u^h(x_j)$ as data. §3.2 examines whether the techniques of §2 can be carried across and concludes that this is possible for the singular perturbation case $p \ll q$. The general O.D.E. bilinear form, for which local recovery is inappropriate, is considered in §3.3-3.5 and here we use global recovery, together with ideas similar to defect and deferred correction, to produce higher-order approximations efficiently. In both §2 and §3 numerical examples are given to illustrate the theoretical results.

2. RECOVERY FROM WEIGHTED L^2 BEST FITS

2.1 The Moment and Point Functionals

With the interval $[a,b]$ partitioned into elements with nodes $a = x_0 < x_1 < \ldots x_{J-1} < x_J = b$ we denote by S^h the space of continuous piecewise linear functions spanned by $\{\phi_j(x)\}$, where $\phi_j(x)$ is a standard basis function satisfying $\phi_j(x_k) = \delta_{jk}$, the Kronecker delta: we also denote by $(\cdot,\cdot)$ the standard $L^2(a,b)$ inner product and by $\|\cdot\|$ the corresponding norm. Given the following scaled moment functionals of an unknown function u :-

$$F_j^M(u) := (qu,\phi_j)/(q,\phi_j) \qquad \forall\phi_j \in S^h , \qquad (2.1.1)$$

where $q(x)$ is a smooth positive weighting function, we wish to estimate the value of $F(u)$, some other bounded linear functional of u. More specifically we will be interested in estimating point values of u and its derivative under given smoothness assumptions.

Under the minimum smoothness requirement on u, $\|u\|_q^2 \equiv (qu,u)$ bounded, it follows that the optimal approximation to any functional $F(u)$ bounded in L^2 is $F(u^h)$, where u^h is the best fit from S^h to u in $\|\cdot\|_q$; that is, where u^h is given by

$$(qu^h,\phi_j) = (qu,\phi_j) \qquad \forall\phi_j \in S^h \qquad (2.1.2)$$

(see Golomb & Weinberger (1959) p. 131). Suppose $g_F \in L^2(a,b)$ is the Riesz representer of $F(\cdot)$ with respect to $\|\cdot\|_q$, that is, $F(v) = (qv,g_F)$ for every $v \in L^2(a,b)$ and $g_F^h \in S^h$ its corresponding $\|\cdot\|_q$ best fit. Then we have

$$\begin{aligned}|F(u) - F(u^h)| &= |(q(u-u^h),g_F)|\\ &= |(q(u-u^h),\ g_F-g_F^h)|\\ &\le \|u-u^h\|_q\ \|g_F-g_F^h\|_q\\ &= |F(g_F-g_F^h)|^{\frac{1}{2}}\left[\|u\|_q^2 - \|u^h\|_q^2\right]^{\frac{1}{2}} \qquad (2.1.3)\end{aligned}$$

and this bound is sharp as it is attained for $u = g_F$.

The key to the above is that u^h is the best fit to u in $\|\cdot\|_q$ from the span of the $\|\cdot\|_q$ representers of the known moment functionals. Unfortunately, due to the lack of smoothness assumed, this framework as it stands cannot be used for estimating point values of u. In assuming more smoothness on u, say $\|u^{(2)}\|$ bounded, the Golomb & Weinberger theory requires one to construct a higher order piecewise polynomial approximation to u; namely, the best fit to u in $\|\cdot^{(2)}\|$ from the span of the $\|\cdot^{(2)}\|$ representers of the known moment functionals. This requirement moves away from our aims as we are interested in estimating point values of an unknown solution u to a differential equation, where in doing so we have already computed a piecewise linear, or some other low order, approximation to u. Thus we wish to estimate the point values of u and its derivative either directly from the given moment functionals or from a low order piecewise polynomial approximation. We will return to the ordinary differential equation problem and its connections to Golomb & Weinberger theory in §3. In this section we will concentrate on the pure approximation problem in the weighted L^2 norm.

Given the data (2.1.1), then we wish to estimate point values of u from a small number of either these moment functionals or the point functionals

$$F_j^P(u) := u^h(x_j) \qquad j = 0 \rightarrow J, \tag{2.1.4}$$

where u^h is defined by (2.1.2). Consider the case of a uniform mesh of length h and $q(x) \equiv 1$. Under these circumstances we have for $u \in C^4(a,b)$ that

$$F_j^M(u) \equiv h^{-1}(u,\phi_j) = \left(1 + \frac{\delta^2}{12}\right) u(x_j) + O(h^4) \qquad j = 1 \rightarrow J-1, \tag{2.1.5}$$

where $\delta u(x_j) := u(\frac{1}{2}[x_j + x_{j+1}]) - u(\frac{1}{2}[x_{j-1} + x_j])$. Operating on both sides of (2.1.5) by $\left(1 - \frac{\delta^2}{12}\right)$ we obtain

$$\left(1 - \frac{\delta^2}{12}\right) F^M_j(u) = u(x_j) + O(h^4) \qquad j = 2 \to J - 2 \quad . \qquad (2.1.6)$$

Therefore a local average of 3 moment functionals, $(-F^M_{j-1}(u) + 14F^M_j(u) - F^M_{j+1}(u))/12$, yields an $O(h^4)$ approximation to $u(x_j)$. Clearly the above approach can be generalised to obtain an arbitrarily high order approximation to $u(x_j)$, provided u is sufficiently smooth and x_j is far enough away from the boundaries, by taking a local average over a larger number of moment functionals.

In addition from (2.1.2) we have that

$$\left(1 + \frac{\delta^2}{6}\right) u^h(x_j) = F^M_j(u) \qquad j = 1 \to J - 1 \qquad (2.1.7)$$

and so local averages of the point functionals yield higher order approximations to $u(x_j)$. For example, combining (2.1.6) and (2.1.7) yields

$$(-F^P_{j-2}(u) + 10F^P_{j-1}(u) + 54F^P_j(u) + 10F^P_{j+1}(u) - F^P_{j+2}(u))/72$$

as an $O(h^4)$ approximation to $u(x_j)$. This result is equivalent to that obtained by the local averaging technique given by Bramble & Schatz (1976). In that paper they present various averaging rules for the best L^2 spline approximations on a uniform mesh. In particular they show that presented with $2m + 1$ point functionals centred about the mesh point x_k, then one can take a local average of them yielding an $O(h^{2m+2})$ approximation to $u(x_k)$. In subsection 2.2 we extend their results for piecewise constants and continuous piecewise linears as follows:- given n successive moment functionals $\{F^M_{k+j}(u)\}^n_{j=1}$, we show that for any $x \in [x_k, x_{k+n+1}]$ there exists a local average of them, depending on x, which yields an $O(h^n)$ approximation to $u(x_k)$. In addition, we show that there exist n points of superconvergence where this local average yields an $O(h^{n+1})$ approximation. Moreover, our results are valid for a general $q \in C[a,b]$ and a non-uniform mesh. We believe our approach to be more transparent and note that it can be generalised to higher order approximation spaces.

Although local averaging formulae based on the point functionals can be obtained from those based on the moment functionals via (2.1.7) or its analogue in the case of a non-uniform mesh and general q, more compact formulae may be obtained by considering the point functionals directly. It is very instructive to ignore boundary effects for the present and consider the case of u^h being the best piecewise linear L^2 fit on a uniform mesh to u over the infinite real line. In this case for $u \in C^4(\mathbb{R})$, combining (2.1.5) and (2.1.7) and operating on both sides by $(1 + \delta^2/6)^{-1}$ yields

$$u^h(x_j) = \left(1 - \frac{\delta^2}{12}\right) u(x_j) + O(h^4) \quad . \qquad (2.1.8)$$

Although u^h is only an $O(h^2)$ approximation to u at the nodes x_j, from (2.1.8) we may first deduce that there exist points of superconvergence:-

$$\begin{aligned} u^h(x_j + \lambda h) &= (1 - \lambda)u^h(x_j) + \lambda u^h(x_{j+1}) \\ &= (1 - \lambda)\left(1 - \frac{\delta^2}{12}\right) u(x_j) + \lambda\left(1 - \frac{\delta^2}{12}\right)u(x_{j+1}) + O(h^4) \\ &= u(x_j + \lambda h) - \frac{h^2}{12}[6\lambda^2 - 6\lambda + 1]u''(x_j + \lambda h) + O(h^3); \end{aligned} \qquad (2.1.9)$$

that is, u^h is $O(h^3)$ accurate where $\lambda = \frac{1}{2}[1 \pm \frac{1}{\sqrt{3}}]$, the Gauss points corresponding to the roots of the second degree Legendre polynomial. Secondly, the relationship (2.1.8) also implies that the local average

$$(F^P_{j-1}(u) + 10F^P_j(u) + F^P_{j+1}(u))/12$$

is an $O(h^4)$ approximation to $u(x_j)$ differing from the earlier formula by $\delta^4 F^P_j(u)/72$. Once again this result can be generalised to obtain an arbitrarily high order approximation to $u(x_j)$, provided u is sufficiently smooth. Thus one can see that different recovery formulae are obtained if one works directly with the point rather than the moment functionals. Note that in the second case the piecewise linear best fit does not need to be computed, although the results can always be given in terms of the $\{F_j^P(u)\}$.

Unfortunately, recovering from the point functionals is not a robust procedure and many of the results are lost as one moves away from a uniform grid over the infinite real line and $q \equiv 1$. This is not surprising because of the link between the best L^2 approximation by continuous piecewise linears and the cubic spline interpolation problem. It is well known for the latter that if the end conditions are not chosen appropriately many superconvergence results in the interior are lost, see Lucas (1974). We discuss this connection and present our limited results for point functionals in subsection 2.3.

One should also note that for some approximation spaces working with the moment functionals is equivalent to working with the point functionals. An example of this occurs when S^h is the space of piecewise constant functions spanned by $\{\chi_j(x)\}$, where $\chi_j(x)$ is the characteristic function for the element $(x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}})$. For $u^h \in S^h$ the moment functional is defined by

$$F_j^M(u^h) = F_j^M(u) := (qu,\chi_j)/(q,\chi_j) \qquad \forall \chi_j \in S^h \tag{2.1.10}$$

and because $F_j^M(\chi_i) = \delta_{ij}$ we have

$$F_j^P(u) := u^h(x_j) = F_j^M(u). \tag{2.1.11}$$

Thus the moment and point functionals are equivalent.

The recovery procedures using the moment and point functionals described above and in subsections 2.2 and 2.3 assume that the underlying solution u is sufficiently smooth. Finally, in subsection 2.4 we address the problem of non-smooth and rapidly varying functions, e.g. shocks and boundary layers.

2.2 Recovery from the Moment Functionals

2.2.1 General mesh

Although the framework we adopt is quite general, for ease of exposition we restrict ourselves to studying only moment functionals arising from either piecewise constant or continuous piecewise linear approximation spaces. These two choices enable us to convey the main ideas which can be generalised to higher order approximation spaces. With the notation of the previous section we are presented with either n consecutive piecewise constant moment functionals

$$F^M_{k+i}(u) = (qu,\chi_{k+i})/(q,\chi_{k+i}) \quad i = 1 \to n \tag{2.2.1a}$$

or n consecutive continuous piecewise linear moment functionals

$$F^M_{k+i}(u) = (qu,\phi_{k+i})/(q,\phi_{k+i}) \quad i = 1 \to n \tag{2.2.1b}$$

where $q \in C[a,b]$ and $q(x) > 0 \quad \forall x \in [a,b]$. In addition, we introduce the following notation. We denote by $I_{n,k}$ the union of support of the basis functions for $k+1 \to k+n$ and by $h_{n,k}$ the length of the maximum element. Thus we have $I_{n,k} \equiv [x_{k+\frac{1}{2}},x_{k+n+\frac{1}{2}}]$ and $h_{n,k} = \max\limits_{i=1\to n} (x_{k+i+\frac{1}{2}}-x_{k+i-\frac{1}{2}})$ for piecewise constants, and $I_{n,k} \equiv [x_k,x_{k+n+1}]$ and $h_{n,k} = \max\limits_{i=0\to n} (x_{k+i+1}-x_{k+i})$ for piecewise linears. We shall also denote by $W^{m,p}(I_{n,k})$ the usual Sobolev norm involving derivatives up to order m in the $L^p(I_{n,k})$ norm.

Lemma 2.1

For any $v \in C[a,b]$, then $F^M_{k+i}(v) = 0 \quad i = 1 \to n$ implies that v has n zeroes in $I_{n,k}$.

Proof: For piecewise constants introduce w such that $w' = qv$ on $I_{n,k}$ and $w(x_{k+\frac{1}{2}}) = 0$. Then $F^M_{k+i}(v) = 0$ implies that $(w',\chi_{k+i}) = w(x_{k+i+\frac{1}{2}}) - w(x_{k+i-\frac{1}{2}}) = 0 \quad i = 1 \to n$. Therefore, we have $w(x_{k+i+\frac{1}{2}}) = 0 \quad i = 0 \to n$ implying that w' has n zeroes in $I_{n,k}$, which yields the desired result as $q > 0$.

For piecewise linears introduce w such that $w'' = qv$ on $I_{n,k}$ with $w(x_k) = w(x_{k+n+1}) = 0$, assuming the moment functionals are in the interior; that is, $k \neq -1$ and $k \neq J - n$. Then $F^M_{k+i}(v) = 0$ implies that $(w'', \phi_{k+i}) = \Delta_-(\Delta_+ w(x_{k+i})/\Delta_+ x_{k+i}) = 0$ $i = 1 \rightarrow n$, where Δ_-, Δ_+ are the usual backward and forward difference operators. Therefore, imposing the boundary conditions, we have $w(x_{k+i}) = 0$ $i = 0 \rightarrow n+1$ implying that w'' and hence v has n zeroes in $I_{n,k}$. The above proof can be adapted by replacing the boundary condition $w(x_k) = 0$ by $w'(x_{k+1}) = 0$ if $k = -1$ and $w(x_{k+n+1}) = 0$ by $w'(x_{k+n}) = 0$ if $k = J - n$. ■

Corollary

There exist unique $(n-1)^{th}$ degree polynomials $c^M_{n,k+j}(x)$ $j = 1 \rightarrow n$ such that

$$F^M_{k+i}\ (c^M_{n,k+j}) = \delta_{ij} \qquad i = 1 \rightarrow n \quad ; \tag{2.2.2}$$

and thus any $(n-1)^{th}$ degree polynomial $p_{n-1}(x)$ can be written in the form

$$p_{n-1}(x) = \sum_{j=1}^{n} F^M_{k+j}\ (p_{n-1}) c^M_{n,k+j}(x). \tag{2.2.3}$$

Proof: With $\{\psi_\ell(x)\}^n_{\ell=1}$ a basis for polynomials of degree $n-1$ we need to show that there exist unique coefficients $b^M_{\ell,j}$ such that

$$c^M_{n,k+j}(x) = \sum_{\ell=1}^{n} b^M_{\ell,j}\psi_\ell(x) \qquad j = 1 \rightarrow n.$$

The equations (2.2.2) imply that $A^M B^M = I_n$, where A^M and B^M are nxn matrices with entries $a^M_{i,\ell} \equiv F^M_{k+i}(\psi_\ell)$ and $b^M_{\ell,j}$, respectively; and I_n is the nxn identity. From Lemma 2.1 we deduce that if n consecutive moment functionals of a $(n-1)^{th}$ degree polynomial $p_{n-1}(x)$ are zero then $p_{n-1}(x) \equiv 0$. This implies the invertibility of the matrix A^M and hence the existence and uniqueness of the polynomials $c^M_{n,k+j}(x)$ $j = 1 \rightarrow n$. As these polynomials are linearly independent the result (2.2.3) follows immediately. ■

We now have the main result of this subsection:-

Theorem 2.1

Given n consecutive moment functionals, $F^M_{k+i}(u)$ $i = 1 \to n$, of an unknown function u. Then if $u \in W^{n,\infty}(I_{n,k})$ it follows that

$$|u(x) - \sum_{i=1}^{n} F^M_{k+i}(u)c^M_{n,k+i}(x)| \leq Ch^n_{n,k} \; \|u^{(n)}\|_{L^\infty(I_{n,k})} \qquad \forall x \in I_{n,k} \; . \tag{2.2.4}$$

Moreover, there exist n points of superconvergence $z^M_1, z^M_2, \ldots, z^M_n$ in $I_{n,k}$, roots of the n^{th} degree monic polynomial $Z^M_{n,k}(x)$ such that

$$F^M_{k+i}(Z^M_{n,k}) = 0 \qquad i = 1 \to n \; , \tag{2.2.5}$$

at which if $u \in W^{n+1,\infty}(I_{n,k})$ we have

$$|u(z^M_\ell) - \sum_{i=1}^{n} F^M_{k+i}(u)c^M_{n,k+i}(z^M_\ell)| \leq Ch^{n+1}_{n,k} \; \|u^{(n+1)}\|_{L^\infty(I_{n,k})} \quad \ell = 1 \to n. \tag{2.2.6}$$

Proof: The $(n-1)^{th}$ degree polynomial

$$p_{n-1}(x) = \sum_{j=1}^{n} F^M_{k+j}(u) \; c^M_{n,k+j}(x) \tag{2.2.7}$$

is such that $F^M_{k+i}(u - p_{n-1}) = 0$ $i = 1 \to n$. From Lemma 2.1 it follows that p_{n-1} interpolates u at n distinct points $\{\xi_i\}^n_{i=1}$ in $I_{n,k}$. Therefore the error between u and p_{n-1} can be bounded using the standard Cauchy remainder for polynomial interpolation to yield

$$|u(x) - \sum_{i=1}^{n} F^M_{k+i}(u) \; c^M_{n,k+i}(x)| \leq \frac{1}{n!} | \prod_{i=1}^{n}(x-\xi_i)| \quad \|u^{(n)}\|_{L^\infty(I_{n,k})}$$

$$\leq Ch^n_{n,k} \quad \|u^{(n)}\|_{L^\infty(I_{n,k})} \qquad \forall x \in I_{n,k} \; .$$

From the proof of Lemma 2.1 we deduce that $\xi_i \in [x_{k+i-\frac{1}{2}}, x_{k+i+\frac{1}{2}}]$ for piecewise constants and $\xi_i \in [x_{k+i-1}, x_{k+i+1}]$ for piecewise linears implying that the constant C in (2.2.4) is bounded above by 1 and $n+1$ in the respective cases.

