Report no. 01/14

Analysis of Adjoint Error Correction for Superconvergent Functional Estimates

Michael B. Giles
Oxford University Computing Laboratory, Oxford, UK.
giles@comlab.ox.ac.uk

Niles A. Pierce
Applied & Computational Mathematics, California Institute of Technology,
Pasadena, CA 91125, USA.
niles@caltech.edu

Earlier work introduced the notion of adjoint error correction for obtaining superconvergent estimates of functional outputs from approximate PDE solutions. This idea is based on *a posteriori* error analysis suggesting that the leading order error term in the functional estimate can be removed by using an adjoint PDE solution to reveal the sensitivity of the functional to the residual error in the original PDE solution. The present work provides *a priori* error analysis that correctly predicts the behavior of the remaining leading order error term. Furthermore, the discussion is extended from the case of homogeneous boundary conditions and bulk functionals, to encompass the possibilities of inhomogeneous boundary conditions and boundary functionals. Numerical illustrations are provided for both linear and nonlinear problems.

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

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

This research was supported by EPSRC under grant GR/K91149, and by NASA/Ames Cooperative Agreement No. NCC 2-5431.

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

August, 2001

1 Introduction

Consider the drag on an aircraft at transonic cruise, the radar cross-section of a glider, the electrostatic free energy of a biomolecule in water, or the flux of fossil fuels through a porous medium. These and many other problems of engineering and scientific interest may be studied quantitatively by computing integral functionals of PDE solutions.

In attempting to improve the accuracy of functional estimates, the numerical practitioner is confronted with an array of possibilities. Reasonable alternatives include refining the computational mesh (mandating larger faster computers), increasing the order of accuracy of the discretization (when practical for the geometry under consideration), iterative refinement of the numerical solution via defect correction (again requiring a higher order discretization), or Richardson extrapolation of the solution or the functional (when the asymptotic convergence rate is reliably known). All of these approaches can be used to improve global solution accuracy, yielding corresponding increases in functional accuracy. However, for problems in which the value of a functional is the most interesting quantitative output of a simulation, there is significant motivation to devise reliable and efficient numerical techniques that specifically enhance the accuracy of the functional estimates without seeking to improve the accuracy of the underlying solution.

The purpose of this paper is to analyze a method for obtaining superconvergent functional estimates for arbitrary underlying numerical discretizations. The key is the use of the solution of the adjoint PDE (the dual of the linearized form of the original primal PDE) which reveals the influence of local solution errors on the functional value of interest. Recent work has demonstrated that the leading order error term in the functional can be estimated and removed by smoothly reconstructing the primal solution, differentiating to approximate the primal residual error, and evaluating the inner product with the reconstructed adjoint solution [24, 10]. Alternatively, the local contributions to this inner product can be employed as an optimal adaptivity criterion for improving functional accuracy [28, 29, 21].

The significance of the adjoint PDE for error analysis and adaptivity has long been realized within the finite element community [1, 2, 3, 4, 7, 15, 19, 22, 23, 27, 26], where it is well known that many finite element methods enjoy natural superconvergence for functional estimates. The adjoint error correction technique extends these results to approximate solutions obtained by any discretization method (or other means of approximation) as well as illustrating the potential for further improvement of the inherent finite element superconvergence.

Our earlier work in adjoint error correction emphasized the *a posteriori* error analysis necessary to motivate the approach for linear and nonlinear problems, with bulk functionals and homogeneous boundary conditions [24, 9]. The present work extends these results in two ways, firstly, by increasing the scope of the analysis to encompass boundary functionals and inhomogeneous boundary conditions, and secondly, by performing *a priori* error analysis to predict the rate of superconvergence of the remaining error terms after correction.

We begin the paper by formulating the approach for general linear PDEs and functionals. Numerical demonstrations are provided for a one-dimensional (1D) Poisson problem and a two-dimensional (2D) Laplace problem with curved boundaries and a geometric singularity in the domain. In the 1D setting, an *a priori* analysis is then performed to evaluate the primal, dual, and cubic spline reconstruction errors, correctly predicting a doubling in the convergence rate of the functional value over that of the baseline scheme.

Next, the approach is formulated for general nonlinear PDEs and functionals. Numerical demonstrations are provided for the system of quasi-1D Euler equations and for a 2D nonlinear thermal diffusion problem. An *a priori* analysis then follows for the 1D case, elucidating the relationships between errors in the nonlinear primal problem, the linear dual problem and the reconstruction errors in both. This analysis demonstrates that the same order doubling phenomenon is also predicted for the nonlinear case. For a first order differential operator in 1D, the somewhat surprising point is made that a delicate cancellation effect allows linear reconstruction to yield the same order of accuracy as the smoother cubic splines.

To demonstrate the generality of adjoint error correction, different numerical experiments are performed with finite difference, finite volume, and finite element methods. The latter examples serve to illustrate the improvement that is achievable over the natural finite element superconvergence.

2 Linear adjoint error correction

Let u be the solution of the linear differential equation

$$Lu = f,$$

in the domain Ω, subject to the linear boundary conditions

$$Bu = e,$$

on the boundary $\partial\Omega$. In general, the dimension of the operator B may be different on different sections of the boundary (e.g. inflow and outflow sections for the convection p.d.e.).

The output functional of interest is taken to be

$$J = (g, u) + (h, Cu)_{\partial\Omega},$$

where the notation $(.,.)$ denotes an integral inner product over the domain Ω, and $(.,.)_{\partial\Omega}$ represents an integral inner product over the boundary $\partial\Omega$. The boundary operator C may be algebraic (e.g. $Cu \equiv u$) or differential (e.g. $Cu \equiv \frac{\partial u}{\partial n}$), but must have the same dimension as the adjoint boundary condition operator B^* to be defined shortly. Note that either g or h may be set to zero if the functional does not have an interior or boundary integral contribution, respectively.

The corresponding linear adjoint problem is

$$L^* v = g,$$

in Ω, subject to the boundary conditions

$$B^* v = h,$$

on the boundary $\partial\Omega$. The fundamental identity defining L^*, B^* and the boundary operator C^* is

$$(L^* w, z) + (B^* w, Cz)_{\partial\Omega} = (w, Lz) + (C^* w, Bz)_{\partial\Omega},$$

for all sufficiently differentiable functions w, z. This identity is obtained by integration by parts, and in a previous paper we describe the construction of the appropriate adjoint operators for the linearized Euler and Navier-Stokes equations [8].

Using the adjoint identity, one immediately obtains the equivalent dual form of the output functional,

$$J = (v, f) + (C^* v, e)_{\partial\Omega}.$$

Suppose that u_h and v_h are approximations to u and v. The subscript h indicates that the approximate solutions are derived from a numerical computation 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 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.

