Report no. 98/18

Adjoint recovery of superconvergent functionals from approximate solutions of partial differential equations

Niles A. Pierce Michael B. Giles

Motivated by applications in computational fluid dynamics, we present a method for obtaining estimates of integral functionals, such as lift or drag, that have twice the order of accuracy of the computed flow solution on which they are based. This is achieved through error analysis which uses an adjoint p.d.e. to relate the local errors in approximating the flow solution to the corresponding global errors in the functional of interest. Numerical evaluation of the local residual error together with an approximation solution to the adjoint equations may thus be combined to produce a correction for the computed functional value that yields twice the order of accuracy.

Numerical results are presented for the Poisson equation in one and two dimensions, and the nonlinear quasi-one-dimensional Euler equations. The superconvergence in these cases is as predicted by the *a priori* error analysis presented in the appendix. The theory is equally applicable to nonlinear equations in complex domains in multiple dimensions, and the technique has great potential for application in a range of engineering disciplines in which a few integral quantities are a key output of numerical approximations.

Subject classifications: AMS(MOS): 65G99,76N15

Key words and phrases: partial differential equations, adjoint equations, error analysis

This research was supported by EPSRC under grant GR/K91149.

Oxford University Computing Laboratory
Numerical Analysis Group
Wolfson Building
Parks Road
Oxford, England OX1 3QD

December, 1998

Contents

1 Introduction

In aeronautical applications of computational fluid dynamics (CFD), engineers desire very accurate predictions of the lift and drag, which are defined by integrals over the entire surface of the wing or aircraft being considered [24]. They are also interested in the details of the flow field in general, but to a lesser degree of accuracy since the main purpose is to understand the qualitative nature of the flow (e.g. is there a strong shock which is producing an extensive flow separation?) in order to make design changes which will improve the lift or drag. Other areas of CFD analysis also have a particular interest in a few key integral quantities, such as total production of nitrous oxides in combustion modeling, or the net seepage of a pollutant into an aquifer when modeling soil contamination.

Integral quantities are important in other disciplines as well. In electrochemical simulations of the behavior of sensors, the quantity of interest is the total current flowing into an electrode [1]. In electromagnetics, radar cross-section calculations are concerned with the scattered field emanating from an aircraft. The amplitude of the wave propagating in a particular direction can be evaluated by a convolution integral over a closed surface surrounding the aircraft [7, 23]. Similar convolution integrals are used in the analysis of multi-port electromagnetic devices such as microwave ovens and EMR body scanners to evaluate radiation, transmission and reflection coefficients which characterize the behavior of the device.

In structural mechanics, one is sometimes concerned with the total force or moment exerted on a surface [26], but more often the quantities of most concern are point quantities such as the maximum stress or temperature. Because integral quantities can be approximated with much greater accuracy, Babuška and Miller developed a technique using an auxiliary function to represent a point quantity by an equivalent integral [3]. The same technique could be used in other applications in which it is point quantities, rather than integral quantities, which are of most importance.

Regardless of the area of application, when integral functionals based on approximate p.d.e. solutions are of significant interest, it is worth considering approaches for enhancing the accuracy of these functional approximations. The question to be addressed in the present work is the following: given an approximate solution to a p.d.e. with boundary conditions, how do errors in the solution affect the accuracy of an integral functional, and how can these functional errors be estimated and used to obtain a more accurate functional approximation?

The key is the solution of the adjoint p.d.e. with inhomogeneous terms appropriate to the functional of interest. We show that it is the adjoint solution which relates the residual error in the primal p.d.e. solution (as measured by the extent to which the numerical solution fails to satisfy the p.d.e.) to the consequent error in the computed value of the functional. Numerical approximations of the adjoint solution and primal residual errors can then be used to correct the error in the functional and obtain a new estimate which is *superconvergent* in that the remaining error is proportional to the product of the errors in the primal and adjoint solutions.

The analysis is closely related to superconvergence results in the finite element literature [3, 4, 5, 6, 12, 23, 25, 26, 29]. The key distinction is that the adjoint error correction term that is evaluated to obtain superconvergence is zero in a large class of finite element methods. Thus, these methods automatically produce superconvergent results for any integral functional without requiring the computation of an approximate adjoint solution. From a finite element perspective, this paper can therefore be viewed as extending superconvergence theory to cover numerical results obtained by any numerical method: finite difference, finite volume or finite element without natural superconvergence. Moreover, we show that the adjoint recovery technique in this paper can also be used to improve the order of accuracy of the superconvergent functionals obtained from finite element methods.

We begin by describing the approach for linear problems including simple examples of its application to the Poisson equation in one and two dimensions. To illustrate the applicability of the theory to numerical results obtained by any discretization method, the one-dimensional solutions are obtained using a finite difference method, whereas the two-dimensional results are based on a finite element discretization. These two-dimensional results demonstrate both the natural superconvergence of the finite element method and the additional orders of accuracy resulting from adjoint error analysis.

Next, we present the approach for nonlinear problems with examples of its use for the quasi-1D Euler equations, a coupled system of three nonlinear o.d.e.'s describing inviscid compressible flow in a variable area duct. The examples include cases with a sonic point at which there is a change in direction of one of the hyperbolic characteristics, and a shock at which there is a discontinuity in the flow field. For these cases, numerical results are obtained using a finite volume method typical of those used in aeronautical CFD calculations.

We conclude by discussing the difficulties and prospects for extending the theory and its implementation to nonlinear p.d.e.'s in multiple dimensions on domains of arbitrary shape.

2 Linear analysis

Let u be the solution of the linear differential equation

$$Lu = f,$$

on the domain Ω, subject to homogeneous boundary conditions for which the problem is well-posed when $f \in L_2(\Omega)$. The adjoint differential operator L^* and associated homogeneous boundary conditions are defined by the identity

$$(v, Lu) = (L^* v, u),$$

for all u, v satisfying the respective boundary conditions. Here the notation $(.,.)$ denotes an integral inner product over the domain Ω.

Suppose now that we are concerned with the value of the functional $J=(g,u)$, for a given function $g \in L_2(\Omega)$. An equivalent dual formulation of the problem is to evaluate the functional $J=(v,f)$, where v satisfies the adjoint equation

$$L^* v = g,$$

subject to the homogeneous adjoint boundary conditions. The equivalence of the two forms of the problem follows immediately from the definition of the adjoint operator.

$$(v,f) = (v,Lu) = (L^* v,u) = (g,u).$$

Digressing slightly, we note that the dual formulation of the problem is exploited in optimal design [17, 18], in which there is only one function g, corresponding to the objective function in the design optimization, but there are multiple functions f, each corresponding to a different geometric design parameter. Therefore, the dual approach is computationally much more efficient since since each design cycle requires just one adjoint calculation whereas the direct approach would require one calculation for each design variable. The existence of adjoint solution methods for design purposes [2, 9, 19, 27] means that in many cases, the building blocks are already in place for rapid exploitation of the error correction ideas in this paper.

Returning to the subject at hand, suppose that u_h and v_h are approximations to u and v, respectively, and satisfy the homogeneous boundary conditions. The subscript h denotes that the approximate solutions are derived from numerical computations using a grid with average spacing h. When using finite difference or finite volume methods, u_h and v_h might be created by interpolation through computed values at grid nodes. With finite element solutions, one might more naturally use the finite element solutions themselves, or one could again use an interpolation through nodal values. A last comment is that u_h and v_h do not have to come from a numerical computation; they could, for example, come from an asymptotic analysis yielding a uniformly valid asymptotic approximation to the solution.

Let the functions f_h and g_h be defined by

$$Lu_h = f_h, \quad L^* v_h = g_h.$$

It is assumed that u_h and v_h are sufficiently smooth that f_h and g_h lie in $L_2(\Omega)$. If u_h and v_h were equal to u and v, then f_h and g_h would be equal to f and g. Thus, the *residual errors* $f_h - f$ and $g_h - g$ are a computable indication of the extent to which u_h and v_h are not the true solutions.

Now, using the definitions and identities given above, we obtain the following expression for the functional:

$$\begin{aligned}
(g,u) &= (g,u_h) - (g_h, u_h - u) + (g_h - g, u_h - u) \\
&= (g,u_h) - (L^* v_h, u_h - u) + (g_h - g, u_h - u) \\
&= (g,u_h) - (v_h, L(u_h - u)) + (g_h - g, u_h - u) \\
&= (g,u_h) - (v_h, f_h - f) + (g_h - g, u_h - u).
\end{aligned}$$

The first term in the final expression is the value of the functional obtained from the approximate solution u_h. The second term is an inner product of the residual error $f_h - f$ and the approximate adjoint solution v_h. The adjoint solution gives the weighting of the contribution of the local residual error to the overall error in the computed functional. Therefore, by evaluating and subtracting this adjoint error term we obtain a more accurate value for the functional.

The third term is the remaining error after making the adjoint correction. If $g_h - g$ is of the same order of magnitude as $v_h - v$ then the remaining error has a bound which is proportional to the product $\|u_h - u\| \, \|v_h - v\|$ (using L_2 norms), and thus the corrected functional value is superconvergent. If the solution errors u_h–u and v_h–v are both $O(h^p)$, so that halving the grid spacing leads to a 2^p reduction in the errors, then the error in the functional is $O(h^{2p})$. Furthermore, the remaining error term can be expressed as $(g - g_h, L^{-1}(f - f_h))$ and so has the computable *a posteriori* error bound,

$$|\text{Error}| \leq \|L^{-1}\| \, \|f_h - f\| \, \|g_h - g\|,$$

with $\|L^{-1}\|$ being assumed to be finite due to wellposedness.

