Defect and Adjoint Error Correction

Michael B. Giles

Oxford University Computing Laboratory, Oxford OX1 3QD, U.K.

Abstract. Motivated by applications in aero-acoustics and electromagnetics, this paper discusses the combined use of defect correction to improve the order of accuracy of numerical solutions, and adjoint error correction to improve the order of accuracy of derived output functionals such as far-field boundary integrals. Numerical results for the 1D Helmholtz equation on an irregular grid show fourth order accuracy for the numerical solution, and sixth order accuracy for the boundary value.

1 Introduction

The primary motivation for the work in this paper is the need for high order accuracy for aeroacoustic and electromagnetics calculations. In steady CFD calculations, grid adaptation can be used to provide high grid resolution in the limited areas which require it. However, using standard second order accurate methods, the wave-like nature of aeroacoustic and electromagnetic solutions would lead to grid refinement throughout the computational domain in order to reduce the wave dispersion and dissipation to acceptable levels. The preferable alternative is to use higher order methods, allowing one to use fewer points per wavelength, which can lead to a very substantial reduction in the total number of grid points for 3D calculations. The difficulty with this is that one often wants to use unstructured grids because of their geometric flexibility, and the construction of higher order approximations on unstructured grids is complicated and computationally expensive.

The current research also followed from previous research by Pierce and Giles on the use of adjoint error correction to obtain improved values for output functionals [6]. The relevance of this to aero-acoustic and electromagnetics is that one is often interested in the value of a far-field boundary integral giving the radiated acoustic energy in aeroacoustics, or the radar cross-section in electromagnetics [5]. Pierce and Giles achieved superconvergent results by using a reconstruction process to formulate a smooth approximate numerical solution. The residual error in approximating the original p.d.e. was then evaluated, and an approximate adjoint solution was used to relate this residual error to the consequential error in the output functional of interest. Removing this estimate of the error gave a doubling of the order of accuracy of the functional in a number of test cases, including the 2D Laplace and quasi-1D Euler equations [3]. An alternative use for the reconstruction and residual error evaluation would have been to use it to improve the whole solution through the well-known established of defect correction (e.g. [1,4,7,8]). However, defect correction and adjoint error correction are not mutually exclusive; the best accuracy is to be achieved through the simultaneous use of both techniques.

Accordingly, in this paper we examine the use of both to improve the accuracy in approximating the scalar Helmholtz equation on an irregular 1D grid. The first section describes the model problem and the simple second-order accurate, piecewise linear, Galerkin finite element method which is used as the basic approximation. The second section describes the defect correction in which a smooth solution is reconstructed by cubic spline interpolation. The residual error then produces the source term in a calculation of a correction using the Galerkin solver; this step is repeated if necessary. The third section very briefly recaps the adjoint error correction procedure. The final section presents the numerical results, showing global fourth order accuracy for the solution obtained with defect correction. For the output functional, which in this case in the solution at one end of the domain, fourth order accuracy is achieved using either defect or adjoint error correction on their own, but sixth order accuracy is obtained when using both.

2 Problem description and Galerkin method

The model problem to be solved is the 1D Helmholtz equation

$$u'' + \pi^2 u = 0, \quad 0 < x < 10,$$

subject to the Dirichlet boundary condition $u = 1$ at $x = 0$ and the radiation boundary condition $u' - \mathrm{i}\pi u = 0$ at $x = 10$. The analytic solution is $u = \exp(\mathrm{i}\pi x)$ and the domain contains precisely five wavelengths. The output functional of interest is the value $u(10)$ at the right hand boundary. This can be viewed as a model of a far-field boundary integral giving the radiated acoustic energy in aeroacoustics, or the radar cross-section in electromagnetics [5].

Integrating by parts, the weak form of the inhomogeneous equation

$$u'' + \pi^2 u = f, \quad 0 < x < 10,$$

subject to the same boundary conditions is

$$-(w', u') + \pi^2 (w, u) + \mathrm{i}\pi w^*(10)\, u(10) = (w, f),$$

for any differentiable $w(x)$ with $w(0) = 0$. Here the inner product (w, u) is defined as

$$(w, u) \equiv \int_0^{10} w^* u \;\mathrm{d}x,$$

with w^* denoting the complex conjugate of w.

