AN ASSESSMENT OF THE FINITE TERMINATION PROPERTY OF THE DEFECT CORRECTION METHOD

G. J. SHAW AND P. I. CRUMPTON

Oxford University Computing Laboratory, Numerical Analysis Group, 11 Keble Road, Oxford, OX1 3QD, U.K.

SUMMARY

This paper investigates the use of defect correction procedures for the solution of finite volume approximations to systems of conservation laws. Particular emphasis is laid on the order of accuracy obtained after a fixed finite number of iterations. It is shown that a high order of accuracy may be achieved after only one defect correction iteration, involving two inversions of a stable lower-order-accurate operator. However, this result is found to be critically dependent on the consistency of the lower-order operator, a property which does not always hold for conservative finite volume discretizations. Through numerical experiments, the lack of consistency of these schemes is found to inhibit severely the finite termination property of the defect correction process. Results are presented for linear advection, Poisson's equation, and the Euler equations.

KEY WORDS Defect correction Conservation laws

1. INTRODUCTION

The discretization of conservation laws almost inevitably results in a compromise between accuracy and efficiency. The linear algebraic systems derived from low-order approximations are generally stable and may be solved very efficiently by methods such as multigrid[1,25] or preconditioned conjugate gradients.[14,18] However, the accuracy obtained from such discretizations is often not satisfactory until the mesh is so fine that the order of the algebraic system becomes prohibitive. An alternative approach is to limit the refinement of the mesh and use more accurate discretizations. The difficulty in this case is that these discretizations are generally less compact and require more complicated procedures for their solution. More importantly, they are often less stable and may lead to matrices with very large condition numbers, which standard preconditioning cannot rectify. These factors can seriously limit the range of efficient solution procedures available and result in unacceptable computer time requirements.

The defect correction method provides a partial answer to some of these problems, producing accurate solutions but requiring only the inversion of a stable lower-order operator at each step of an iterative process. For this purpose, the efficient algorithms mentioned above may be employed. This paper is concerned with the use of defect correction for the solution of systems of conservation laws, such as the Euler equations describing the flow of an inviscid, compressible fluid. The method has been adopted for the solution of these equations by many authors,[12,13,15] who use a highly efficient multigrid method for the inversion of the lower-order operator.

Although defect correction is designed to converge to a particular higher-order discrete solution, it is not always necessary to converge the iteration fully in order to obtain an approximate solution of the required accuracy. This paper examines, in particular, the accuracy

0271–2091/93/030199–17$13.50

Received May 1992
Revised September 1992

obtained after a single defect correction iteration, showing that, for linear advection, a second-order approximation may be obtained at the cost of only two inversions of a first-order operator. It is found that, for this property to hold, the consistency of the first-order scheme is of paramount importance. For the Euler equations, conservative finite volume schemes are the natural discretization. However, such schemes are not always consistent on distorted meshes, achieving their accuracy through supraconvergence. The finite termination property of the defect correction iteration is shown not to hold in this case.

The paper begins with a description of defect correction for linear partial differential equations. Some straightforward results are given concerning the accuracy obtained after a single iteration. These are illustrated by the numerical results presented in Sections 3–5, for linear advection, Poisson's equation and the non-linear Euler equations, respectively.

2. DEFECT CORRECTION

This section introduces the defect correction procedure as applied to linear partial differential equations. The method has several possible generalizations to non-linear problems, which are described, for example, in Reference 7. One particular non-linear variant is used for the Euler equations in Section 5.

Consider the linear partial differential equation

$$Lu=f, \quad u=u(\mathbf{x}), \ \mathbf{x}\in\Omega\subset\mathbb{R}^d, \tag{1}$$

subject to suitable conditions on the boundary $\delta\Omega$ of Ω. Define a mesh

$$\Omega_h=\{\mathbf{x}_i\colon i=1,2,\ldots,n\}\subset\bar{\Omega}, \tag{2}$$

with characteristic steplength h. Let $\mathbf{U}\in\mathbb{R}^n$ denote the restriction of u to Ω_h, with ith element $u(\mathbf{x}_i)$, and similarly let $\mathbf{F}\in\mathbb{R}^n$ be the restriction of f.

Consider two distinct approximations to the differential problem (1), defined on the mesh Ω_h and given by

$$L_1\mathbf{U}_1=\mathbf{F}_1, \qquad L_2\mathbf{U}_2=\mathbf{F}_2. \tag{3}$$

It is assumed that the $n\times n$ matrix $L_1:\mathbb{R}^n\rightarrow\mathbb{R}^n$ is easily invertible, but has a lower consistency order than $L_2:\mathbb{R}^n\rightarrow\mathbb{R}^n$. Thus,

$$L_1\mathbf{U}=\mathbf{F}_1+\mathbf{T}_1, \qquad L_2\mathbf{U}=\mathbf{F}_2+\mathbf{T}_2, \tag{4}$$

where the truncation error vectors $\mathbf{T}_1$ and $\mathbf{T}_2$ satisfy

$$\|\mathbf{T}_1\|\leq C_1h^{p_1}, \quad \|\mathbf{T}_2\|\leq C_2h^{p_2}, \qquad p_2>p_1, \tag{5}$$

for some vector norm $\|\cdot\|$ and constants C_1, C_2. It is further assumed that the boundary conditions have been discretized and incorporated into the linear systems (3), so that $\mathbf{F}_1, \mathbf{F}_2$ represent approximations to $\mathbf{F}$ with appropriate boundary data included.

Defect correction is an iterative process

$$\begin{aligned}&\mathbf{U}^{(0)}=L_1^{-1}\mathbf{F}_1,\\ &\mathbf{U}^{(k+1)}=\mathbf{U}^{(k)}+L_1^{-1}(\mathbf{F}_2-L_2\mathbf{U}^{(k)}), \quad k=0,1,2,\ldots,\end{aligned} \tag{6}$$

which involves inversion of the lower-order operator L_1 only. The method evolved from the deferred correction scheme of Fox[6] and has since then been studied by many authors; examples include References 21, 22, 10, 11 and 15.

Defining the discrete error $\mathbf{E}^{(k)}=\mathbf{U}_2-\mathbf{U}^{(k)}$, equation (6) gives

$$\mathbf{E}^{(k+1)}=G\mathbf{E}^{(k)}, \tag{7}$$

where

$$G=I-L_1^{-1}L_2 \tag{8}$$

is the iteration matrix. The process, therefore, converges to the higher-order solution $\mathbf{U}_2$ provided the spectral radius $\rho(G)$ satisfies $\rho(G)<1$, the asymptotic rate of convergence being governed by the value of the spectral radius.

However, Pereyra[22] proved that, under certain conditions, a higher-order solution is indeed obtained long before the iteration has converged, after a small finite number of iterations. The following single iteration theorem exemplifies Pereyra's results.