If u_h and v_h are taken to be the finite element solutions obtained from a Galerkin finite element method (or more generally any finite element method for which the test and trial spaces for the primal problem are interchanged to become the trial and test spaces for the adjoint problem) the adjoint correction term is always zero because of the orthogonality arising from the weak formulation of the finite element discretization. Thus, the values of all integral functionals are automatically superconvergent. However, if the operator L involves derivatives of up to degree m, then usually $f_h - f = O(h^{p-m})$ and hence the error in any functional is $O(h^{2p-m})$. This loss of accuracy is due to a lack of smoothness in the finite element solution. If a smoother interpolated solution can be recovered from the finite element solution, then there is a possibility of using the adjoint error correction to recover an improved functional estimate whose error is $O(h^{2p})$. This will be demonstrated in the second of the two examples to follow.

To conclude this section, we return to the topic of boundary conditions. For simplicity in presenting the analysis, we have assumed that the primal problem has homogeneous boundary conditions, and that the functional is simply an inner product of the whole domain and does not have a boundary integral term. More generally, inhomogeneous boundary conditions and boundary integrals in the functional are both permissible. Inhomogeneous boundary conditions for the primal problem lead to a boundary integral term for the adjoint formulation, and similarly a boundary integral in the primal form of the functional leads to inhomogeneous adjoint boundary conditions. Although the analysis is slightly more complicated, the final form of the adjoint error correction is exactly the same as before, provided the approximate solutions u_h and v_h still exactly satisfy the inhomogeneous boundary conditions. If they do not, then there is an additional correction term to account for this error.

3 Two linear examples

3.1 1D finite difference calculation

The first example is the one-dimensional Poisson equation,

$$\frac{d^2u}{dx^2} = f,$$

on the unit interval $[0,1]$ subject to homogeneous boundary conditions $u(0)=u(1)=0$. This is approximated numerically on a uniform grid, with spacing h, using a second order finite difference discretization,

$$h^{-2}\delta_x^2 u_j = f(x_j).$$

The approximate solution $u_h(x)$ is then defined by cubic spline interpolation through the nodal values u_j.

The dual problem is the Poisson equation,

$$\frac{d^2v}{dx^2} = g,$$

subject to the same homogeneous boundary conditions, and the approximate adjoint solution v_h is obtained in exactly the same manner.

Numerical results have been obtained for the case

$$f = x^3(1-x)^3, \qquad g = \sin(\pi x).$$

Figure 1 shows the residual error $f_h - f$ when $h = \frac{1}{32}$, as well as the three Gaussian quadrature points on each sub-interval which are used in the numerical integration of the inner product $(v_h, f_h - f)$. Since u_h is a cubic spline, $f_h \equiv \frac{d^2u_h}{dx^2}$ is continuous and piecewise linear. The best piecewise linear approximation to f has an error whose dominant term is quadratic on each sub-interval; this explains the scalloped shape of the residual error. Figure 2 shows the approximate adjoint solution v_h, which reveals that the residual error in the center of the domain contributes most to the overall error in the functional.

Figure 3 depicts a log-log plot of three quantities versus the number of cells: the error in the base value of the functional (g, u_h); the remaining error after subtracting the adjoint correction term $(v_h, f_f{-}f)$; the *a posteriori* error bound $\|L^{-1}\|\,\|f_h{-}f\|\,\|g_h{-}g\|$. The superimposed lines have slopes of -2 and -4, confirming that the base solution is second order accurate while the error in the corrected functional and the error bound are both fourth order. It is also worth noting that on a grid with 16 cells, which might be a reasonable choice for practical computations, the error in the corrected value of the functional is over 200 times smaller than the uncorrected error.

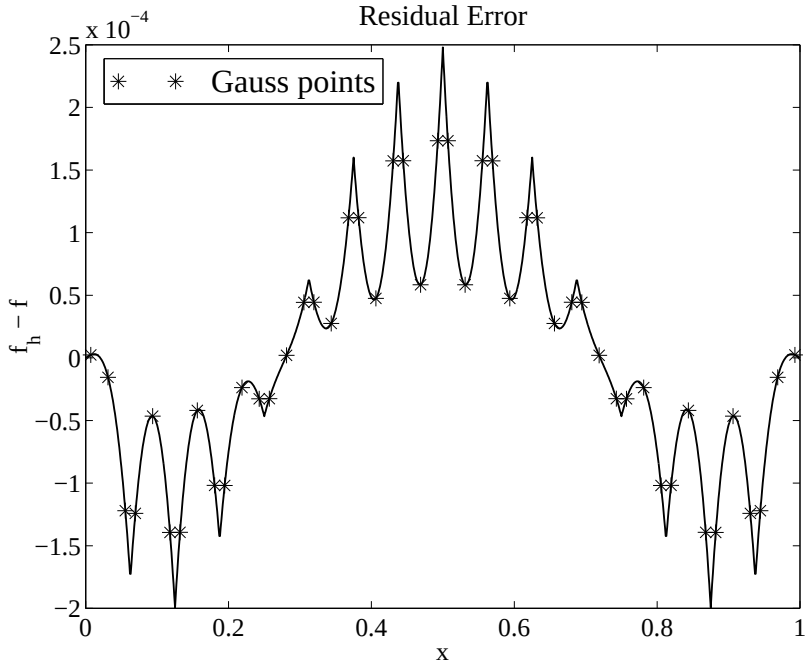

Figure 1: Residual error for 1D Poisson equation.

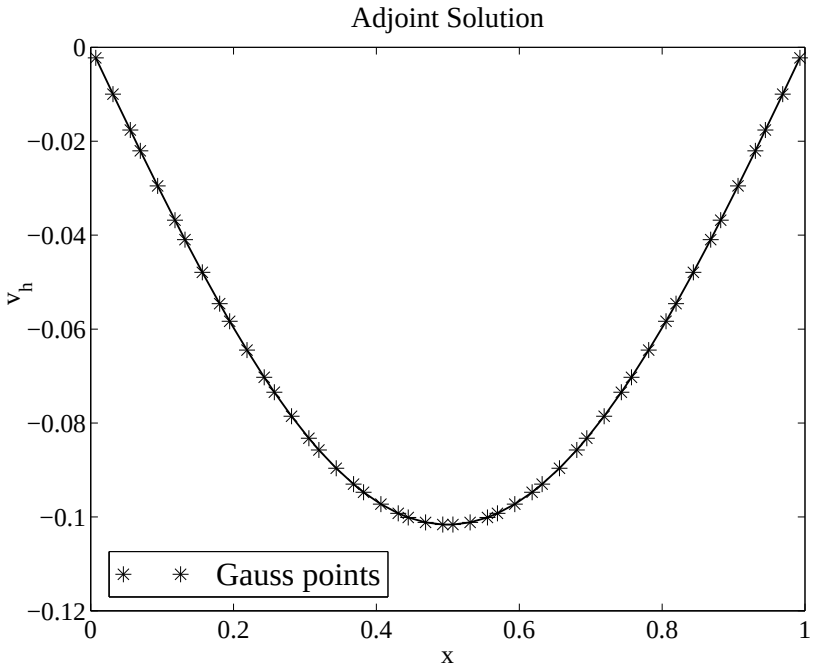

Figure 2: Adjoint solution for 1D Poisson equation.

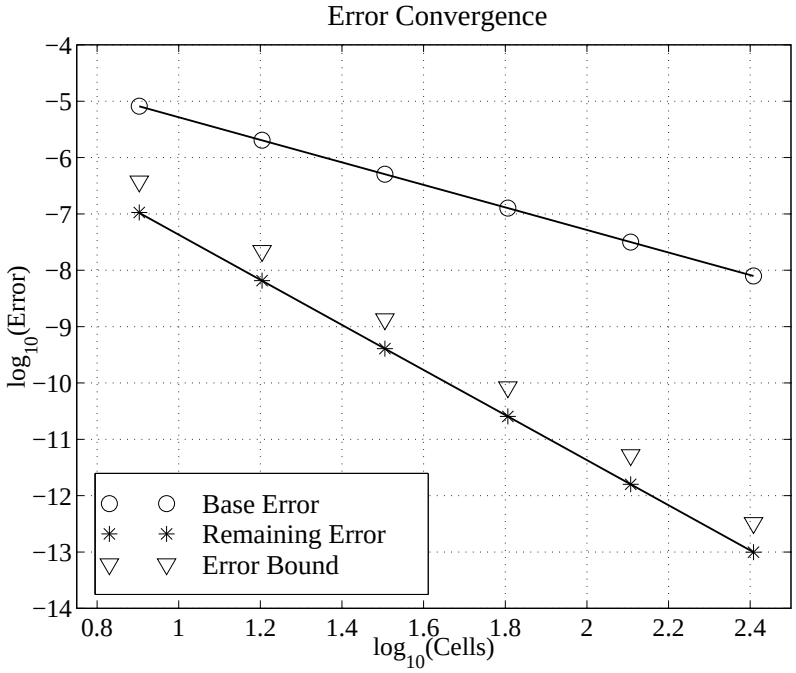

Figure 3: Functional error convergence for 1D Poisson equation.