The n^{th} degree monic polynomial

$$Z^M_{n,k}(x) = x^n - \sum_{j=1}^{n} F^M_{k+j}(x^n)c^M_{n,k+j}(x) \tag{2.2.8}$$

is such that $F^M_{k+i}(Z^M_{n,k}) = 0 \quad i = 1 \to n$. Hence the polynomial p_{n-1}, defined by (2.2.7), interpolates $u - \beta Z^M_{n,k}$ for any $\beta \in \mathbb{R}$ at n distinct points $\{\eta_i(\beta)\}^n_{i=1}$ in $I_{n,k}$ and so

$$|u(x) - \beta Z^M_{n,k}(x) - p_{n-1}(x)| \leq \frac{1}{n!}\left|\prod_{i=1}^{n} (x-\eta_i)\right| \; \|u^{(n)}-\beta n!\|_{L^\infty(I_{n,k})}$$

$$\forall x \in I_{n,k} \; . \tag{2.2.9}$$

Choosing $\beta = u^{(n)}(x_{k+\frac{1}{2}(n+1)})/n!$ and noting from Lemma 2.1 that $Z^M_{n,k}$ has n zeroes $\{z^M_\ell\}^n_{\ell=1}$ in $I_{n,k}$ yields the desired result (2.2.6) for $u \in W^{n+1,\infty}(I_{n,k})$. ∎

Corollary

Since u' is interpolated by p'_{n-1}, where p_{n-1} is given by (2.2.7) at $n-1$ distinct points it follows that

$$|u'(x) - \sum_{i=1}^{n} F^M_{k+i}(u)c^{M'}_{n,k+i}(x)| \leq Ch^{n-1}_{n,k} \; \|u^{(n)}\|_{L^\infty(I_{n,k})} \qquad \forall x \in I_{n,k} \; . \tag{2.2.10}$$

Moreover, there exist $n-1$ points of superconvergence $z^{M'}_1, z^{M'}_2, \ldots, z^{M'}_{n-1}$ in $I_{n,k}$, roots of $Z^{M'}_{n,k}(x)$ where $Z^M_{n,k}(x)$ is given by (2.2.8), at which the order of approximation is increased to n if $u \in W^{n+1,\infty}(I_{n,k})$.

2.2.2 Uniform mesh, $q \equiv 1$ and linear elements

We illustrate the recovery procedure using moment functionals by explicitly generating the formulae for the case of continuous piecewise linears on a uniform mesh of length h with the weighting function q chosen to be identically one. Taking an odd number of interior moment functionals,

$n = 2m + 1$, implies that the mid-point of the interval $I_{n,k}$ is by symmetry a point of superconvergence for approximating the function. Thus to exploit symmetry it is desirable to consider a collection of moment functionals $\{F^M_{k+i}(u) \equiv h^{-1}(u, \phi_{k+i})\}^m_{i=-m}$ centred about the fixed, superconvergent, point x_k. In practice the Lagrangian basis $\{c^M_{n,k+j}(x)\}^m_{j=-m}$ satisfying

$$F^M_{k+i}(c^M_{n,k+j}) = \delta_{ij} \qquad i = -m \rightarrow m \tag{2.2.11}$$

is not ideal, since if m is increased one has to recompute the complete set of basis functions. It is far better to use a Newton type basis $\{d^M_{k+j}(x)\}$, where $d^M_{k+j}(x)$ is a polynomial of degree $2j$ if $j \geqq 0$, and of degree $-2j - 1$ if $j < 0$ satisfying .

$$d^M_k(x) \equiv 1 \quad ,$$

$$F^M_{k+i}(d^M_{k+j}) = 0 \qquad i = -j \rightarrow j-1 \qquad j \geqq 1 \tag{2.2.12}$$

and

$$F^M_{k+i}(d^M_{k+j}) = 0 \qquad i = -(j+1) \rightarrow j+1 \qquad j \leqq -1.$$

For a uniform mesh and $q \equiv 1$ the polynomials $d^M_{k+j}(x)$ are of the form

$$d^M_{k+j}(x) \equiv w^M_j\left(\frac{x - x_k}{h}\right) \quad . \tag{2.2.13}$$

Clearly through symmetry $w^M_j(t)$ is an odd function about $t = 0$ for j negative and an even function about $t = -\frac{1}{2}$ for j positive. For what follows though it is convenient to express them in the following way

$$w^M_0(t) \equiv 1$$

$$w^M_{-j}(t) = t^{2j-1} + \sum_{\ell=-(j-1)}^{j-1} \alpha^M_{-j,\ell}\, w^M_\ell(t) \qquad j \geqq 1$$

and

$$w^M_j(t) = (t+\tfrac{1}{2})^{2j} + \sum_{\ell=-j}^{j-1} \alpha^M_{j,\ell} w^M_\ell(t) \qquad j \geqq 1 \; , \tag{2.2.14}$$

where the coefficients $\alpha^M_{j,\ell}$ are defined by the relationships (2.2.12).

Since $\delta^2 F^M_k(v) \equiv F^M_k(\delta^2 v)$ and $\delta F^M_{k-\frac{1}{2}}(v) \equiv F^M_k(\delta T_{-\frac{1}{2}} v)$, where $T_{-\frac{1}{2}} v(x) = v\ (x + \frac{1}{2}h)$, it follows that

$$\alpha^M_{j,\ell} = \begin{cases} -\left[\delta^{2\ell} F^M_0\left(\left(\frac{x}{h}\right)^{2j}\right)\right] \Big/ (2\ell)! & \ell = 0 \to j-1 \quad j \geq 1, \\ -\left[\delta^{2|\ell|-1} F^M_{-\frac{1}{2}}\left(\left(\frac{x}{h}\right)^{2j}\right)\right] \Big/ (2|\ell| - 1)! & \ell = -j \to -1 \quad j \geq 1; \end{cases}$$

$$\alpha^M_{-j,\ell} = \begin{cases} -\left[\delta^{2\ell} F^M_0\left(\left(\frac{x}{h}+\frac{1}{2}\right)^{2j-1}\right)\right] \Big/ (2\ell)! & \ell = 0 \to j-1 \quad j \geq 1, \\ -\left[\delta^{2|\ell|-1} F^M_{-\frac{1}{2}}\left(\left(\frac{x}{h}+\frac{1}{2}\right)^{2j-1}\right)\right] \Big/ (2|\ell|-1)! & \ell = -(j-1) \to -1 \quad j \geq 2. \end{cases} \tag{2.2.15}$$

As our recovery formula is exact for all polynomials of degree $2m$ it is a simple matter to show that it takes the form

$$\sum_{j=0}^{m} w^M_j(t) \frac{\delta^{2j} F^M_k(u)}{(2j)!} + \sum_{j=1}^{m} w^M_{-j}(t) \frac{\delta^{2j-1} F^M_{k-\frac{1}{2}}(u)}{(2j-1)!} \quad , \tag{2.2.16}$$

yielding an $O(h^{2m+1})$ approximation to $u(x_k + th)$ for $|t| \leq m+1$.

Introducing $C^M_{2j} := w^M_j(0)/(2j)!$ $j \geq 0$ we deduce from (2.2.14) and (2.2.15) that (2.2.16) evaluated at the superconvergent point $t=0$, $O(h^{2m+2})$, reduces to

$$\sum_{j=0}^{m} C^M_{2j}\ \delta^{2j} F^M_k(u) \quad , \tag{2.2.17}$$

where

$$C^M_{2j} = -\left\{ \sum_{\ell=0}^{j-1} \delta^{2\ell} F^M_0 \left(\left(\frac{x}{h}\right)^{2j}\right) C^M_{2\ell} \right\} \Big/ (2j)! \qquad j \geq 1$$

and

$$C^M_0 = 1 \ . \tag{2.2.18}$$

A simple calculation yields that

$$\delta^{2\ell} F^M_0\left(\left(\frac{x}{h}\right)^{2j}\right) = \frac{\delta^{2\ell+2}\left[\left(\frac{x}{h}\right)^{2j+2}\right]_{x=0}}{(2j+2)(2j+1)} \qquad \ell = 0 \to j-1, \quad j \geq 1. \tag{2.2.19}$$

Thus using the results (2.2.18) and (2.2.19) it is very simple to generate the coefficients C_{2j}^M and we obtain

$$C_2^M = -\frac{1}{12}, \quad C_4^M = \frac{1}{90}, \quad C_6^M = -\frac{1}{560} \quad \ldots \quad . \tag{2.2.20}$$

The error in the approximation (2.2.16) can be expressed in the form

$$E_{2m+1,k}^M(u,t) := u(x_k+th) - \int_{-(m+1)}^{m+1} g_{2m+1}^M(t,s)u(x_k+sh)ds, \tag{2.2.21}$$

where

$$g_{2m+1}^M(t,s) := \sum_{j=0}^{m} w_j^M(t)\frac{\delta^{2j}\phi_0(sh)}{(2j)!} + \sum_{j=1}^{m} w_{-j}^M(t)\frac{\delta^{2j-2}[\phi_0(sh)-\phi_{-1}(sh)]}{(2j-1)!}. \tag{2.2.22}$$

Expanding $u(x)$ in a Taylor series about x_k we find

$$u(x) = \sum_{j=0}^{2m+1} \frac{(x-x_k)^j}{j!} u^{(j)}(x_k) + \int_{x_k}^{x} \frac{(x-s)^{2m+1}}{(2m+1)!} u^{(2m+2)}(s)ds$$

$$= P_{2m+1}(x) + R_{2m+1}(x).$$

Since $E_{2m+1,k}^M(P_{2m+1},0) = 0$ and $R_{2m+1}(x_k) = 0$ we find (by interchanging the order of integration in the double integral for the error) that the error at the superconvergent point x_k is

$$E_{2m+1,k}^M(u,0) = -\int_{-(m+1)}^{m+1} g_{2m+1}^M(0,s)\, R_{2m+1}(x_k+sh)\, ds$$

$$= \int_{-(m+1)}^{m+1} G_{2m+1}^M(0,s)u^{(2m+2)}(x_k+sh)\, ds\ , \tag{2.2.23}$$

where

$$G_{2m+1}^M(0,s) := \frac{h^{2m+2}}{(2m+1)!}\int_{-(m+1)}^{s} g_{2m+1}^M(0,y)(s-y)^{2m+1}\, dy \qquad -(m+1) \leqq s \leqq 0 \tag{2.2.24}$$

and

$$G_{2m+1}^M(0,s) := G_{2m+1}^M(0,-s) \qquad 0 \leqq s \leqq m+1\ .$$

Thus we have

$$|E^M_{2m+1,k}(u,0)| \le \left[\int_{-(m+1)}^{m+1} |G^M_{2m+1}(0,s)|\; ds\right] \; \|u^{(2m+2)}\|_{L^\infty[x_{k-m-1},x_{k+m+1}]} \; . \qquad (2.2.25)$$

If the kernel $G^M_{2m+1}(0,s)$ is of one sign it becomes a simple matter to calculate the error constant, as we have using (2.2.23)

$$\begin{aligned}\int_{-(m+1)}^{m+1} |G^M_{2m+1}(0,s)|\; ds &= \left|\int_{-(m+1)}^{m+1} G^M_{2m+1}(0,s)\; ds\right| \\ &= \frac{h^{2m+2}}{(2m+2)!}\left|\int_{-(m+1)}^{m+1} g^M_{2m+1}(0,s)s^{2m+2}ds\right| \quad . \qquad (2.2.26)\end{aligned}$$

A simple calculation using (2.2.22) and (2.2.19) yields that the right-hand side of (2.2.26) is simply $|C^M_{2m+2}|\; h^{2m+2}$. Therefore if $G^M_{2m+1}(0,s)$ is of one sign we have that

$$|u(x_k) - \sum_{j=0}^{m} C^M_{2j}\; \delta^2 F^M_k(u)| \le |C^M_{2m+2}|\; h^{2m+2} \; \|u^{(2m+2)}\|_{L^\infty[x_{k-m-1},x_{k+m+1}]} . \qquad (2.2.27)$$

We have checked explicitly that $G^M_{2m+1}(0,s)$ is of one sign for $m = 0 \to 2$ and we conjecture that it is true for all m. As an example we have from (2.2.20)

$$|u(x_k) - \left(1 - \frac{\delta^2}{12} + \frac{\delta^4}{90}\right) F^M_k(u) \;| \le \frac{h^6}{560}\|u^{(6)}\|_{L^\infty[x_{k-3},x_{k+3}]} \; . \qquad (2.2.28)$$

In a similar manner to the above one can obtain a simple recovery formula for the derivative of u. It follows from the corollary to Theorem 2.1 that

$$\sum_{j=0}^{m-1} w^{M'}_j(t)\,\frac{\delta^{2j}F^M_k(u)}{(2j)!\; h} \quad + \quad \sum_{j=1}^{m} w^{M'}_{-j}\;(t)\;\frac{\delta^{2j-1}F^M_{k-\frac{1}{2}}(u)}{(2j-1)!h} \qquad (2.2.29)$$

is an $O(h^{2m-1})$ approximation to $u'(x_k+th)$ for $-(m+1) \le t \le m$.

Introducing $C^M_{2j-1} := w^{M'}_{-j}(-\frac{1}{2})/(2j-1)!$ $\quad j \geqq 1$ it follows from (2.2.14) and (2.2.15) that (2.2.29) evaluated at the superconvergent point $t = -\frac{1}{2}$, $O(h^{2m})$, reduces to

$$h^{-1} \sum_{j=1}^{m} C^M_{2j-1} \, \delta^{2j-1} \, F^M_{k-\frac{1}{2}}(u) \quad , \tag{2.2.30}$$

where

$$C^M_{2j-1} = -\left\{ \sum_{\ell=1}^{j-1} \delta^{2\ell-1} \, F^M_{-\frac{1}{2}} \left(\left(\frac{x}{h} + \frac{1}{2} \right)^{2j-1} \right) C^M_{2\ell-1} \right\} /(2j-1)! \qquad j \geqq 2 \tag{2.2.31}$$

and

$$C^M_1 = 1.$$

A similar calculation to (2.2.19) yields that

$$\delta^{2\ell-1} \, F^M_{-\frac{1}{2}} \left(\left(\frac{x}{h} + \frac{1}{2} \right)^{2j-1} \right) = \frac{\delta^{2\ell+1} \left[\left(\frac{x}{h} \right)^{2j+1} \right]_{x=0}}{(2j+1)2j} \qquad \ell = 1 \to j-1, \quad j \geqq 2 \; . \tag{2.2.32}$$

Once again if the corresponding kernel is of one sign, which we conjecture to be true for all m, we have

$$\left| u'(x_{k-\frac{1}{2}}) - h^{-1} \sum_{j=1}^{m} C^M_{2j-1} \, \delta^{2j-1} \, F^M_{k-\frac{1}{2}}(u) \right| \leqq \left| C^M_{2m+1} \right| h^{2m} \left\| u^{(2m+1)} \right\|_{L^\infty[x_{k-m-1}, x_{k+m}]}. \tag{2.2.33}$$

As an example we have from (2.2.31) and (2.2.32) that $C^M_3 = -\frac{1}{8}$ and $C^M_5 = \frac{37}{1920}$ implying that

$$\left| u'(x_{k-\frac{1}{2}}) - h^{-1} \left(1 - \frac{\delta^2}{8} \right) \delta F^M_{k-\frac{1}{2}}(u) \right| \leqq \frac{37}{1920} h^4 \left\| u^{(5)} \right\|_{L[x_{k-3}, x_{k+2}]}. \tag{2.2.34}$$

In addition, from the construction (2.2.12) and (2.2.13) of the polynomials w^M_j we deduce that all zeroes of $w^M_{-(m+1)}(t)$ are points of superconvergence for the recovery formula (2.2.16) and zeroes of $w^{M'}_m(t)$ are superconvergence points of the recovery formula (2.2.29). Therefore it is useful to tabulate these polynomials. Using (2.2.14) and (2.2.15) we obtain for example

$$\begin{aligned} w_{-1}^{M}(t) &= t \\ w_{1}^{M}(t) &= t^2 + t - \frac{1}{6} \\ w_{-2}^{M}(t) &= t^3 - \frac{3}{2}t \\ w_{2}^{M}(t) &= t^4 + 2t^3 - 2t^2 - 3t + \frac{4}{15} \ . \end{aligned} \tag{2.2.35}$$

As an example we have $(1 + \lambda)F_k^M(u) - \lambda F_{k-1}^M(u)$ is an $O(h^3)$ approximation to $u(x_k + \lambda h)$ if $\lambda = \frac{1}{2}(-1 \pm \sqrt{\frac{5}{3}})$, roots of $w_1^M(t)$. This is the moment functional analogue of the result (2.1.9).

2.2.3 Numerical Examples

We now describe various numerical examples to illustrate how the techniques described in this section perform in practice. Throughout we consider moment functionals generated from continuous piecewise linear functions and take $u(x)=e^x$. All the moment functionals have been evaluated using a 4-point Gauss rule on each element; this is of sufficiently high accuracy that we may ignore this quadrature error.

In Table 2.1, where we adopt the notation $0.35(-4) \equiv 0.35 \times 10^{-4}$, we present results in the case of a uniform mesh of length $h = 0.1$ and $q \equiv 1$. We compare the performance of recovery formulae based on 2, 3 and 4 moment functionals, in all cases being centred about the origin. For each formula we state all the superconvergent points for recovery of the function and the derivative and give the error between the true and recovered solution at these points. For comparison we give the error between the true and recovered solution at a non-superconvergent point.

Number of Moment Functions	x	Comparing Function or Derivative	Expected Rate of Convergence at x	(True - Recovered Solution) at x
2	-0.64550 (-1)	F	3	0.35 (-4)
	0.0		2	-0.21 (-2)
	0.64550 (-1)		3	-0.37 (-4)
	0.0	D	2	-0.13 (-2)
3	-0.12247	F	4	-0.19 (-5)
	0.0		4	0.11 (-5)
	0.12247		4	-0.21 (-5)
	-0.70711 (-1)	D	3	0.58 (-4)
	0.0		2	-0.25 (-2)
	0.70711 (-1)		3	-0.60 (-4)
4	-0.17772	F	5	0.12 (-6)
	-0.58454 (-1)		5	-0.48 (-7)
	0.0		4	0.45 (-5)
	0.58454 (-1)		5	0.50 (-7)
	0.17772		5	-0.14 (-6)
	-0.13229	D	4	-0.34 (-5)
	0.0		4	0.19 (-5)
	0.13229		4	-0.37 (-5)

TABLE 2.1 : Uniform mesh $h = 0.1$ and $q \equiv 1$

From Table 2.1 we see clearly the advantage of sampling our recovery formulae at the superconvergent points. To link up these results with the formulae of section 2.2 we consider for example the case of 3 moment functionals. The superconvergent points of the recovery formula (2.2.16) for the function value $u(th)$ occur at the roots of $w^{M}_{-2}(t)$; that is, $t = \pm\sqrt{\frac{3}{2}}$, 0, as can be seen from (2.2.35). Hence we obtain the superconvergent points 0.0, $\pm\sqrt{\frac{3}{2}}h$ ($\approx \pm 0.12247$) in Table 2.1. The approximation to $u(0)$ is then given by

$(1 - \delta^2/12)\ F_0^M(u)$ and from (2.2.27) we have a bound for the error:-

$$|u(0) - (1 - \delta^2/12)F_0^M(u)| \leq \frac{(0.1)^4}{90} \ ||u^{(iv)}||_{L^\infty[-0.2,0.2]} \quad .$$

Therefore the bound $(0.1)^4\ e^{0.2}/90 \approx 0.13571 \times 10^{-5}$ is in good agreement with the true error, $\approx 0.11 \times 10^{-5}$, given in Table 2.1.