Given approximate solutions u_h, v_h we define e_h, f_h, g_h, h_h by

$$Lu_h = f_h, \quad L^* v_h = g_h, \quad Bu_h = e_h, \quad B^* v_h = h_h.$$

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 e_h, f_h, g_h, h_h would be equal to e, f, g, h, respectively. Thus, the *residual errors* $e_h - e, f_h - f, g_h - g, h_h - h$ 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:

$$\begin{aligned}
(g, u) + (h, Cu)_{\partial\Omega} \;&=\; (g, u_h) + (h, Cu_h)_{\partial\Omega} \\
&\quad -(g_h, u_h - u) - (h_h, C(u_h - u))_{\partial\Omega} \\
&\quad +(g_h - g, u_h - u) + (h_h - h, C(u_h - u))_{\partial\Omega} \\
&=\; (g, u_h) + (h, Cu_h)_{\partial\Omega} \\
&\quad -(L^* v_h, u_h - u) - (B^* v_h, C(u_h - u))_{\partial\Omega} \\
&\quad +(g_h - g, u_h - u) + (h_h - h, C(u_h - u))_{\partial\Omega} \\
&=\; (g, u_h) + (h, Cu_h)_{\partial\Omega} \\
&\quad -(v_h, L(u_h - u)) - (C^* v_h, B(u_h - u))_{\partial\Omega}
\end{aligned}$$

$$
\begin{aligned}
& +(g_h - g, u_h - u) + (h_h - h, C(u_h - u))_{\partial\Omega} \\
= \; & (g, u_h) + (h, Cu_h)_{\partial\Omega} \\
& -(v_h, f_h - f) - (C^* v_h, e_h - e)_{\partial\Omega} \\
& +(g_h - g, u_h - u) + (h_h - h, C(u_h - u))_{\partial\Omega}.
\end{aligned}
$$

In the final result, the first line is the functional based on the approximate solution u_h. The second line contains two computable terms. The first 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. The second term performs a similar task for the error $e_h - e$ in satisfying the boundary conditions. Together, these two terms form what we label the *adjoint error correction*, giving the leading order effect of the residual errors on the functional of interest. Adding this correction to the quantity in the first line gives a more accurate approximation to the value of the functional.

The third line is the remaining error after making the adjoint correction. Each of the terms is an inner product of two quantities, the first being a function of $v_h - v$ and the second being a function of $u_h - u$. If each quantity is $O(h^p)$ so that halving the grid spacing results in a 2^p reduction, then the remaining error for the functional is $O(h^{2p})$. Furthermore, if we assume that the approximate solutions u_h, v_h exactly satisfy the boundary conditions, so that $e_h - e = h_h - h = 0$, then the second term is zero. The first term can be expressed as $(g_h - g, L^{-1}(f_h - f))$, where the operator L^{-1} is defined subject to homogeneous boundary conditions. There is therefore a computable *a posteriori* error bound $\|L^{-1}\| \, \|f_h - f\| \, \|g_h - g\|$.

In Galerkin finite element methods (or any finite element method in which the test and trial spaces are interchanged for the primal and dual problems) the adjoint correction term is always zero due to the inherent orthogonality of the discretization. This desirable property yields automatic superconvergence for all integral functionals. However, if p is the order of accuracy of the finite element solution u_h, and the operator L involves derivatives of up to degree m, then usually the residual error satisfies $f_h - f = O(h^{p-m})$ and hence the error in any smoothly weighted functional is $O(h^{2p-m})$. Replacing the finite element solution with a smoother reconstructed solution allows adjoint error correction to recover an improved functional estimate with an error that is $O(h^{2p})$. This procedure will be demonstrated in the second of the two linear examples to follow. In this, in which a piecewise linear finite element solution of the Poisson equation (for which the functional accuracy is $O(h^2)$) is replaced by a cubic spline interpolation leading to an improved accuracy of $O(h^4)$.

3 Linear examples

3.1 1D Poisson equation with finite differences

The first example is the one-dimensional equation,

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

on the unit interval $[0,1]$ subject to homogeneous boundary conditions $u(0)=u(1)=0$. This example has appeared previously [24], and is included here to serve as the basis for a detailed *a priori* error analysis. The problem 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 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 that are used in the numerical integration of the inner product $(v_h, f_h - f)$. Figure 2 is 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 $(v_h, f_h - f)$; 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.

3.2 2D Laplace equation with finite elements

The second example is a much more testing 2D problem, with a boundary functional over a curved boundary with a cusp at one point. The domain and functional mimic the challenges that arise in computational fluid dynamics in considering the flow around two-dimensional airfoils.

The test case is constructed with the aid of a conformal mapping. Starting with the region in the complex Z-plane between two circles centered at $(X,Y) = (-0.1, 0)$ with radii of $R_1 = 1.1$ and $R_2 = 3.0$, application of the Joukowski mapping

$$z = Z + \frac{1}{Z},$$

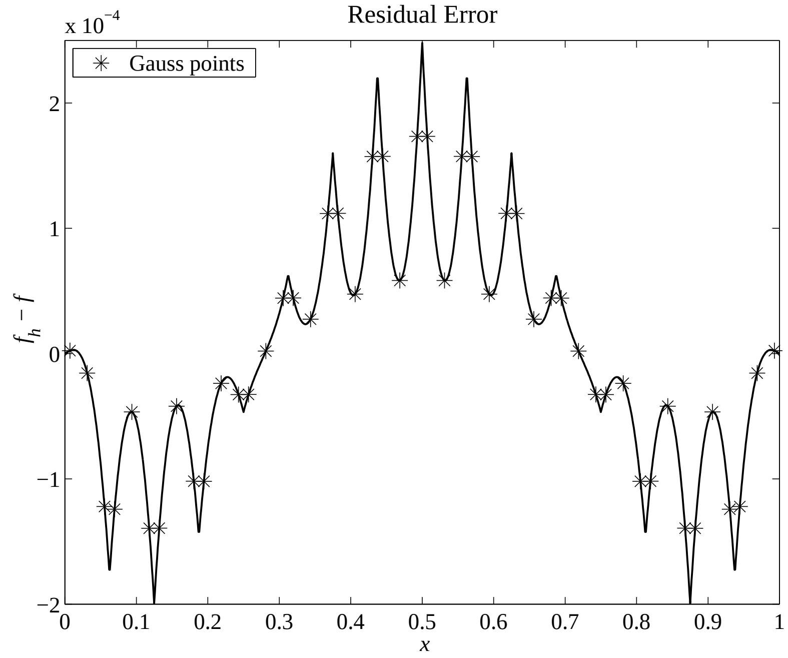

Figure 1: Residual error for a 1D Poisson problem.

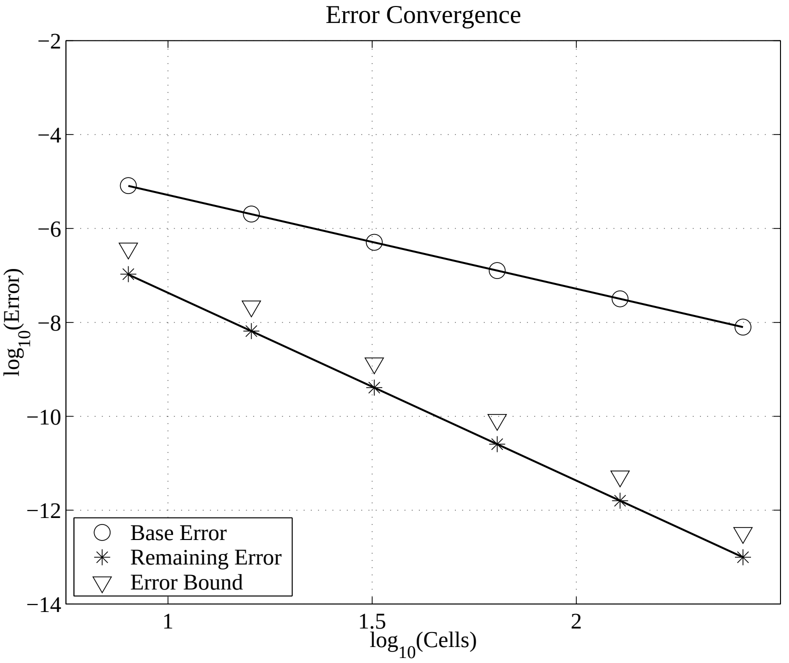

Figure 2: Functional error convergence for a 1D Poisson problem.

then produces a computational domain Ω lying between a cusped airfoil ($\partial\Omega_{z1}$) and a smooth outer boundary ($\partial\Omega_{z2}$).