Theorem 1

Suppose that $L_1U=\mathbf{F}_1+\mathbf{T}_1$, $L_2\mathbf{U}=\mathbf{F}_2+\mathbf{T}_2$ and define the iterates

$$\mathbf{U}^{(0)}=L_1^{-1}\mathbf{F}_1, \qquad \mathbf{U}^{(1)}=\mathbf{U}^{(0)}+L_1^{-1}(\mathbf{F}_2-L_2\mathbf{U}^{(0)}), \tag{9}$$

where the truncation error vectors $\mathbf{T}_1$, $\mathbf{T}_2$ satisfy the conditions $\|\mathbf{T}_1\|\leq C_1h^{p_1}$, $\|\mathbf{T}_2\|\leq C_2h^{p_2}$ for some norm $\|\cdot\|$. Let $\mathscr{E}^{(k)}=\mathbf{U}-\mathbf{U}^{(k)}$, $k=0, 1$.

Then $\|\mathscr{E}^{(1)}\|\leq Ch^r$, where $r=\min(p_1+q, p_2)$, provided

1. $\|L_1^{-1}\|\leq C_S$
2. $\|(I-L_1^{-1}L_2)L_1^{-1}\|\leq C_Ah^q$,

for some positive real constants C_1, C_2, C_S, C_A, C independent of h.

Proof. From the definition of the truncation error $\mathbf{T}_2$, it is clear that

$$\mathbf{U}=\mathbf{U}+L_1^{-1}(\mathbf{F}_2-L_2\mathbf{U}+\mathbf{T}_2). \tag{10}$$

Subtracting the equation for $\mathbf{U}^{(1)}$ in (9) from (10) gives

$$\mathscr{E}^{(1)}=(I-L_1^{-1}L_2)\mathscr{E}^{(0)}+L_1^{-1}\mathbf{T}_2. \tag{11}$$

But

$$\mathscr{E}^{(0)}=\mathbf{U}-L_1^{-1}\mathbf{F}_1=L_1^{-1}\mathbf{T}_1. \tag{12}$$

Thus,

$$\mathscr{E}^{(1)}=(I-L_1^{-1}L_2)L_1^{-1}\mathbf{T}_1+L_1^{-1}\mathbf{T}_2, \tag{13}$$

and, hence,

$$\|\mathscr{E}^{(1)}\|\leq C_1C_Ah^{(p_1+q)}+C_2C_Sh^{p_2}. \quad \square \tag{14}$$

Remarks

1. Note that the error $\mathscr{E}^{(k)}$ is between the analytical solution of (1) and the *kth* iterate $\mathbf{U}^{(k)}$, whereas $\mathbf{E}^{(k)}$ denotes the error between the fixed point of (6) and $\mathbf{U}^{(k)}$.
2. Condition 1 is a common stability condition on the lower-order operator, which may be expected for many standard discretizations. This condition ensures that the global error of an approximation is of at least the same order as the truncation error.

3. Typically, $||I-L_1^{-1}L_2||=O(1)$, and it is only by taking the product with L_1^{-1} that a value $q>0$ can be obtained for condition 2, i.e.

$$||(I-L_1^{-1}L_2)L_1^{-1}||\leqslant||I-L_1^{-1}L_2||\,||L_1^{-1}||. \tag{15}$$

As an example of Theorem 1, consider the case where $||\mathbf{T}_1||$ is first-order and $||\mathbf{T}_2||$ second-order, i.e. $p_1=1$ and $p_2=2$. If L_1 has a bounded inverse and $||GL_1^{-1}||=O(h)$, the analytical error of the first iterate $||\mathscr{E}^{(1)}||$ is already second-order. Results of this type are illustrated in Section 3. In this situation there may be little advantage in continuing the iteration to convergence. Indeed, examples will be presented for which the first iterate is more accurate than the fixed point of the iteration.

It is important to note that on a non-uniform mesh it is quite possible to have $||\mathbf{T}_1||=O(1)$ but $||L_1^{-1}\mathbf{T}_1||=O(h)$, in which case the method achieves first-order global accuracy despite being inconsistent, and is said to be supraconvergent. Suppose now that $\mathbf{T}_2$ is second-order and condition 2 holds with $q=1$. Under these conditions, the first iterate cannot be proved to be second-order accurate since the term $||GL_1^{-1}\mathbf{T}_1||$ may be split either as $||G||\,||L_1^{-1}\mathbf{T}_1||$ or $||GL_1^{-1}||\,||\mathbf{T}_1||$, which are both generally only of first order. The results of Section 3 will demonstrate that second-order accuracy is, in fact, generally not obtained for these inconsistent supraconvergent lower-order schemes.

The single iteration theorem is readily generalized to deal with multiple iterations of defect correction. Following the manner in which (11) was derived, it is easy to show from (4) and (6) that

$$\mathscr{E}^{(k+1)}=G\mathscr{E}^{(k)}+L_1^{-1}\mathbf{T}_2,\quad k=0,1,2,\dots . \tag{16}$$

As a consequence, it follows that

$$\begin{aligned}\mathscr{E}^{(k)}&=G^k\mathscr{E}^{(0)}+\sum_{j=0}^{k-1}G^jL_1^{-1}\mathbf{T}_2\\&=G^kL_1^{-1}\mathbf{T}_1+\sum_{j=0}^{k-1}G^jL_1^{-1}\mathbf{T}_2,\quad k=1,2,\dots .\end{aligned} \tag{17}$$

Given a stable lower-order operator and a convergent defect correction process, the accuracy of the kth iterate is then dependent on the order of $||G^kL_1^{-1}||$. However, it is generally more difficult to obtain an accurate bound for this quantity.

Hackbusch[7] presents the more general result that the kth iterate is of order $\min\{p_2,(k+1)p_1\}$ provided L_1 is stable, and a relative consistency condition bounding the norm of L_2-L_1 holds. This result is based on the use of norms on scales of Banach spaces. Subject to these conditions, optimal accuracy is obtained as soon as $k\geq p_2/p_1-1$.

3. DEFECT CORRECTION FOR LINEAR ADVECTION

In this section numerical results are presented for the constant-coefficient linear advection equation

$$au_x+bu_y=0,\quad a>0,\ b>0,\qquad \mathbf{x}=(x,y)\in[0,1]^2, \tag{18}$$

discretized on successively distorted and stretched grids. These are generated by first defining a uniform $N\times N$ mesh

$$\Omega_h=\{(x_{ij},y_{ij})\colon x_{ij}=(i-1)h,\ y_{ij}=(j-1)h,\ i,j=1,\dots,N\}, \tag{19}$$

where $h=1/(N-1)$, which is then distorted by randomly perturbing each interior node within a circle of radius $\sigma h/200$. The parameter σ, therefore, represents the percentage of distortion. The extreme case of $\sigma=100$ permits coincident nodes. The stretching is achieved using the function

$$w(\xi,\mu)=\begin{cases}\dfrac{\exp(-\mu\xi)-1}{\exp(-\mu)-1}, & \mu\neq 0\\ 0, & \mu=0,\end{cases} \tag{20}$$

by redefining

$$x_{ij}:=w(x_{ij},\mu),\quad y_{ij}:=w(y_{ij},\mu),\qquad i,j=1,\dots,N, \tag{21}$$

for some stretching parameter μ. The aim of the experiments is to establish whether the finite termination property of defect correction holds in practice on such meshes.