The Galerkin solution on the irregular grid $x_j, j = 0, 1, 2, \ldots, N$, is defined as

$$U(x) = \sum_{j=0}^{N} U_j \phi_j(x)$$

where the $\phi_j(x)$ are the usual piecewise linear 'hat' functions for which $\phi_j(x_i) = \delta_{ij}$. The value U_0 is given by the Dirichlet boundary condition. The values of the other coefficients U_j for $j > 0$ are obtained from the equations

$$-(\phi_i', U') + \pi^2(\phi_i, U) + \mathrm{i}\pi\phi_i(10)\, U(10) = 0, \quad i = 1, 2, \ldots, N.$$

It is well established that this discretisation is second order accurate, producing dispersion but no dissipation on a uniform grid.

3 Defect correction

The first step in the defect correction is to define a new approximate solution $u_h(x)$ by cubic spline interpolation of the nodal values U_j. The choice of end conditions for the cubic spline is very important. A natural cubic spline would have $u_h'' = 0$ at both ends, but this would introduce small errors at each end since $u'' \neq 0$ for the analytic solution. Instead, at $x = 10$ we require the splined solution to satisfy the analytic boundary condition by imposing $u_h' - i\pi u_h = 0$. At $x = 0$, the analytic boundary condition is already imposed through having the correct value for the end point $U(0)$. Therefore, here we require that $u_h'' + \pi^2 u_h = 0$ so the splined solution satisfies the o.d.e. at the boundary.

The solution error, $e = u(x) - u_h(x)$ satisfies the inhomogeneous Helmholtz equation

$$e'' + \pi^2 e = -(u_h'' + \pi^2 u_h), \quad 0 < x < 10,$$

the right-hand-side of which is the residual error of the approximation $u_h(x)$. Given the homogeneous Dirichlet boundary condition at $x = 0$, and the same radiation boundary condition at $x = 10$, the Galerkin approximation to the error is given by the equations

$$-(\phi_i', E') + \pi^2(\phi_i, E) + \mathrm{i}\pi\phi_i(10)\, E(10) = -(\phi_i, u_h'' + \pi^2 u_h), \quad i = 1, 2, \ldots, N.$$

Adding the nodal corrections E_j to the original nodal values U_j gives a corrected solution. The whole procedure can then be repeated to improve the accuracy. This follows the procedure described by Barrett *et al* who also showed that it converges to a solution of an appropriately defined Petrov-Galerkin discretisation, with the trial space being the space of cubic splines, while the test space is the space of piecewise linear functions [1].

4 Adjoint error correction

We begin with a presentation of the linear theory for adjoint error correction in applications with inhomogeneous boundary conditions and boundary functionals; for the nonlinear theory, see [3].

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 operator B may be different on different parts of the boundary, and in some applications (e.g. inflow and outflow sections for the convection p.d.e.) even its dimension may differ.

The output functional of interest is taken to be

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

where $(.,.)$ represents an integral inner product over the domain Ω and $(.,.)_{\partial\Omega}$ represents an integral inner product over the boundary $\partial\Omega$. In [3], the theory was presented for real variables, but here we are considering complex variables and so for the general case of vector variables u, v the inner product (v, u) is defined as

$$(v, u) \equiv \int_\Omega v^H u \ \mathrm{d}V$$

with v^H being the Hermitian (complex conjugate transpose) of v.

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 the components of h may be set to zero if the functional does not have a boundary integral contribution.

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

$$(L^* v, u) + (B^* v, Cu)_{\partial\Omega} = (v, Lu) + (C^* v, Bu)_{\partial\Omega},$$

for all u, v. 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 [2]. We will follow the same process later to construct the adjoint boundary operators for the Helmholtz equation.