Using cylindrical coordinates R, θ defined by

$$X + 0.1 = R\cos\theta, \qquad Y = R\sin\theta,$$

the function

$$U(X,Y) = \frac{R^2 - R_1^2}{R}\sin\theta,$$

is a solution of the Laplace equation subject to the boundary conditions $U=0$ on the inner circle, and $U = [(R_2^2 - R_1^2)/R_2]\sin\theta$ on the outer cylinder.

In the z-plane, the function $u(x,y) = U(X,Y)$ is the solution of the Laplace equation

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0,$$

subject to $u=0$ on the airfoil, and the appropriate Dirichlet boundary conditions on the far-field boundary. As illustrated in Figure 3, this solution corresponds to the stream function for incompressible inviscid flow around the airfoil, with zero circulation.

The boundary functional in the Z-plane is defined to be

$$\int_0^{2\pi} \sin\theta \left.\frac{\partial U}{\partial n}\right|_{R=R_1} d\theta,$$

so the analytic value is -2π. When mapped into the z-plane, the corresponding expression for the functional is

$$\left(\frac{\sin\theta}{R_1}, \frac{\partial u}{\partial n}\right)_{\partial\Omega_{z1}},$$

and hence the dual problem is the Laplace equation subject to the inhomogeneous Dirichlet condition $v=\sin\theta/R_1$ on the airfoil surface and $v = 0$ on the far-field boundary.

The problem in the z-plane is solved numerically using a bi-linear Galerkin finite element method on a structured grid with quadrilateral elements. The approximate reconstructed solutions u_h and v_h are then obtained from the nodal values by bi-cubic spline interpolation. The computational coordinates at the grid points are also splined, so that x, y, u_h and v_h are all expressed parametrically as a function of spline coordinates ξ, η. By differentiating these functions one can then obtain the necessary derived quantities such as f_h. The inner products are evaluated using 3-point Gaussian quadrature for the boundary integrals and 3×3 Gaussian quadrature for interior integrals.

Figure 4 shows the numerical values obtained for the functional, with and without the adjoint error correction, plotted versus the square root of the number of cells, which is a measure of $1/h$, an average inverse grid spacing. Again the superimposed lines of slope -2 and -4 show that the base solution is second order accurate whereas the corrected value for the functional is fourth order accurate. Note that on a 128×32 mesh (the third data points), the increase in accuracy is greater than a factor of 6×10^4. This improvement is achieved despite the cusped trailing edge and the added complication of curved boundaries. The computable error bound does not appear to be useful for this case due to a singularity in the adjoint residual near the trailing edge.

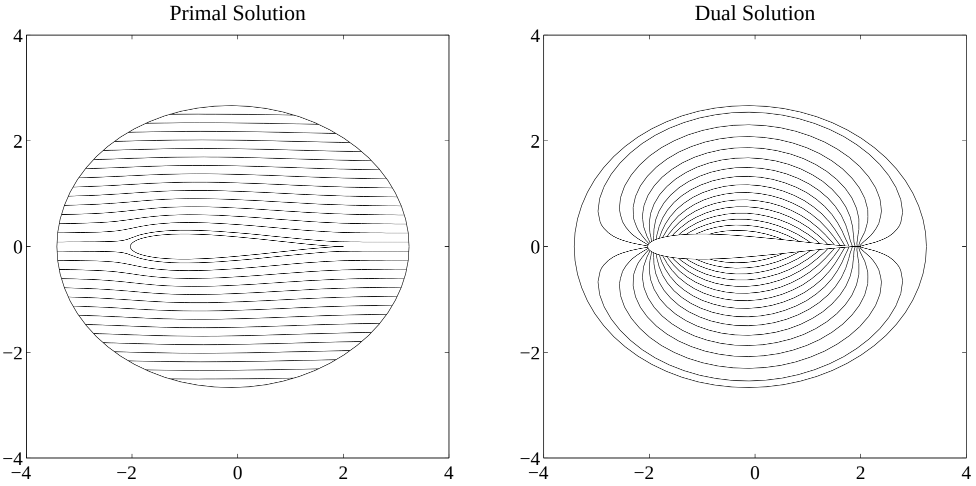

Figure 3: The primal and dual solutions for a 2D Laplace problem around a Joukowski airfoil with a cusped trailing edge.

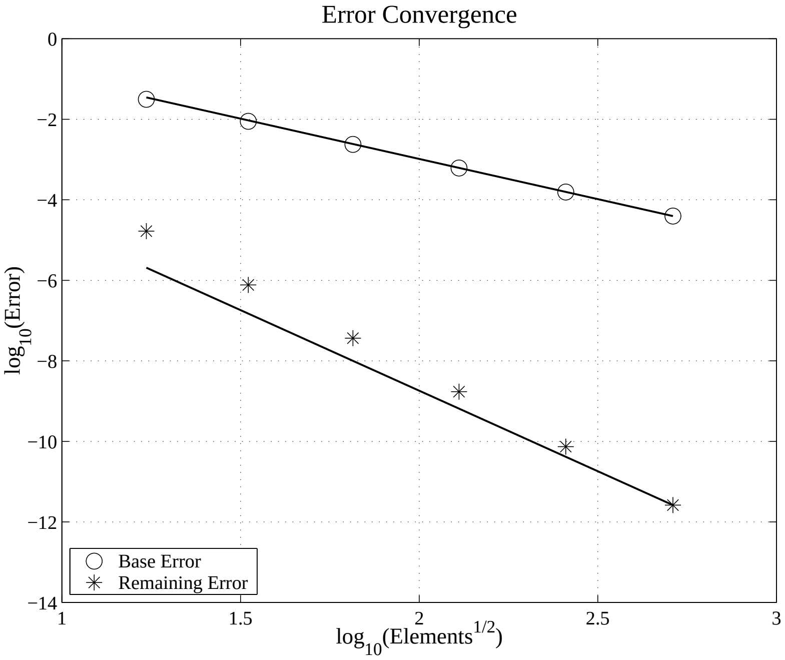

Figure 4: Error convergence of a boundary functional for a 2D Laplace problem around a Joukowski airfoil.

4 Linear *a priori* error analysis

In this section we analyze the accuracy of the approximate primal and adjoint solutions for the first of the linear examples, and derive an *a priori* error estimate proving that the error in the functional after applying the adjoint error correction is fourth order. The proof is intended to serve as a template for the *a priori* analysis of other applications, and so it is written in a more general form than is necessary for the particular problem being considered.

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_∞ 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

$$\|\boldsymbol{u}_h - u(\boldsymbol{x}_h)\| \le c\, h^p, \tag{4.1}$$

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

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

Turning now to the analysis of numerical example 3.1, the differential equation

$$Lu = f,$$

subject to homogeneous boundary conditions, is approximated on a uniform grid with spacing h by the finite difference equation,

$$\boldsymbol{L}_h \boldsymbol{u}_h = \boldsymbol{f}_h.$$

The purpose of the first part of the analysis is to bound the discrete solution error, $\|\boldsymbol{u}_h - u(\boldsymbol{x}_h)\|$.

The first two lemmas concern the accuracy and stability of the discretization.

Lemma 1 *For $f \in C^4[0,1]$, there exists a function $\tau \in C^2[0,1]$ and constant c_1, both independent of h, such that*

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

and

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

where $w \in C^4[0,1]$ is the solution of

$$Lw = \tau,$$

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.