The higher-order scheme is derived from the cell vertex discretization.[20,4,19] It is obtained by taking an area-weighted linear combination of four neighbouring cell residuals, as in the update procedure of Reference 8. Let U_r denote the approximate solution at the node r as depicted in Figure 1. Similarly, let (x_r, y_r) be the co-ordinates of that node and define $\delta x_{rs}=x_r-x_s$, $\delta y_{rs}=y_r-y_s$. The higher-order scheme may then be written as

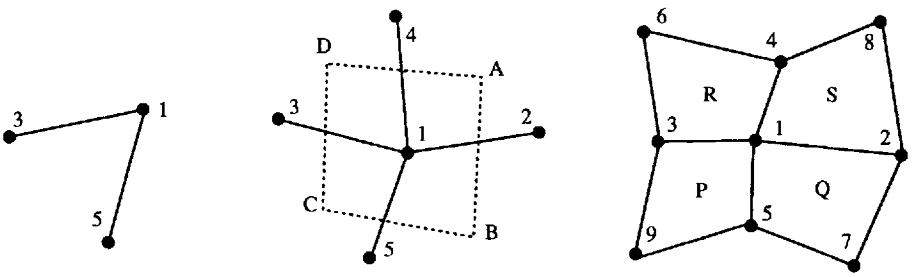
$$(au_x+bu_y)\approx\frac{1}{2V}\left[\begin{array}{l}(a\delta y_{87}-b\delta x_{87})U_2+(a\delta y_{42}-b\delta x_{42})U_8\\ \quad+(a\delta y_{68}-b\delta x_{68})U_4+(a\delta y_{34}-b\delta x_{34})U_6\\ \quad+(a\delta y_{96}-b\delta x_{96})U_3+(a\delta y_{53}-b\delta x_{53})U_9\\ \quad+(a\delta y_{79}-b\delta x_{79})U_5+(a\delta y_{25}-b\delta x_{25})U_7\end{array}\right], \tag{22}$$

where $V=V_P+V_Q+V_R+V_S$ is the sum of the respective control volumes surrounding node 1. Due to the presence of spurious error modes in this discretization, an iterative solution procedure will converge only very slowly.[2]

Two distinct lower-order operators are used, the first is the unique three-point consistent upwind finite difference method given by

$$(au_x+bu_y)\approx\frac{1}{\delta x_{31}\delta y_{51}-\delta x_{51}\delta y_{31}}\left[\begin{array}{l}(a\delta y_{51}-b\delta x_{51})(U_3-U_1)\\ \quad+(a\delta y_{31}-b\delta x_{31})(U_1-U_5)\end{array}\right], \tag{23}$$

which has a first-order truncation error on any mesh. The second is a conservative vertex-centred



First Order
Finite Difference

First Order
Finite Volume

Second Order
Finite Volume

Figure 1. Stencil of operators

upwind finite volume scheme

$$(au_x+bu_y)\approx\frac{1}{V}\left[\begin{array}{l}(a\delta y_{\mathrm{AB}}-b\delta x_{\mathrm{AB}})U_1+(a\delta y_{\mathrm{DA}}-b\delta x_{\mathrm{DA}})U_1\\ +(a\delta y_{\mathrm{CD}}-b\delta x_{\mathrm{CD}})U_3+(a\delta y_{\mathrm{BC}}-b\delta x_{\mathrm{BC}})U_5\end{array}\right], \tag{24}$$

where V is the area of the quadrilateral ABCD, whose vertices are the centroids of the mesh cells surrounding node 1. This method loses consistency on non-uniform meshes, but obtains first-order accuracy through supraconvergence. It has been adopted for solution of the Euler equations.[5,9] The stencils for (22), (23) and (24) are represented in Figure 1.

The convergence behaviour of a defect correction method using these definitions of L_1 and L_2 may be investigated by means of Fourier analysis on a uniform mesh $\mathbf{x}=(x, y)\in\Omega_h$ as defined by (19), in which case the lower-order operators (23) and (24) are identical. The Fourier symbol $G(\boldsymbol{\theta})$ of the iteration matrix, defined by

$$G\exp(i\boldsymbol{\theta}\cdot\mathbf{x}/h)=G(\boldsymbol{\theta})\exp(i\boldsymbol{\theta}\cdot\mathbf{x}/h),\quad \boldsymbol{\theta}=(\theta_1,\theta_2), \tag{25}$$

is shown in Figure 2 for the case $a=1$, $b=2$. No damping is observed on the lines $\boldsymbol{\theta}=(\pm\pi, \theta_2)$, $\boldsymbol{\theta}=(\theta_1, \pm\pi)$, or in their neighbourhoods. This is a consequence of the fact that these modes lie in the nullspace of L_2 but not of L_1. There is also no damping near the origin for characteristic modes such that $a\theta_1+b\theta_2=0$. Clearly, the iterates will not converge rapidly to the fixed point $\mathbf{U}_2$. However, the general conclusion is that the iteration damps smooth modes far more effectively than high-frequency modes.

Figure 3 shows experimental results on successively randomized meshes, with $\sigma=10, 25, 50$ and 90. The graph shows $\log(\|\mathscr{E}^{(k)}\|)$ against $\log(1/N)$ for increasing N; the slope, therefore, represents the order of convergence. The first ten defect correction iterations $k=0,\ldots, 10$ are shown on each graph. Clearly, the finite volume scheme achieves only first-order accuracy after several defect correction iterations, whereas the consistent finite difference scheme gives second-order accuracy after only one iteration. This is so despite the fact that even the solution of the higher-order scheme, which is the fixed point of the iteration, is not second-order on these distorted meshes.[24] This is illustrated by Figure 4, which shows the accuracy of the fixed-point