In Table 2.2 we present the results of using 4 moment functionals centred about the origin on a uniform mesh of length $h = 0.1$ with $q = 1 + x$. We see there is no deterioration in the results due to q being non-uniform and again observe the advantage of sampling at the predicted superconvergent points.

Number of Moment Functions	x	Comparing Function or Derivative	Expected Rate of Convergence at x	(True - Recovered Solution) at x
4	-0.17588	F	5	0.13 (-6)
	-0.56567 (-1)		5	-0.48 (-7)
	0.0		4	0.45 (-5)
	0.60163 (-1)		5	0.49 (-7)
	0.17907		5	-0.13 (-6)
	-0.2	D	3	-0.76 (-3)
	-0.13045		4	-0.34 (-5)
	0.17775 (-2)		4	0.19 (-5)
	0.13376		4	-0.36 (-5)

TABLE 2.2 : Uniform mesh and $q = 1 + x$

In Table 2.3 we present the results of using 4 moment functionals on a non-uniform mesh with nodes -0.25, -0.1, 0.025, 0.125, 0.2, 0.25 and $q \equiv 1$. Once again there is no deterioration in the results.

Number of Moment Functions	x	Comparing Function or Derivative	Expected Rate of Convergence at x	(True - Recovered Solution) at x
4	-0.2	F	4	0.56 (-4)
	-0.15073		5	0.36 (-6)
	0.36293 (-2)		5	-0.73 (-7)
	0.12195		5	0.35 (-7)
	0.20848		5	-0.35 (-7)
	-0.2	D	3	-0.17 (-2)
	-0.95217 (-1)		4	-0.66 (-5)
	0.59391 (-1)		4	0.21 (-5)
	0.17333		4	-0.19 (-5)

TABLE 2.3 : Non-uniform mesh and $q \equiv 1$

2.3 Recovery from the Point Functionals

2.3.1 Uniform mesh over the infinite real line and $q \equiv 1$

As we have seen in 2.1, for piecewise constants recovering from the point functionals is equivalent to recovering from the moment functionals. For continuous piecewise linears these procedures are different and this is the case that we consider here. We are presented with n consecutive point functionals

$$F^{p}_{k+i}(u) = u^{h}(x_{k+i}) \qquad i = 1 \to n, \qquad (2.3.1)$$

where $u^h \in S^h$ is such that

$$(q(u - u^h), \phi_j) = 0 \qquad \forall \phi_j \in S^h. \qquad (2.3.2)$$

It is convenient to introduce the discrete Green's function $g^h_i \in S^h$ such that

$$(q\ g^h_i, \phi_j) = \delta_{ij} \qquad \forall \phi_j \in S^h \quad . \qquad (2.3.3)$$

Then we have

$$F_i^p(u) \equiv u^h(x_i) = (q\ g_i^h,\ u^h) = (q\ g_i^h,\ u) = (qu,\ g_i^h) \quad . \tag{2.3.4}$$

Unfortunately unlike recovering from the moment functionals, recovering from the point functionals is not a robust procedure and many results that one may expect to hold do not. For example $F_{k+i}^p(p_{n-1}) = 0 \quad i = 1 \to n$ in general does not imply that the $(n - 1)^{th}$ degree polynomial $p_{n-1}(x) \equiv 0$. To illustrate this consider the following example.' With the interval $[-1,1]$ partitioned into four elements with nodes $\{-1,-h,0,h,1\}$, a simple calculation yields that the best continuous piecewise linear fit u^h to the quadratic $6x^2 - h^2$ in the L^2 norm (that is, $q \equiv 1$) is such that $u^h(-h) = u^h(0) = u^h(h) = 0$ if h is chosen to be $[(17)^{\frac{1}{2}} - 3]/4 \approx 0.2808$. Thus this fundamental result does not hold on a general mesh for $n = 3$. One may ask under what restrictions does it hold. We have the following result.

Lemma 2.2

If u^h is the best piecewise linear fit on a uniform mesh of length h over the infinite real line to a $(n - 1)^{th}$ degree polynomial p_{n-1} in the L^2 norm (that is, $q \equiv 1$), then $F_{k+i}^p(p_{n-1}) \equiv u^h(x_{k+i}) = 0 \quad i = 1 \to n$ implies that $p_{n-1}(x) \equiv 0$.

Proof: From (2.3.4) we have in general

$$\begin{aligned} F_{k+i}^p(p_{n-1}) = (qp_{n-1},\ g_{k+i}^h) &= \sum_{j=0}^{n-1} [(q(x-x_{k+i})^j,\ g_{k+i}^h)p_{n-1}^{(j)}(x_{k+i})/j!] \\ &= \sum_{j=0}^{n-1} [\sigma_{ij}p_{n-1}^{(j)}\ (x_{k+i})/j!], \end{aligned} \tag{2.3.5}$$

where

$$\sigma_{ij} = (q(x-x_{k+i})^j,\ g_{k+i}^h) \ . \tag{2.3.6}$$

If $q \equiv 1$ and S^h is defined on a uniform mesh over the infinite real line, then σ_{ij} is independent of i because

$$\begin{aligned}\sigma_{ij} &= \int_{-\infty}^{\infty} (x - x_{k+i})^j \, g^h_{k+i}(x) \, dx \\ &= \int_{-\infty}^{\infty} x^j \, g^h_{k+i}(x + x_{k+i}) \, dx \\ &= \int_{-\infty}^{\infty} x^j g^h_0(x) \, dx = \sigma_j \qquad \forall i \in Z \quad . \end{aligned} \qquad (2.3.7)$$

We note that if any of the above constraints are relaxed then σ_{ij} is no longer independent of i. Defining the $(n - 1)^{th}$ degree polynomial $Q_{n-1}(x)$ by

$$Q_{n-1}(x) = \sum_{j=0}^{n-1} [\sigma_j p^{(j)}_{n-1}(x)/j!] \quad ,$$

we deduce that $Q_{n-1}(x) \equiv 0$ since we are given that $Q_{n-1}(x_{k+i}) = F^p_{k+i}(p_{n-1}) = 0$ $i = 1 \rightarrow n$. This implies that $p_{n-1}(x) \equiv 0$ as $\sigma_0 = 1$. ■

Corollary

With $q \equiv 1$ and a uniform mesh over the infinite real line there exist unique $(n - 1)^{th}$ degree polynomials $c^p_{n,j}(x)$ $j = 1 \rightarrow n$ such that

$$F^p_i(c^p_{n,j}) = \delta_{ij} \qquad i = 1 \rightarrow n \qquad (2.3.8)$$

and thus any $(n - 1)^{th}$ degree polynomial $p_{n-1}(x)$ can be written in the form

$$p_{n-1}(x) = \sum_{j=1}^{n} F^p_{k+j}(p_{n-1}) c^p_{n,j}(x - x_k) . \qquad (2.3.9)$$

Proof: An explicit calculation yields that

$$g^h_0(x) = h^{-1}\sqrt{3} \sum_{j=-\infty}^{\infty} (\sqrt{3}-2)^{|j|} \phi_j(x)$$

and so

$$g^h_i(x) = g^h_0(x - x_i) = h^{-1}\sqrt{3} \sum_{j=-\infty}^{\infty} (\sqrt{3}-2)^{|j-i|} \phi_j(x) \quad . \qquad (2.3.10)$$

We express $c^P_{n,j}(x)$ in terms of $\{\hat{\psi}_\ell(x)\}^n_{\ell=1}$, where $\hat{\psi}_\ell(x) = \psi_\ell\left(\frac{x}{h}\right)$ and $\{\psi_\ell(x)\}^n_{\ell=1}$ is any basis independent of h for polynomials of degree $n - 1$,

$$c^P_{n,j}(x) = \sum_{\ell=1}^{n} b^P_{\ell,j}\, \hat{\psi}_\ell(x) \qquad j = 1 \rightarrow n. \tag{2.3.11}$$

The equations (2.3.8) are equivalent to $A^P B^P = I_n$, where A^P and B^P are $n \times n$ matrices independent of h with entries $a^P_{i,\ell} \equiv F^P_i(\hat{\psi}_\ell) = (\hat{\psi}_\ell, g^h_i)$ and the unknown coefficients $b^P_{\ell,j}$, respectively. The invertibility of A^P and hence the existence and uniqueness of $c^P_{n,j}(x)$ $j = 1 \rightarrow n$ follow from Lemma 2.2. Setting $c^P_{n,k+j}(x) = c^P_{n,j}(x - x_k)$ it follows that $F^P_{k+i}(c^P_{n,k+j}) = \delta_{ij}$, yielding the desired result (2.3.9). ■

Even under the restrictive conditions above of a uniform mesh over the infinite real line and $q \equiv 1$ the analogue of Lemma 2.1 does not holds, that is, given $v \in C(-\infty,\infty)$ $F^P_{k+i}(v) = 0$ $i = 1 \rightarrow n$ does not imply that v has n zeroes. For example the quartic polynomial $15x^4 - h^4$ has only two zeroes, but a simple calculation yields that its corresponding best piecewise linear fit u^h is such that $u^h(-h) = u^h(0) = u^h(h) = 0$. This means that one cannot bound the error

$$u(x) - \sum_{i=1}^{n} F^P_{k+i}(u)\, c^P_{n,i}(x - x_k) \tag{2.3.12}$$

using interpolation theory. Furthermore, one cannot deduce that there exist n points of superconvergence. Suboptimal results can be obtained by relating the point functionals to the moment functionals; that is, $F^P_{k+i}(v) = 0$ $i = 1 \rightarrow n$ implies that $F^M_{k+i}(v) = 0$ $i = 2 \rightarrow n - 1$. Thus using interpolation theory the error (2.3.12) can be shown to be at least $O(h^{n-2})$ and that there exist $n - 2$ points of superconvergence, where the error is at least $O(h^{n-1})$. However, this does not identify any advantage in using the point functionals over the moment functionals.

An alternative approach as we are only considering the case of a uniform mesh and $q \equiv 1$ is to explicitly generate the recovery formulae as we did for the moment functionals in subsection 2.2.2. Choosing $n \equiv 2m + 1$ to exploit symmetry we consider a collection of point functionals $\{F^p_{k+i}(u)\}^m_{i=-m}$ centred about x_k. Adopting analogous notation to that in 2.2.2. it follows that our recovery formula approximating $u(x_k + th)$ is the analogue of (2.2.16):-

$$\sum_{j=0}^{m} w^p_j(t) \frac{\delta^{2j} F^p_k(u)}{(2j)!} + \sum_{j=1}^{m} w^p_{-j}(t) \frac{\delta^{2j-1} F^p_{k-\frac{1}{2}}(u)}{(2j-1)!} \tag{2.3.13}$$

$$= \int_{-\infty}^{\infty} g^p_{2m+1}(t,s) u(x_k + sh)\, ds \quad ,$$

where

$$g^p_{2m+1}(t,s) := \sum_{j=0}^{m} w^p_j(t) \frac{\delta^{2j} g^h_0(sh)}{(2j)!} + \sum_{j=1}^{m} w^p_{-j}(t) \frac{\delta^{2j-2}[g^h_0(sh)-g^h_{-1}(sh)]}{(2j-1)!} \quad ; \tag{2.3.14}$$

the analogue of (2.2.22). Expanding $u(x)$ in a Taylor series as in 2.2.2 we have that the error in the approximation (2.3.13) is given by

$$E^p_{2m+1,k}(u,t) := \int_{-\infty}^{\infty} H^p_{2m+1}(t,s) u^{(2m+1)}(x_k + sh)\, ds \quad , \tag{2.3.15}$$

where

$$H^p_{2m+1}(t,s) := \frac{h^{2m+1}}{(2m)!} \begin{cases} \displaystyle\int_{-\infty}^{s} g^p_{2m+1}(t,y)(s-y)^{2m} dy & -\infty < s < t \\ \\ \displaystyle -\int_{s}^{\infty} g^p_{2m+1}(t,y)(y-s)^{2m} dy & t < s < \infty \quad . \end{cases} \tag{2.3.16}$$

Hence we have

$$|E^p_{2m+1,k}(u,t)| \leqq \int_{-\infty}^{\infty} |H^p_{2m+1}(t,s)| ds \; \|u^{(2m+1)}\|_{L^\infty(-\infty,\infty)}$$

$$\leqq Ch^{2m+1} \; \|u^{(2m+1)}\|_{L^\infty(-\infty,\infty)} \qquad \text{for} \quad |t| \leqq m + 1. \tag{2.3.17}$$

At the superconvergent point $t = 0$ (2.3.13) reduces to

$$\sum_{j=0}^{m} C_{2j}^{p}\ \delta^{2j} F_{k}^{p}(u) \quad , \tag{2.3.18}$$

where

$$C_{2j}^{p} = -\left\{\sum_{\ell=0}^{j-1} \delta^{2\ell} F_{0}^{p}\left(\left(\frac{x}{h}\right)^{2j}\right) C_{2\ell}^{p}\right\} /(2j)! \qquad j \geqq 1 \tag{2.3.19}$$

and

$$C_{0}^{p} = 1 \ ;$$

and the error is

$$E_{2m+1,k}^{p}(u,0) := \int_{-\infty}^{\infty} G_{2m+1}^{p}(0,s) u^{(2m+2)}(x_k + sh) ds, \tag{2.3.20}$$

where

$$G_{2m+1}^{p}(0,s) := \frac{h^{2m+2}}{(2m+1)!} \int_{-\infty}^{s} g_{2m+1}^{p}(0,y)(s-y)^{2m+1} dy \qquad -\infty < s \leqq 0$$

(2.3.21)

and

$$G_{2m+1}^{p}(0,s) := G_{2m+1}^{p}(0,-s) \qquad 0 \leqq s < \infty .$$

Hence we have

$$\begin{aligned} |E_{2m+1,k}^{p}(u,0)| &\leq \int_{-\infty}^{\infty} |G_{2m+1}^{p}(0,s)| ds \quad \|u^{(2m+2)}\|_{L^{\infty}(-\infty,\infty)} \\ &\leqq Ch^{2m+2} \quad \|u^{(2m+2)}\|_{L^{\infty}(-\infty,\infty)} \ . \end{aligned} \tag{2.3.22}$$

After some tedious calculations due to the global nature of the approximation one can evaluate the coefficients C_{2j}^{p} from (2.3.19) to obtain

$$C_{2}^{p} = \frac{1}{12} , \qquad C_{4}^{p} = -\frac{1}{360} \qquad \ldots \quad . \tag{2.3.23}$$

Unfortunately, unlike with the moment functions, the kernel $G_{2m+1}^{p}(0,s)$ is highly oscillatory thus making it difficult to isolate the constant in the error. It is obviously simple to obtain a lower bound for this constant, since $\int_{-\infty}^{\infty} |G_{2m+1}^{p}(0,s)| ds \geq \left|\int_{-\infty}^{\infty} G_{2m+1}^{p}(0,s) ds\right| = |C_{2m+2}^{p}|\ h^{2m+2}$. As an example we have

$$\left|u(x_k) - \left(1 + \frac{\delta^2}{12}\right) F^p_k(u)\right| \leq Ch^4 \ \|u^{(4)}\|_{L^\infty(-\infty,\infty)} \ ; \qquad (2.3.24)$$

where we know the constant C is no smaller than $\frac{1}{360}$, which it would be if u were a quartic polynomial. Clearly, one can obtain corresponding formulae for the derivatives of u. It is also useful to tabulate the polynomials $w^p_j(t)$, whose roots determine the points of superconvergence, and we obtain for example

$$w^p_{-1}(t) = t$$

$$w^p_1(t) = t^2 + t + \frac{1}{6}$$

$$w^p_{-2}(t) = t^3 - \frac{1}{2}t$$

$$w^p_2(t) = t^4 + 2t^3 - t - \frac{1}{15} \ .$$

As an example we have $(1 + \lambda)F^p_k(u) - \lambda F^p_{k-1}(u)$ is an $O(h^3)$ approximation to $u(x_k + \lambda h)$ if $\lambda = \frac{1}{2}(-1 \pm \sqrt{\frac{1}{3}})$, roots of $w^p_1(t)$, as we derived previously in (2.1.9).

2.3.2 General case

We now look to see how the above results are affected if we consider the case of $q \equiv 1$ and a uniform mesh on a finite domain $[a,b]$. To do this we follow Chandler (1980) and exploit the connection between the best L^2 approximation by continuous piecewise linears and interpolation by cubic splines. For a function v, its cubic spline interpolate v_c satisfies the system of equations

$$h\left(1 + \frac{\delta^2}{6}\right) v''_c(x_j) = \frac{\delta^2}{h} v(x_j) \qquad j = 1 \rightarrow J - 1, \qquad (2.3.25)$$

see Schultz (1973) for example. This system can be rewritten, since v''_c is a continuous piecewise linear function, in the form

$$(v''_c - v'', \phi_j) = 0 \qquad j = 1 \rightarrow J - 1. \qquad (2.3.26)$$

To define the cubic spline interpolate uniquely, the system (2.3.26) has to be supplemented with two extra conditions. As is well-known these have to be chosen carefully so as not to reduce the rate of convergence in the interior, see for example Behforooz and Papamichael (1979) and Lucas (1974). One can relate the piecewise linear L^2 fit u^h to u to a cubic spline interpolation problem by noting that $u^{h^{(-2)}}$ is the cubic spline interpolate of $u^{(-2)}$, where $u^{(-m)}$ denotes the m-fold indefinite integral of u. The two extra conditions used in defining u^h can then be related to cubic spline end conditions. At the left hand end we would either impose $u^h(a) = u(a)$, corresponding to $v_c''(a) = v''(a)$, or $(u - u^h, \phi_0) = 0$, corresponding to $v_c'(a) = v'(a)$ from the minimality property of cubic splines; similarly at the right hand end.