This results in

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

and

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

■

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

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

Proof Standard convergence analysis for elliptic operators based on a maximum principle and a comparison function gives $c_2 = \frac{1}{8}$ (*e.g.* see page 165 in [20]). ■

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*

$$\boldsymbol{u}_h = u(\boldsymbol{x}_h) - h^2 w(\boldsymbol{x}_h) + \boldsymbol{r}_h^{(3)}, \tag{4.2}$$

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

$$\|\boldsymbol{r}_h^{(3)}\| \leq c_1 c_2 h^4, \tag{4.3}$$

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

Proof Lemma 1 gives

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

and the result then follows from the bounds in Lemmas 1 and 2. ■

The second part of the analysis considers the errors introduced by the cubic spline interpolation of the discrete solution $\boldsymbol{u}_h$ and the corresponding discrete adjoint solution $\boldsymbol{v}_h$. A cubic spline interpolates the data with a C^2 piecewise cubic polynomial. One boundary condition is required at each end, and here we discuss the case of complete splines for which the second derivative is specified.

As with the discretization of the differential equation, we need results concerning the accuracy and stability of cubic spline reconstruction.

Lemma 4 *For a given function $u(x) \in C^4[0,1]$, the cubic spline defined by the knot conditions $s(\boldsymbol{x}_j) = u(\boldsymbol{x}_j)$ and the end conditions $s''(0) = u''(0)$, $s''(1) = u''(1)$ satisfies the bounds*

$$\begin{aligned} \|s-u\| &\leq \tfrac{5}{384}h^4 \,\|u''''\|\,, \\ \|s''-u''\| &\leq \tfrac{1}{2}h^2 \,\|u''''\|\,. \end{aligned}$$

(Proof: see page 68 in [5]).

Lemma 5 *The cubic spline $s(x)$ defined by the knot conditions $s(\boldsymbol{x}_j) = h^4\boldsymbol{u}_j$ and the end conditions $s''(0) = h^2U_0$, $s''(1) = h^2U_1$, satisfies the following bounds:*

$$\begin{aligned} \|s\| &\leq h^4 \max\left(\tfrac{5}{2}\,\|\boldsymbol{u}\|\,,\ \tfrac{5}{24}|U_0|,\ \tfrac{5}{24}|U_1|\right), \\ \|s''\| &\leq h^2 \max\left(12\,\|\boldsymbol{u}\|\,,\ |U_0|,\ |U_1|\right). \end{aligned}$$

Proof In the interior of the domain, the equation used to determine the second derivatives of the spline at the mesh points $s''(\boldsymbol{x}_j)$ is

$$s''(\boldsymbol{x}_j) = \tfrac{3}{2}h^2\left(\boldsymbol{u}_{j+1} - 2\boldsymbol{u}_j + \boldsymbol{u}_{j-1})\right) - \tfrac{1}{4}(s''(\boldsymbol{x}_{j+1}) + s''(\boldsymbol{x}_{j-1})),$$

so therefore

$$|s''(\boldsymbol{x}_j)| \leq 6h^2\|\boldsymbol{u}\| + \tfrac{1}{2}\|s''\|.$$

Including the end conditions, and the fact that within each mesh interval the second derivative is a linear interpolation of the two values at either end, gives

$$\|s''\| \leq \max\left(6h^2\|\boldsymbol{u}\| + \tfrac{1}{2}\|s''\|,\ h^2|U_0|,\ h^2|U_1|\right)$$

and hence

$$\|s''\| \leq h^2 \max\left(12\|\boldsymbol{u}\|,\ |U_0|,\ |U_1|\right).$$

The bound for $\|s\|$ follows from the reconstruction formula within each interval.

$$\|s\| \leq h^4\|\boldsymbol{u}\| + \tfrac{1}{8}h^2\|s''\| \leq h^4 \max\left(\tfrac{5}{2}\|\boldsymbol{u}\|,\ \tfrac{5}{24}|U_0|,\ \tfrac{5}{24}|U_1|\right).$$

■

Lemma 6 *If $f, g \in C^4[0,1]$, then the approximate solutions $u_h(x)$ and $v_h(x)$ obtained by cubic spline interpolation of $\boldsymbol{u}_h$ and $\boldsymbol{v}_h$, with end conditions $u_h''(0) = f(0), u_h''(1) = f(1), v_h''(0) = g(0), v_h''(1) = g(1)$, are second order approximations to $u(x)$ and $v(x)$, respectively. Furthermore, the residual errors $f_h - f$ and $g_h - g$ are both $O(h^2)$ and*

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

Proof The cubic spline reconstruction $u_h(x)$ can be written as the sum of 3 parts:
i) $s_1(x)$ defined by $s_1(\boldsymbol{x}_j) = u(\boldsymbol{x}_j)$ with end conditions $s_1''(0) = f(0), s_1''(1) = f(1)$;
ii) $h^2 s_2(x)$ where $s_2(\boldsymbol{x}_j) = -w(\boldsymbol{x}_j)$ and $s_2''(0) = -w''(0), s_2''(1) = -w''(1)$;
iii) $s_3(x)$ satisfying $s_3(\boldsymbol{x}_j) = \boldsymbol{r}_j^{(3)}$ and $s_3''(0) = h^2 w''(0), s_3''(1) = h^2 w''(1)$.

If $f \in C^4[0,1]$, then $u \in C^6[0,1]$ and $w \in C^4[0,1]$. Hence, using the triangle inequality and applying Lemma 4 to i) and ii), and Lemma 5 to iii), gives

$$\begin{aligned} \|u_h - u\| &= \|s_1 + h^2 s_2 + s_3 - u\| \\ &\le \|s_1 - u\| + h^2 \|s_2 + w\| + h^2 \|w\| + \|s_3\| \\ &\le \tfrac{5}{384} h^4 \left(\|u''''\| + h^2 \|w''''\| \right) + h^2 \|w\| \\ &\quad + \max \left(\tfrac{5}{2} \|\boldsymbol{r}_h^{(3)}\|, \tfrac{5}{24} h^4 |w''(0)|, \tfrac{5}{24} h^4 |w''(1)| \right), \end{aligned}$$

and likewise

$$\begin{aligned} \|u_h'' - u''\| &\le \tfrac{1}{2} h^2 \left(\|u''''\| + h^2 \|w''''\| \right) + h^2 \|w''\| \\ &\quad + \max \left(12 h^{-2} \|\boldsymbol{r}_h^{(3)}\|, h^2 |w''(0)|, h^2 |w''(1)| \right). \end{aligned}$$

Introducing the bounds on $\|\boldsymbol{r}_h^{(3)}\|$ from Lemma 3 gives the conclusion that $u_h - u$ and $f_h - f$ are both $O(h^2)$. The same argument applies to the adjoint solution, and the final result that $(g_g - g, u_h - u) = O(h^4)$ follows immediately. ■

As well as proving the fourth order accuracy of the corrected functional in this particular case, this proof provides guidelines for proving superconvergence in other applications with linear p.d.e.'s. Proving a property corresponding to Lemma 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 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 *a priori* error analysis for the quasi-1D Euler equations using piecewise linear interpolation, presented later in this paper, we will see the difficulties that can arise.

5 Nonlinear adjoint error correction

Let u be the solution of the nonlinear differential equation

$$N(u) = 0,$$

in the domain Ω, subject to the nonlinear boundary conditions

$$D(u) = 0,$$

on the boundary $\partial\Omega$.