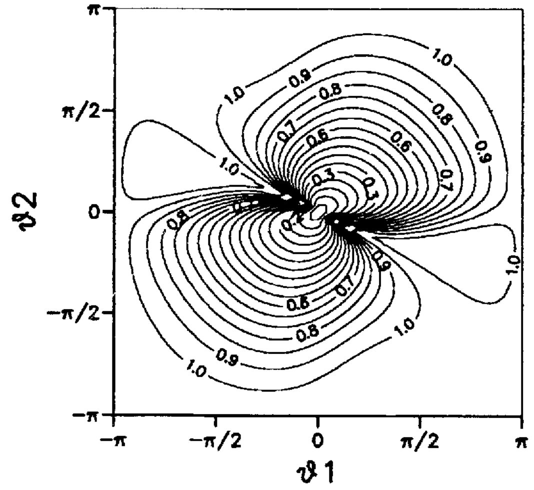
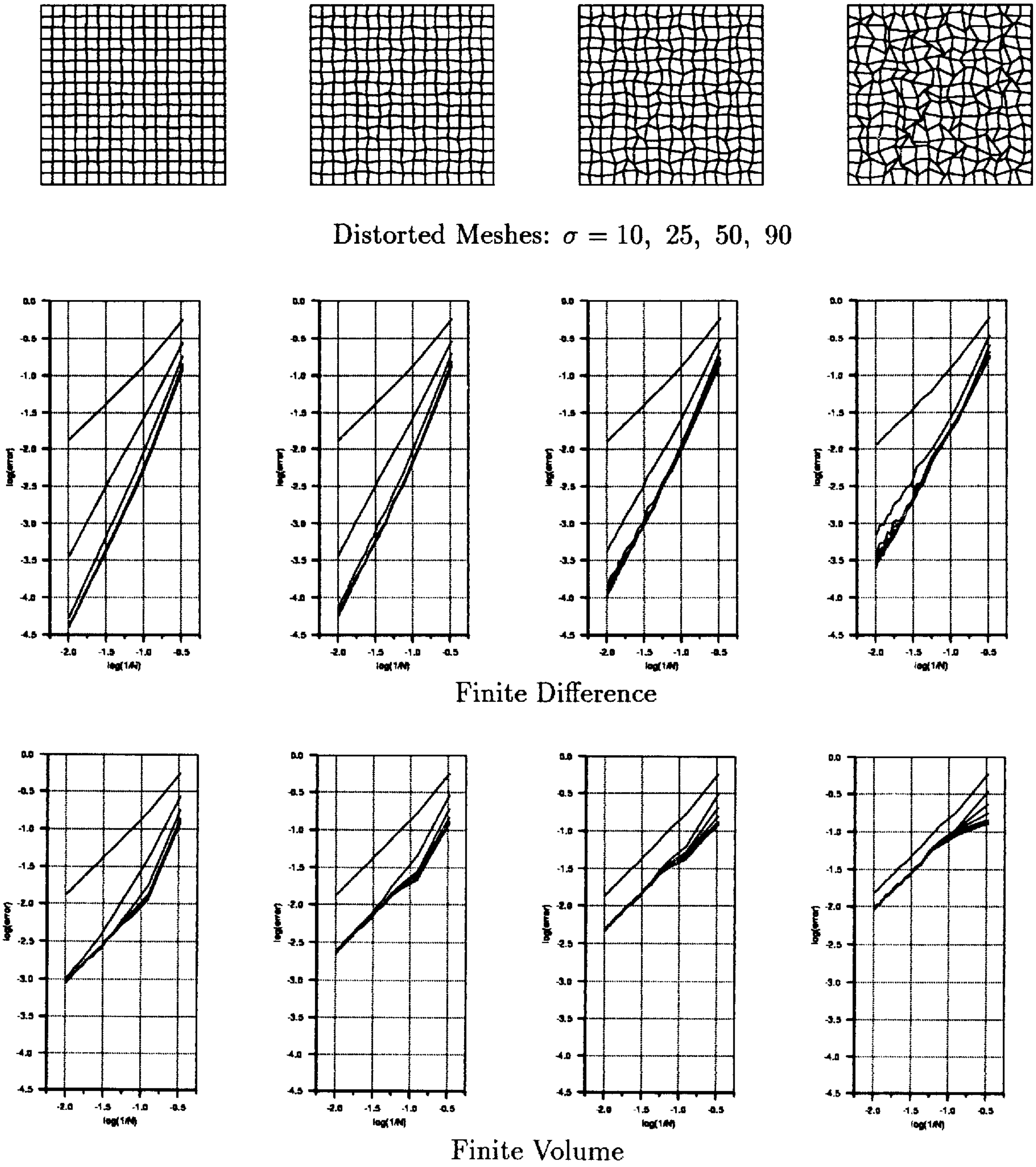


Figure 2. Symbol of iteration matrix for linear advection (18)

solution of $L_2\,\mathbf{U}_2 = \mathbf{F}_2$ on the same sequence of meshes. Note that on distorted meshes the first iterate is significantly more accurate than the limit of the iteration.

Some insight into the behaviour of the method for the two definitions of L_1 may be gained from Figure 5. This shows the errors $\mathscr{E}^{(0)}$ and $\mathscr{E}^{(1)}$, on a randomly distorted mesh, before and after