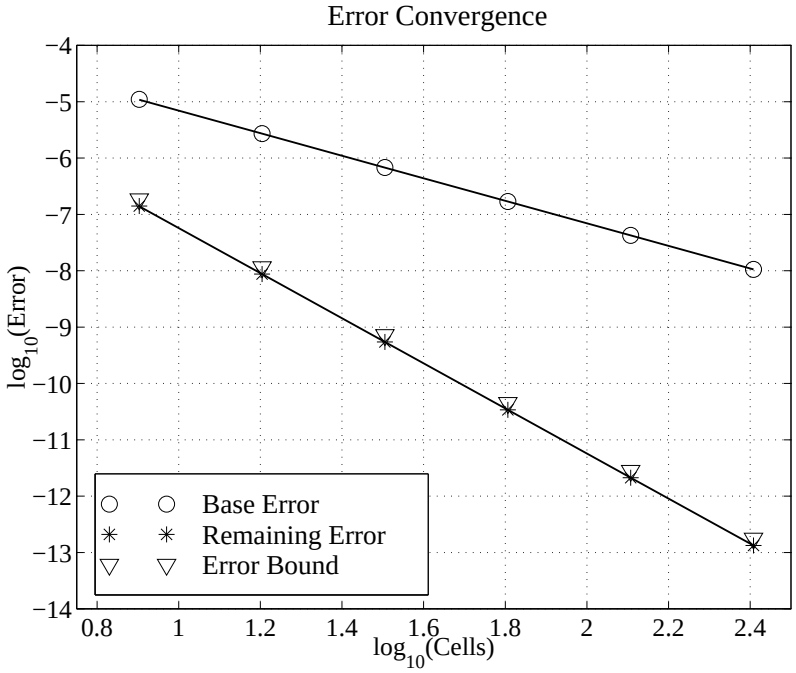

Figure 4: Functional error convergence for 2D Poisson equation.

3.2 2D finite element calculation

The second example is the two-dimensional Poisson equation,

$$\nabla^2 u = f,$$

on the unit square $[0,1] \times [0,1]$ subject to homogeneous Dirichlet boundary conditions. The dual problem is

$$\nabla^2 v = g,$$

with the same boundary conditions.

For this example, the equations are approximated using a Galerkin finite element method with piecewise bilinear elements on a uniform Cartesian grid. Recalling that in the present case, $p = m = 2$, finite element error analysis reveals that the solution error for the primal problem is $O(h^2)$ with a corresponding residual error that is $O(1)$. The inherent superconvergence of the finite element method thus yields a computed functional that is $O(h^2)$. However, by using bi-cubic spline interpolation through the computed nodal values, one can reconstruct an improved approximate solution $u_h(x,y)$ with an error which is $O(h^2)$ in the H^2 Sobolev norm, and hence has a residual error which is also $O(h^2)$. Using a similarly reconstructed approximate adjoint solution $v_h(x,y)$, the adjoint error correction term then produces a corrected functional whose accuracy is $O(h^4)$. All inner product integrals are approximated by 3×3 Gaussian quadrature on each square cell to ensure that the numerical quadrature errors are of a higher order.

Figure 4 shows the numerical results obtained for the functions

$$f(x,y) = x(1-x)\, y(1-y), \qquad g(x,y) = \sin(\pi x)\, \sin(\pi y).$$

The ordinate is the log of the number of cells in each dimension, and lines of slope -2 and -4 are again superimposed. As predicted by the analysis, the base error in the functional is second order while the corrected functional error and the error bound are again both fourth order.

4 Nonlinear analysis

For nonlinear problems, the conceptual approach is very similar, but the mathematical presentation becomes somewhat more involved. Let u be the solution of the nonlinear differential equation

$$N(u) = f,$$

subject to appropriate boundary conditions, and let the functional of interest, $J(u)$, be an integral of a nonlinear algebraic function of u over the domain Ω.

The linear differential operator L_u is defined to be the Fréchet derivative of N,

$$L_u\, \tilde{u} \equiv \lim_{\epsilon \to 0} \frac{N(u + \epsilon \tilde{u}) - N(u)}{\epsilon},$$

and, similarly, the function $g(u)$ is defined by the identity

$$(g(u), \tilde{u}) \equiv \lim_{\epsilon \to 0} \frac{J(u + \epsilon \tilde{u}) - J(u)}{\epsilon}.$$

The corresponding linear adjoint problem is then

$$L_u^* v = g,$$

subject to the appropriate homogeneous adjoint boundary conditions.

We now consider approximate solutions u_h, v_h and define f_h, g_h by

$$N(u_h) = f_h, \qquad L_{u_h}^* v_h = g_h.$$

Note the use of $L_{u_h}^*$, the Fréchet derivative based on u_h which is known, instead of L_u^* based on u which is not known.

In addition, the analysis requires averaged Fréchet derivatives $\overline{L}_{(u,u_h)}$ and $\overline{g}(u, u_h)$ defined by

$$\begin{aligned} \overline{L}_{(u,u_h)} &= \int_0^1 L|_{u+\theta(u_h - u)} \, d\theta, \\ \overline{g}(u, u_h) &= \int_0^1 g(u + \theta(u_h - u)) d\theta, \end{aligned}$$

so that

$$\begin{aligned} N(u_h) - N(u) &= \int_0^1 \frac{\partial}{\partial \theta} N(u + \theta(u_h - u)) \, d\theta \\ &= \overline{L}_{(u,u_h)} \, (u_h - u), \end{aligned}$$

and similarly

$$J(u_h) - J(u) = (\overline{g}(u, u_h), u_h - u).$$

We then obtain the following result:

$$\begin{aligned} J(u) &= J(u_h) - (\overline{g}(u, u_h), u_h - u) \\ &= J(u_h) - (g_h, u_h - u) + (g_h - \overline{g}(u, u_h), u_h - u) \\ &= J(u_h) - (L_{u_h}^* v_h, u_h - u) + (g_h - \overline{g}(u, u_h), u_h - u) \\ &= J(u_h) - (v_h, L_{u_h}(u_h - u)) + (g_h - \overline{g}(u, u_h), u_h - u) \\ &= J(u_h) - (v_h, \overline{L}_{(u,u_h)}(u_h - u)) + (g_h - \overline{g}(u, u_h), u_h - u) \\ &\qquad - (v_h, (L_{u_h} - \overline{L}_{(u,u_h)})(u_h - u)) \\ &= J(u_h) - (v_h, N(u_h) - N(u)) + (g_h - \overline{g}(u, u_h), u_h - u) \\ &\qquad - ((L_{u_h}^* - \overline{L}_{(u,u_h)}^*) v_h, u_h - u) \\ &= J(u_h) - (v_h, f_h - f) + (g_h - \overline{g}(u, u_h), u_h - u) - ((L_{u_h}^* - \overline{L}_{(u,u_h)}^*) v_h, u_h - u) \end{aligned}$$

The first term in the final line is again the functional evaluated using the approximate solution u_h. The second term is again a computable adjoint error correction term which is an inner product of the residual error and the approximate adjoint solution. The last two terms form the remaining error in the corrected functional.

The third term is similar to the remaining error term in the linear case, while the fourth term is associated with the nonlinearity in the operator $N(u)$. If the solution errors for the nonlinear primal problem and the linear adjoint problem are of the same order, and they are both sufficiently smooth that the corresponding residual errors are also of the same order, then the order of accuracy of the functional approximation after making the adjoint correction is twice the order of the primal and adjoint solutions.

An *a posteriori* error bound is harder to construct than in the linear case. Splitting the remaining error into three pieces,

$$\text{Error} = (g_h - g(u_h), u_h - u) + (g(u_h) - \overline{g}(u, u_h), u_h - u) - ((L^*_{u_h} - \overline{L}^*_{(u,u_h)})v_h, u_h - u),$$

we can obtain asymptotic error bounds by converting each inner product into an alternative representation which is asymptotically equivalent and has a computable bound. With the first inner product we have

$$(g_h - g(u_h), u_h - u) \approx (g_h - g(u_h), L_u^{-1}(f_h - f)).$$

For the second, we define G_u to be the Fréchet derivative of $g(u)$,

$$G_u \tilde{u} = \lim_{\epsilon \to 0} \frac{g(u + \epsilon \tilde{u}) - g(u)}{\epsilon},$$

and then obtain

$$\begin{aligned} (g(u_h) - \overline{g}(u, u_h), u_h - u) &\approx \tfrac{1}{2}(G_u(u_h - u), u_h - u) \\ &\approx \tfrac{1}{2}(L_u^{*\,-1} G_u L_u^{-1}(f_h - f), f_h - f). \end{aligned}$$

For the third inner product, we define the operator $H_{u,v}$ as

$$H_{u,v}\tilde{u} = \lim_{\epsilon \to 0} \frac{L^*_{u+\epsilon\tilde{u}} v - L^*_u v}{\epsilon},$$

so that

$$\begin{aligned} ((\overline{L}^*_{(u,u_h)} - L^*_{u_h})v_h, u - u_h) &\approx \tfrac{1}{2}(H_{u,v}(u - u_h), u - u_h) \\ &\approx \tfrac{1}{2}(L_u^{*\,-1} H_{u,v} L_u^{-1}(f - f_h), f - f_h). \end{aligned}$$

Together, these give the approximate asymptotic bound

$$|\ \text{Error}\ | \le c_1 \|f_h - f\|\, \|g_h - g(u_h)\| + c_2 \|f_h - f\|^2,$$

where

$$c_1 = \|L_u^{-1}\|, \qquad c_2 = \tfrac{1}{2}\left\|L_u^{*\,-1}(H_{u,v} - G_u)L_u^{-1}\right\|.$$

The problem in evaluating this *a posteriori* error bound is that c_1 and c_2 will not be known in general, and may be hard to bound analytically. A more practical option may be to estimate them computationally based on the corresponding discrete operators.

5 Nonlinear finite volume examples

The steady quasi-1D Euler equations describe the flow of an inviscid, compressible ideal gas in a variable area duct. The functional of interest is the integral of the pressure along the duct, which serves as a model for the computation of lift and drag on airfoils in two dimensions, and wings and aircraft in three dimensions.