The linear differential operators L_u and B_u are defined to be the Fréchet derivatives of N and D, respectively,

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

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

It is assumed that the nonlinear functional of interest, $J(u)$, has a Fréchet derivative of the following form,

$$\lim_{\epsilon\to 0}\frac{J(u+\epsilon\tilde{u})-J(u)}{\epsilon} = (g(u),\tilde{u}) + (h, C_u\tilde{u})_{\partial\Omega}.$$

Here the dimension of the operator C_u (which may be differential) is required to equal the dimension of the adjoint boundary operator B_u^*, to be defined shortly.

The corresponding linear adjoint problem is

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

in Ω, subject to the boundary conditions

$$B_u^* v = h$$

on the boundary $\partial\Omega$. The identity defining L_u^*, B_u^* and the boundary operator C_u^* is

$$(L_u^* w, \tilde{u}) + (B_u^* w, C_u\tilde{u})_{\partial\Omega} = (w, L_u\tilde{u}) + (C_u^* w, B_u\tilde{u})_{\partial\Omega},$$

for all $\tilde{u}, w$.

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

$$L_{u_h}^* v_h = g_h, \qquad B_{u_h}^* v_h = h_h.$$

Note the use of the Fréchet derivatives based on u_h which is known, instead of u which is not.

The analysis also requires averaged Fréchet derivatives defined by

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

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

$$\begin{aligned} D(u_h)-D(u) &= \overline{B}_{(u,u_h)}\,(u_h-u), \\ J(u_h)-J(u) &= (\overline{g}(u,u_h), u_h-u) + (h, \overline{C}_{(u,u_h)}(u_h-u))_{\partial\Omega}. \end{aligned}$$

We now obtain the following:

$$\begin{aligned} J(u_h) - J(u) &= (\overline{g}(u,u_h), u_h-u) + (h, \overline{C}_{(u,u_h)}(u_h-u))_{\partial\Omega} \\ &= (g_h, u_h-u) + (h_h, C_{u_h}(u_h-u))_{\partial\Omega} \\ &\quad -(g_h-\overline{g}(u,u_h), u_h-u) - (h, (C_{u_h}-\overline{C}_{(u,u_h)})(u_h-u))_{\partial\Omega} \\ &\quad -(h_h-h, C_{u_h}(u_h-u))_{\partial\Omega} \\ &= (L^*_{u_h}v_h, u_h-u) + (B^*_{u_h}v_h, C_{u_h}(u_h-u))_{\partial\Omega} \\ &\quad -(g_h-\overline{g}(u,u_h), u_h-u) - (h, (C_{u_h}-\overline{C}_{(u,u_h)})(u_h-u))_{\partial\Omega} \\ &\quad -(h_h-h, C_{u_h}(u_h-u))_{\partial\Omega} \\ &= (v_h, L_{u_h}(u_h-u)) + (C^*_{u_h}v_h, B_{u_h}(u_h-u))_{\partial\Omega} \\ &\quad -(g_h-\overline{g}(u,u_h), u_h-u) - (h, (C_{u_h}-\overline{C}_{(u,u_h)})(u_h-u))_{\partial\Omega} \\ &\quad -(h_h-h, C_{u_h}(u_h-u))_{\partial\Omega} \\ &= (v_h, \overline{L}_{(u,u_h)}(u_h-u)) + (C^*_{u_h}v_h, \overline{B}_{(u,u_h)}(u_h-u))_{\partial\Omega} \\ &\quad -(g_h-\overline{g}(u,u_h), u_h-u) - (h, (C_{u_h}-\overline{C}_{(u,u_h)})(u_h-u))_{\partial\Omega} \\ &\quad -(h_h-h, C_{u_h}(u_h-u))_{\partial\Omega} + (v_h, (L_{u_h}-\overline{L}_{(u,u_h)})(u_h-u)) \\ &\quad +(C^*_{u_h}v_h, (B_{u_h}-\overline{B}_{(u,u_h)})(u_h-u))_{\partial\Omega} \\ &= (v_h, N(u_h)) + (C^*_{u_h}v_h, D(u_h))_{\partial\Omega} \\ &\quad -(g_h-\overline{g}(u,u_h), u_h-u) - (h, (C_{u_h}-\overline{C}_{(u,u_h)})(u_h-u))_{\partial\Omega} \\ &\quad -(h_h-h, C_{u_h}(u_h-u))_{\partial\Omega} + (v_h, (L_{u_h}-\overline{L}_{(u,u_h)})(u_h-u)) \\ &\quad +(C^*_{u_h}v_h, (B_{u_h}-\overline{B}_{(u,u_h)})(u_h-u))_{\partial\Omega}. \qquad (5.1) \end{aligned}$$

In the final result, the first line is the adjoint correction term taking into account residual errors in approximating both the p.d.e. and the boundary conditions. The other lines are the remaining errors, which include the consequences of nonlinearity in L, B, C and g as well as residual errors in approximating the adjoint problem.

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. If we again assume that the boundary integral terms are zero, or at least negligible, then the two interior inner products can be split into the following three pieces:

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

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

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

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

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

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

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

Together, these give the approximate asymptotic bound

$$|\ \text{Error}\ | < c_1 \|N(u_h)\|\, \|g_h - g(u_h)\| + c_2 \|N(u_h)\|^2,$$

where

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

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.

6 Nonlinear examples

6.1 Quasi-1D Euler equations

The steady quasi-1D Euler equations for the flow of an ideal compressible fluid in a variable area duct are

$$\frac{d}{dx}(AF) - \frac{dA}{dx}P = 0,$$

where $A(x)$ is the cross-sectional area of the duct and u, $F(u)$ and $P(u)$ 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.

The functional of interest is the 'lift'

$$J = \int p \; dx,$$

and the corresponding adjoint equations are

$$L_u^* v \;\equiv\; -A\left(\frac{\partial F}{\partial u}\right)^T v' - \frac{dA}{dx}\left(\frac{\partial P}{\partial u}\right)^T v \;=\; \left(\frac{\partial p}{\partial u}\right)^T.$$

The nonlinear equations are approximated using a standard second order finite volume method with characteristic smoothing on a uniform computational grid. The linear adjoint problem is approximated by the so-called 'continuous' method, which involves discretizing the analytic adjoint equations on the same uniform grid as the flow solution [14]. This approach produces consistent approximations to the analytic adjoint solution [9], which has been determined in closed form for the quasi-1D Euler equations [11].

Two different reconstruction methods have been investigated: cubic splines and linear interpolation. Each component of the primal solution u_h and the dual solution v_h is independently reconstructed from the nodal values. The integrals that form the base value for the functional and the adjoint correction are then approximated by 3-point Gaussian quadrature.

6.1.1 Subsonic flow

The first case is smooth subsonic flow in a converging-diverging duct corresponding to the Mach number distribution depicted in Figure 5. Figure 6 shows the error convergence for the computed functional. The superimposed lines of slope -2 and -4 show that the base error is second order using either linear interpolation or cubic splines for the reconstruction, and the error in the corrected functional is fourth order, again using either method of reconstruction.

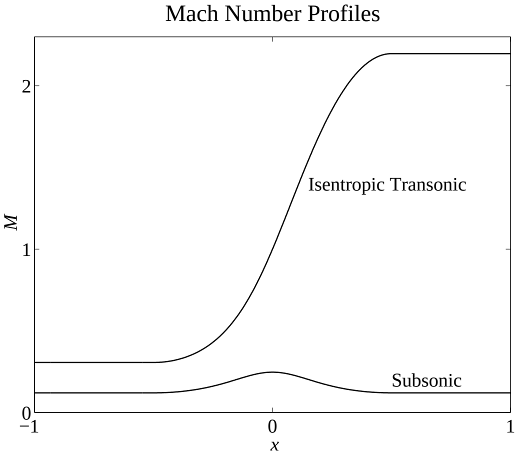

Figure 5: Mach number profiles for quasi-1D Euler equation test cases.