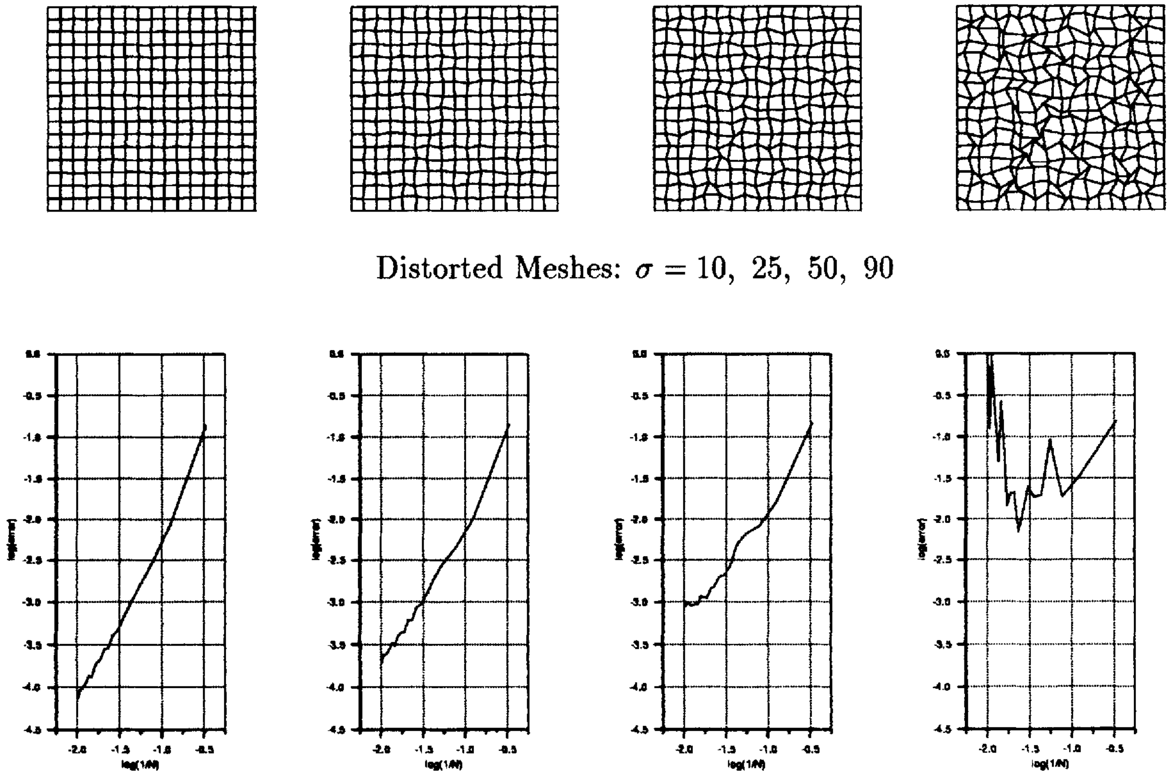


Distorted Meshes: $\sigma = 10,\ 25,\ 50,\ 90$

Finite Difference

Finite Volume

Figure 3. Convergence of ten defect correction iterations on distorted meshes



Distorted Meshes: $\sigma = 10,\ 25,\ 50,\ 90$

Fixed Point Soluton

Figure 4. Convergence of $\mathbf{U}_2$ on distorted meshes

a defect correction iteration, for the two methods. For both methods, we expect $\mathscr{E}^{(0)}$ to be first-order, by virtue of supraconvergence for the finite volume scheme and the combination of stability and consistency for the finite difference method. This indeed is the case, as may be observed in Figure 3. However, $\mathscr{E}^{(0)}$ for the consistent finite difference method may be seen to be significantly smoother than for the conservative finite volume scheme. Since

$$\mathscr{E}^{(1)}=G\mathscr{E}^{(0)}+L_1^{-1}\mathbf{T}_2 \tag{26}$$

from (11), the accuracy of $\mathbf{U}^{(1)}$ is influenced largely by the effect of the iteration matrix on $\mathscr{E}^{(0)}$. As predicted by the Fourier analysis, this matrix damps smooth modes more effectively than oscillatory ones. The result is that the defect correction iteration is more effective for the smooth errors produced by (23) than for those of (24), and second-order accuracy is achieved in the former case after one step, as shown in Figure 3.

Figure 6 shows the results for (18) discretized on stretched meshes, in the same format as for Figure 3. For these meshes the two lower-order schemes give identical results, with second-order accuracy being achieved after one iteration for both schemes. In this case the first-order finite volume scheme maintains consistency through the 'smooth' stretching function (20).

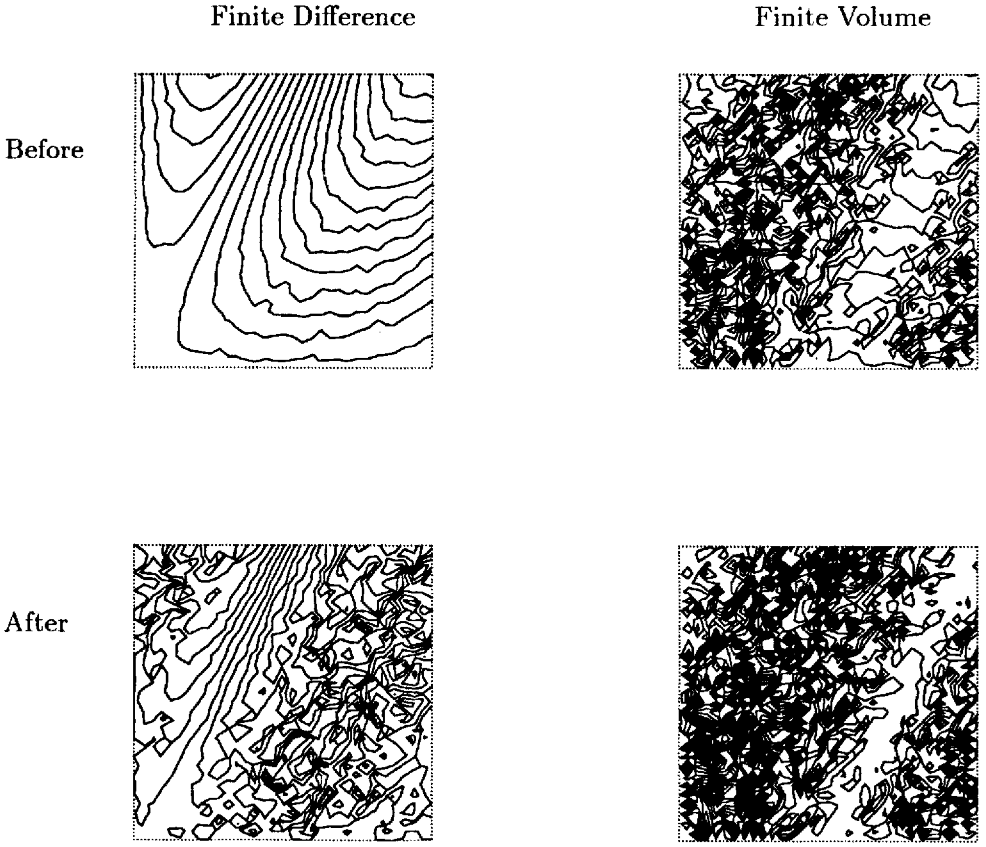


Figure 5. Contour plot of the error before and after a defect correction iteration with finite difference and finite volume lower-order operators on a randomly distorted mesh: $\sigma=25$

4. DEFECT CORRECTION FOR POISSON'S EQUATION

The results of the previous section illustrate that second-order accuracy may be obtained at the cost of only two inversions of a first-order operator. As a further demonstration of the finite termination property, it is shown in this section that fourth-order accuracy may be achieved after two inversions of a stable second-order approximation to the Poisson's equation