Now that this link has been made, results concerning u^h and u can be lifted from the well-developed theory of cubic spline interpolation. In fact the result (2.3.24) was given by Curtis and Powell (1967) for the second derivative of a function and its corresponding cubic spline interpolate, when ignoring boundary effects. The effect of the boundary conditions considered above has been studied by Lucas (1974) among others for cubic splines and unfortunately it is shown that the result (2.3.24) is lost, in the sense that it does not hold uniformly for $k = 1 \rightarrow J - 1$, but clearly still holds for a fixed node x_k as h tends to zero. However, the result (2.1.9) remains if we use $(u - u^h, \phi_j) = 0$ $j = 0$ and J as our end conditions. The latter result has been proved independently by Richter (1978), when studying integral equations where the L^2 norm plays an important role. We now show that this result generalises to a quasi-uniform mesh and for a general smooth weighting function q.

Theorem 2.2

For $u \in W^{3,\infty}(a,b)$ and u^h defined by (2.3.2) we have that

$$|(u - u^h)(x^{G\pm}_{j+\frac{1}{2}})| \leq Ch^3 \qquad j = 0 \rightarrow J - 1 \;, \qquad (2.3.27)$$

where

$$x^{G\pm}_{j+\frac{1}{2}} = x_j + \frac{1}{2}\left(1 \pm \frac{1}{\sqrt{3}}\right)h_{j+\frac{1}{2}} \quad , \qquad h_{j+\frac{1}{2}} = x_{j+1} - x_j \qquad (2.3.28a)$$

and

$$h = \max_{j=0\to J-1} h_{j+\frac{1}{2}}$$

provided the mesh is quasi-uniform; that is,

$$h_{j+\frac{1}{2}}/h_{j-\frac{1}{2}} = 1 + O(h) \qquad j = 1 \to J - 1. \qquad (2.3.28b)$$

In addition we have that

$$\left|u'(\tfrac{1}{2}(x_j+x_{j+1})) - h^{-1}_{j+\frac{1}{2}}[u^h(x_{j+1}) - u^h(x_j)]\right| \leqq Ch^2 \qquad j = 0 \to J - 1. \qquad (2.3.29)$$

Proof: On each element (x_j,x_{j+1}) let u^I denote the linear interpolate of u at the Gauss points $x^{G\pm}_{j+\frac{1}{2}}$ so that we have

$$u(x) = u^I(x) + \tfrac{1}{2}(x - x^{G-}_{j+\frac{1}{2}})(x - x^{G+}_{j+\frac{1}{2}})u''(\tfrac{1}{2}(x_j + x_{j+1})) + R \quad ,$$

$$x \in (x_j,x_{j+1}) \qquad (2.3.30)$$

where R is $o(h^3_{j+\frac{1}{2}})$. Over the interval $[a,b]$ u^I is a discontinuous piecewise linear function. We now introduce another discontinuous piecewise linear function τ defined by

$$\tau(x_j\pm) = \mp \tfrac{1}{2}(u^I(x_j+) - u^I(x_j-)) \qquad j = 1 \to J - 1$$

$$\tau(x_0) = \tau(x_J) = 0 \qquad (2.3.31)$$

so that $u^I + \tau$ is a continuous piecewise linear function over $[a,b]$. A straightforward calculation yields that

$$\tau(x_j\pm) = \mp \frac{1}{24}(h_{j+\frac{1}{2}} + h_{j-\frac{1}{2}})(h_{j-\frac{1}{2}} - h_{j+\frac{1}{2}})u''(x_j) + O(h^3) \qquad (2.3.32)$$

$$j = 1 \to J - 1.$$

Therefore under the quasi-uniformity assumption on the mesh we have

$$|\tau(x)| \leqq Ch^3 \qquad \forall x \in [a,b] \quad . \qquad (2.3.33)$$

On each element (x_j, x_{j+1}) approximate the weighting function q by the constant $q(\frac{1}{2}(x_j + x_{j+1}))$. Denoting this piecewise constant approximation by $\bar{q}$, we introduce $\bar{u}^h \in S^h$ such that

$$(\bar{q}(u - \bar{u}^h), \phi_i) = 0 \qquad i = 0 \rightarrow J \quad . \qquad (2.3.34)$$

Since

$$\int_{x_j}^{x_{j+1}} (x - x^{G-}_{j+\frac{1}{2}})(x - x^{G+}_{j+\frac{1}{2}})(\alpha x + \beta)\ dx = 0 \qquad \forall\alpha,\beta \in \mathbb{R} \quad j = 0 \rightarrow J - 1$$

it follows from (2.3.30) that

$$(\bar{q}\bar{u}^h, \phi_i) = (\bar{q}[(u^I + \tau) + (R - \tau)], \phi_i) \qquad i = 0 \rightarrow J \; .$$

From the diagonal dominance of the mass matrix M with entries $M_{ij} = (\bar{q}\phi_j,\phi_i)/(\bar{q},\phi_i)$ $\quad i,j = 0 \rightarrow J$ we have

$$|\bar{u}^h(x) - [u^I(x) + \tau(x)]| \leqq |M^{-1} \{(\bar{q}(R - \tau),\phi_i)/(\bar{q},\phi_i)\}^{J}_{i=0}|_\infty$$
$$\leqq Ch^3 \quad .$$

The above with (2.3.33) implies that

$$|(u - \bar{u}^h)(x^{G\pm}_{j+\frac{1}{2}})| \leqq Ch^3 \qquad j = 0 \rightarrow J - 1. \qquad (2.3.35)$$

In addition we have

$$\begin{aligned}\bar{u}^h(x_{j+1}) - \bar{u}^h(x_j) &= u^I(x_{j+1}-) - u^I(x_j+) + O(h^3)\\ &= \sqrt{3}\ [u(x^{G+}_{j+\frac{1}{2}}) - u(x^{G-}_{j+\frac{1}{2}})] + O(h^3)\\ &= h_{j+\frac{1}{2}}\ u'(\tfrac{1}{2}(x_j + x_{j+1})) + O(h^3) \qquad j = 0 \rightarrow J - 1 \; .\end{aligned} \qquad (2.3.36)$$

The desired results (2.3.27) and (2.3.29) follow from (2.3.35) and (2.3.36), respectively, if we can show that

$$|\bar{u}^h(x) - u^h(x)| \leqq Ch^3 \qquad \forall x \in [a,b]. \qquad (2.3.37)$$

Since $|q(x) - \bar{q}(x)| \leqq Ch$ and $|u(x) - u^h(x)| \leqq Ch^2$ $\quad \forall x \in [a,b]$, we have that

$$\left|(\bar{q}(\bar{u}^h - u^h),\ \phi_i)/(\bar{q},\ \phi_i)\right|$$

$$= \left|((\bar{q} - q)(u - u^h),\ \phi_i)/(\bar{q},\phi_i)\right| \leqq Ch^3.$$

The desired result (2.3.37) then follows from the diagonal dominance of M. ■

2.3.3 Numerical Results

We now describe various numerical examples to illustrate how the techniques described in this section perform in practice. Throughout we consider point functionals generated from continuous piecewise linear functions over the interval $[0,1]$ and take $u = \cos x$. Once again a 4 point Gauss rule was used on each element to evaluate the necessary integrals.

In Table 2.4 we present results in the case of a uniform mesh of length h and $q \equiv 1$. We see that the predicted rates of convergence occur. The superconvergence of u^h, $O(h^3)$, to u at the Gauss points of degree 2 is seen from row 2, whereas row 1 shows that u^h is only $O(h^2)$ at the nodes. The superconvergence of $u^{h'}$, $O(h^2)$, to u' at the midpoints of the elements is seen from row 3. The fact that $\left(1 + \frac{\delta^2}{12}\right) u^h(x_j)$ is an $O(h^4)$ approximation to $u(x_j)$ in the interior is seen from row 5, whereas row 4 shows that this does not hold uniformly over the interval $[0,1]$, as predicted.

In Table 2.5 we see how these results are affected by choosing a non-constant q, $q \equiv (x + 1)^{-1}$. As predicted by Theorem 2.2 we still have the superconvergence of u^h at the Gauss points and of $u^{h'}$ at the midpoints. However, in addition we see that $\left(1 + \frac{\delta^2}{12}\right) u^h(x_j)$ still yields an $O(h^4)$ approximation to $u(x_j)$ in the interior.

We next look at the effect of a non-uniform mesh. In Table 2.6 we present results for $q \equiv 1$ with

$$h_{2j+\frac{1}{2}} = \frac{2}{3}h \qquad j = 0 \rightarrow \frac{J}{2} - 1$$

and

$$h_{2j+\frac{3}{2}} = \frac{4}{3}h \qquad j = 0 \rightarrow \frac{J}{2} - 1.$$

We note that this mesh is not quasi-uniform; that is, it does not satisfy (2.3.28b). Thus as expected we lose the superconvergence of E2 and E3.

In Table 2.7 we present results for $q \equiv 1$ with

$$h_{j+\frac{1}{2}} = [1 - \tfrac{1}{2}h(J)]h_{j-\frac{1}{2}} \qquad j = 1 \rightarrow J - 1$$

with

$$h_{\frac{1}{2}} = h(J) \equiv 2(1 - (\tfrac{1}{2})^{\frac{1}{J}}).$$

We note that this mesh is quasi-uniform and so we regain the superconvergence of E2 and E3. Note that $h_{max} = h(J)$ and $h(J)/h(2J) = 1 + (\frac{1}{2})^{\frac{1}{2J}}$.

Key to the tables:-

$$E1 = \{\tfrac{1}{2}h_{\frac{1}{2}}[e(0)]^2 + \sum_{j=1}^{J-1} \tfrac{1}{2}[h_{j-\frac{1}{2}}+h_{j+\frac{1}{2}}][e(x_j)]^2 + \tfrac{1}{2}h_{J-\frac{1}{2}}[e(1)]^2\}^{\frac{1}{2}} ,$$

$$E2 = \{ \sum_{j=0}^{J-1} \tfrac{1}{2} h_{j+\frac{1}{2}}[(e(x^{G-}_{j+\frac{1}{2}}))]^2 + [e(x^{G+}_{x_{j+\frac{1}{2}}})]^2]\}^{\frac{1}{2}}$$

$$E3 = \{ \sum_{j=0}^{J-1} h_{j+\frac{1}{2}} \left[\frac{e(x_{j+1})-e(x_j)}{h_{j+\frac{1}{2}}}\right]^2 \}^{\frac{1}{2}} ,$$

$$E4 = \{ \sum_{j=1}^{J-1} (h_{j-\frac{1}{2}} + h_{j+\frac{1}{2}})(u(x_j) - \frac{1}{12} [u^h(x_{j-1}) + 10u^h(x_j) + u^h(x_{j+1})])^2 \}^{\frac{1}{2}},$$

and

$$E5 =\{ \sum_{j=(J+4)/4}^{(3J-4)/4} (h_{j-\frac{1}{2}} + h_{j+\frac{1}{2}})(u(x_j) - \frac{1}{12} [u^h(x_{j-1}) + 10u^h(x_j) + u^h(x_{j+1})])^2 \}^{\frac{1}{2}} ,$$

where $e(x) = u(x) - u^h(x)$, $h_{j+\frac{1}{2}} = x_{j+1} - x_j$ and $x^{G\pm}_{j+\frac{1}{2}} = x_j + \frac{1}{2}\left(1 \pm \frac{1}{\sqrt{3}}\right) h_{j+\frac{1}{2}}$.

$h=\frac{1}{J}$, J=	16	32	64
E1	0.28 (-3)	0.59 (-4)	0.17 (-4)
E2	0.20 (-5)	0.25 (-6)	0.32 (-7)
E3	0.16 (-3)	0.42 (-4)	0.11 (-4)
E4	0.19 (-6)	0.17 (-7)	0.15 (-8)
E5	0.40 (-7)	0.24 (-8)	0.15 (-9)

TABLE 2.4 : Uniform mesh and $q \equiv 1$

$h=\frac{1}{J}$, J=	16	32	64
E1	0.28 (-3)	0.69 (-4)	0.17 (-4)
E2	0.20 (-5)	0.25 (-6)	0.32 (-7)
E3	0.16 (-3)	0.42 (-4)	0.11 (-4)
E4	0.27 (-6)	0.23 (-7)	0.20 (-8)
E5	0.72 (-7)	0.43 (-8)	0.27 (-9)

TABLE 2.5 : Uniform mesh and $q = (1+x)^{-1}$

$h=\frac{1}{J}$, J=	16	32	64
E1	0.37 (-3)	0.93 (-4)	0.23 (-4)
E2	0.17 (-3)	0.43 (-4)	0.11 (-4)
E3	0.30 (-2)	0.10 (-2)	0.37 (-3)

TABLE 2.6 : Non-quasi-uniform mesh and $q \equiv 1$

J =	16	32	64
E1	0.36 (-3)	0.91 (-4)	0.23 (-4)
E2	0.10 (-4)	0.13 (-5)	0.16 (-6)
E3	0.54 (-3)	0.14 (-3)	0.34 (-4)

TABLE 2.7 : Quasi-uniform mesh and $q \equiv 1$

2.4 Recovery Techniques for Non-Smooth Functions

The recovery procedures described in subsections 2.2 and 2.3 assume that the underlying solution u is sufficiently smooth and hence it can be well approximated by polynomials. In this subsection we discuss briefly the problem of non-smooth and rapidly varying functions, e.g. shocks and boundary layers. As we have seen recovering from the moment functionals for piecewise linear approximations is far more robust than recovering from the point functionals and so we attempt to generalise the former to cope with non-smooth functions.

Presented with n consecutive moment functionals $\{F^m_{k+i}(u)\}^n_{i=1}$, in subsection 2.2 we constructed a $(n-1)^{th}$ polynomial $p_{n-1}(x)$ as our recovery function whose coefficients were determined by requiring

$$F^M_{k+i}(u - p_{n-1}) = 0 \qquad i = 1 \rightarrow n \quad ; \qquad (2.4.1)$$

this yielded.

$$p_{n-1}(x) = \sum_{j=1}^{n} F^M_{k+j}(u)\, c^M_{n,k+j}(x) \quad , \qquad (2.4.2)$$

where the $(n-1)^{th}$ degree polynomials $c^M_{n,k+j}(x)$ $j = 1 \rightarrow n$ satisfy (2.2.2). From Theorem 2.1 we see that if u is smooth in $I_{n,k}$ then $p_{n-1}(x)$ is a good approximation to $u(x)$. However, if u is non-smooth or rapidly varying then an alternative form of recovery function may be more appropriate.

Consider the more general form of recovery function u_R:-

$$u_R(x) = \sum_{j=1}^{n} \alpha_j g_j(x) \quad , \qquad (2.4.3)$$

which still depends linearly on parameters α_j but where the basis functions $g_j \in C[a,b]$ are chosen to incorporate any a priori knowledge of the form of u that one might have. The coefficients α_j are then determined from requiring the generalisation of (2.4.1) to hold:-

$$F^M_{k+i}(u - u_R) = 0 \qquad i = 1 \rightarrow n. \qquad (2.4.4)$$

A constraint on the choice of basis $\{g_j(x)\}_{j=1}^n$ is that there should exist a unique solution u_R to the set of equations (2.4.4). Thus we have to show that $F_{k+i}^M(u_R) = 0 \quad i = 1 \to n$ implies that $u_R \equiv 0$. From Lemma 2.1 we see that this is equivalent to showing that if u_R has n zeroes in $I_{n,k}$ then $u_R \equiv 0$. Therefore to guarantee the existence and uniqueness of u_R to (2.4.4) we require the set of basis functions $\{g_j(x)\}_{j=1}^n$ to be unisolvent on $I_{n,k}$.

As an example if u exhibits a boundary layer it may be more appropriate to use an exponential rather than a polynomial basis; that is, for $n = 3$ choose $g_1(x) \equiv 1$, $g_2(x) = x$ and $g_3(x) = e^{\sigma x}$ for some given constant instead of $g_3(x) = x^2$. It is a simple matter to show that $\{1,x,e^{\sigma x}\}$ is unisolvent for any $\sigma \neq 0$ and hence the recovery procedure (2.4.4) is well-posed. If a good choice of σ is used, e.g. the boundary layer width is known a priori from an asymptotic analysis, then this exponential basis performs far better than the polynomial basis.

However, in general, a good estimate of σ may not be known a priori and indeed it may be the most important aspect of the underlying boundary layer that one wishes to recover. Incorporating it as a parameter we could, for example with $n = 3$, take

$$u_R(x) = \alpha_1 + \alpha_2 \, e^{\alpha_3 x} \qquad (2.4.5)$$

as an appropriate recovery function. Once again the parameters $\{\alpha_j\}_{j=1}^3$ are determined from requiring (2.4.4) to hold. The crucial difference now is that we have a non-linear recovery problem as a non-linear equation for the parameter α_3 has to be solved. Hence it is difficult to guarantee the existence and uniqueness of u_R since this depends on the given data $\{F_{k+i}^M(u)\}_{i=1}^3$. Despite the lack of theory the form (2.4.5) of recovery function is easy to implement and has worked well in practice when boundary layers have been present. Several numerical examples are given in Barrett & Morton (1980,1984) and Morton & Scotney (1985) for the continuous piecewise linear moment functionals

$$F^M_{k+i}(u) = (pu',\phi'_{k+i}) + (qu,\phi_{k+i})$$

with $q \gg p$, which arise from a symmetrizing Petrov-Galerkin approximation of diffusion-convection problems.

Another situation where a non-linear recovery procedure is appropriate arises in the approximation of hyperbolic conservation laws. The underlying function u may well contain shock discontinuities or at least discontinuities of gradient: these will generally be separated by regions of smooth variation. If a finite element approximation yields an L^2 best fit on a fixed mesh at each time step, point values will need to be recovered for the calculation of flux functions to be used in following the solution through the next time step. For such problems there is much to be said for using piecewise constant approximations rather than piecewise linears: we have already seen in sub-section 2.2 that recovery of smooth functions is just as easy and indeed, the coefficients in the uniform mesh case decrease more rapidly; there is no distinction between point and moment functionals so the recovery is always a robust procedure; and the greater compactness of the basis functions is important in recovering rapidly varying functions.

Applications of these ideas may be found in Morton (1982b,1984). In the vicinity of a shock a recovery function of the form

$$u_R(x) = \begin{cases} \alpha_1 & x < \alpha_3 \\ \alpha_2 & x > \alpha_3 \end{cases} \qquad (2.4.6)$$

works well in practice and is very easily implemented: for piecewise constant approximations it is merely a matter of replacing three fixed elements by two elements with a free boundary, the shock position. The presence of shocks is detected by an appropriate shock recognition criterion and smoother recovery is used between them. For example, piecewise constants can be recovered by quadratic splines or by piecewise linears. In the latter case this may be achieved by an adaptive procedure which maintains monotonicity - see Morton (1984).