The unsteady quasi-1D Euler equations in conservative form are

$$A\frac{\partial U}{\partial t} + \frac{\partial}{\partial x}(AF) - \frac{dA}{dx}P = 0,$$

where $A(x)$ is the cross-sectional area of the duct and U, F and P are defined as

$$U = \begin{pmatrix} \rho \\ \rho q \\ \rho E \end{pmatrix}, \qquad F = \begin{pmatrix} \rho q \\ \rho q^2 + p \\ \rho q H \end{pmatrix}, \qquad P = \begin{pmatrix} 0 \\ p \\ 0 \end{pmatrix}.$$

Here, ρ is the density, q is the velocity, p is the pressure, E is the total energy and H is the stagnation enthalpy. The system is closed by the equation of state for an ideal gas

$$H = E + \frac{p}{\rho} = \frac{\gamma p}{(\gamma - 1)\rho} + \tfrac{1}{2}q^2,$$

where γ is the ratio of specific heats.

The unsteady quasi-1D Euler equations are a hyperbolic system with three characteristic wave speeds, q, $q+c$ and $q-c$, with $c = \sqrt{\gamma p/\rho}$ being the local speed of sound. Accordingly, the nature of the steady flow solution varies depending on whether the flow is subsonic ($M<1$) or supersonic ($M>1$), where $M \equiv q/c$ is the Mach number. In order of increasing difficulty, we will consider the subsonic, isentropic transonic and shocked transonic flows depicted in terms of Mach number in Fig. 5.

Steady flow solutions are obtained by marching the nonlinear unsteady system to a steady state using a standard second order finite volume method with characteristic smoothing on a uniform computational grid. The linear adjoint problem is approximated by linearizing the nonlinear flow equations, constructing the analytic adjoint equations and boundary conditions, and then forming a discrete approximation to these on the same uniform grid as the flow solution [18, 2]. Previous research has confirmed that this produces a consistent approximation to the analytic adjoint solution which has been determined in closed form for the quasi-1D Euler equations [14].

The approximate solution $u_h(x)$ is constructed from the discrete flow solution by cubic spline interpolation of the cell-centered values of the three components of the state vector U (except in the shocked case to be described later). The flow residual f_h is then formed using analytic derivatives of this reconstruction. The approximate adjoint solution $v_h(x)$ is also obtained by cubic spline interpolation of the cell-centered values of the three components of the discrete adjoint solution. The integrals which form the base value for the functional and the adjoint correction are then approximated using 3-point Gaussian quadrature.

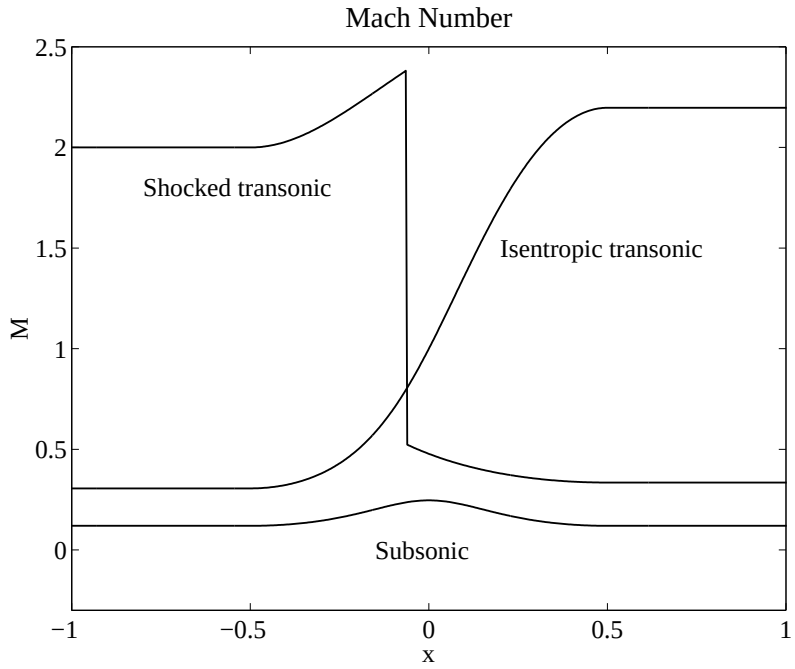

Figure 5: Mach number distributions for quasi-1D Euler equation test cases.

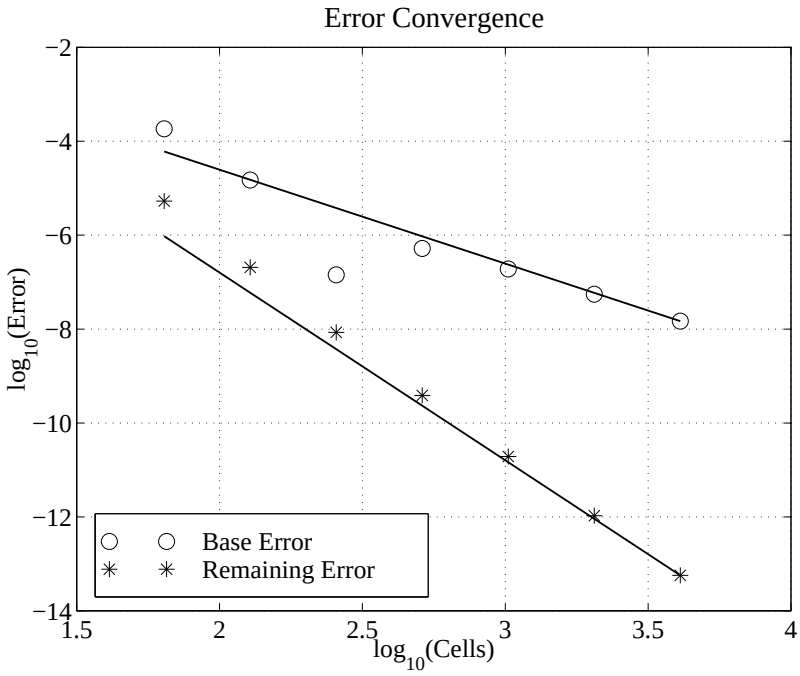

Figure 6: Error convergence for quasi-1D subsonic flow.

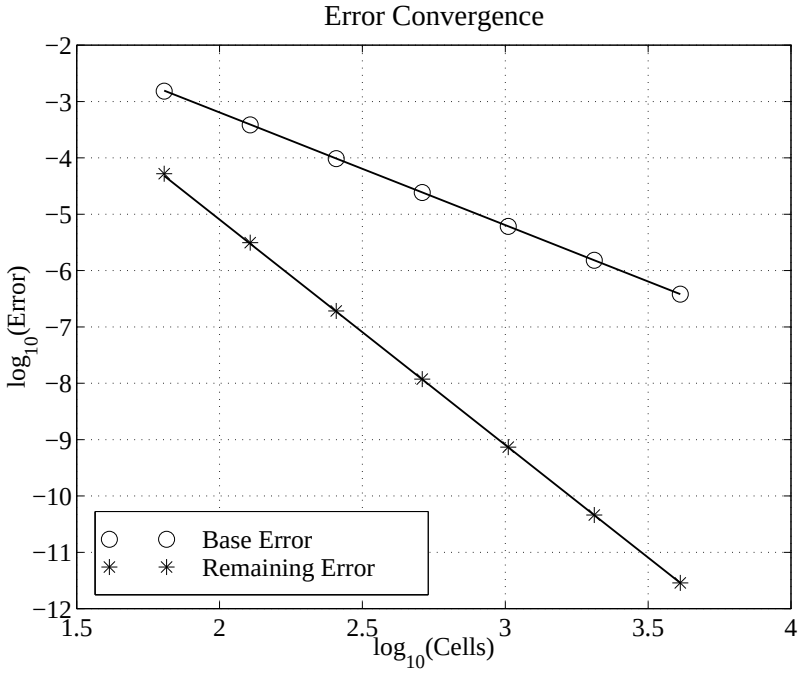

Figure 7: Error convergence for quasi-1D shock-free transonic flow.

5.1 Subsonic flow

As a first case, consider smooth subsonic flow in a converging-diverging duct. The error convergence of the computed functional is shown in Figure 6, where the superimposed lines of slope -2 and -4 demonstrate that the base error is second order and the error in the corrected functional is fourth order. This is in agreement with the *a priori* error analysis in the Appendix, based on the nonlinear convergence theory of Keller [20] and Sanz-Serna [22, 28] and stability bounds of Kreiss [21], which proves that $u_h - u$, $v_h - v$ and their first derivatives are all $O(h^2)$ for the present finite volume scheme, and hence the error in the corrected functional is $O(h^4)$.

5.2 Isentropic transonic flow

The error convergence for a transonic flow in a converging-diverging duct is shown in Figure 7. The flow is subsonic upstream of the throat and supersonic downstream of the throat. Again the results show that the base error is second order while the remaining error after the adjoint correction is fourth order.