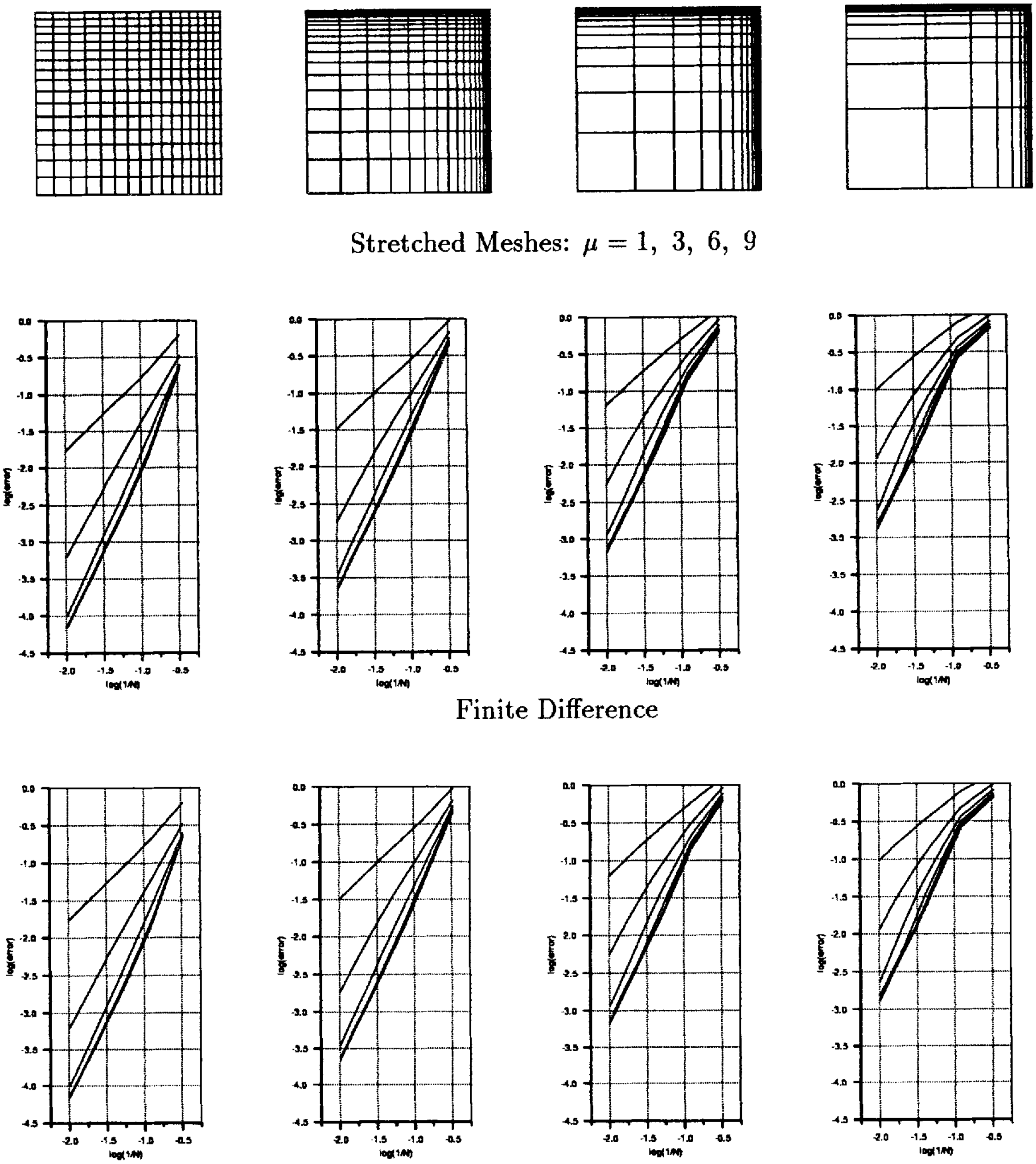
$$\nabla^2 u=f, \quad u=u(\mathbf{x}), \quad \mathbf{x}=(x, y)\in\Omega=[0, 1]^2, \tag{27}$$

with u given by Dirichlet data on $\delta\Omega$.

Equation (27) is discretized on the $N\times N$ uniform mesh Ω_h defined by (19). Let U_{ij} denote the approximation to $u(x_{ij}, y_{ij})$. The lower-order operator L_1 is the usual five-point scheme obtained by

$$u_{xx}(x, y)\approx[U_{i-1,j}-2U_{i,j}+U_{i+1,j}]/h^2, \tag{28}$$

with a similar approximation for u_{yy}. This method has a consistency order of two on the uniform mesh Ω_h, and can be inverted using the optimally efficient multigrid method.[1] The discretization



Stretched Meshes: $\mu = 1,\ 3,\ 6,\ 9$

Finite Difference

Finite Volume

Figure 6. Convergence of ten defect correction iterations on stretched meshes

Table I. Error norms and orders of convergence for defect correction iterations $k=0,\dots,5$ applied to Poisson's equation. L_1 is defined using (28) and L_2 using (29) and (30)

	Norm of $\mathscr{E}^{(k)}$				Order of convergence			
k	$N=9$	$N=17$	$N=33$	$N=65$	k	9–17	17–33	33–65
0	$0{\cdot}26\times10^{-4}$	$0{\cdot}64\times10^{-5}$	$0{\cdot}16\times10^{-5}$	$0{\cdot}39\times10^{-6}$	0	2·04	2·03	2·02
1	$0{\cdot}14\times10^{-5}$	$0{\cdot}87\times10^{-7}$	$0{\cdot}53\times10^{-8}$	$0{\cdot}33\times10^{-9}$	1	3·99	4·03	4·02
2	$0{\cdot}63\times10^{-7}$	$0{\cdot}13\times10^{-8}$	$0{\cdot}10\times10^{-9}$	$0{\cdot}72\times10^{-11}$	2	5·65	3·65	3·80
3	$0{\cdot}40\times10^{-7}$	$0{\cdot}14\times10^{-8}$	$0{\cdot}11\times10^{-9}$	$0{\cdot}75\times10^{-11}$	3	4·90	3·61	3·89
4	$0{\cdot}40\times10^{-7}$	$0{\cdot}13\times10^{-8}$	$0{\cdot}11\times10^{-9}$	$0{\cdot}75\times10^{-11}$	4	4·89	3·61	3·89
5	$0{\cdot}40\times10^{-7}$	$0{\cdot}13\times10^{-8}$	$0{\cdot}11\times10^{-9}$	$0{\cdot}75\times10^{-11}$	5	4·89	3·61	3·89

can be regarded either as a finite difference scheme, or as a finite volume method using the control volume ABCD shown in Figure 1.

The higher-order operator is defined by replacing (28) in the interior by the approximation

$$u_{xx}(x, y) \approx [-U_{i-2,j} + 16U_{i-1,j} - 30U_{ij} + 16U_{i+1,j} - U_{i+2,j}]/12h^2. \tag{29}$$

Near the boundary, at $x=h$, (28) is replaced instead by

$$u_{xx}(x, y) \approx [10U_{i-1,j} - 15U_{ij} - 4U_{i+1,j} + 14U_{i+2,j} - 6U_{i+3,j} + U_{i+4,j}]/12h^2, \tag{30}$$

with a similar modification at $x=1-h$. With these definitions, and a similar treatment of u_{yy}, the approximation L_2 is consistent to order four.

This example differs substantially from those discussed in Section 3. The higher-order operator L_2 does not in this case admit spurious high-frequency error modes, and, as a consequence, the defect correction process converges rapidly. Table I shows the result of applying five defect corrections using these definitions of L_1 and L_2. Clearly, $\mathscr{E}^{(0)}$ is second-order, which reflects the fact that (28) has a second-order truncation error and L_1 is stable. Fourth-order accuracy is obtained after the first iteration, although the error norms at this stage are significantly larger than those for subsequent iterations. After the second iteration, the error norms change very little and there is clearly no advantage in continuing the defect correction process. These results suggest that the conditions of the single-iteration theorem hold, with $p_1=2$, $p_2=4$ and $q=2$. Thus, the order of accuracy of the first iterate is given as $r=\min(p_1+q, p_2)=4$. Alternatively, the result proved by Hackbusch[7] suggests the order $r=\min\{p_2, (k+1)p_1\}=4$.