3. RECOVERY FROM O.D.E. BEST FITS

The continuous piecewise linear Galerkin approximation u^h to the solution of the two-point boundary value problem

$$\begin{aligned} Lu \equiv -(pu')' + qu = f \qquad & p > 0,\ q \geq 0 \\ u(a) = u(b) = 0 \qquad & \end{aligned} \tag{3.1}$$

satisfies

$$\begin{aligned} a(u^h,\phi_j) &= a(u,\phi_j) \\ &= (f,\phi_j) \end{aligned} \tag{3.2}$$

for $j = 1,\ldots,J - 1$. We use the same notation as section 2 with the symmetric bilinear form

$$a(v,w) \equiv \int_a^b (pv'w' + qvw)\, dx \quad , \tag{3.3}$$

and also let $I = [a,b]$, $I_{j+\frac{1}{2}} = [x_j,x_{j+1}]$, $h_{j+\frac{1}{2}} = x_{j+1} - x_j$ and $h = \max\{h_{j+\frac{1}{2}}\}$ $j = 0 \rightarrow J - 1$. For the moment it is only assumed that $p,q \in L^\infty(I)$ and $f \in H^{-1}(I)$ which is sufficient for (3.2) to be well-defined and for $a(\cdot,\cdot)$ to be an energy norm equivalent to the usual norm over the space $H_0^1(I)$ of functions in H^1 satisfying homogeneous Dirichlet boundary conditions; thus

$$\|v\|_E^2 \equiv a(v,v) \qquad v \in H_0^1(I) \ . \tag{3.4}$$

3.1 Limited applicability of local recovery

In the previous section, on recovery from weighted L^2 best fits, it was found that particular sampling points and local averages of u^h could be determined which gave more accurate approximations to u and its derivative. For the case of O.D.E. best fits, however, such results are very limited and in general an alternative approach, developed in 3.3, is necessary. Nevertheless for the important class of problems with $p \ll q$, usually termed singularly perturbed problems, we can extend some of the weighted L^2 best fit analysis and this is given in 3.2. In this subsection we shall comment on the O.D.E.

best fit problem for general p,q and present those results which are possible.

If the framework of Golomb & Weinberger (1959) is used and it is assumed that data functionals $a(u,\phi_j)$ $j = 1 \rightarrow J-1$ are given together with a bound on the energy norm $a(u,u)^{\frac{1}{2}}$ then, because the ϕ_j are the representers of the data functionals with respect to the energy inner product, we immediately obtain the result that u^h is the optimal approximation to u in this sense. Hence any other information required of u, i.e. any linear functional $F(\cdot)$ of u bounded w.r.t. the energy norm, can best be approximated by forming $F(u^h)$. If we consider at which points one ought to sample u^h, i.e. $F(v) = v(\bar{x})$ and $\bar{x}$ is chosen to minimise the optimal error bound, then we have

$$\begin{aligned} |u(\bar{x}) - u^h(\bar{x})| &\leq \|G(\bar{x},\cdot) - G^h(\bar{x},\cdot)\|_E \{\|u\|_E^2 - \|u^h\|_E^2\}^{\frac{1}{2}} \\ &= \{G(\bar{x},\bar{x}) - G^h(\bar{x},\bar{x})\}^{\frac{1}{2}} \{\|u\|_E^2 - \|u^h\|_E^2\}^{\frac{1}{2}} \quad , \end{aligned} \tag{3.1.1}$$

where $G(\cdot,\cdot)$ is the Green's function for L and $G^h(\bar{x},\cdot)$ the $\|\cdot\|_E$ Galerkin approximation to $G(\bar{x},\cdot)$. (This bound is achieved when $u = G(\bar{x},\cdot)$.) Now if $p \in W^{1,\infty}(I)$ it is clear that asymptotically $G^h(\bar{x},\cdot)$ will give a better approximation to $G(\bar{x},\cdot)$ when $\bar{x}$ is a node because only then can the derivative discontinuity in $G(\bar{x},\cdot)$ be accurately approximated by continuous piecewise linear functions. Thus the "best" points at which to sample u^h are at the nodes. (In fact it can easily be shown, using the explicit representation of $G(\cdot,\cdot)$ and $G^h(\cdot,\cdot)$, that for contant p.q on a uniform mesh $\{G(\bar{x},\bar{x}) - G^h(\bar{x},\bar{x})\}^{\frac{1}{2}}$ is $O(h)$ at the nodes but only $O(h^{\frac{1}{2}})$ at other points.) Note that our nonlinear data is a bound on the energy norm of u and so only $u \in H_0^1(I)$ is assumed, later in this subsection we shall look at the ideas suggested by the Golomb & Weinberger approach when greater smoothness on u is introduced.

A more straightforward method for deriving accuracy results for point functionals of u is to use the Green's function idea directly, as in Douglas &

Dupont (1974). Thus

$$u(x_j) - u^h(x_j) = a(G(x_j,\bullet),\ u - u^h) \qquad (3.1.2)$$
$$= a(G(x_j,\bullet) - v^h,\ u - u^h),$$

where v^h is an arbitrary continuous piecewise linear function and so

$$|u(x_j) - u^h(x_j)| \leqq \inf_{v^h \in S^h} \|G(x_j,\bullet) - v^h\|_E\ \|u - u^h\|_E. \qquad (3.1.3)$$

If $p \in W^{1,\infty}(I)$, and so $G(x_j,\bullet) \in H^1(I) \cap H^2([a,x_j]) \cap H^2([x_j,b])$. this immediately gives the standard result that for $u \in H^2(I)$ $u - u^h$ is $O(h^2)$ at the nodes. In fact

$$\max_{1 \leqq j \leqq J-1} |(u - u^h)(x_j)| \leqq Ch^2\ \|u''\|_{L^2(I)} \qquad (3.1.4)$$

is a superconvergence result in itself since $O(h^2)$ error bounds at points normally require $u \in W^{2,\infty}(I)$. An improved asymptotic convergence rate only appears, however, when higher order approximating subspaces are used. Nevertheless even for continuous piecewise linear functions we are able to obtain superconvergent results for the derivatives of u. Thus with the standard notation for divided differences we have

$$(u - u^h)[x_{j-1},x_j] = a(G([x_{j-1},x_j],\bullet),\ u - u^h) \qquad (3.1.5)$$
$$= a(G([x_{j-1},x_j],\bullet) - v^h,\ u - u^h)$$

where v^h is again arbitrary. The Green's function can be written

$$G(x,\xi) \begin{cases} s_1(x)s_2(\xi)/t(x) & x \leqq \xi \\ s_1(\xi)s_2(x)/t(x) & x \geqq \xi \end{cases} \qquad (3.1.6)$$

with $t = p(s_1's_2 - s_1s_2')$, where s_1 and s_2 are linearly independent solutions of $Lv = 0$, s_1 satisfying the boundary condition at a and s_2 the boundary condition at b; hence v^h can be chosen so that

$$\|G([x_{j-1},x_j],\bullet) - v^h\|_{H^1(I/I_{j-\frac{1}{2}})} \leqq C_1h$$

and $\qquad (3.1.7)$

$$\|G([x_{j-1},x_j],\bullet) - v^h\|_{W^{1,1}(I_{j-\frac{1}{2}})} \leqq C_2h_{j-\frac{1}{2}}$$

with C_1 and C_2 depending on $\|p\|_{W^{1,\infty}(I)}$ and $\|q\|_{L^{\infty}(I)}$ through $\|s_i\|_{W^{2,\infty}(I)}$ $i = 1,2$. Combined with (3.1.5) this gives

$$|(u - u^h)[x_{j-1},x_j]| \leqq Ch(\|u - u^h\|_{H^1(I/I_{j-\frac{1}{2}})} + \|u - u^h\|_{W^{1,\infty}(I_{j-\frac{1}{2}})}). \tag{3.1.8}$$

Thus if $u \in W^{3,\infty}(I)$, so that $\|u - u^h\|_{W^{1,\infty}(I)}$ is $O(h)$ (see Douglas & Dupont (1974)) and $u[x_{j-1},x_j]$ is an $O(h^2)$ approximation to $u'(x_{j-\frac{1}{2}})$, then

$$\max_{1\leqq j\leqq J} |(u - u^h)'(x_{j-\frac{1}{2}})| \leqq Ch^2 \|u\|_{W^{3,\infty}(I)} . \tag{3.1.9}$$

In certain circumstances similar arguments can be used to produce $O(h^2)$ approximations to the second derivatives of u from

$$(u - u^h)[x_{j-1},x_j,x_{j+1}] = a(G([x_{j-1},x_j,x_{j+1}],\cdot) - v^h, u - u^h). \tag{3.1.10}$$

Thus we may choose v^h so that

$$\|G([x_{j-1},x_j,x_{j+1}],\cdot) - v^h\|_{H^1(I/(I_{j-\frac{1}{2}}UI_{j+\frac{1}{2}}))} \leqq C_1 h \tag{3.1.11}$$

and

$$\begin{aligned} &\left|\int_{x_{j-1}}^{x_{j+1}} p(G([x_{j-1},x_j,x_{j+1}],\cdot) - v^h)'(u - u^h)' \right.\\ &\quad \left. + q(G([x_{j-1},x_j,x_{j+1}],\cdot) - v^h)(u - u^h)\right| \\ &\quad \leqq C_2\{h \|u - u^h\|_{W^{1,\infty}(I_{j-\frac{1}{2}}UI_{j+\frac{1}{2}})} + |(u - u^h)[x_{j-1},x_j]| \\ &\quad + |(u - u^h)[x_j,x_{j+1}]| + |h_{j+\frac{1}{2}} - h_{j-\frac{1}{2}}|\ |u''(x_j)| \\ &\quad + h^2 \|u'''\|_{L^{\infty}(I_{j-\frac{1}{2}}UI_{j+\frac{1}{2}})}\} \end{aligned} \tag{3.1.12}$$

where C_1 and C_2 depend on $\|p\|_{W^{2,\infty}(I)}$ and $\|q\|_{W^{1,\infty}(I)}$ through $\|s_i\|_{W^{3,\infty}(I)}$ $i = 1,2$. Hence if the mesh is uniform and $u \in W^{4,\infty}(I)$, so that twice $u[x_{j-1},x_j,x_{j+1}]$ gives an $O(h^2)$ approximation to $u''(x_j)$, we have

$$\max_{1\leq j\leq J-1} |u''(x_j) - 2u^h[x_{j-1},x_j,x_{j+1}]| \leq Ch^2 \, \|u\|_{W^{4,\infty}(I)} \, . \qquad (3.1.13)$$

The above convergence results for divided differences of u and u^h may be compared with those derived by a different method in 3.3. Note that even on a non-uniform mesh it is possible to obtain $O(h^2)$ approximations to u'' at the midpoint of each sub-interval by using the differential equation.

At the beginning of this subsection it was stated that difficulties arise if one tries to extend the techniques of section 2, i.e. local recovery using a small number of functionals of the form

$$\begin{aligned} &\text{(i)} \quad F_j^M(u) \equiv a(u,\phi_j)/(h_{j-\frac{1}{2}} + h_{j+\frac{1}{2}}) \\ \text{or} \quad &\text{(ii)} \quad F_j^P(u) \equiv u^h(x_j) \end{aligned} \qquad (3.1.14)$$

and we now investigate why this is so.

Suppose, for example, that n consecutive moment functionals, $F_{k+i}^M(u)$ $i = 1 \to n$ say, are used and we try to form a linear combination $\sum_{i=1}^{n} \alpha_i F_{k+i}^M(u)$ so that

$$p_{n-1}(\overline{x}) = \sum_{i=1}^{n} \alpha_i F_{k+i}^M(p_{n-1}) \qquad (3.1.15)$$

for all polynomials p_{n-1} of degree $n - 1$, where $\overline{x}$ is some chosen point in $I_{n,k} \equiv [x_k,x_{k+n+1}]$. It is immediately clear, however, that this is impossible if q is identically zero over $I_{n,k}$ since then the r.h.s. of (3.1.15) is zero for constants. Even if the assumption $q > 0$ on $I_{n,k}$ is made, so that we could now define

$$F_j^M(u) \equiv a(u,\phi_j)/(q,\phi_j) \qquad (3.1.16)$$

to agree with Section 2, there is no guarantee that $\alpha_1,\dots,\alpha_n$ satisfying (3.1.15) can be found. For example if p_{n-1} is a polynomial such that $p_{n-1}^? > 0$ and $p'_{n-1} < 0$ over $I_{n,k}$, it is possible for a given q to choose $p > 0$ on each subinterval in turn so that

$$h^{-1}_{j+\frac{1}{2}} \int_{x_j}^{x_{j+1}} p\, p'_{n-1} = h^{-1}_{j-\frac{1}{2}} \int_{x_{j-1}}^{x_j} p\, p'_{n-1} + \int_{x_{j-1}}^{x_{j+1}} q\, p_{n-1}\phi_j \qquad (3.1.17)$$

for $j = k+1 \to k+n$. Thus $F^M_{k+i}(p_{n-1}) = 0 \quad i = 1 \to n$ but $p_{n-1} \not\equiv 0$. Finally even if (3.1.15) were solvable for $\alpha_1,\ldots,\alpha_n$; e.g. p,q constant allows the techniques of Section 2 to be used to prove this result; we would still face disappointment. Thus

$$u(\bar{x}) - \sum_{i=1}^{n} \alpha_i F^M_{k+i}(u) = u(\bar{x}) - p_{n-1}(\bar{x}) - \sum_{i=1}^{n} \alpha_i F^M_{k+i}(u - p_{n-1}) \qquad (3.1.18)$$

with arbitrary p_{n-1}, but although p_{n-1} could be chosen so that $\|u - p_{n-1}\|_{L^\infty(I_{n,k})} = O(h^n)$ the form of the functionals F^M_j means that $u(\bar{x}) - \sum_{i=1}^{n} \alpha_i F^M_{k+i}(u)$ is only $O(h^{n-2})$. The only type of O.D.E. for which such problems can be avoided are those of singularly perturbed form with q of order unity and $p \ll 1$. Then the question of solving (3.1.15) for $\alpha_1,\ldots,\alpha_n$ may be analysed by regarding it as a perturbation of the corresponding weighted L^2 problem. Also the error bounds for $u(\bar{x}) - \sum_{i=1}^{n} \alpha_i F^M_{k+i}(u)$ are of the form $O(h^n) + \|p\|_{L^\infty(I_{n,k})} h^{n-2}$ and thus if p is comparable with h^2 there is no loss of accuracy. Some results in this direction are given in 3.2.

Similar difficulties arise if we try to recover locally using the point functionals $F^P_j(u)$. The global nature of these functionals,

$$F^P_j(u) \equiv a(G^h(x_j,\cdot),u) \quad , \qquad (3.1.19)$$

together with the fact that the error $u - u^h$ does not better $O(h^2)$ in any negative Sobolev norm for piecewise linear approximating functions, implies that local recovery by polynomials to the same accuracy as for the weighted L^2 best fit is not possible, except in the trivial case p constant and q zero. If p is comparable with qh^2, however, $G^h(x_j,\xi)$ is exponentially decreasing as ξ moves away from x_j, for example if p and q are constant then

$$G^h(x_j,x_k) \approx s^{-|j-k|\sqrt{qh^2/p}} \qquad (3.1.20)$$

with $s > 1$.

As was mentioned at the beginning of this subsection, the optimal approximation to u in the sense of Golomb & Weinberger (1959), when data functionals $a(u,\phi_j)$ $j = 1 \to J - 1$ and a bound on the energy norm $a(u,u)^{\frac{1}{2}}$ are given, is the Galerkin approximation u^h. Now we try to generalise by using the same data functionals but assuming more smoothness on u, i.e. using a new energy norm containing higher derivatives of u, and it is clear that different choices will lead to different optimal approximations. For example if we first consider the special case of p and q constant, we may define an energy norm

$$A(v,w) \equiv \int_a^b (pv''w'' + qv'w')dx \qquad (3.1.21)^2$$

and rewrite the data in terms of this norm, viz.

$$a(u,\phi_j) = A(u,\chi_j) \qquad j = 1 \to J - 1 \qquad (3.1.22)$$

where $\chi_j'' = -\phi_j$ and $\chi_j(a) = \chi_j(b) = 0$. Thus the χ_j are natural cubic splines and the optimal approximation for the data $a(u,\phi_j)$ $j = 1 \to J - 1$ and the bound $A(u,u)^{\frac{1}{2}}$ is the natural cubic spline χ^h defined by the Petrov-Galerkin equations

$$a(u - \chi^h,\phi_j) = 0 \qquad J = 1 \to J - 1 \quad . \qquad (3.1.23)$$

Since this may also be written as a Galerkin approximation w.r.t. the new energy norm,

$$a(u - \chi^h,\phi_j) = A(u - \chi^h,\chi_j) \quad , \qquad (3.1.24)$$

it is clearly well-defined. For the general case of variable p and q, with u assumed to have k derivatives in $L^2(I)$ we introduce the energy inner-products for $k \geqq 2$:-

$$A_k(v,w) \equiv \begin{cases} (L^{k/2}v, L^{k/2}w) + \sum_{\ell=1}^{k/2-1} \{L^\ell v(a) L^\ell w(a) + L^\ell v(b) L^\ell w(b)\} & k \text{ even} \\ a(L^{(k-1)/2}v, L^{(k-1)/2}w) + \sum_{\ell=1}^{(k-1)/2} \{L^\ell v(a) L^\ell w(a) + L^\ell v(b) L^\ell w(b)\} & k \text{ odd} \end{cases} \tag{3.1.25}$$

and rewrite the data as

$$a(u,\phi_j) = A_k(u,\psi_j), \tag{3.1.26}$$

where $\psi_j = \theta_j + \xi_j$ with the two components satisfying

(i) $L^{k-1}\theta_j = \phi_j$ and $L^\ell\theta_j(a) = L^\ell\theta_j(b) = 0$ $\ell = 0 \to k - 2$

(ii) $L^{k/2}\xi_j = 0$ if k even or $a(L^{(k-1)/2}\xi_j, L^{(k-1)/2}\xi_j) = 0$ if k odd.