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

$$\begin{aligned} Lu_h &= f_h, \qquad & L^* v_h &= g_h, \\ Bu_h &= e_h, \qquad & B^* v_h &= h_h, \end{aligned}$$

and hence obtain

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

In the final result, the first line is the functional based on the approximate solution u_h. The second line is the adjoint correction term which now includes a term related to the extent to which the primal solution does not correctly satisfy the boundary conditions. The third line is the remaining error for which an *a posteriori* error bound can be found, in principle [6].

To apply this theory to the Helmholtz problem, the first step is to construct the appropriate adjoint problem. Integration by parts reveals that the Helmholtz equation is self-adjoint, so

$$L^* v \equiv v'' + \pi^2 v,$$

and

$$(v,\, Lu) - (L^* v,\, u) = \left[\boldsymbol{v}^H \boldsymbol{A} \boldsymbol{u}\right]_0^{10},$$

where

$$\boldsymbol{u} = \begin{pmatrix} u \\ \frac{du}{dx} \end{pmatrix}, \quad \boldsymbol{v} = \begin{pmatrix} v \\ \frac{dv}{dx} \end{pmatrix},$$

and

$$\boldsymbol{A} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

At $x = 10$ we have

$$Bu \equiv u' - \mathrm{i}\pi u \equiv \boldsymbol{B}\boldsymbol{u}, \quad \boldsymbol{B} = \begin{pmatrix} -\mathrm{i}\pi & 1 \end{pmatrix},$$

and

$$Cu \equiv u \equiv \boldsymbol{C}\boldsymbol{u}, \quad \boldsymbol{C} = \begin{pmatrix} 1 & 0 \end{pmatrix}.$$

To satisfy the adjoint identity [2], we require $\boldsymbol{B}^*$ and $\boldsymbol{C}^*$ such that

$$\boldsymbol{A} = \left(\frac{-\boldsymbol{C}^*}{\boldsymbol{B}^*} \right)^H \left(\frac{\boldsymbol{B}}{\boldsymbol{C}} \right).$$

Solving this gives

$$\begin{pmatrix} -\boldsymbol{C}^* \\ \hline \boldsymbol{B}^* \end{pmatrix} = \begin{pmatrix} \boldsymbol{B} \\ \hline \boldsymbol{C} \end{pmatrix}^{-H} \boldsymbol{A}^H = \begin{pmatrix} 1 & 0 \\ -\mathrm{i}\pi & -1 \end{pmatrix}$$

and hence $B^*v \equiv -v' - \mathrm{i}\pi v$ and $C^*v \equiv -v$. Similarly, at $x=0$, we obtain $B^*v = v$ and $C^*v = v'$.

Now, noting that in our application $f=g=0$, and h has value 0 at $x=0$ and 1 at $x=10$, then the full specification of the adjoint problem is

$$v'' + \pi^2 v = 0, \quad 0 < x < 10,$$

with $v=0$ at $x=0$ and $-v' - \mathrm{i}\pi v = 1$ at $x=10$.

Let v_h be an approximate solution of this problem, obtained by the same Galerkin and cubic spline reconstruction approach as u_h, with or without defect correction. Noting that the cubic spline reconstruction ensures that the boundary conditions are satisfied exactly, the corrected approximation to the value $u(10)$ is

$$u_h(10) - (v_h, u_h'' + \pi^2 u_h).$$

The theory gives the error in this corrected functional as being

$$(v_h - v, u_h'' + \pi^2 u_h).$$

In the absence of defect correction, both terms in this inner product are second order in the average grid spacing and so the error is fourth order. With defect correction, the first term is fourth order while the second term remains second order. Therefore, the error remaining after the adjoint error correction is sixth order.

5 Numerical results

Numerical results have been obtained for grids with 4, 8, 16, 32, 64 and 128 points per wavelength. To test the ability to cope with irregular grids, the coordinates for the grid with N intervals are defined as

$$x_0 = 0, \quad x_N = 10, \quad x_j = \frac{10}{N}\left(j + \sigma_j\right), \; 0<j<N,$$

where σ_j is a uniformly distributed random variable in the range $[-0.3, 0.3]$.