5. DEFECT CORRECTION FOR THE EULER EQUATIONS

This section considers the use of defect correction techniques for solution of the steady two-dimensional Euler equations

$$\mathbf{f}_x + \mathbf{g}_y = 0, \quad \mathbf{x} = (x, y) \in \Omega, \tag{31}$$

where the fluxes $\mathbf{f}$ and $\mathbf{g}$ are defined in terms of the state vector $\mathbf{w}$ by

$$\mathbf{f(w)} = \begin{pmatrix} \rho u \\ \rho u^2 + p \\ \rho uv \\ (E+p)u \end{pmatrix}, \quad \mathbf{g(w)} = \begin{pmatrix} \rho v \\ \rho uv \\ \rho v^2 + p \\ (E+p)v \end{pmatrix}, \quad \mathbf{w} = \begin{pmatrix} \rho \\ \rho u \\ \rho v \\ E \end{pmatrix}. \tag{32}$$

These equations describe conservation of mass, momentum and energy for an inviscid compressible fluid. In (32) ρ is the density, (u, v) are the Cartesian components of velocity and E is the total energy per unit volume. It is assumed that the fluid is a perfect gas. In this context, the pressure p is defined by the equation of state

$$p=(\gamma-1)[E-\tfrac{1}{2}\rho(u^2+v^2)], \qquad (33)$$

where γ is the ratio of specific heats.

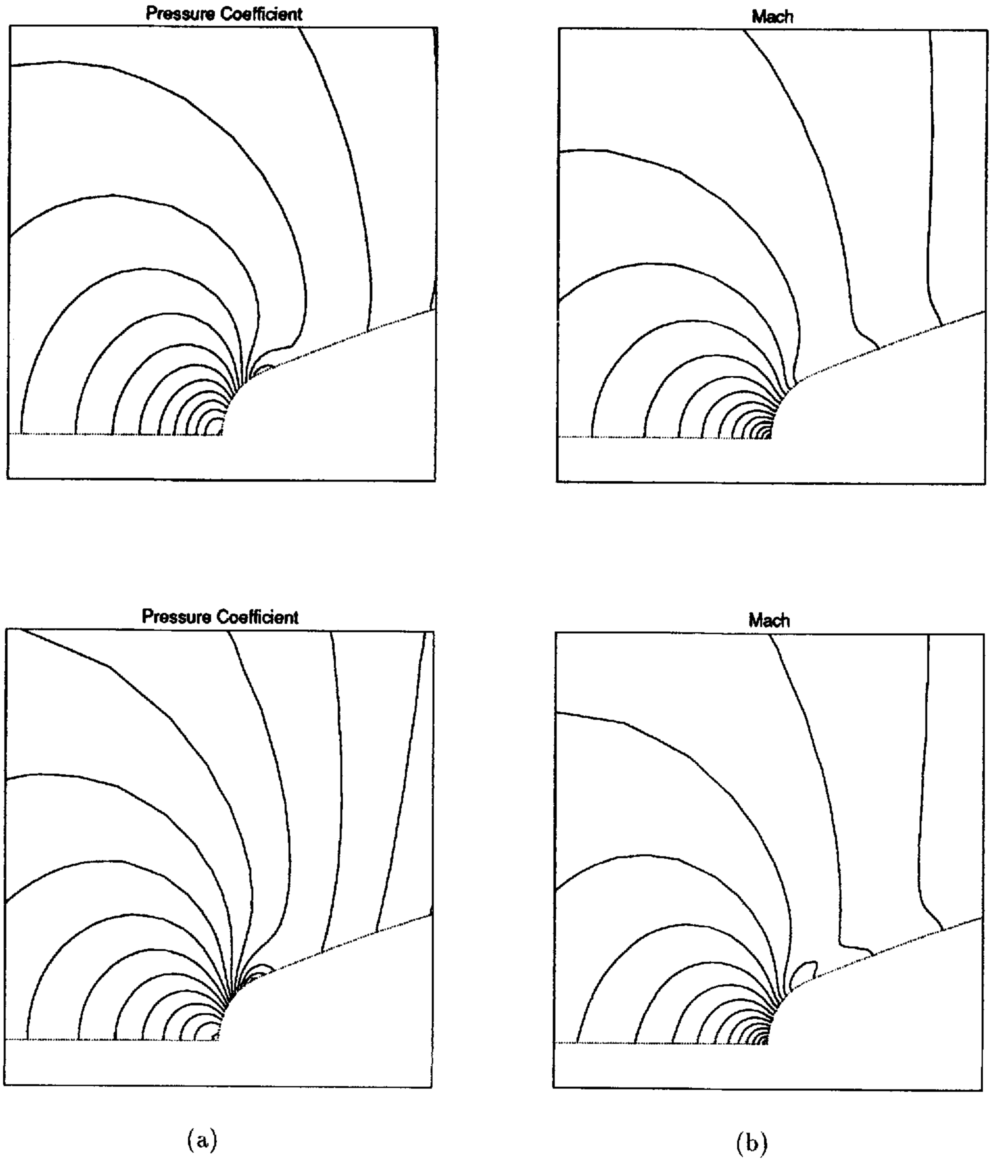


Figure 7. Contours of (a) pressure coefficient and (b) Mach number for subsonic flow past a blunt forebody, $M_\infty=0{\cdot}5$: (top) first-order solution; (bottom) after one defect correction

For the Euler equations, a conservative discretization is crucial for correct shock location.[16] The lower-order scheme used here is a generalization of (24), upwinded by van Leer flux vector splitting,[17] solved using a highly efficient multigrid procedure,[23] with alternating line Gauss–Seidel smoothing. The higher-order scheme is obtained by taking an area-weighted linear combination of cell residuals,[8] as in Section 3. Note that no artificial diffusion terms are used to lend stability to this discretization, which will consequently be liable to produce unphysical oscillatory solutions, especially in the neighbourhood of shocks. Furthermore, the presence of spurious error modes in this discretization leads to very slow convergence of an iterative solution procedure, including defect correction; hence, the desirability of the finite termination property. However, assuming the finite termination property to hold, the residuals of the higher-order scheme will not be zero after only a few iterations. It is, therefore, unclear precisely what discrete equations have been solved; this prohibits straightforward error analysis.