It is particularly noteworthy that the linear interpolation gives fourth order accuracy for the corrected functional. Linear interpolation gives a primal solution error that is second order, and an adjoint residual error that is first order, so one might expect that the error remaining after adjoint error correction would be third order. This point is analyzed later in Section 7, where it is proved that a cancellation effect zeroes this leading order term in the inner product of the primal residual error and the adjoint solution error.

6.1.2 Isentropic transonic flow

Figure 7 shows the error convergence for a transonic flow in a converging-diverging duct corresponding to the Mach number distribution of Figure 5. The flow is subsonic at the inflow boundary and upstream of the throat (located at $x = 0$), and supersonic downstream of the throat and at the outflow boundary. Again the results show that the base error is second order while the remaining error after the adjoint correction is fourth order, regardless of the choice of reconstruction method. This result is obtained despite the fact that there is a logarithmic singularity in the adjoint solution at the throat [11].

6.2 Nonlinear thermal diffusion

The computational domain is the circular annulus $1 \leq r \leq 3$ and the p.d.e. is the nonlinear diffusion equation

$$\nabla \cdot (u \nabla u) = 0,$$

subject to the requirement that u is positive. Dirichlet boundary conditions are specified at the inner and outer boundaries so as to agree with the analytic solution

$$u(r, \theta) = \left(1 + \left(\frac{r}{4} - \frac{1}{r}\right) \cos\theta\right)^{1/2}.$$

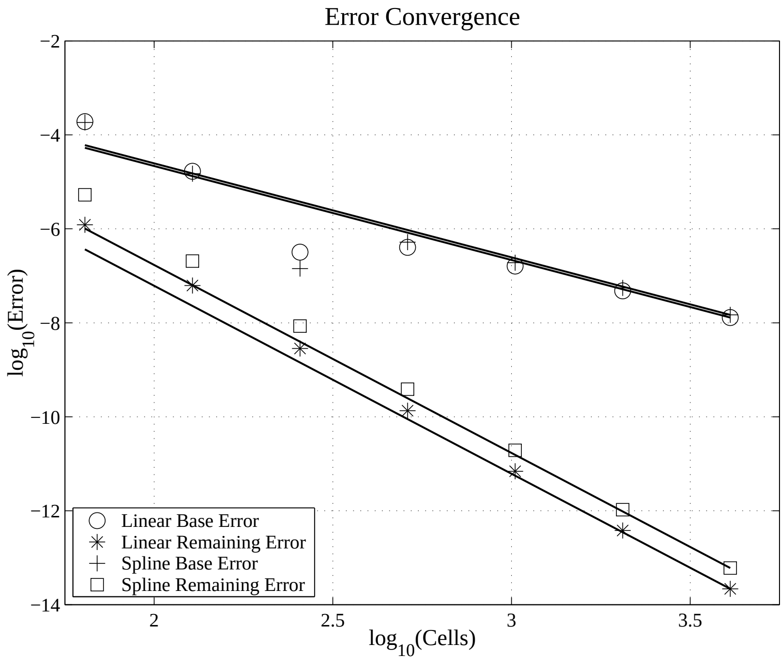

Figure 6: Error convergence for quasi-1D subsonic flow.

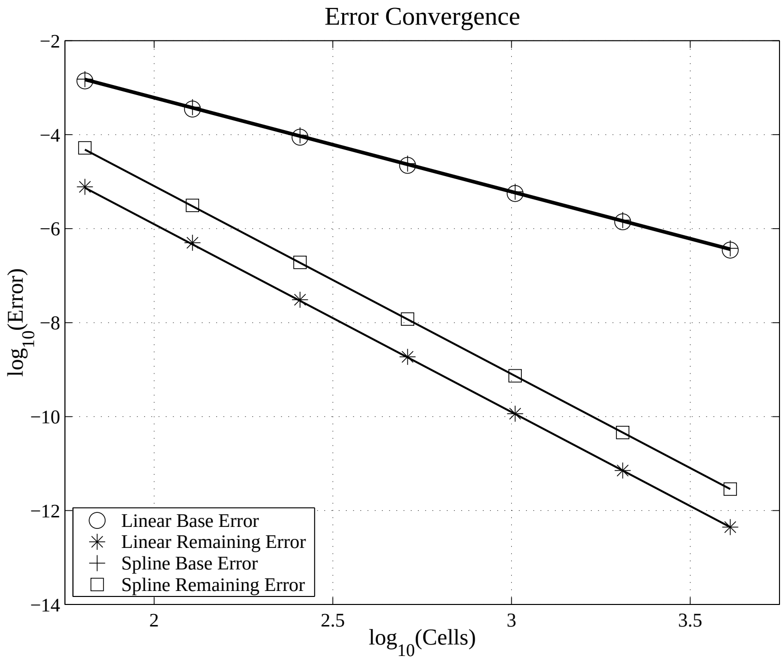

Figure 7: Error convergence for quasi-1D shock-free transonic flow.

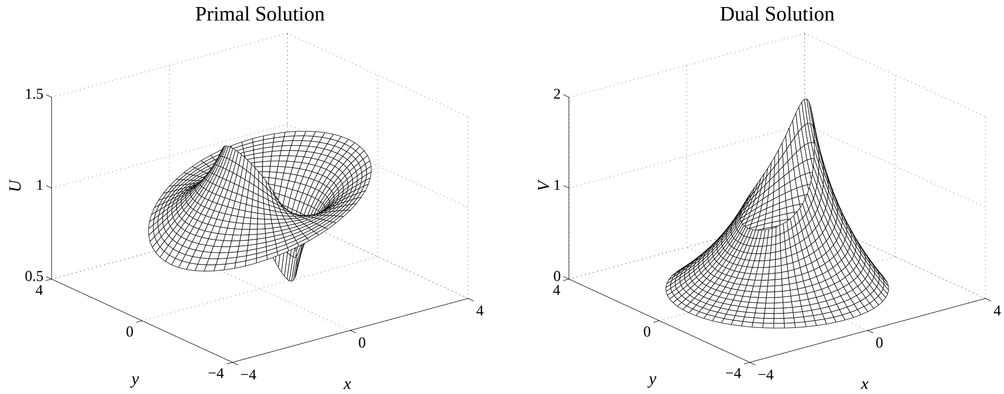

Figure 8: The primal and dual solutions for a 2D nonlinear thermal diffusion problem.

The functional of interest is

$$J(u) = \int_0^{2\pi} \left. \frac{\partial u}{\partial n} \right|_{r=1} d\theta,$$

and the corresponding dual problem is

$$L_u^* v \equiv u \nabla^2 v = 0,$$

with Dirichlet boundary conditions of $1/u$ and 0 on the inner and outer boundaries, respectively.

The primal and dual solutions shown in Figure 8 are obtained by a bi-linear Galerkin finite element formulation using 3×3 Gaussian quadrature to evaluate the mass and stiffness matrices. The nonlinear equations are solved using a full approximation scheme multigrid method. Bi-cubic spline interpolation and 3×3 Gaussian quadrature are then used to calculate the functional with and without the adjoint correction. The results are plotted in Figure 9 with superimposed lines of slopes -2 and -4, showing second order accuracy for the basic finite element solution and fourth order accuracy after the inclusion of the adjoint error correction. For a 128×32 mesh, the error decreases by more than a factor of 10^5.

7 Nonlinear *a priori* error analysis

In this section we consider the subsonic quasi-1D Euler test case and establish the fourth order accuracy of the corrected functional using both cubic spline and piecewise linear interpolation. For subsonic flow, the analytic primal and dual solutions are both known to be smooth for smooth duct geometries [11].