The accuracy of the corrected functional in this case is a little puzzling because the adjoint solution has a logarithmic singularity at the throat [14], as shown in Figure 8. Therefore, $v_h - v$ is $O(1)$ in a small region of size $O(h)$ on either side of the throat. Based on this, one might expect that the remaining error would be $O(h^3)$ since the numerical results confirm that the solution error $u_h - u$ and the consequent residual error for the nonlinear equations are both $O(h^2)$. The explanation for the fourth order convergence

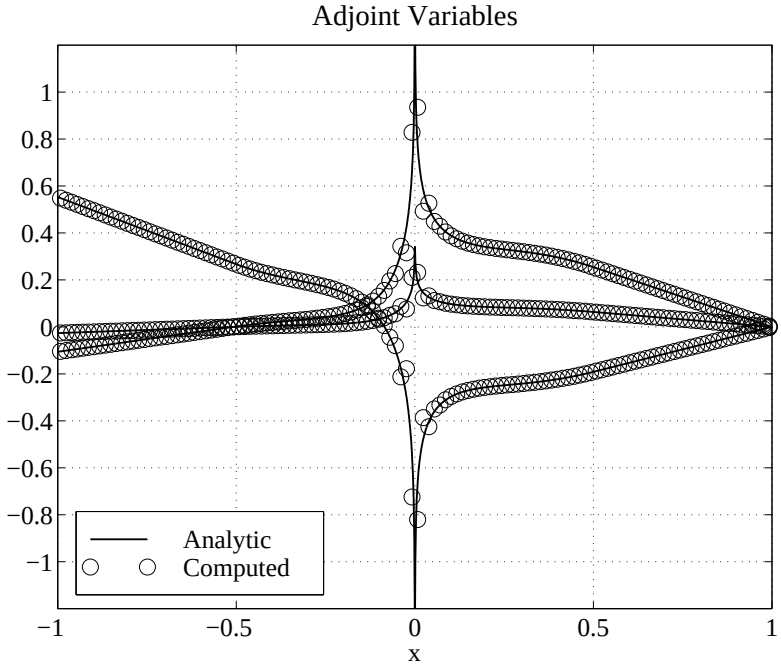

Figure 8: Adjoint solution for quasi-1D shock-free transonic flow.

must lie in a cancellation of the leading order terms within the remaining error, but the reason for this is not yet understood.

5.3 Shocked transonic flow

The final case is a shocked flow in a diverging duct where the flow is wholly supersonic upstream of the shock and subsonic downstream of it. At the shock, the analytic adjoint solution is continuous and has zero gradient [14], and so the adjoint variables pose no special difficulty in this case.

The challenge is the reconstruction of the approximate solution $u_h(x)$ from the cell-centered quantities produced by the finite volume calculation. The analytic solution is discontinuous at the shock and satisfies the Rankine-Hugoniot shock jump relations which require that there is no discontinuity in the nonlinear flux F. To recover a discontinuous approximate solution $u_h(x)$, we first interpolate the computed values of F which is known to be continuous across a shock. From these one can deduce the conservation variables U by solving a quadratic equation, one branch of which gives a subsonic flow solution, the other being supersonic. Therefore, given a shock position, one can reconstruct a supersonic solution on the upstream side, a subsonic solution on the downstream side, and automatically satisfy the Rankine-Hugoniot shock jump conditions at the shock itself.

To determine the shock position, we rely upon the fact that the integrated pressure along the duct is correct to second order when using a finite volume method which

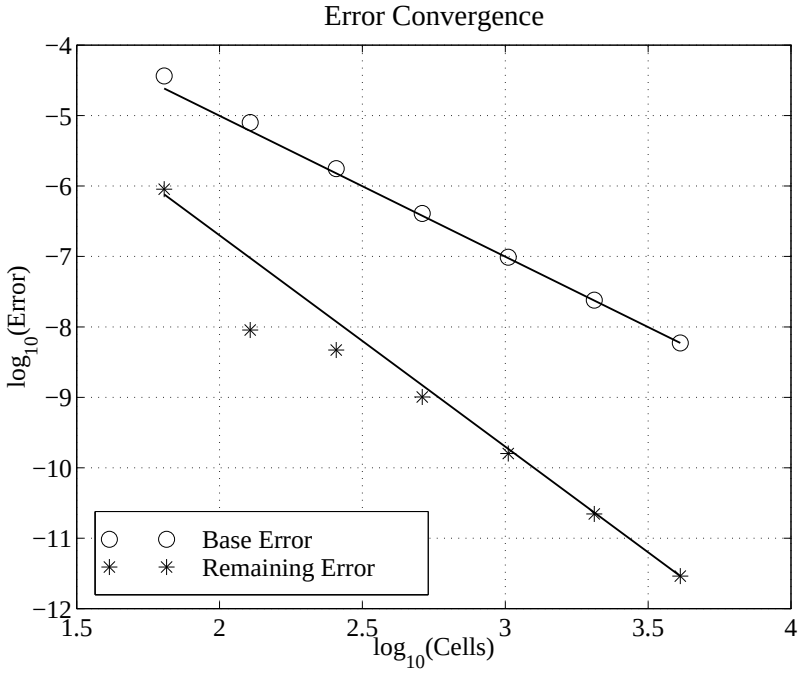

Figure 9: Error convergence for quasi-1D shocked flow.

is conservative and second order accurate in smooth flow regions [10]. Therefore, we iteratively adjust the position of the shock until the reconstructed solution has the same base functional value (i.e. without the adjoint correction) as the original numerical approximation, thereby obtaining the correct shock position to second order.

The form of the adjoint error correction term is exactly the same as before. This conclusion follows from a slight extension of the nonlinear formulation to take the shock into account as an internal boundary. The corresponding adjoint linearization includes perturbations to the shock position which lead to an internal boundary condition for the adjoint equations [14].

The baseline error is expected to remain second order for shocked flow, but in the neighborhood of the shock, there is an $O(h)$ error in $u_h(x)$, so the corrected error is expected to be third order rather than fourth. This behavior is confirmed by the error convergence results shown in Figure 9, where the superimposed lines have slopes of -2 and -3.

6 Conclusions and future challenges

This paper presents a method for estimating the value of an integral functional with twice the order of accuracy of the numerical p.d.e. solution on which the functional is based. The additional cost is the computation of an approximate solution to the associated adjoint problem. The formulation of the method for linear and nonlinear problems is relatively simple although some further complications arise when considering inhomo-

geneous boundary conditions and boundary integral terms in the functional. Given that many researchers are developing adjoint solvers because of their importance in optimal design, there is the potential for rapid exploitation of this error correction technique in a variety of engineering applications in which integral functionals are of interest.

In cases where the functional is a point quantity, the above theory could be applied using a distribution function for g. However, the loss of smoothness in g will often result in a poorer order of accuracy for the approximate adjoint solution v_h, leading to a consequent reduction in the order of accuracy of the corrected functional. To circumvent this difficulty, it may be possible to follow the approach of inner and outer matched asymptotic expansions, to combine an approximate analytic near-field solution with a computed far-field solution. Alternatively, one could use the technique of Babuška and Miller [3] to convert the point quantity into an equivalent integral representation.

A number of challenges will arise in applying the theory to more complex nonlinear problems in multiple dimensions. For curved boundaries, the computational domain covered by a grid is only an approximation to the true domain and so there may be complications in extending the numerical solution to cover the full domain. Likewise, there is the problem of ensuring that the approximate solution u_h exactly satisfies the boundary conditions imposed on the analytic solution u. Otherwise, an additional adjoint correction term associated with the residual error in the boundary conditions must be computed. In multiple dimensions, the functionals of interest are often boundary integrals rather than integrals over the whole domain. The formulation must then be modified by the introduction of inhomogeneous boundary conditions in the adjoint problem [13].

Possibly the biggest challenge in multiple dimensions will be the treatment of discontinuities and singularities. In one dimension, reconstructive shock-fitting is relatively straightforward, but in multiple dimensions it will likely prove infeasible, especially when there are shock junctions. A more practical approach may be to use local grid refinement at the shock to reduce its width to $O(h^2)$ where h is the average cell width in the rest of the grid. In this way, it may be possible to ensure that the error from the shock region is of the same order as the error from the rest of the domain. Similarly, there are singularities in the adjoint Euler equations in multiple dimensions [13] which will need to be well resolved to achieve the desired superconvergence in the corrected functional.

This leads to the whole topic of optimal grid adaptation [6, 25, 29, 11]. The magnitude of the adjoint error correction term $(v_h, f_h - f)$ is minimized by adapting the grid in the regions in which the product $v_h^T(f_h - f)$ is largest. Alternatively, if grid adaptation is to be used in conjunction with adjoint error correction then the remaining error is perhaps best minimized by adapting the grid where the residual errors $f_h - f$ and $g_h - g$ are largest.

We conclude this paper with an open question. As discussed earlier in presenting the linear approach, Galerkin finite element methods automatically provide superconvergent estimates of order h^{2p-m} for integral functionals with sufficiently smooth weighting functions. From this result it can be deduced that the solution error is $O(h^{2p-m})$ when measured in an appropriate negative Sobolev norm. The question is under what conditions it is possible to reconstruct from the finite element solution a smoother approximate solution u_h for which the residual error is also $O(h^{2p-m})$, leading to an adjoint error cor-

rection that produces functionals with accuracy $O(h^{4p-2m})$! The numerical results for the 2D Poisson equation confirm that it is possible in the case when $p=m=2$, but it is not clear to what extent this result can be generalized.

7 Acknowledgments

During this research, we have benefited from many discussions with our colleagues Dr. Endre Süli and Dr. Paul Houston who are also researching adjoint error analysis and optimal grid adaptation for finite element methods.

References

[1] J. Alden and R. Compton, *A general method for elctochemical simulations. Part 1: formulation of the strategy for two-dimensional simulations*, J. Phys. Chem. B, 101 (1997), pp. 8941–8954.

[2] W. Anderson and V. Venkatakrishnan, *Aerodynamic design optimization on unstructured grids with a continuous adjoint formulation.* AIAA Paper 97-0643, 1997.

[3] I. Babuška and A. Miller, *The post-processing approach in the finite element method – Part 1: calculation of displacements, stresses and other higher derivatives of the displacements*, Intern. J. Numer. Methods Engrg., 20 (1984), pp. 1085–1109.

[4] ——, *The post-processing approach in the finite element method – part 2: the calculation of stress intensity factors*, Intern. J. Numer. Methods Engrg., 20 (1984), pp. 1111–1129.