Thus the ψ_j are well-defined for any k and the optimal approximation $\psi^h = \sum_{j=1}^{J-1} \alpha_j\psi_j$ is defined by the Petrov-Galerkin equations

$$a(u - \psi^h, \phi_j) = 0 \qquad j = 1 \to J - 1 \tag{3.1.27}$$

or the Galerkin equations

$$A_k(u - \psi^h, \psi_j) = 0 \qquad j = 1 \to J - 1. \tag{3.1.28}$$

The function ψ^h is a generalised natural spline, although not quite fitting into the theory of Varga (1971), and if p and q are sufficiently smooth, e.g. $p \in C^{2(k-1)}(I)$ and $q \in C^{(2k-3)}(I)$, it has a continuous $2(k-1)^{th}$ derivative with jump discontinuities in the $(2k-1)^{th}$ derivative at the nodes. Like ψ_j above, ψ^h may be split into two components θ^h and ξ^h, the latter ensuring that $L^\ell\psi^h(a) = L^\ell u(a)$ and $L^\ell\psi^h(b) = L^\ell u(b)$ for $\ell = 1 \to [(k-1)/2]$ and the former giving the standard natural spline minimisation property w.r.t. the integral part of the energy norm $A_k(\cdot,\cdot)$. The approximating power of the ψ_j may be derived by adaptations of the following argument, here given for $k = 2$ and the ∞-norm. For $z \in W^{4,\infty}(I)$ let $\phi^h = \sum_{j=1}^{J-1} \beta_j\phi_j$ be the continuous piecewise linear L^2 best fit to Lz, then

$$z(x) - \sum_{j=1}^{J-1} \beta_j\psi_j(x) = (G(x,\cdot), Lz - \phi^h) \qquad (3.1.29)$$
$$= (G(x,\cdot) - G^I(x,\cdot), Lz - \phi^h)$$

where $G^I(x,\cdot)$ is the continuous piecewise linear interpolate of $G(x,\cdot)$. Thus

$$|z(x) - \sum_{j=1}^{J-1} \beta_j\psi_j(x)| \leqq \|G(x,\cdot) - G^I(x,\cdot)\|_{L^1(I)} \|Lz - \phi^h\|_{L^\infty(I)} \qquad (3.1.30)$$
$$\leqq \quad Ch^2 \quad \|Lz - \phi^h\|_{L^\infty(I)} \quad .$$

Now if Lz were zero at the boundary

$$\|Lz - \phi^h\|_{L^\infty(I)} \leq Ch^2 \quad \|Lx\|_{W^{2,\infty}(I)} \qquad (3.1.31)$$
$$\leq Ch^2 \quad \|z\|_{W^{4,\infty}(I)} \quad ,$$

but since this need not hold we have in general the standard natural spline result that $\|z - \sum_{j=1}^{J-1} \beta_j\psi_j\|_{L^\infty(I)}$ is only $O(h^2)$ over I but $O(h^4)$ over any interior interval.

Of course in practice one would not be able to use these optimal generalised splines, but this viewpoint shows how to proceed. A higher-order piecewise polynomial space is chosen and we find the element of this space whose Galerkin approximation is the same as that of u itself, i.e. a Petrov-Galerkin formulation as used above. Efficient methods for implementing this idea are examined in 3.3.

3.2 Recovery for Singularly Perturbed Problems

In this section we shall examine the possibility of recovery using moment functionals when $p << q$ in (3.1). Thus we shall assume that n consecutive moment functionals

$$F^m_{k+i}(u) = a(u,\phi_{k+i})/(q,\phi_{k+i}) \qquad i = 1 \to n \qquad (3.2.1)$$

are given and we wish to prove that there is a unique polynomial q_{n-1} of degree $n - 1$ or less satisfying

$$F^m_{k+i}(q_{n-1}) = F^m_{k+i}(u) \qquad i = 1 \to n \tag{3.2.2}$$

and to show how well q_{n-1} approximates u. We shall try to proceed as in Section 2 and regard the $(pu', \phi_j')/(q,\phi_j)$ term in (3.2.1) as a perturbation.

Let P_{n-1} denote the space of $(n-1)^{th}$ degree (or less) polynomials on $I_{n,k}$ and T_2 the linear mapping from P_{n-1} to R^n defined by

$$T_2 p_{n-1} = \underset{\sim}{b} \tag{3.2.3}$$

where

$$b_i = (pp'_{n-1}, \phi'_{k+i})/(q,\phi_{k+i}) \qquad i = 1 \to n. \tag{3.2.4}$$

Using the -norm on both spaces together with the fact that

$$\|p'_{n-1}\|_{L^\infty(I_{n,k})} \leq C_2 h^{-1}_{n,k} \|p_{n-1}\|_{L^\infty(I_{n,k})} \tag{3.2.5}$$

for any element of P_{n-1}, we have

$$\|T_2\|_\infty \leq 2C_2 \|p\|_{L^\infty(I_{n,k})} \|q^{-1}\|_{L^\infty(I_{n,k})} /(h_{n,k}\overline{h}_{n,k}) \tag{3.2.6}$$

where $\overline{h}_{n,k} = \min\limits_{0 \leqq j \leqq n} h_{k+j+\frac{1}{2}}$. We already know from Section 2 that the mapping $T_1 : P_{n-1} \to R^n$ defined by

$$T_1\, p_{n-1} = \underset{\sim}{c} \quad , \tag{3.2.7}$$

where

$$c_i = (qp_{n-1}, \phi_{k+i})/(q,\phi_{k+i}) \qquad i = 1 \to n \quad , \tag{3.2.8}$$

has a bounded inverse and thus if

$$\|p\|_{L^\infty(I_{n,k})} < h_{n,k}\overline{h}_{n,k} /(2C_2 \|q^{-1}\|_{L^\infty(I_{n,k})} \|T_1^{-1}\|_\infty \tag{3.2.9}$$

then $(T_1 + T_2)^{-1}$ exists with

$$\|(T_1 + T_2)^{-1}\|_\infty \leq \|T_1^{-1}\|_\infty /(1 - \delta) \tag{3.2.10}$$

where

$$2C_2 \|p\|_{L^\infty(I_{n,k})} \|q^{-1}\|_{L\infty(I_{n,k})} \|T_1^{-1}\|_\infty /(h_{n,k}\overline{h}_{n,k}) \leq \delta < 1. \tag{3.2.11}$$

Thus if p is sufficiently small to satisfy (3.2.9), there is a unique $(n-1)^{th}$ degree polynomial q_{n-1} which satisfies (3.2.2). To bound the error $q_{n-1} - u$, let r_{n-1} be any element of P_{n-1}, so that

$$(T_1 + T_2)\,(r_{n-1} - q_{n-1}) = (T_1 + T_2)\,(r_{n-1} - u). \tag{3.2.12}$$

Regarding $(T_1 + T_2)\,(r_{n-1} - u)$ as an element of R^n, we have

$$\|(T_1 + T_2)(r_{n-1} - u)\|_\infty \leq \|q^{-1}\|_{L^\infty(I_{n,k})} \,(2\,\|p\|_{L^\infty(I_{n,k})} \|(r_{n-1} - u)'\|_{L^\infty(I_{n,k})} \,\bar{h}^{-1}_{n,k} + \|q\|_{L^\infty(I_{n,k})} \|r_{n-1} - u\|_{L^\infty(I_{n,k})}) \tag{3.2.13}$$

and thus

$$\|r_{n-1} - u\|_{L^\infty(I_{n,k})} \leq \|(T_1 + T_2)^{-1}\|_\infty \|q^{-1}\|_{L^\infty(I_{n,k})} (2\,\|p\|_{L^\infty(I_{n,k})} \,\bar{h}^{-1}_{n,k} \|(r_{n-1} - u)'\|_{L^\infty(I_{n,k})} + \|q\|_{L^\infty(I_{n,k})} \|r_{n-1} - u\|_{L^\infty(I_{n,k})}) \tag{3.2.14}$$

Therefore

$$\begin{aligned} \|q_{n-1} - u\|_{L^\infty(I_{n,k})} &\leq \|q_{n-1} - r_{n-1}\|_{L^\infty(I_{n,k})} + \|r_{n-1} - u\|_{L^\infty(I_{n,k})} \\ &\leq (1 + \|(T_1 + T_2)^{-1}\|_\infty \|q^{-1}\|_{L^\infty(I_{n,k})} \|q\|_{L^\infty(I_{n,k})})\, \|r_{n-1} - u\|_{L^\infty(I_{n,k})} \\ &\quad + 2\,\|(T_1 + T_2)^{-1}\|_\infty \,\bar{h}^{-1}_{n,k} \|p\|_{L^\infty(I_{n,k})} \|q^{-1}\|_{L^\infty(I_{n,k})} \|(r_{n-1} - u)'\|_{L^\infty(I_{n,k})}. \end{aligned} \tag{3.2.15}$$

Since p is restricted by (3.2.9) the above inequality implies that q_{n-1} approximates u to optimal order provided that the various inverse operators are bounded independently of the mesh size.

The problem that we have not answered either here or in Section 2 is to bound T_1^{-1} in terms of q and the mesh spacing. In general this seems to be a difficult problem and we restrict ourselves to the following comments.

(a) If q is constant and the mesh is uniform one may use the linear transformation $y = (x - x_k)/h$ to show that T_1^{-1} can be bounded independently of h and q, i.e. $\|T_1^{-1}\|_\infty \leq C_1$. Thus the restriction on the size of p in (3.2.9) can be written

$$\|p\|_{L^\infty(I_{n,k})} < qh^2/(2C_1C_2) \tag{3.2.16}$$

(b) If q is constant and the mesh non-uniform then the size of T_1^{-1} depends on the mesh spacing. If

$$v^h \equiv \sum_{i=1}^{n} \alpha_i\phi_{k+i} \tag{3.2.17}$$

satisfies.

$$T_1v^h = \underset{\sim}{c} \tag{3.2.18}$$

(c.f. (3.2.7)) and w is the cubic spline defined by

$$w'' = v^h \tag{3.2.19}$$

and $w(x_k) = w(x_{k+n+1}) = 0$, then the $(n+1)^{th}$ degree polynomial p_{n+1} defined by

$$p''_{n+1} = p_{n-1} \tag{3.2.20}$$

and $p_{n+1}(x_k) = p_{n+1}(x_{k+n+1}) = 0$ interpolates w at $x_k,\ldots,x_{k+n+1}$. Thus the error in polynomial interpolation gives

$$p_{n-1}(x) = v^h(x) + \frac{d^2}{dx^2}\{(x-x_k)\ldots\ldots(x-x_{k+n+1})\; w\,[x_k,\ldots,x_{k+n+1},x]\} \tag{3.2.21}$$

and from this we would like to be able to deduce that

$$\|p_{n-1}\|_{L^\infty(I_{n,k})} \leq C\,\|v^h\|_{L^\infty(I_{n,k})} \tag{3.2.22}$$

where C is a function of the mesh-spacing. If this is so then the diagonal dominance of the matrix $A_{ij} \equiv \int \phi_i \phi_j$ allows us to write

$$\|v^h\|_{L^\infty(I_{n,k})} \leq \frac{3}{q} \|\underset{\sim}{c}\|_\infty \tag{3.2.23}$$

and hence

$$\|T_1^{-1}\|_\infty \leq \frac{3C}{q} \quad . \tag{3.2.24}$$

(c) We do not know how to tackle the case of variable q although of course one could always write

$$a(p_{n-1},\phi_j) = q(x_j)\left\{(p_{n-1},\phi_j) + \left(\frac{q - q(x_j)}{q(x_j)} \quad p_{n-1},\phi_j\right) + \frac{(pp'_{n-1},\phi'_j)}{q(x_j)}\right\} \tag{3.2.25}$$

and regard the last two terms as perturbations. However this would involve unnecessary restrictions on q of the form of bounds on

$$\sup_{x\in[x_{j-1},x_{j+1}]} \left| \frac{q(x) - q(x_j)}{q(x_j)} \right| \quad . \tag{3.2.26}$$

3.2.1 Numerical Results

We give below three tables of results which are completely analogous to the tables in section 2.2.3 except now we are recovering from moment functionals containing a derivative term with small coefficient p. The addition of this term does not seem to diminish the accuracy of the recovered solution but note that the alternation in sign of the error at consecutive superconvergent points does not now occur.

Number of Moment Functions	x	Comparing Function or Derivative	Expected rate of Convergence at x	(True - Recovered Solution) at x
2	-0.46547 (-1)	F	3	-0.45 (-5)
	0.0		2	-0.11 (-2)
	0.46547 (-1)		3	0.59 (-5)
	0.0	D	2	-0.25 (-3)
3	-0.94871 (-1)	F	4	0.75 (-6)
	0.0		4	0.44 (-6)
	0.94871 (-1)		4	0.90 (-6)
	-0.54772 (-1)	D	3	0.95 (-5)
	0.0		2	-0.15 (-2)
	0.54772		3	-0.87 (-5)
4	-0.14102	F	5	-0.72 (-7)
	-0.55799 (-1)		5	-0.20 (-7)
	0.0		4	0.26 (-5)
	0.55799 (-1)		5	0.21 (-7)
	0.14102		5	0.84 (-7)
	-0.10741	D	4	-0.56 (-6)
	0.0		4	0.84 (-6)
	0.10741		4	0.85 (-6)

TABLE 3.1 : Uniform mesh $h = 0.1$, $p \equiv 0.1\ (-1)$ and $q \equiv 1$

Number of Moment Functions	x	Comparing Function or Derivative	Expected rate of Convergence at x	(True - Recovered Solution) at x
4	-0.13049	F	5	-0.10 (-6)
	-0.56281 (-1)		5	-0.27 (-7)
	0.0		4	—
	0.56083 (-1)		5	0.16 (-7)
	0.14956		5	0.49 (-7)
	-0.2	D	3	-0.10 (-2)
	-0.10029		4	0.67 (-6)
	0.12403 (-2)		4	0.86 (-6)
	0.11321		4	-0.38 (-6)

TABLE 3.2 : Uniform mesh h = 0.1, p = 0.1 (-2) cos x and q = 1 + x.

4	-0.2	F	4	0.77 (-4)
	-0.12314		5	0.11 (-7)
			5	
			5	
			5	
	-0.2	D	3	-0.18 (-2)
	-0.75042 (-1)		4	-0.18 (-5)
	0.64484 (-1)		4	0.97 (-6)
	0.13320		4	0.19 (-5)

TABLE 3.3 : Non-uniform mesh, p ≡ 0.1 (-2) and q ≡ 1

3.3 Recovery through Defect Correction

Following the discussion at the end of section 3.1, we now consider how to obtain a more accurate global approximation to u by using a higher-order piecewise polynomial space T^h and finding the element of this space whose continuous piecewise linear Galerkin approximation is the same as that of u itself. Thus, denoting the span of continuous piecewise linear functions by S^h again, we wish to solve

$$a(\tilde{u}^h, v^h) = a(u, v^h) \qquad \forall v^h \in S^h \qquad (3.3.1)$$

for $\tilde{u}^h \in T^h$. (Note that u^h plays the same role here as u_R in section 2.4). This can only be a well-posed problem if T^h and S^h have the same dimension and we then have a Petrov-Galerkin formulation

$$a(\tilde{u}^h, v^h) = (f, v^h) \qquad \forall v^h \in S^h \qquad (3.3.2)$$

with continuous piecewise linear test functions and higher-order trial functions. The basic results concerning existence, uniqueness and approximation for this type of method are contained in the following theorem adapted from Morton (1982a).

Theorem 3.1

If

(i) $$\inf_{\tilde{v}^h \in T^h} \sup_{v^h \in S^h} |a(\tilde{v}^h, v^h)|^2 \geq \gamma^2 a(\tilde{v}^h, \tilde{v}^h)\, a(v^h, v^h) \qquad \gamma > 0$$

and (ii) $$\sup_{\tilde{v}^h \in T^h} a(\tilde{v}^h, v^h) > 0 \qquad \forall\, v^h \in S^h, \quad v^h \neq 0,$$

then (3.3.2) has a unique solution $\tilde{u}^h \in T^h$ with

$$a(\tilde{u}^h - u, \tilde{u}^h - u) \leq \gamma^{-2} \inf_{\tilde{v}^h \in T^h} a(v^h - u, v^h - u) \quad ■$$

The natural choice for T^h would be either

(a) higher-order splines with appropriate end conditions

or (b) higher-order continuous piecewise Lagrange polynomials.

It is hardly efficient however to first obtain the Galerkin approximation u^h and then the Petrov-Galerkin approximation $\tilde{u}^h$. In fact we wish to avoid solving any system of algebraic equations other than those with the symmetric coefficient matrix $\{a(\phi_i, \phi_j)\}$. This can be achieved by solving (3.3.2) in Stetter's (1978) defect correction iterative form

$$a(u^h_{i+1}, v^h) = a(u^h_i, v^h) - \{a(Pu^h_i, v^h) - (f, v^h)\} \qquad \forall v^h \in S^h \qquad (3.3.3)$$

with $u^h_0 \equiv u^h$. Here P is the nodal interpolatory mapping between S^h and T^h which we assume to be bijective. Thus we obtain a sequence of functions $\{u^h_i\} \subset S^h$ which hopefully converge to u^h_∞ where $Pu^h_\infty = \tilde{u}^h$. By writing (3.3.3) as

$$a(u^h_{i+1} - u^h_\infty, v^h) = a(u^h_i - u^h_\infty - P(u^h_i - u^h_\infty), v^h) \qquad (3.3.4)$$

we can clearly see that convergence of $\{u^h_i - u^h_\infty\}$ to zero is to be expected if the sequence remains "smooth enough" for the piecewise linear interpolation error projection I-P to exert its contracting power - see Skeel (1981).

We shall now prove some convergence results about the iteration (3.3.3). This will be done, however, in a more general setting by not assuming the existence of a Petrov-Galerkin approximation. We shall merely let P be a bijective nodal interpolatory mapping between S^h and a higher-order continuous approximating space T^h, and compute the iterates u^h_i by (3.3.3). Thus the analogue of (3.3.4) is

$$a(u^h_{i+1} - u^h_I, v^h) = a(u^h_i - u^h_I - P(u^h_i - u^h_I), v^h) + a(u - Pu^h_I, v^h) \qquad (3.3.5)$$

with u^h_I the continuous piecewise linear interpolate of u; the additional final term shows that the smallness of $\{u^h_i - u^h_I\}$ is restricted by the approximating power of the space T^h.

First we show that the iteration (3.3.3) will converge with at least an $O(h)$ rate.