Figure 1 shows the L_2 norm of the error in the reconstructed cubic spline solution before and after defect correction. Without defect correction, the error is second order, while with defect correction it is fourth order. Note that a second application of defect correction makes a significant reduction in the error even though it remains fourth order. This is because one application of the defect correction procedure gives a correction which is second order in magnitude, with a corresponding error which is second order in relative magnitude and therefore

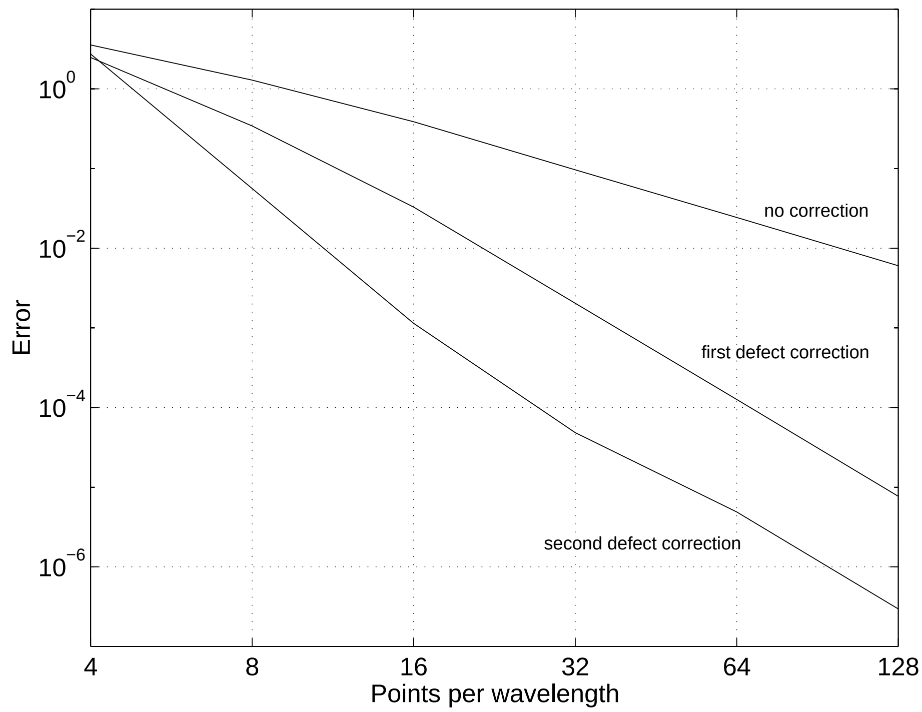

Fig. 1. L_2 error in the numerical approximation to $u(x)$

fourth order in absolute magnitude. It is this error which is corrected by a second application of the defect correction procedure.

Figure 2 shows the error in the numerical value for the output functional $u(10)$. Without any correction, the error is second order. Using either defect correction or adjoint error correction on their own increases the order of accuracy to fourth order, but using them both increases the accuracy to sixth order. Note that the calculation with 8 points per wavelength plus both defect and adjoint error correction gives an error which is approximately 2×10^{-3}. This is more accurate than the calculation with 128 points per wavelength and no corrections, and comparable to the results using 14 points and defect correction, or 30 points with adjoint error correction.

In 3D, the computational cost is proportional to the cube of the number of points per wavelength, so this indicates the potentially huge savings offered by the combination of defect and adjoint error correction. The cost of computing the corrections is five times the cost of the original calculation, due to the additional two calculations for the defect correction, and the one adjoint calculation plus its two defect corrections. In practice, the second defect correction for the primal and adjoint calculations make negligible difference to the value obtained after