[5] J. Barrett and C. Elliott, *Total flux estimates for a finite-element approximation of elliptic equations*, IMA J. Numer. Anal., 7 (1987), pp. 129–148.

[6] R. Becker and R. Rannacher, *Weighted a posteriori error control in finite element methods*, tech. report, Universitat Heidelberg, 1996. Preprint No. 96-1.

[7] D. Colton and R. Kress, *Inverse acoustic and electromagnetic scattering theory*, Number 93 in Applied Mathematical Sciences, Springer-Verlag, 1991.

[8] J. Dieudonné, *Foundations of Modern Analysis*, Academic Press, 1969.

[9] J. Elliott and J. Peraire, *Practical 3D aerodynamic design and optimization using unstructured meshes*, AIAA J., 35 (1997), pp. 1479–1485.

[10] M. Giles, *Analysis of the accuracy of shock-capturing in the steady quasi-1D Euler equations*, Comput. Fluid Dynamics J., 5 (1996), pp. 247–258.

[11] M. Giles, *On adjoint equations for error analysis and optimal grid adaptation in CFD*, in Advances and Prospects in Computational Aerodynamics, M. Hafez, ed., World Scientific, 1998.

[12] M. Giles, M. Larson, J. Levenstam, and E. Süli, *Adaptive error control for finite element approximations of the lift and drag in viscous flow*, Tech. Report NA97/06, Oxford University Computing Laboratory, 1997.

[13] M. Giles and N. Pierce, *Adjoint equations in CFD: duality, boundary conditions and solution behaviour.* AIAA Paper 97-1850, 1997.

[14] M. Giles and N. Pierce, *On the properties of solutions of the adjoint Euler equations*, in Numerical Methods for Fluid Dynamics VI, M. Baines, ed., 1998.

[15] S. Godunov and V. Ryabenkii, *Special criteria of stability of boundary-value problems for non-self-adjoint difference equations*, Uspekhi Mat. Nauk, 18 (1963), p. 3.

[16] ——, *The Theory of Difference Schemes–An Introduction*, North Holland, Amsterdam, 1964.

[17] A. Jameson, *Aerodynamic design via control theory*, J. Sci. Comput., 3 (1988), pp. 233–260.

[18] ——, *Optimum aerodynamic design using control theory*, Comput. Fluid Dynam. Rev., (1995), pp. 495–528.

[19] A. Jameson, N. Pierce, and L. Martinelli, *Optimum aerodynamic design using the Navier–Stokes equations.* AIAA Paper 97-0101, 1997.

[20] H. Keller, *Approximation methods for nonlinear problems with application to two-point boundary value problems*, Math. Comp., 29 (1975), pp. 464–474.

[21] H.-O. Kreiss, *Difference approximations for boundary and eigenvalue problems for ordinary differential equations*, Mathematics of Computation, 26 (1972), p. 605.

[22] J. López-Marcos and J. Sanz-Serna, *Stability and convergence in numerical analysis iii: linear investigation of nonlinear stability*, IMA J. Numer. Anal., 8 (1988), pp. 71–84.

[23] P. Monk and E. Süli, *The adaptive computation of far field patterns by a posteriori error estimation of linear functionals*, Tech. Report NA98/02, Oxford University Computing Laboratory, 1998. To appear in SIAM J. Num. Anal.

[24] J. Moran, *An Introduction to Theoretical and Computational Aerodynamics*, John Wiley, 1984.

[25] M. Paraschivoiu, J. Peraire, and A. Patera, *A posteriori finite element bounds for linear-functional outputs of elliptic partial differential equations*, Comput. Methods Appl. Mech. Engrg., 150 (1997), pp. 289–312.

[26] J. Peraire and A. Patera, *Bounds for linear-functional outputs of coercive partial differential equations: local indicators and adaptive refinement*, in New Advances in Adaptive Computational Methods in Mechanics, P. Ladeveze and J. Oden, eds., Elsevier, 1997.

[27] J. Reuther, A. Jameson, J. Farmer, L. Martinelli, and D. Saunders, *Aerodynamic shape optimization of complex aircraft configurations via an adjoint formulation.* AIAA Paper 96-0094, 1996.

[28] J. Sanz-Serna, *Two topics in nonlinear stability*, in Advances in Numerical Analysis, vol. 1, Claredon Press, 1991, pp. 147–174.

[29] E. Süli, *A posteriori error analysis and adaptivity for finite element approximations of hyperbolic problems*, Tech. Report NA97/21, Oxford University Computing Laboratory, 1997.

Appendix A *A priori* error analysis

Before beginning the analysis, we begin with a few comments on notation. Bold type (e.g. $\boldsymbol{u}$) denotes a vector of discrete quantities at the nodes of a computational grid, and discrete operators acting on such data. Regular type is used for continuous functions and differential operators. $u(\boldsymbol{x}_h)$ denotes the discrete data obtained by evaluating the function $u(x)$ at the grid nodes whose coordinates are $\boldsymbol{x}_h$.

All norms, both discrete and continuous, are L_2 norms. In addition, the notation $O(h^p)$ when used in a context such as

$$\boldsymbol{u}_h = u(\boldsymbol{x}_h) + O(h^p),$$

means that there exists a constant c such that

$$\|\boldsymbol{u}_h - u(\boldsymbol{x}_h)\| \leq c\, h^p,$$

or, equivalently, $\boldsymbol{u}_h \in B(u(\boldsymbol{x}_h), ch^p)$, where the ball $B(\boldsymbol{u}, \epsilon)$ is defined as

$$B(\boldsymbol{u}, \epsilon) = \{\boldsymbol{w} : \|\boldsymbol{w} - \boldsymbol{u}\| \leq \epsilon\}\,.$$

A.1 Linear Poisson equations

In this section we analyse the accuracy of the approximate primal and adjoint solutions for the 1D and 2D Poisson equations, and derive an *a priori* error estimate proving that the error in the functionals after applying the adjoint error correction is fourth order.

In both problems there is a linear p.d.e,

$$Lu = f,$$

which is approximated on a uniform grid with spacing h by a finite difference equation,

$$\boldsymbol{L}_h \boldsymbol{u}_h = \mathbf{h}.$$

The purpose of the first part of the analysis is to bound the discrete solution error, $\|\boldsymbol{u}_h - u(\boldsymbol{x}_h)\|$, where $u(\boldsymbol{x}_h)$ denotes the discrete data obtained by evaluating the analytic solution $u(x)$ at the coordinates of the computational grid.

The first two lemmas establish key properties of both the 1D finite difference approximation and the 2D finite element discretisation.

Lemma 1 *For $f \in C^4(\Omega)$, there exists a function $\tau \in C^2(\Omega)$ and constant c_1, both independent of h, such that*

$$\boldsymbol{L}_h u(\boldsymbol{x}_h) - \mathbf{h} = \mathbf{h^2}\tau(\boldsymbol{x}_\mathbf{h}) + \boldsymbol{r}_\mathbf{h}^{(1)}, \qquad \|\boldsymbol{r}_\mathbf{h}^{(1)}\| \leq \mathbf{c_1 h^4},$$

and

$$\boldsymbol{L}_h \left(u(\boldsymbol{x}_h) - h^2 w(\boldsymbol{x}_h)\right) - \mathbf{h} = \boldsymbol{r}_\mathbf{h}^{(2)}, \qquad \|\boldsymbol{r}_\mathbf{h}^{(2)}\| \leq \mathbf{c_1 h^4},$$

where $w \in C^4(\Omega)$ is the solution to the p.d.e.

$$Lw = \tau,$$

subject to the given homogeneous boundary conditions.

Proof The function τ is easily found through a Taylor series expansion of the solution u about the central node in the discrete operator. The bounds on $\boldsymbol{r}_h^{(1)}$ and $\boldsymbol{r}_h^{(2)}$ are then found using appropriate truncated Taylor series expansions.

In the 1D case, this results in

$$\tau = -\frac{1}{12}\frac{d^2 f}{dx^2},$$

and

$$c_1 = \frac{7}{720}\max\left|\frac{d^4 f}{dx^4}\right|,$$

■

Lemma 2 *There exists a constant c_2, independent of h, such that*

$$\|\boldsymbol{L}_h^{-1}\| \leq c_2$$

Proof For uniform grids on a unit interval in 1D, or a unit square in 2D, the eigenfunctions of $\boldsymbol{L}_h$ are Fourier modes and so a precise bound is easily established.

In the 1D case, $h \leq \frac{1}{2}$ and hence

$$\|\boldsymbol{L}_h^{-1}\| \;=\; \frac{\left(\frac{h}{2}\right)^2}{\sin^2\left(\frac{\pi h}{2}\right)} \;\leq\; \tfrac{1}{8}.$$

■

From these two results we can prove the following lemma regarding the error in the numerical solution.

Lemma 3 *The discrete solution $\boldsymbol{u}_h$ can be written as*

$$\boldsymbol{u}_h = u(\boldsymbol{x}_h) - h^2 w(\boldsymbol{x}_h) + \boldsymbol{r}_h^{(3)},$$

where the function $w(x)$ is as defined in Lemma A.1 and the remainder term $\boldsymbol{r}_h^{(3)}$ is bounded by

$$\|\boldsymbol{r}_h^{(3)}\| \leq c_1 c_2 h^4,$$

with the constants c_1, c_2 as defined in Lemmas A.1 and A.2.

Proof Lemma A.1 gives

$$\boldsymbol{L}_h \boldsymbol{r}_h^{(3)} = -\boldsymbol{r}_h^{(2)},$$

and the result then follows immediately from the definitions of c_1 and c_2. ■

The final step is to consider the errors introduced by the cubic spline interpolation of the discrete solution $\boldsymbol{u}_h$ and the corresponding discrete adjoint solution $\boldsymbol{v}_h$, and thereby establish the order of accuracy of the corrected functional.