Theorem 3.2

If $p \in W^{1,\infty}(I)$ and P is bounded in the $H^1(I)$ norm uniformly in h then

$$\|u^h_{i+1} - u^h_I\|_{H^1(I)} \leqq C\{h\,\|u^h_i - u^h_I\|_{H^1(I)} + \|u - Pu^h_I\|_{L^2(I)}\} \, . \qquad (3.3.6)$$

Proof

Setting $v^h = u^h_{i+1} - u^h_I$ in (3.3.5), integrating by parts and using the fact that P is interpolatory yields

$$\|u^h_{i+1} - u^h_I\|^2_{H^1(I)} \leqq Ca(u^h_{i+1} - u^h_I, u^h_{i+1} - u^h_I) \qquad \text{by coercivity of } a\,(\bullet,\bullet)$$

$$= C\int_a^b \{q(u^h_{i+1}-u^h_I) - p'(u^h_{i+1}-u^h_I)'\}\{u^h_i-u^h_I-P(u^h_i-u^h_I) + (u - Pu^h_I)\}\,dx$$

$$\leqq C\,\|u^h_{i+1} - u^h_I\|_{H^1(I)}\{\,\|u^h_i - u^h_I - P(u^h_i - u^h_I)\|_{L^2(I)} + \|u - Pu^h_I\|_{L^2(I)}\} \, .$$

Thus

$$\|u^h_{i+1} - u^h_I\|_{H^1(I)} \leqq C\{h\,\|P(u^h_i - u^h_I)\|_{H^1(I)} + \|u - Pu^h_I\|_{L^2(I)}\}$$

$$\leqq C\{h\,\|u^h_i - u^h_I\|_{H^1(I)} + \|u - Pu^h_I\|_{L^2(I)}\}$$

by the stability of P. ■

Using

$$\|u^h_0 - u^h_I\|^2_{H^1(I)} \leqq C\,a(u^h_0 - u^h_I,\, u^h_0 - u^h_I) = C\,a(u - u^h_I, u^h_0 - u^h_I) \qquad (3.3.7)$$

we can show, by applying the same integration by parts as in the above proof, that

$$\|u^h_0-u^h_I\|_{H^1(I)} \leqq C\,\|u-u^h_I\|_{L^2(I)} \leqq Ch^2\,\|u''\|_{L^2(I)} \, . \qquad (3.3.8)$$

(This is equivalent to the $O(h^2)$ convergence of the first divided differences of u^h and u, cf. section 3.1). It follows that, if u is sufficiently smooth and the approximating power of the space T^h is sufficiently great,

$$\|u_i^h - u_I^h\|_{H^1(I)} \leqq Ch^{2+i} \quad . \tag{3.3.9}$$

For example if T^h consisted of cubic splines with suitable end conditions and $u \in H^4(I)$ we should reach the asymptotic accuracy limit with u_2^h, when

$$\|u_2^h - u_I^h\|_{H^1(I)} \leqq Ch^4 \|u\|_{H^4(I)} \quad , \tag{3.3.10}$$

and further iterations would not improve this power of h. The significance of convergence in the H^1 norm is that first derivative approximations to the same accuracy will usually be possible. Thus, continuing the example, if Δ_1^h is a nodal finite-difference functional which is bounded in the H^1 norm,

$$|\Delta_1^h v| \leqq C \|v\|_{H^1(x)} \quad , \tag{3.3.11}$$

and such that

$$|\Delta_1^h v - v'(\bar{x})| \leqq C h^4 \|v\|_{H^5(I)} \tag{3.3.12}$$

for some $\bar{x} \in I$, then we should have

$$\begin{aligned} |u'(\bar{x}) - \Delta_1^h u_2^h| &\leqq |u'(\bar{x}) - \Delta_1^h u_x^h| + |\Delta_1^h (u_I^h - u_2^h)| \\ &\leqq Ch^4 \|u\|_{H^5(I)} \quad . \end{aligned} \tag{3.3.13}$$

In certain circumstances, however, this convergence rate can be improved. To analyse these situations we need a measure of the "smoothness" of functions which are only in $H^1(I)$ and so semi-norms based on the divided differences of nodal values are used. Denoting $w(x_j)$ be w_j we have the standard divided difference notation

$$\begin{aligned} &\text{(i)} \quad w(x_j, x_{j+1}] \equiv (w_{j+1} - w_j)/h_{j+\frac{1}{2}} \\ &\text{(ii)} \quad w[x_{j-1}, x_j, x_{j+1}] = (w[x_j, x_{j+1}] - w[x_{j-1}, x_j])/(h_{j-\frac{1}{2}} + h_{j+\frac{1}{2}}) \end{aligned} \tag{3.3.14}$$

and we define

$$\text{(i)} \quad \|w\|_{D^1} \equiv \left\{ \sum_{j=0}^{J-1} h_{j+\frac{1}{2}} (w[x_j, x_{j+1}])^2 \right\}^{\frac{1}{2}}$$

$$\text{(ii)} \quad \|w\|_{D^2} \equiv \left\{ \sum_{j=1}^{J-1} (h_{j-\frac{1}{2}} + h_{j+\frac{1}{2}})(w[x_{j-1}, x_j, x_{j+1}])^2 \right\}^{\frac{1}{2}}. \qquad (3.3.15)$$

(Note that these are norms on S^h and we could have used (i) in Theorem 3.2, $\|(v^h)'\|_{L^2(I)} = \|v^h\|_{D^1}$ for all $v^h \in S^h$, but this was unnecessary since $S^h \subset H^1(I)$.)

We can now show that if the coefficient p in the differential equation is constant then, by considering convergence in the D^2 norm, the iterates of (3.3.3) converge at an $O(h^2)$ rate.

Lemma 3.1

If p is constant and v^h is the piecewise linear Galerkin approximation to v,

$$\|v - v^h\|_{D^2} \leqq C \|v - v^h\|_{L^2(I)} \quad . \qquad (3.3.16)$$

Proof

$$\begin{aligned}(v - v^h)[x_{j-1}, x_j, x_{j+1}] &= -(h_{j-\frac{1}{2}} + h_{j+\frac{1}{2}})^{-1} \int_a^b (v - v^h)\, \phi_j \, dx \\ &= (h_{j-\frac{1}{2}} + h_{j+\frac{1}{2}})^{-1}\, p^{-1} \int_a^b q(v - v^h)\phi_j \, dx\end{aligned}$$

and hence

$$|(v - v^h)[x_{j-1}, x_j, x_{j+1}]| \leqq C(h_{j-\frac{1}{2}} + h_{j+\frac{1}{2}})^{-\frac{1}{2}} \left\{ \int_{x_{j-1}}^{x_{j+1}} (v - v^h)^2 dx \right\}^{\frac{1}{2}}$$

and

$$\|v - v^h\|_{D^2} \leqq C \|v - v^h\|_{L^2(I)} \quad ■$$

Theorem 3.3

If p is constant and

$$\left\{ \sum_{j=0}^{J-1} \|(\tilde{v}^h)''\|^2_{L^2(I_{j+\frac{1}{2}})} \right\}^{\frac{1}{2}} \leqq \alpha \|\tilde{v}^h\|_{D^2} \qquad \forall \tilde{v}^h \in T^h \qquad (3.3.17)$$

with α independent of h, then

$$\|u^h_{i+1}-u^h_I\|_{D^2} \leqq C\{h^2\|u^h_i-u^h_I\|_{D^2} + h\|u - pu^h_I\|_{H^1(I)}\} \quad . \qquad (3.3.18)$$

Proof

Since $u^h_i - u^h_I - P(u^h_i-u^h_I)$ and $u - pu^h_I$ are both zero at the nodes

$$\|u^h_{i+1}-u^h_I\|_{D^2} = \|u^h_{i+1}-u^h_I - \{u^h_i-u^h_I-P(u^h_i-u^h_I)\} - (u-Pu^h_I)\|_{D^2}$$

and Lemma 3.1 using (3.3.5) shows that

$$\|u^h_{i+1}-u^h_I\|_{D^2} \leqq C\|u^h_{i+1}-u^h_I - \{u^h_i-u^h_I-P(u^h_i-u^h_I)\} - (u-Pu^h_I)\|_{L^2(I)} \quad .$$

Then using standard Galerkin error analysis and (3.3.17) we have

$$\|u^h_{i+1}-u^h_I\|_{D^2} \leqq C\{h^2(\sum_{j=0}^{J-1}\|(P(u^h_i-u^h_I))''\|^2_{L^2(I_{j+\frac{1}{2}})})^{\frac{1}{2}} + h\|u-Pu^h_I\|_{H^1(I)}\}$$

$$\leqq C\{h^2\|u^h_i-u^h_I\|_{D^2} + h\|u-Pu^h_I\|_{H^1(I)}\} \qquad \blacksquare$$

We regard (3.3.17) as a natural stability assumption and it is satisfied easily if T^h is a span of cubic splines for example. It corresponds to the boundedness of P in the H^1 norm assumed in Theorem 3.2. From Lemma 3.1 it follows that

$$\|u^h_0 - u^h_I\|_{D^2} \leqq Ch^2\|u\|_{H^2(I)} \qquad (3.3.19)$$

and so

$$\|u^h_i - u^h_I\|_{D^2} \leqq Ch^{2(i+1)} \qquad (3.3.20)$$

if u is sufficiently smooth and T^h is of sufficient approximating power. Convergence in the D^2 norm means that $O(h^{2(i+1)})$ approximations to u'' may be obtained by applying suitable difference stencils to u^h_i provided that u is smooth enough.

If p is not constant the iterates will only converge at a rate higher than $O(h)$ if the mesh is smooth.

Lemma 3.2

If $u \in H^3(I)$ and $p \in W^{2,\infty}(I)$ then

$$\|u-u^h\|_{D^2} \leq C\{(h^2 + \delta h)\,\|u\|_{H^2(I)} + h^2\,\|u\|_{H^3(I)}\} \tag{3.3.21}$$

where $\delta h \equiv \max\limits_{1 \leq j \leq J-1} \{|h_{j+\frac{1}{2}} - h_{j-\frac{1}{2}}|\}$.

Proof

$$\begin{aligned}
(u-u^h)[x_{j-1},x_j,x_{j+1}] &= -(h_{j-\frac{1}{2}}+h_{j+\frac{1}{2}})^{-1} \int_{x_{j-1}}^{x_{j+1}} (u-u^h)'\phi_j' \,dx \\
&= (h_{j-\frac{1}{2}}+h_{j+\frac{1}{2}})^{-1} p(x_j)^{-1} \int_{x_{j-1}}^{x_{j+1}} [(p-p(x_j))(u-u^h)'\phi_j' + q(u-u^h)\phi_j]dx \\
&= (h_{j-\frac{1}{2}}+h_{j+\frac{1}{2}})^{-1} p(x_j)^{-1} \int_{x_{j-1}}^{x_{j+1}} [q(u-u^h)\phi_j - p'(u-u^h)\phi_j']dx \\
&\quad + h_{j+\frac{1}{2}}\, p[x_j,x_{j+1}](u_I^h-u^h)[x_j,x_{j+1}] \\
&\quad + h_{j-\frac{1}{2}}\, p[x_{j-1},x_j](u_I^h-u^h)[x_{j-1},x_j] \\
&\quad + (h_{j-\frac{1}{2}}+h_{j+\frac{1}{2}})p[x_{j-1},x_j,x_{j+1}](u_I^h-u^h)(x_j)\ \} .
\end{aligned}$$

Each of the components of the r.h.s. is easily bounded apart from

$$\begin{aligned}
\left|\int_{x_{j-1}}^{x_{j+1}} p'(u-u^h)\phi_j' \,dx\right| &= \left|h_{j+\frac{1}{2}}^{-1} \int_{x_j}^{x_{j+1}} p'(u-u^h)dx - h_{j-\frac{1}{2}}^{-1} \int_{x_{j-1}}^{x_j} p'(u-u^h)dx\right| \\
&\leq \left|h_{j+\frac{1}{2}}^{-1} \int_{x_j}^{x_{j+1}} (p'-p'(x_j))(u-u^h)\,dx \right. \\
&\qquad \left. - h_{j-\frac{1}{2}}^{-1} \int_{x_{j-1}}^{x_j} (p'-p'(x_j))(u-u^h)dx\right| \\
&\quad + |p'(x_j)|\ \left|h_{j+\frac{1}{2}}^{-1} \int_{x_j}^{x_{j+1}} (u_I^h-u^h)dx - h_{j-\frac{1}{2}}^{-1} \int_{x_{j-1}}^{x_j} (u_I^h-u^h)dx\right| \\
&\quad + |p'(x_j)|\ \left|h_{j+\frac{1}{2}}^{-1} \int_{x_j}^{x_{j+1}} (u-u_I^h)dx - h_{j-\frac{1}{2}}^{-1} \int_{x_{j-1}}^{x_j} (u-u_I^h)dx\right| \\
&\leq C(h_{j-\frac{1}{2}}+h_{j+\frac{1}{2}})^{\frac{1}{2}} \{\|u-u^h\|_{L^2(I_{j-\frac{1}{2}} \cup I_{j+\frac{1}{2}})} \\
&\quad + \|u^h-u_I^h\|_{H^1(I_{j-\frac{1}{2}} \cup I_{j+\frac{1}{2}})} + \delta h\,\|u\|_{H^2(I_{j-\frac{1}{2}} \cup I_{j+\frac{1}{2}})} \\
&\quad + h^2\,\|u\|_{H^3(I_{j-\frac{1}{2}} \cup I_{j+\frac{1}{2}})}\ \} .
\end{aligned}$$

Hence

$$\|u-u^h\|_{D^2} \equiv \{ \sum_{j=1}^{J-1} (h_{j-\frac{1}{2}}+h_{j+\frac{1}{2}})((u-u^h)[x_{j-1},x_j,x_{j+1}])^2\}^{\frac{1}{2}}$$

$$\leqq C\{(h^2 + \delta h) \|u\|_{H^2(I)} + h^2 \|u\|_{H^3(I)}\} \quad ■$$

This lemma enables us to obtain a convergence result for the first iterate u_1^h .

Theorem 3.4

If $u \in H^3(I)$, $p \in W^{2,\infty}(I)$, $\delta h \leqq Ch^2$ uniformly and (3.3.17) holds then

$$\|u_1^h - u_I^h\|_{L^2(I)} \leqq C\{h^4 + h\|u - Pu_I^h\|_{H^1(I)}\} \quad . \qquad (3.3.22)$$

Proof

Proceeding as in Theorem 3.3

$$\|u_1^h-u_I^h\|_{L^2(I)} \leqq \|u_1^h-u_I^h - \{u_0^h-u_I^h-P(u_0^h-u_I^h)\} - (u-Pu_I^h)\|_{L^2(I)}$$

$$+ \|u_0^h-u_I^h-P(u_0^h-u_I^h) + (u-Pu_I^h)\|_{L^2(I)}$$

$$\leqq C\{h^2\|u_0^h - u_I^h\|_{D^2} + h\|u-Pu_I^h\|_{H^1(I)}\}$$

and thus the result follows from Lemma 3.2 ■

Therefore, if T^h is capable of $O(h^3)$ approximation in the H^1 norm for $u \in H^4$, we would have

$$\|u_1^h - u_I^h\|_{L^2(I)} = O(h^4) \quad . \qquad (3.3.23)$$

To show that subsequent iterates can converge at an $O(h^2)$ rate demands a more detailed analysis and to simplify this we assume henceforth that the mesh is uniform i.e. $h_{j+\frac{1}{2}} \equiv h$. In any case:-

(a) the smoothness required of a non-uniform mesh would be so great that it would normally only be satisfied by uniformity,

(b) u is assumed to be very smooth so there is no reason for using a non-uniform mesh.

We also need the following additional notation for higher-order divided difference semi-norms:-

$$\|w\|_{D^n} \equiv \{\sum_{j=0}^{J-n} nh\, \partial_n\, (w_j)^2\}^{\frac{1}{2}} \qquad (3.3.24)$$

where

$$\partial_n(w_j) \equiv w[x_j, x_{j+1}, \ldots, x_{j+n}] \quad . \qquad (3.3.25)$$

Now it can be shown that the higher-order differences of $u - u^h$ are also $O(h^2)$.

Lemma 3.3

If $n \geqq 2$, $u \in H^{n+1}(I)$, $p \in W^{n,\infty}(I)$ and $q \in W^{n-2,\infty}(I)$ then

$$\|u - u^h\|_{D^n} \leqq C(h^2 \|u\|_{H^{n+1}(I)} + \sum_{k=0}^{n-1} \|u - u^h\|_{D^k}). \qquad (3.3.26)$$

Proof

$$\begin{aligned}\partial_n((u-u^h)_{j-1}) &= (n(n-1)h)^{-1}\partial_{n-2}(p(x_j)^{-1}\int_{x_{j-1}}^{x_{j+1}}(p-p(x_j))(u-u^h)'\phi_j' \\ &\qquad + q(u-u^h)\phi_j) \\ &= (n(n-1)h)^{-1}\partial_{n-2}(p(x_j)^{-1}\{\int_{x_{j-1}}^{x_{j+1}} q(u-u^h)\phi_j - p'(u-u^h)\phi_j' \\ &\quad + h\partial_1(p_j)\partial_1((u-u^h)_j) + h\partial_1(p_{j-1})\partial_1((u-u^h)_{j-1}) \\ &\quad + 2h\partial_2(p_{j-1})(u-u^h)(x_j)\}\end{aligned}$$

as in Lemma 3.2 and then (3.3.26) follows by repeated use of Leibnitz's formula for divided differences of products. ■

Since we already know that $\|u - u^h\|_{D^k}$ $k = 0,1,2$ are $O(h^2)$, the higher divided differences will also be $O(h^2)$ by induction, provided that u, p and q are sufficiently smooth. This leads to the $O(h^2)$ rate of convergence for the iterates in (3.3.3).

Theorem 3.5

If $m \geqq 2$, $u \in H^{2m}(I)$, $p \in H^m(I)$, $q \in H^{m-2}(I)$ and

$$\left\{ \sum_{j=0}^{J-k} \|\partial_K((\tilde{v}^h)'')\|^2_{L^2(I_{j+\frac{1}{2}})} \right\}^{\frac{1}{2}} \leqq \alpha \|\tilde{v}^h\|_{D^{k+2}} \qquad \forall \tilde{v}^h \in T^h \tag{3.3.27}$$

for $0 \leqq k \leqq m-2$ with α independent of h, then

$$\|u_i^h - u_I^h\|_{D^k} \leqq Ch^{2(i+1)} \|u\|_{H^{2m}(I)} \tag{3.3.28}$$

for $0 \leqq i \leqq m-1$, $0 \leqq k \leqq m - i$.