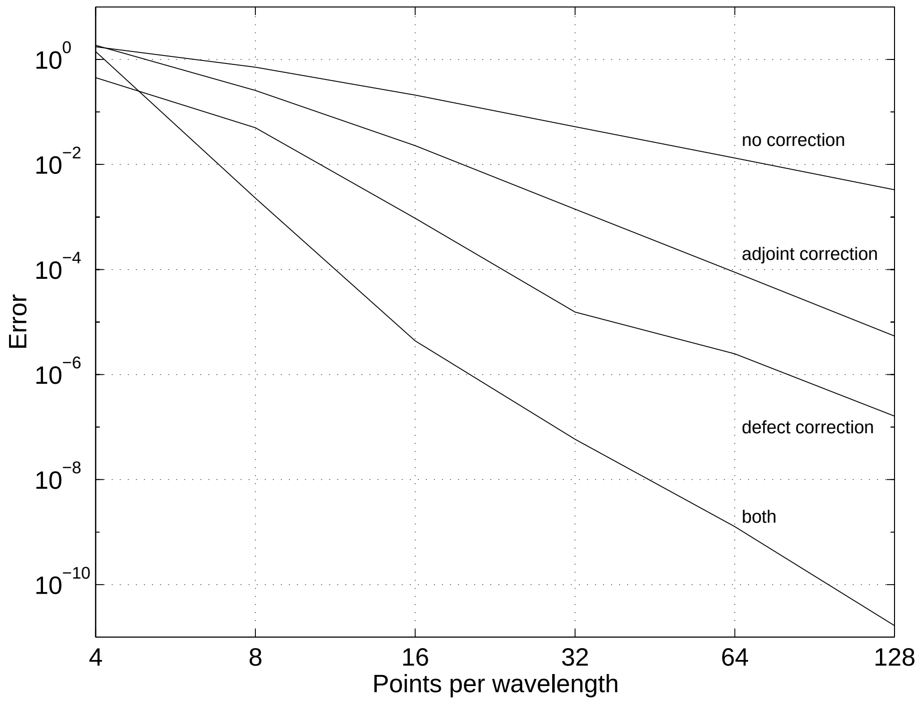

Fig. 2. Error in the numerical approximation to $u(10)$

the adjoint error correction, so these can be omitted, reducing the cost of the corrections to just three times the cost of the original calculation.

6 Concluding remarks

The numerical results which have been presented show the potential offered by defect correction and adjoint error correction, but there is much work to be done to achieve this potential for multi-dimensional applications. There will be some problems in the representation and approximation of curved boundaries and boundary integrals, but the key issue is likely to be the smooth reconstruction of a numerical solution from nodal data. On a structured grid, cubic spline interpolation can be used in each direction, but on an unstructured grid one would need a suitable generalisation of cubic spline interpolation to produce a reconstructed solution of sufficient smoothness. This will be the main challenge in trying to reproduce similar improvements in accuracy for aeroacoustic and electromagnetic calculations on 3D unstructured grids.

Acknowledgements

This research was supported by EPSRC grants GR/K91149 and GR/L95700, and by Rolls-Royce plc.

References

1. J.W. Barrett, G. Moore, and K.W. Morton. Optimal recovery in the finite element method, part 2: Defect correction for ordinary differential equations. *IMA J. Numer. Anal.*, 8:527–540, 1988.
2. M.B. Giles and N.A. Pierce. Adjoint equations in CFD: duality, boundary conditions and solution behaviour. AIAA Paper 97-1850, 1997.
3. M.B. Giles and N.A. Pierce. Improved lift and drag estimates using adjoint Euler equations. AIAA Paper 99-3293, 1999.
4. B. Koren. Defect correction and multigrid for an efficient and accurate computation of airfoil flows. *J. Comput. Phys.*, 76, 1988.
5. 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(1):251–274, 1998.
6. N.A. Pierce and M.B. Giles. Adjoint recovery of superconvergent functionals from PDE approximations. *SIAM Rev.*, 42(2):247–264, 2000.
7. R.D. Skeel. A theoretical framework for proving accuracy results for deferred corrections. *SIAM J. Numer. Anal.*, 19:171–196, 1981.
8. H.J. Stetter. The defect correction principle and discretization methods. *Numer. Math.*, 29:425–443, 1978.