Lemma 4 *For the approximate solutions $u_h(x)$ and $v_h(x)$ obtained by cubic spline interpolation of $\boldsymbol{u}_h$ and $\boldsymbol{v}_h$, respectively, the solution errors $u_h - u$ and $v_h - v$ and the residual errors $f_h - f$ and $g_h - g$ are all $O(h^2)$ and*

$$(g_g - g, u_h - u) = O(h^4).$$

Proof Using standard results from the theory of cubic spline interpolation one can prove that

$$u_h(x) = u(x) - h^2 w(x) + O(h^4),$$

and

$$\nabla^2 u_h = \nabla^2 u - h^2 \nabla^2 w + O(h^2),$$

and so the solution error $u_h - u$ and the residual error $f_h - f$ are both second order. The same argument applies to the adjoint problem and the final result then follows from the Cauchy-Schwartz inequality. ■

As well as proving the fourth order accuracy of the corrected functional in these two particular cases, this proof serves as a template for proving superconvergence in other applications with linear p.d.e.'s. Proving a property corresponding to Lemma A.1 with the appropriate powers of h will usually be relatively easy; note that this will require f (and g in the adjoint problem) to satisfy certain smoothness constraints. Establishing a uniform bound on the inverse operator, as in Lemma A.2, will usually be a much harder task, similar to proving coercivity in finite element analyses. The final step of interpolation error analysis may also be troublesome in some cases; in the error analysis for the quasi-1D Euler equations using piecewise linear interpolation we will see the difficulties that can arise.

A.2 Nonlinear quasi-1D Euler equations

In this section we consider the subsonic test case and establish the fourth order accuracy of the corrected functional using both cubic spline and piecewise linear interpolation.

The analysis is split into three parts. In the first, the existence and uniqueness of second order accurate solutions to the nonlinear equations is proved under specified conditions, following the theoretical approach of Keller [20] and Sanz-Serna *et al* [22, 28]. The second part analyses the error in the adjoint solution, and then the final part considers the errors introduced by the interpolation and proves that each of the terms in the remaining error for the functional is fourth order in magnitude.

A.2.1 Nonlinear *a priori* error analysis

The nonlinear quasi-1D Euler equations,

$$N(u) = 0,$$

with appropriate boundary conditions, are approximated by the nonlinear finite difference equations

$$\boldsymbol{N}_h(\boldsymbol{u}_h) = 0.$$

We define the differential operator L_w to be the Fréchet derivative of N evaluated at w, and the discrete operator $\boldsymbol{L}_{\boldsymbol{w}}$ to be the Fréchet derivative of $\boldsymbol{N}_h$ evaluated at $\boldsymbol{w}$. We also, for convenience, use the shorthand $\boldsymbol{L}_u$ to represent $\boldsymbol{L}_{u(\boldsymbol{x}_h)}$.

We will assume that the nonlinear discretisation has the following properties:

Property 1: there exists a constant c_1, independent of h, such that

$$\|\boldsymbol{N}_h(u(\boldsymbol{x}_h))\| \leq c_1 h^2$$

Property 2: There exists a constant c_2, independent of h, such that

$$\left\|\boldsymbol{L}_u^{-1}\right\| \leq c_2.$$

Property 3: There exists a constant c_3, independent of h, such that

$$\|\boldsymbol{L}_{\boldsymbol{w}} - \boldsymbol{L}_u\| \leq \frac{1}{2c_2},$$

for any $\boldsymbol{w} \in B(u(\boldsymbol{x}_h), c_3 h)$.

If the duct area $A(x)$ is sufficiently smooth ($A \in C^2(\Omega)$), Properties 1 and 2 are easily proved for the finite volume scheme which was used to obtain the numerical results in this paper. Property 2 is much more difficult to establish, but it is thought it could be proved by following the approach of Kreiss [21]. The key steps in the proof would be to show that the p.d.e. is wellposed, the discretisation of the p.d.e. is consistent and strictly dissipative on the interior, and the discretisation of the boundary conditions is stable in the sense of Godunov and Ryabenkii [15, 16].

Given these three properties, the objective is to use the Fixed Point Theorem to prove that the nonlinear finite difference equations have a unique solution $\boldsymbol{u}_h$ which is second order accurate. A preliminary lemma is required to prepare for the main theorem.

Lemma 5 *If* $\boldsymbol{w}_h^{(1)}, \boldsymbol{w}_h^{(2)} \in B(u(\boldsymbol{x}_h), c_3 h)$, *then*

$$\left\|\boldsymbol{L}_u^{-1}\left(\boldsymbol{N}_h(\boldsymbol{w}_h^{(2)}) - \boldsymbol{N}_h(\boldsymbol{w}_h^{(1)}) - \boldsymbol{L}_u(\boldsymbol{w}_h^{(2)} - \boldsymbol{w}_h^{(1)})\right)\right\| \leq \tfrac{1}{2}\left\|\boldsymbol{w}_h^{(2)} - \boldsymbol{w}_h^{(1)}\right\|.$$

Proof

$$\frac{d}{d\theta}\left\{\boldsymbol{N}_h\left(\boldsymbol{w}_h^{(1)}+\theta(\boldsymbol{w}_h^{(2)}-\boldsymbol{w}_h^{(1)})\right)\right\}=\boldsymbol{L}_{\boldsymbol{w}_h^{(1)}+\theta(\boldsymbol{w}_h^{(2)}-\boldsymbol{w}_h^{(1)})}(\boldsymbol{w}_h^{(2)}-\boldsymbol{w}_h^{(1)}),$$

and so,

$$\begin{aligned}\boldsymbol{N}_h(\boldsymbol{w}_h^{(2)})-\boldsymbol{N}_h(\boldsymbol{w}_h^{(1)})-\boldsymbol{L}_u(\boldsymbol{w}_h^{(2)}-\boldsymbol{w}_h^{(1)}) \;&=\\ &\left\{\int_0^1\left(\boldsymbol{L}_{\boldsymbol{w}_h^{(1)}+\theta(\boldsymbol{w}_h^{(2)}-\boldsymbol{w}_h^{(1)})}-\boldsymbol{L}_u\right)d\theta\right\}(\boldsymbol{w}_h^{(2)}-\boldsymbol{w}_h^{(1)}).\end{aligned}$$

The desired result then follows from Properties 2 and 3 and the convexity of the ball $B(u(\boldsymbol{x}_h),c_3h)$. ■

Theorem 6 *There exist constants $c_4,h_0>0$, such that if $h<h_0$ then the nonlinear equations*

$$\boldsymbol{N}_h(\boldsymbol{u}_h)=0,$$

have a unique solution $\boldsymbol{u}_h\in B(u(\boldsymbol{x}_h),c_4h^2)$.

Proof Let $c_4=2c_1c_2$, $h_0=c_3/c_4$, and consider the mapping

$$\boldsymbol{T}:\boldsymbol{T}(\boldsymbol{u})=\boldsymbol{u}-\boldsymbol{L}_u^{-1}\boldsymbol{N}_h(\boldsymbol{u}).$$

If $h<h_0$ then $c_4h^2\le c_3h$ and so, from Lemma 5

$$\left\|\boldsymbol{T}(\boldsymbol{w}^{(2)})-\boldsymbol{T}(\boldsymbol{w}^{(1)})\right\|\;\le\;\tfrac{1}{2}\|\boldsymbol{w}^{(2)}-\boldsymbol{w}^{(2)}\|,$$

for any $\boldsymbol{w}^{(1)},\boldsymbol{w}^{(2)}\in B(u(\boldsymbol{x}_h),c_4h^2)$. Also, for any $\boldsymbol{w}\in B(u(\boldsymbol{x}_h),c_4h^2)$,

$$\begin{aligned}\|\boldsymbol{T}(\boldsymbol{w})-u(\boldsymbol{x}_h)\| &\le \|\boldsymbol{T}(\boldsymbol{w})-\boldsymbol{T}(u(\boldsymbol{x}_h))\|+\|\boldsymbol{T}(u(\boldsymbol{x}_h))-u(\boldsymbol{x}_h)\|\\ &\le \tfrac{1}{2}\|\boldsymbol{w}-u(\boldsymbol{x}_h)\|+\|\boldsymbol{L}_u^{-1}\boldsymbol{N}_h(u(\boldsymbol{x}_h))\|\\ &\le c_4h^2,\end{aligned}$$

using Properties 1 and 2.

Thus, $\boldsymbol{T}$ is a contraction mapping of $B(u(\boldsymbol{x}_h),c_4h^2)$ into itself, and so by the Fixed Point Theorem [8] there exists a unique fixed point $\boldsymbol{u}_h\in B(u(\boldsymbol{x}_h),c_4h^2)$ for which $\boldsymbol{T}(\boldsymbol{u}_h)=\boldsymbol{u}_h$ and hence $\boldsymbol{N}_h(\boldsymbol{u}_h)=0$. ■

A.2.2 Error analysis for adjoint solution

In this section we will assume throughout that $h\le h_0$ so that the nonlinear solution $\boldsymbol{u}_h$ is known to exist and satisfy the error bounds given in the last section.