The discussion is split into three parts. The first examines the conditions required to ensure second order convergence of the nonlinear solution and the second analyzes the error in the adjoint solution. The objective of these first two parts is to describe a

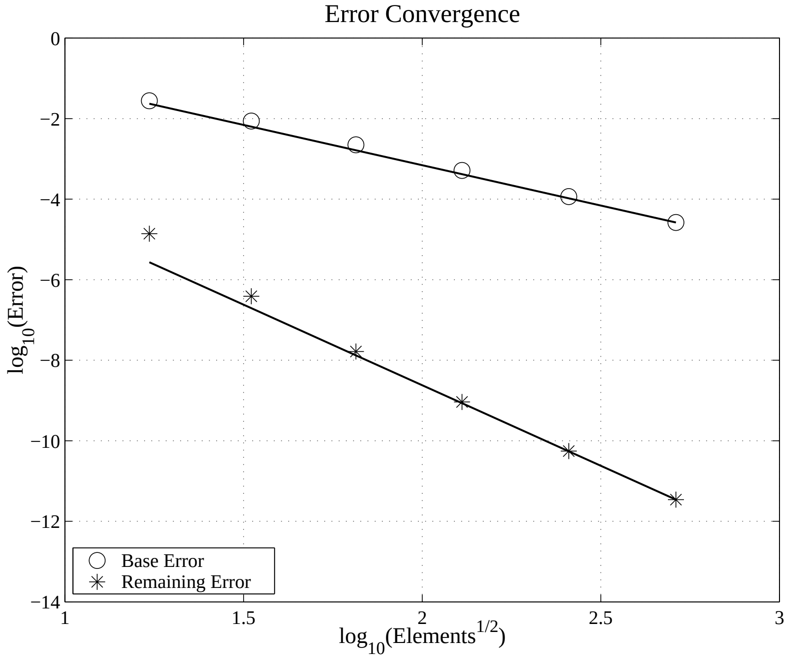

Figure 9: Error convergence of a boundary functional for a 2D nonlinear thermal diffusion problem.

minimal set of discretization properties that are sufficient to ensure the desired superconvergence behavior. The third and 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.

Analysis of the primal solution

Building on the ideas of Keller [16], Sanz-Serna and Lopez-Marcos [18, 25] have developed a powerful framework for analyzing discretizations of nonlinear PDEs. The thrust of their work is that, with appropriate definitions, consistency and stability are sufficient to ensure existence, local uniqueness and convergence. In particular, it is possible to identify the order of convergence of the global solution error, which is of paramount importance to the present discussion.

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 discretization 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

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

Property 3: There exists a constant c_3, independent of h, such that, for any $\boldsymbol{w} \in B(u(\boldsymbol{x}_h), c_3 h)$,

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

Property 1 is a local consistency condition on the nonlinear operator. Property 2 is equivalent to requiring stability of the linearized operator. These conditions mirror Lemmas 1 and 2 of the earlier linear analysis. Property 3 is a new smoothness condition on the nonlinear operator (e.g. Lipschitz or continuous Fréchet differentiability). Sanz-Serna and Lopez-Marcos have shown that these conditions are sufficient to guarantee that the numerical solution $\boldsymbol{u}_h$ lies in the ball $B(u(\boldsymbol{x}_h), c_4 h^2)$ for some positive c_4 independent of h [18].

If the duct area $A(x)$ is sufficiently smooth ($A \in C^2(\Omega)$), Properties 1 and 3 may be verified by Taylor series substitutions into the nonlinear and linearized discretization operators. Property 2 is more difficult to establish, but it appears that a proof could be constructed following the approach of Kreiss [17]. Essentially, the matrix $\boldsymbol{L}_u^{-1}$ is a discrete approximation to the Green's function for the continuous problem. The Green's function is bounded due to the well-posedness of the p.d.e., and a uniform bound for the difference between the Green's function and $\boldsymbol{L}_u^{-1}$ follows from using a discretization that is consistent and strictly dissipative on the interior, with discrete boundary conditions that are stable in the sense of Godunov and Ryabenkii [12, 13].

Analysis of the dual solution

In this section we will assume throughout that $h \leq h_0$ 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. (i.e. an interpolation of the discrete solution $\boldsymbol{u}_h$), our objective in this section is to analyze 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 previously. $\boldsymbol{v}_h$ 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 also involves the discrete operator $\boldsymbol{L}^*_u$, which again is a shorthand for $\boldsymbol{L}^*_{u(\boldsymbol{x}_h)}$

We will assume that the adjoint discretization has the following three properties:

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

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

and

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

where $w \in C^1(\Omega)$ 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

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

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

Conditions corresponding to Properties 1 and 2 were previously proved for the 1D linear analysis, and Property 3 was already encountered in the discussion of the primal discretization. For the present linear analysis, Properties 1 and 3 could be verified by Taylor series substitutions into the finite volume scheme used to obtain the numerical results (assuming $A \in C^2(\Omega)$). Property 2 would again be the hardest to prove. An estimate of the error in the adjoint solution now requires the following 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 provided $h < h_1$, then $\boldsymbol{u}_h \in B(u(\boldsymbol{x}_h), c_4 h^2) \subset B(u(\boldsymbol{x}_h), c_6 h)$. Hence, using Properties 2 and 3,

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

For any matrix A for which $\|A\| < 1$ we have $\|(I + A)^{-1}\| \leq \sum_{n=0}^{\infty} \|A\|^n = 1/(1 - \|A\|)$. Therefore,

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

From this 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

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

■

The main lemma then follows immediately from Property 1 and Lemma 7.

Lemma 8 *For $h < h_1$, the adjoint solution satisfies*

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

Analysis of reconstruction 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.

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

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

$$\|v_h' - v'\| \leq d_3 h^2.$$

From equation (5.1), the error in the functional after the adjoint error correction is the sum of five terms:

$$\begin{aligned}
&-(g_h - \overline{g}(u, u_h), u_h - u) - (h, (C_{u_h} - \overline{C}_{(u,u_h)})(u_h - u))_{\partial\Omega} \\
&-(h_h - h, C_{u_h}(u_h - u))_{\partial\Omega} + (v_h, (L_{u_h} - \overline{L}_{(u,u_h)})(u_h - u)) \\
&+(C^*_{u_h} v_h, (B_{u_h} - \overline{B}_{(u,u_h)})(u_h - u))_{\partial\Omega}.
\end{aligned}$$

We will now consider each of these in turn. Noting that $g_h = L^*_{u_h} v_h$ and $g(u) = L^*_u v$, the first term can be expressed as the sum of three other terms:

$$\begin{aligned}
&(g_h - \overline{g}(u, u_h), \; u_h - u) \\
&\quad = \; \underbrace{\left((L^*_{u_h} - L^*_u)\, v_h, \; u_h - u\right)}_{\text{(1a)}} + \underbrace{\left(L^*_u(v_h - v), \; u_h - u\right)}_{\text{(1b)}} + \underbrace{\left(g(u) - \overline{g}(u, u_h), \; u_h - u\right)}_{\text{(1c)}}.
\end{aligned}$$

By bounding the differences in the coefficient matrices in the operators $L^*_{u_h}$ and L^*_u, we can obtain a bound of the form

$$\left\|(L^*_{u_h} - L^*_u)\, v_h\right\| \leq d_4 \|u_h - u\|,$$

so therefore term (1a) has the fourth order bound

$$\left|\left((L^*_{u_h} - L^*_u)\, v_h, u_h - u\right)\right| \leq d_4\, d_1\, h^4.$$

For the term (1b), we use the second order bounds on $v_h - v$ and its derivative to obtain a bound of the form

$$\|L_u^*(v_h - v)\| \leq d_5\,(d_2 + d_3)\,h^2,$$

and hence

$$|(L_u^*(v_h - v), u_h - u)| \leq d_5\,(d_2 + d_3)\,d_1 h^4.$$

For (1c), the Fréchet derivative of $g(u)$ is continuous and bounded, so there exists another constant d_6 such that

$$\|g(u) - \overline{g}(u, u_h)\| \leq d_6 \|u_h - u\|,$$

and hence

$$|(g(u) - \overline{g}(u, u_h), u_h - u)| \ \leq \ d_6\, d_1^2\, h^4.$$

The second and third terms are both identically zero because the functional does not have any boundary terms and therefore $h = h_h = 0$. The boundary operators in the fifth term are all algebraic, not differential, and therefore it has a bound similar to that for (1c), involving the maximum errors at the two boundaries, which are not greater than the maximum errors over the whole interval.