Proof

For $i = 0$ Lemma 3.3 gives the result. For $0 < i \leqq m-1$ and $k = 1$ we have

$$\|u_i^h - u_I^h\|_{H^1(I)} \leqq C\left\{h^2 \|u_{i-1}^h - u_I^h\|_{D^2} + \|u - Pu_I^h\|_{L^2(I)}\right\}$$

by adapting the argument of Theorem 3.2. For $0 < i < m-1$ and $k \geqq 2$ we have

$$\begin{aligned}\partial_K\left\{(u_i^h - u_I^h)_{j-1}\right\} = \partial_K\Big\{(u_i^h - u_I^h)_{j-1} - [u_{i-1}^h - u_I^h - P(u_{i-1}^h - u_I^h)]_{j-1} \\ - (u - Pu_I^h)_{j-1}\Big\} \\ = -(k(k-1)h)^{-1} \partial_{K-2}\left(\int w' \phi_j'\right)\end{aligned}$$

where $w = u_i^h - u_I^h - \{u_{i-1}^h - u_I^h - P(u_{i-1}^h - u_I^h)\} - (u - Pu_I^h)$. Hence using

$$-\int w'\phi_j' = p(x_j)^{-1}\int_{x_{j-1}}^{x_{j+1}} (p - p(x_j))w'\phi_j' + qw\phi_j$$

$$= p(x_j)^{-1}\left\{\int_{x_{j-1}}^{x_{j+1}} qw\phi_j - p'w\phi_j' + \partial_1(p_{j-1})(u_i^h - u_I^h)(x_{j-1}) - \partial_1(p_j)(u_i^h - u_I^h)(x_{j+1})\right\}$$

and repeated application of Leibniz's formula gives

$$\|u_i^h - u_I^h\|_{D^K} \leq C\sum_{\ell=1}^{2}\left\{\|u^h - u_I^h\|_{D^{k-\ell}} + h^2\|u_{i-1}^h - u_I^h\|_{D^{k-\ell+2}} + h^2\left(\int_a^{x_{J-k+\ell+1}} (\partial_{K-\ell}(u-pu_I^h))^2\right)^{\frac{1}{2}}\right\} .$$

Combining these results proves the theorem ■

3.3.1 Numerical Results

The two tables below contain results for the differential equation

$$-(\cos x\, u')' + xu = f \qquad u(0) = u(1) = 0,$$

where f is chosen so that the solution is

$$u(x) \equiv x(1-x)e^x \quad .$$

J	4	8	16	32	64
D0	0.59(-2)	0.16(-2)	0.40(-3)	0.10(-3)	0.25(-4)
D1	0.20(-1)	0.58(-2)	0.15(-2)	0.38(-3)	0.95(-4)
D2	0.58(-1)	0.21(-1)	0.61(-2)	0.16(-2)	0.43(-3)
DC	0.47(-3)	0.27(-4)	0.12(-5)	0.46(-7)	0.18(-8)

TABLE 3.4: One iteration of defect-correction

J	8	16	32	64
DC1	0.12(-4)	0.77(-6)	0.48(-7)	0.30(-8)
DC2	0.28(-6)	0.33(-8)	0.41(-10)	0.58(-11)

TABLE 3.5: Two iterations of defect-correction

Approximations to the above problem were computed on a uniform mesh with J sub-divisions. The first three lines in Table 3.4 are the errors in the Galerkin solution point-values, first divided differences and second divided differences respectively; calculated by the formulae

(a) $$\{ 2h \sum_{j=1}^{J-1} (u(x_j)-u^h(x_j))^2\}^{\frac{1}{2}}$$

(b) $$\{ h \sum_{j=1}^{J} ((u-u^h)[x_{j-1},x_j])^2\}^{\frac{1}{2}}$$

(c) $$\{ 2h \sum_{j=1}^{J-1} ((u-u^h)[x_{j-1},x_j,x_{j+1}])^2\}^{\frac{1}{2}} .$$

The results support the theoretical predictions of $O(h^2)$ convergence.

The last line in Table 3.4 contains the error in the defect-correction solution, calculated as in (a) above, when one iteration with piecewise Lagrange cubic polynomials was used. Table 3.5 contains the errors in the defect correction solutions, after one and two iteractions, when piecewise Lagrange quintic polynomials were used. The results support the theoretical predictions of $O(h^4)$, $O(h^4)$ and $O(h^6)$ convergence respectively.

One point Gauss quadrature was used to compute the basic piecewise-linear Galerkin solution, while subsequent defect correction iterations involved the calculation of a right-hand side with two or three Gauss points. The original $O(h^2)$ accurate coefficient matrix however was retained.

3.4 Recovery through Deferred Correction

An alternative method of obtaining $O(h^4)$ approximations to u is to use an idea similar to the deferred correction approach with finite differences. We do not, however, need to use asymptotic expansions directly.

If u_I^h is the piecewise linear interpolate of u

$$u_I^h(x) = u^h(x) - a(G^h(x,.),u-u_I^h) \tag{3.4.1}$$

and on each sub-interval

$$u(x) - u_I^h(x) = \int_{x_j}^{x_{j+1}} g(x,t)u''(t)dt \tag{3.4.2}$$

where g is the Peano kernel for linear interpolation. Thus if z^h is an approximation to u" we may form

$$w(x) = \int_{x_j}^{x_{j+1}} g(x,t)z^h(t)dt \tag{3.4.3}$$

and w^h, the Galerkin approximation to w, will give a correction to subtract from u^h in (3.4.1). Hence an improved estimate of u_I^h is obtained.

Theorem 3.5

If $p \in W^{1,\infty}(I)$ then

$$\|u_I^h - (u^h-w^h)\|_{L^\infty(I)} \leqslant Ch^2\|u''-z^h\|_{L^2(I)} \quad . \tag{3.4.4}$$

Proof

At a node x_j

$$u_I^h(x_j)-(u^h-w^h)(x_j) = a(G^h(x_j,.),\quad w-(u-u_I^h)) \qquad (3.4.5)$$
$$= a(G^h(x_j,.)-G(x_j,.),\quad w-(u-u_I^h))$$

since w and $u-u_I^h$ are zero at the nodes. Thus

$$\|u_I^h-(u^h-w^h)\|_{L^\infty(I)} \leqslant C \max_{1\leqslant j\leqslant J-1} \{\|G(x_j,.)-G^h(x_j,.)\|_{H^1(I)}\}\|w-(u-u_I^h)\|_{H^1(I)}$$
$$\leqslant Ch^2\|u''-z\|_{L^2(I)} \qquad (3.4.6)$$

using the smoothness of the Green's function. ■

There are several ways of generating an $O(h^2)$ piecewise linear approximation to u" and thus computing w^h will give $O(h^4)$ approximations to u at the nodes. If the mesh is uniform then we set

$$z^h(x_j) = (u^h(x_{j+1}) - 2u^h(x_j) + u^h(x_{j-1}))/h^2 \qquad (3.4.7)$$

for $j = 1 \rightarrow J - 1$ and estimate u" at the end points by linear extrapolation i.e.

$$\begin{aligned} &(a) \qquad z^h(x_0) = 2\, z^h(x_1) - z^h(x_2) \\ &(b) \qquad z^h(x_J) = 2z^h(x_{J-1}) - z^h(x_{J-2}) \; . \end{aligned} \qquad (3.4.8)$$

Then

$$\|u''-z^h\|_{L^2(I)} \leqslant \|u''-y^h\|_{L^2(I)} + \|y^h-z^h\|_{L^2(I)} \; , \qquad (3.4.9)$$

where y^h is the piecewise linear function with nodal values $u''(x_j)$

$j = 0 \longrightarrow J$, and so

$$\|u''-y^h\|_{L^2(I)} \leq Ch^2\|u^{iv}\|_{L^2(I)} \tag{3.4.10}$$

and $$\|y^h-z^h\|_{L^2(I)} \leq \{ \frac{h}{2} (y^h(x_0)-z^h(x_0))^2 + h \sum_{j=1}^{J-1} (y^h(x_j)-z^h(x_j))^2 + + \frac{h}{2} (y^h(x_J)-z^h(x_J))^2\}^{\frac{1}{2}} .$$

At the internal nodes the last expression is bounded by

$$\begin{aligned}|y^h(x_j)-z^h(x_j)| &\leq |u''(x_j)-2u[x_{j-1},x_j,x_{j+1}]| + 2|(u-u^h)[x_{j-1},x_j,x_{j+1}]| \\ &\leq C\{h^{3/2}\|u^{iv}\|_{L^2(I_{j-\frac{1}{2}}\cup I_{j+\frac{1}{2}})} + h^{-\frac{1}{2}}\|u-u^h\|_{L^2(I_{j-\frac{1}{2}}\cup I_{j+\frac{1}{2}})}\end{aligned} \tag{3.4.11}$$

using standard approximation results and (3.3.16). At the initial node

$$\begin{aligned}|y^h(x_0)-z^h(x_0)| &\leq |u''(x_0)-(2u''(x_1)-u''(x_2))| \\ &+ |2(u''(x_1)-2u[x_0,x_1,x_2])-(u''(x_2)-2u[x_1,x_2,x_3])| \\ &+ |2(u-u^h)[x_0,x_1,x_2] - (u-u^h)[x_1,x_2,x_3]| \\ &\leq C\{h^{3/2}\|u^{iv}\|_{L^2(I_{\frac{1}{2}}\cup I_{3/2}\cup I_{5/2})} + h^{-\frac{1}{2}}\|u-u^h\|_{L^2(I_{\frac{1}{2}}\cup I_{3/2}\cup I_{5/2})}\}\end{aligned} \tag{3.4.12}$$

using the Peano kernel error formula for linear extrapolation. Inserting (3.4.11), (3.4.12) and the analogous result at the final node into (3.4.10) leads to

$$\|u''-z\|_{L^2(I)} \leq Ch^2\|u\|_{H^4(I)} .$$

If the mesh is not uniform a possible appraoch is to compute for

$j=1 \rightarrow J$

$$\tilde{z}_{j-\frac{1}{2}} = \frac{1}{p(x_{j-\frac{1}{2}})} \quad \{q(x_{j-\frac{1}{2}})u^h(x_{j-\frac{1}{2}})-p'(x_{j-\frac{1}{2}})\; u^h[x_{j-1},x_j] - f(x_{j-\frac{1}{2}})\} \tag{3.4.13}$$

and then obtain the nodal values of z^h by linear interpolation and extrapolation i.e.

$$\begin{aligned} z^h(x_0) &= \{(2h_{\frac{1}{2}}+h_{3/2})\tilde{z}_{\frac{1}{2}} - h_{\frac{1}{2}}\tilde{z}_{3/2}\}/(h_{\frac{1}{2}}+h_{3/2}) \\ z^h(x_j) &= (h_{j+\frac{1}{2}}\tilde{z}_{j-\frac{1}{2}} + h_{j-\frac{1}{2}}\tilde{z}_{j+\frac{1}{2}})/(h_{j-\frac{1}{2}}+h_{j+\frac{1}{2}}) \qquad j = 1 \rightarrow J-1 \\ z^h(x_J) &= \{(2h_{J-\frac{1}{2}}+h_{J-3/2})\tilde{z}_{J-\frac{1}{2}}-h_{J-\frac{1}{2}}\tilde{z}_{J-3/2}\}/(h_{J-\frac{1}{2}}+h_{J-3/2}) \quad . \end{aligned} \tag{3.4.14}$$

To prove that $u'' - z^h$ is $O(h^2)$ we first note that

$$\begin{aligned} u''(x_{j-\frac{1}{2}})-\tilde{z}_{j-\frac{1}{2}} = \frac{1}{p(x_{j-\frac{1}{2}})} \; &\{q(x_{j-\frac{1}{2}})(u-u^h)(x_{j-\frac{1}{2}})-p'(x_{j-\frac{1}{2}})(u-u^h)[x_{j-1},x_j] \\ &-p'(x_{j-\frac{1}{2}})(u'(x_{j-\frac{1}{2}})-u[x_{j-1},x_j])\} \end{aligned} \tag{3.4.15}$$

and so

$$\begin{aligned} |u''(x_{j-\frac{1}{2}})-\tilde{z}_{j-\frac{1}{2}}| \leqslant C\{&|(u-u^h)(x_{j-1})| + |(u-u^h)(x_j)| + |(u-u^h)[x_{j-1},x_j]| \\ &+ h^{3/2}_{j-\frac{1}{2}} \, \|u'''\|_{L^2(I_{j-\frac{1}{2}})} \end{aligned} \tag{3.4.16}$$

and thus

$$\{\sum_{j=1}^{J} h_{j-\frac{1}{2}}|u''(x_{j-\frac{1}{2}})-\tilde{z}_{j-\frac{1}{2}}|^2\}^{\frac{1}{2}} \leqslant Ch^2\|u\|_{H^3(I)} \quad . \tag{3.4.17}$$

If y^h, however, is the piecewise linear function obtained as in (3.4.14) but using the true mid-point values of u" rather than $\tilde{z}$ then

$$\|y^h - z^h\|_{L^2(I)} \leqslant C\{\sum_{j=1}^{J} h_{j-\frac{1}{2}} |u''(x_{j-\frac{1}{2}}) - \tilde{z}_{j-\frac{1}{2}}|^2\}^{\frac{1}{2}} \qquad (3.4.18)$$

provided that the mesh is not too distorted: i.e. there exist positive constants c_1 and c_2 independent of h and j such that

$$c_1 h_{j-\frac{1}{2}} \leqslant h_{j+\frac{1}{2}} \leqslant c_2 h_{j-\frac{1}{2}} \qquad (3.4.19)$$

for j = 1 → J-1.   Hence

$$\|u'' - z^h\|_{L^2(I)} \leqslant \|u'' - y^h\|_{L^2(I)} + \|y^h - z^h\|_{L^2(I)} \leqslant Ch^2 \|u\|_{H^4(I)} \qquad (3.4.20)$$

where the first term on the right-hand side is bounded by the error in linear interpolation and extrapolation.

REFERENCES

BABUSKA, I. & RHEINBOLDT, W.C. (1978) A posteriori error estimates for the finite element method. Int. J. Num. Meth. Eng. 12, 1597-1615.

BARRETT, J.W. & MORTON, K.W. (1980) Optimal finite element solutions to diffusion-convection problems in one dimension. Int. J. Num. Meth. Engng. 15, 1457-1474.

BARRETT, J.W. & MORTON, K.W. (1981) Optimal Petrov-Galerkin methods through approximate symmetrization. IMA J. Numer. Anal. 1, 439-468.

BARRETT, J.W. & MORTON, K.W. (1984) Approximate symmetrization and Petrov-Galerkin methods for diffusion-convection problems. Comp. Meths. in Appl. Mech. Engng. 45, 97-122.

BEHFOROOZ, G.H. & PAPAMICHAEL, N. (1979) End conditions for cubic spline interpolation. J. Inst. Maths. Applics. 23, 355-366.

BRAMBLE, J.H. & SCHATZ, A.H. (1976) Estimates for spline projections. R.A.I.R.O. Analyse Numerique 10, 5-37.

CHANDLER, G.A. (1980) Superconvergence for second kind integral equations. Application and Numerical Solution of Integral Equations (Eds. R.S. Anderssen, F.R. de Hoog & M.A. Lukas) Sijthoff and Nordhoff.

CULLEN, M.J.P. & MORTON, K.W. (1980) Analysis of evolutionary error in finite element and other methods. J. Comp. Phys. 34, 245-267.

CURTIS, A.R. & POWELL, M.J.D. (1967) Using cubic splines to approximate functions of one variable to prescribed accuracy. A.E.R.E. Report 5602 Harwell, England.

DOUGLAS, J. JNR & DUPONT, T. (1974) Galerkin approximations for the two point boundary problem using continuous piecewise polynomial spaces. Numer. Math. 22, 99-109.

GARTLAND, E.C. JNR (1984) Computable pointwise error bounds and the Ritz method in one dimension. SIAM J. Numer. Anal. 21, 84-100.

GOLOMB, M. & WEINBERGER, H.F. (1959) Optimal approximation end error bounds, Symp. on Numerical Approximation (ed. R.E. Langer), Madison, 117-190.

HEMKER, P.W. (1977) A numerical study of stiff two-point boundary problems. Thesis, Mathematisch Centrum, Amsterdam.

LUCAS, T.R. (1974) Error bounds for interpolating cubic splines under various end conditions. SIAM J. Numer. Anal. *11*, 569-584.

MICCHELLI, C.A. & RIVLIN, T.J. (1976) A survey of optimal recovery. *Optimal Estimation in Approximation Theory* (Eds. C.A. Micchelli & T.J. Rivlin), Plenum Press, New York, 1-54.

MORTON, K.W. (1982a) Finite element methods for non-self-adjoint problems. Proc. SERC Summer School, 1981 (Ed. P.R. Turner), Lect. Notes in Maths. 965, Springer-Verlag, Berlin, 113-148.

MORTON, K.W. (1982b) Shock capturing, fitting and recovery. Proc. 8th Int. Conf. on Numerical Methods in Fluid Dynamics, Aachen. (Ed. E. Krause), Lect. Notes in Physics 170, Springer-Verlag, Berlin, 77-93.

MORTON, K.W. (1984) Generalised Galerkin methods for hyperbolic problems. Oxford Univ. Comp. Lab. Report 84/1. Comp. Meths. in Appl. Mech. Engng. (to appear).

MORTON, K.W. & SCOTNEY, B.W. (1985) Petrov-Galerkin methods and diffusion-convection problems in 2D. MAFELAP '84 (Ed. J.R. Whiteman) Academic Press, London (to appear).

REINHARDT, H.J. (1981) A posteriori error estimates for the finite element solution of a singularly perturbed linear ordinary differential equation. SIAM J. Numer. Anal. *18*, 406-430.

RICHTER, G.R. (1978) Superconvergence for piecewise polynomial Galerkin approximations for Fredholm integral equations of the second kind. Numer. Math. *31*, 63-70.

SCHULTZ, M.J. (1973) *Spline Analysis*. Prentice-Hall, New York.

SKEEL, R.D. (1981) A theoretical framework for proving accuracy results for deferred corrections. SIAM J. Numer. Anal. *19*, 171-196.

STETTER, H.J. (1978) The defect correction principle and discretization methods. Numer. Math. 29, 425-443.

STRANG, G. & FIX, G.J. (1973) An Analysis of the Finite Element Method. Prentice-Hall, New York.

VAN LEER, B. (1979) Towards the ultimate conservative difference scheme V: A second order sequel to Godunov's method. J. Comp. Phys. 32, 101-136.

VARGA, R.S. (1971) Functional Analysis and Approximation Theory in Numerical Analysis. SIAM, Philadelphia.