Given an approximate solution u_h of the nonlinear p.d.e. through an interpolation of the discrete solution $\boldsymbol{u}_h$, our objective in this section is to analyse the difference $\boldsymbol{v}_h-v(\boldsymbol{x}_h)$. Here v is the solution of the differential equation

$$L_u^*v=g(u),$$

subject to homogeneous boundary conditions, where L_u and $g(u)$ are the Fréchet derivatives based on u, as defined in the main text. $\boldsymbol{v}_h$ is the solution of the corresponding linear finite difference equations

$$\boldsymbol{L}^*_{\boldsymbol{u}_h}\boldsymbol{v}_h = \boldsymbol{g}_h,$$

with $\boldsymbol{L}^*_{\boldsymbol{u}_h}$ and $\boldsymbol{g}_h$ both based on the discrete solution $\boldsymbol{u}_h$. The analysis will also involve the discrete operator $\boldsymbol{L}^*_u$, which again is a shorthand for $\boldsymbol{L}^*_{u(\boldsymbol{x}_h)}$

We will assume that the adjoint discretisation has the following three properties.

Property 1: There exists a function $\tau \in C^0(\overline{\Omega})$ such that

$$\boldsymbol{L}^*_{\boldsymbol{u}_h} v(\boldsymbol{x}_h) - \boldsymbol{g}_h = h^2\tau(\boldsymbol{x}_h) + O(h^3),$$

and

$$\boldsymbol{L}^*_{\boldsymbol{u}_h}\left(v(\boldsymbol{x}_h) - h^2 w(\boldsymbol{x}_h)\right) - \boldsymbol{g}_h = O(h^3),$$

where $w \in C^1(\overline{\Omega})$ is the solution to the linear p.d.e.

$$L^*_u w = \tau,$$

subject to homogeneous boundary conditions.

Property 2: There exists a uniform bound c_5, independent of h, such that

$$\|\boldsymbol{L}^{*\,-1}_u\| \leq c_5.$$

Property 3: There exists a constant c_6, independent of h, such that

$$\left\|\boldsymbol{L}^*_{\boldsymbol{u}_h} - \boldsymbol{L}^*_u\right\| \leq \frac{1}{2c_5},$$

when $\boldsymbol{u}_h \in B(u(\boldsymbol{x}_h), c_6 h)$.

For the finite volume method which was used to obtain the numerical results, it is fairly easy to derive the function τ in Property 1 if $A \in C^2(\overline{\Omega})$. Establishing the bounds on $\boldsymbol{r}_h^{(1)}$ and $\boldsymbol{r}_h^{(2)}$ is straightforward but tedious. Property 2 is again be the hardest to prove. Property 3 is easily established; its significance is that it is required for the following preparatory lemma.

Lemma 7 *There exists a constant $h_1 > 0$ such that, for $h < h_1$,*

$$\|\boldsymbol{L}^{*\,-1}_{\boldsymbol{u}_h}\| \leq 2c_5.$$

Proof Define $\boldsymbol{D} = \boldsymbol{L}^*_{\boldsymbol{u}_h} - \boldsymbol{L}^*_u$ and let $h_1 = \min\{h_0, c_6/c_4\}$ so that if $h < h_1$, then $\boldsymbol{u}_h \in B(u(\boldsymbol{x}_h), c_4h^2) \subset B(u(\boldsymbol{x}_h), c_6h)$. Hence, using Properties 2 and 3,

$$\|\boldsymbol{D}\| \leq \frac{1}{2c_5}, \qquad \|\boldsymbol{L}^{*\,-1}_u \boldsymbol{D}\| \leq \tfrac{1}{2}.$$

$\boldsymbol{I}+\boldsymbol{L}^{*\,-1}_u \boldsymbol{D}$ is therefore non-singular, and $\|(\boldsymbol{I}+\boldsymbol{L}^{*\,-1}_u \boldsymbol{D})^{-1}\| \leq 2$. It follows that $\boldsymbol{L}^*_{\boldsymbol{u}_h} = \boldsymbol{L}^*_u + \boldsymbol{D} = \boldsymbol{L}^*_u(\boldsymbol{I}+\boldsymbol{L}^{*\,-1}_u \boldsymbol{D})$ is non-singular, and

$$\|\boldsymbol{L}^{*\,-1}_{\boldsymbol{u}_h}\| \leq \|(\boldsymbol{I}+\boldsymbol{L}^{*\,-1}_u \boldsymbol{D})^{-1}\| \, \|\boldsymbol{L}^{*\,-1}_u\| \leq 2c_5.$$

■

Having finished these preliminaries, we come to the main result of this section.

Lemma 8 *Under the conditions of Lemma 7, and for* $h < h_1$,

$$\boldsymbol{v}_h = v(\boldsymbol{x}_h) - h^2 w(\boldsymbol{x}_h) + O(h^3).$$

Proof Follows immediately from Property 1 and Lemma 7. ■

A.2.3 Interpolation and functional errors

If one uses cubic spline interpolation to construct the approximate solutions u_h and v_h, then the analysis from the previous sections together with standard interpolation error analysis for cubic spline interpolation lead to error bounds of the following form.

$$\|u_h - u\| \leq d_1 h^2,$$

$$\|v_h - v\| \leq d_2 h^2,$$

$$\left\|\tfrac{dv_h}{dx} - \tfrac{dv}{dx}\right\| \leq d_3 h^2.$$

The error in the functional after the adjoint error correction can be split into three inner products: $(g(u) - \overline{g}(u, u_h), u_h - u)$, $(g_h - g(u), u_h - u)$, $\left((L^*_{u_h} - \overline{L}^*_{(u,u_h)})v_h, u_h - u\right)$. The Fréchet derivative of $g(u)$ is continuous and finite at all values of the analytic solution $u(x)$, and so there exists a constant d_4 such that

$$\|g(u) - \overline{g}(u, u_h)\| \leq d_4 \|u_h - u\|,$$

and hence

$$|(\overline{g}(u, u_h) - g(u), u - u_h)| \;\leq\; d_4 d_1^2 h^4.$$

Considering the second of the inner products,

$$\begin{aligned} g_h - g(u) &= L^*_{u_h} v_h - L^*_u v \\ &= (L^*_{u_h} - L^*_u)v_h + L^*_u(v_h - v), \end{aligned}$$

by bounding the differences in the coefficient matrices in $L^*_{u_h}$ and $\overline{L}^*_{(u,u_h)}$, which depend on u_h and u, and using the second order bounds on $u_h - u$, $v_h - v$ and $\frac{dv_h}{dx} - \frac{dv}{dx}$, one obtains a bound of the form

$$\|g_h - g(u)\| \leq d_5 h^2,$$

and hence

$$|(g_h - g(u), u_h - u)| \ \leq \ d_5 d_1 h^4.$$

Turning to the final inner product, by again bounding the differences in the coefficient matrices in $L^*_{u_h}$ and $\overline{L}^*_{(u,u_h)}$, one obtains a bound of the form

$$\left\| (L^*_{u_h} - \overline{L}^*_{(u,u_h)}) v_h \right\| \leq d_6 \|u_h - u\|,$$

and hence

$$\left| \left((L^*_{u_h} - \overline{L}^*_{(u,u_h)}) v_h, u_h - u \right) \right| \ \leq \ d_6 d_1^2 h^4.$$

This completes the *a priori* analysis proving the fourth order accuracy of the corrected functionals in the subsonic flow case when using cubic spline interpolation. When using piecewise linear interpolation, some of the above analysis has to be modified because there is now a first order error in $\frac{dv_h}{dx}$. Consequently, at first sight it appears that the bound on the second inner product will become third order rather than fourth. However, numerical results show that the second inner product term remains fourth order. To explain this behaviour requires careful attention to the nature of the error introduced by piecewise linear interpolation.

The starting point is the earlier result that

$$\boldsymbol{v}_h = v(\boldsymbol{x}_h) - h^2 w(\boldsymbol{x}_h) + O(h^3).$$

Defining I_h to be the operator performing piecewise linear interpolation through the nodal values of a continuous function, and defining I to be the identity operator, then

$$v_h = v - h^2 w + (I_h - I) v + O(h^3).$$

Next, we use standard results to express the interpolation error $(I_h - I)v$ as

$$(I_h - I) v = q(x) + O(h^3),$$

where $q(x)$ is a function which on the interval $[x_j, x_{j+1}]$ is

$$q(x) = \tfrac{1}{2} a_j (x - x_j)(x - x_{j+1}),$$

with a_j defined as

$$a_j = -\frac{1}{h} \left(\left. \frac{dv}{dx} \right|_{x_{j+1}} - \left. \frac{dv}{dx} \right|_{x_j} \right).$$

Hence,

$$\frac{dv_h}{dx} = \frac{dv}{dx} + l(x) + O(h^2),$$

where $l(x)$ on the open interval (x_j, x_{j+1}) is

$$l(x) = a_j(x - x_{j+1/2}),$$

so $l(x)$ is antisymmetric about $x_{j+1/2}$, the midpoint of the interval.

When this error representation is substituted into the second inner product error term, the component involving $l(x)$ is of the form

$$(Cl, u_h - u),$$

where $C(x)$ is a matrix function which has a bounded derivative. Therefore, $C(x)$ can be decomposed into a dominant part C_0 which is constant on each subinterval, plus a remainder which is $O(h)$.

Also, the interpolation error $u_h - u$ can be decomposed into a dominant part $r(x)$ which, like $q(x)$, is zero at nodes, piecewise quadratic and $O(h^2)$, plus a remainder which is $O(h^3)$.

The key observation is that the inner product involving all of the leading order terms, $(C_0 l, r)$, is zero because on each subinterval the product $(C_0 l)^T r$ is anti-symmetric about the midpoint of the interval. All of the other inner product contributions involve integrals of products which are $O(h^4)$. Therefore, by bounding all of these one arrives at the result that $|(g_h - g(u), u_h - u)|$ is still $O(h^4)$. This concludes the proof that the corrected functional value is fourth order accurate even when using linear interpolation.