The fourth term is the last to be considered. Integrating it by parts yields

$$((L_{u_h}^* - \overline{L}_{(u,u_h)}^*) v_h,\ u_h - u)$$

(which is fourth order by a similar argument to (1a)), plus some boundary terms that are fourth order by the same argument as for the fifth term.

Thus, the second and third terms are zero and the other three are all fourth order in magnitude, so the remaining error is $O(h^4)$. This completes the outline of an *a priori* analysis proving the fourth order accuracy of the corrected functional in the subsonic flow case using cubic spline interpolation.

When using piecewise linear interpolation, the error in v_h' becomes first order. In the analysis above, this error is important only in considering term (1b), where it initially appears that the degradation in the order of accuracy of v_h' will cause (1b) to become third order rather than fourth order. However, numerical experiments continue to exhibit fourth order functional convergence. An explanation for this behavior 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

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

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

$$a_j = -\frac{1}{h}\left(v'(x_{j+1}) - v'(x_j)\right).$$

Hence,

$$v_h' = v' + l(x) + O(h^2),$$

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

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

Note in particular that $l(x)$ is anti-symmetric about $x_{j+1/2}$, the midpoint of the interval.

When this error representation is substituted into the (1b) inner product error term, the component involving $l(x)$ is of the form

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

where $C(x)$ is a matrix function with bounded derivative. Therefore, $C(x)$ can be decomposed into a dominant part C_0 that is piecewise constant and a remainder that is $O(h)$. The primal interpolation error $u_h - u$ can also 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 critical observation is that the component of the inner product involving all of the leading order terms, $(C_0 l, r)$, is zero because in each sub-interval the product $(C_0 l)^T r$ is anti-symmetric about the midpoint. All of the other inner product contributions involving non-dominant terms are $O(h^4)$. Therefore, this cancellation effect is responsible for producing a functional error that remains $O(h^4)$ even when using linear interpolation.

8 Conclusions and future challenges

This work provides the first comprehensive treatment of adjoint error correction methods for bulk and boundary functional estimates based on linear and nonlinear PDE solutions with homogeneous and inhomogeneous boundary conditions. *A priori* error analysis of one linear and one nonlinear problem correctly predicts the observed superconvergence of the functional estimates. These discussions provide a framework for the analysis of functional estimates of other linear and nonlinear problems. Numerical demonstrations included a linear 1D bulk functional, a linear 2D boundary functional with a geometric singularity in the domain, a bulk functional of a quasi-1D nonlinear system, and a boundary functional of a 2D nonlinear problem.

In the linear case, *a posteriori* analysis leads to a computable bound on the error remaining after adjoint error correction. Further work is required to explore the sharpness and utility of these bounds and to develop reliable computable bounds for functionals of nonlinear problems.

The treatment of geometric singularities and solution discontinuities also requires further investigation. Mesh movement or adaptivity implemented in a carefully prescribed manner is required to yield fourth order functional estimates for primal solutions containing shocks. We are currently working on a theoretical treatment of this problem in 1D that appears to hold promise for multi-dimensional problems.

9 Acknowledgments

We wish to thank our colleagues Prof. Endre Süli and Dr. Paul Houston for many helpful discussions during our time shared in the Oxford University Computing Laboratory.

References

[1] 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.

[2] ——, *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.

[3] 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.

[4] R. Becker and R. Rannacher, *Weighted a posteriori error control in finite element methods*, in Proceedings of ENUMATH-97, *et al.* H.G. Block, ed., World Scientific Publishing, 1998, pp. 621–637.

[5] C. de Boor, *A Practical Guide to Splines*, Springer, 1978.

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

[7] M. Giles, M. Larson, M. Levenstam, and E. Süli, *Adaptive error control for finite element approximations of the lift and drag in a viscous flow*, Tech. Report NA 97/06, Oxford University Computing Laboratory, Wolfson Building, Parks Road, Oxford, OX1 3QD, 1997.

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

[9] ——, *On the properties of solutions of the adjoint Euler equations*, in Numerical Methods for Fluid Dynamics VI, M. Baines, ed., ICFD, Jun 1998.

[10] ——, *Improved lift and drag estimates using adjoint Euler equations.* AIAA Paper 99-3293, 1999.

[11] ——, *Analytic adjoint solutions for the quasi-one-dimensional Euler equations*, J. Fluid Mech., 426 (2001), pp. 327–345.

[12] 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.

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

[14] A. Jameson, *Optimum aerodynamic design using control theory*, in Computational Fluid Dynamics Review 1995, M. Hafez and K. Oshima, eds., John Wiley & Sons, 1995, pp. 495–528.

[15] C. Johnson, R. Rannacher, and M. Boman, *Numerics and hydrodynamic stability – toward error control in computational fluid dynamics*, SIAM J. Numer. Anal., 32 (1995), pp. 1058–1079.

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

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

[18] 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.

[19] P. Monk and E. Süli, *The adaptive computation of far field patterns by a posteriori error estimates of linear functionals*, SIAM J. Numer. Anal., 36 (1998), pp. 251–274.

[20] K. Morton and D. Mayers, *Numerical Solution of Partial Differential Equations – an Introduction*, Cambridge University Press, Cambridge, 1994.

[21] J.-D. Müller and M. Giles, *Solution adaptive mesh refinement using adjoint error analysis.* AIAA Paper 2001-2550, 2001.

[22] 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.

[23] 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.

[24] N. Pierce and M. Giles, *Adjoint recovery of superconvergent functionals from PDE approximations*, SIAM Rev., 42 (2000), pp. 247–264.

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

[26] E. Süli, *A posteriori error analysis and adaptivity for finite element approximations of hyperbolic problems*, in An Introduction to Recent Developments in Theory and Numerics for Conservation Laws, D. Kröner, M. Ohlberger, and C. Rohde, eds., vol. 5 of Lecture Notes in Computational Science and Engineering, Springer-Verlag, 1998, pp. 123–194.

[27] E. Süli and P. Houston, *Finite element methods for hyperbolic problems: a posteriori error analysis and adaptivity*, in The State of the Art in Numerical Analysis, I. Duff and G. Watson, eds., Clarendon Press, 1997, pp. 441–471.

[28] D. Venditti and D. Darmofal, *Adjoint error estimation and grid adaptation for functional outputs: application to quasi-one-dimensional flow*, J. Comput. Phys., 164 (2000), pp. 204–227.

[29] D. Venditti and D. Darmofal, *A grid adaptive methodolgy for functional outputs of compressible flow simulations.* AIAA Paper 01-2659, 2001.