Note that in this case the discrete approximations yield non-linear systems of algebraic equations. As a consequence, the defect correct iteration (6) must be modified to take account of the non-linearity. Denoting the lower-order approximation by $N_1(\mathbf{U}_1)=0$ and the higher-order scheme by $N_2(\mathbf{U}_2)=0$, the defect correction iteration used here is

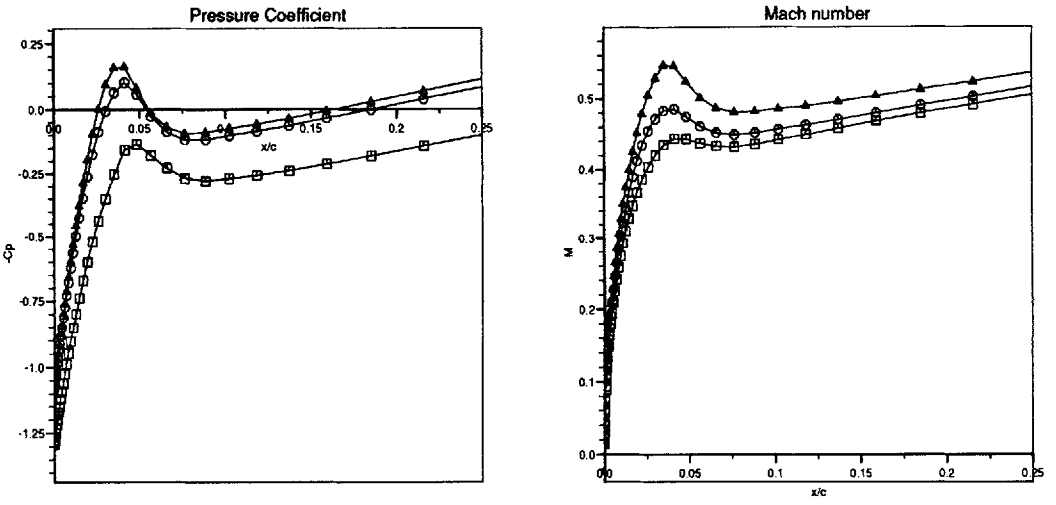
$$\begin{gathered} N_1(\mathbf{U}^{(0)})=0, \\ N_1(\mathbf{U}^{(k+1)})=N_1(\mathbf{U}^{(k)})-N_2(\mathbf{U}^{(k)}), \quad k=0,1,\ldots . \end{gathered} \tag{34}$$

This same non-linear correction procedure is used in the FAS multigrid algorithm of Brandt.[1] Note that defining

$$N_1(\mathbf{U})=L_1\mathbf{U}-\mathbf{F}_1, \qquad N_2(\mathbf{U})=L_2\mathbf{U}-\mathbf{F}_2, \tag{35}$$

where L_1, L_2 are the linear operators of Section 2, the iteration (34) reduces to (6).

Two test problems are considered here. The first is a subsonic flow past a blunt forebody with freestream Mach number $M_\infty=0{\cdot}5$. Figure 7 compares contour plots of the solution of the



(a) (b)

Figure 8. Pressure coefficient (a) and Mach number (b) around leading edge of the blunt forebody: □ first-order scheme; ⊕ after one defect correction; △ second-order vertex-centred scheme

Table II. Comparison of CPU times for defect correction

Method	CPU (s)
First-order scheme	106
One-defect correction	187
Nine-defect corrections	717
Second-order scheme	760

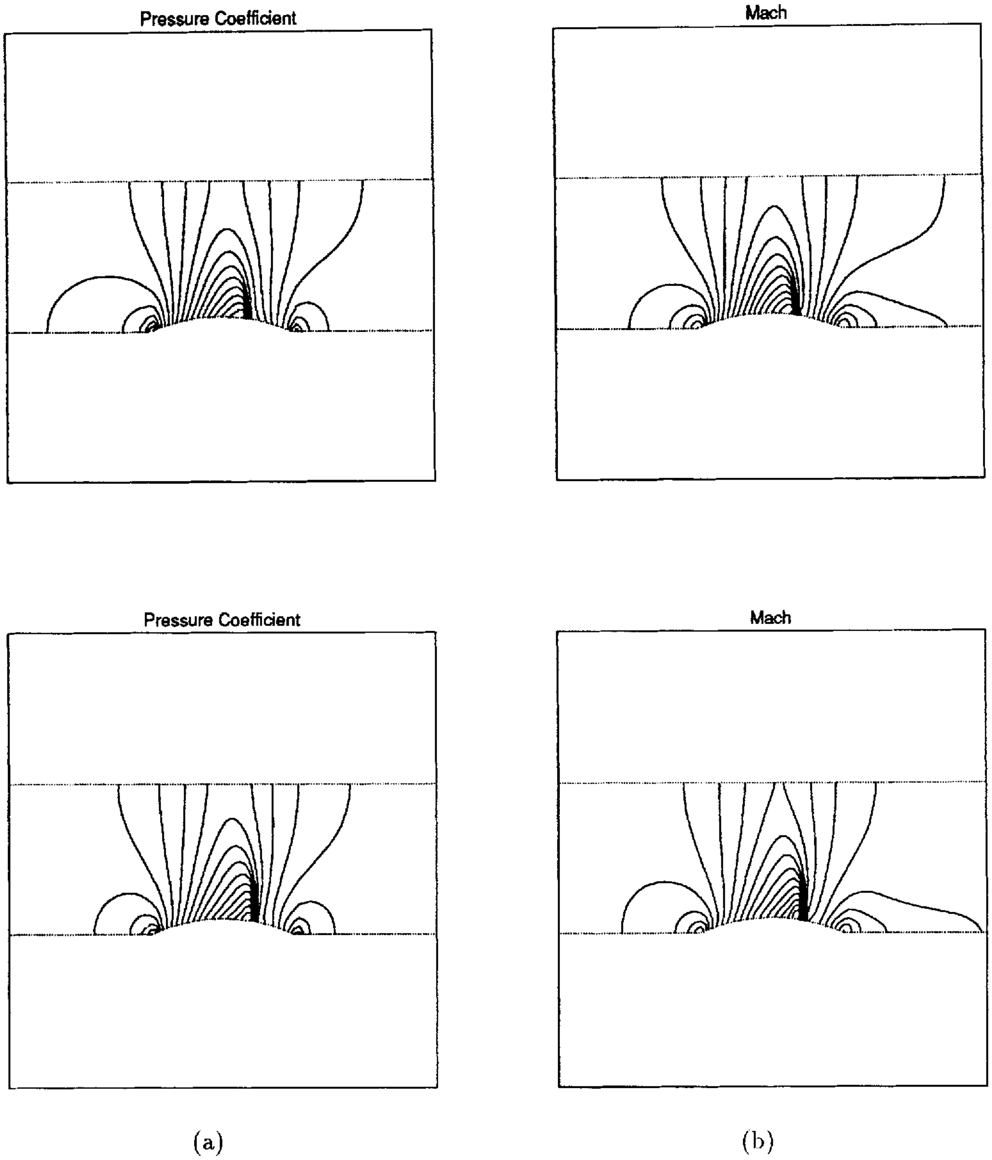


Figure 9. Contours of (a) pressure coefficient and (b) Mach number for transonic channel flow at $M_{\infty}=0{\cdot}675$; (top) first-order solution; (bottom) after one defect correction

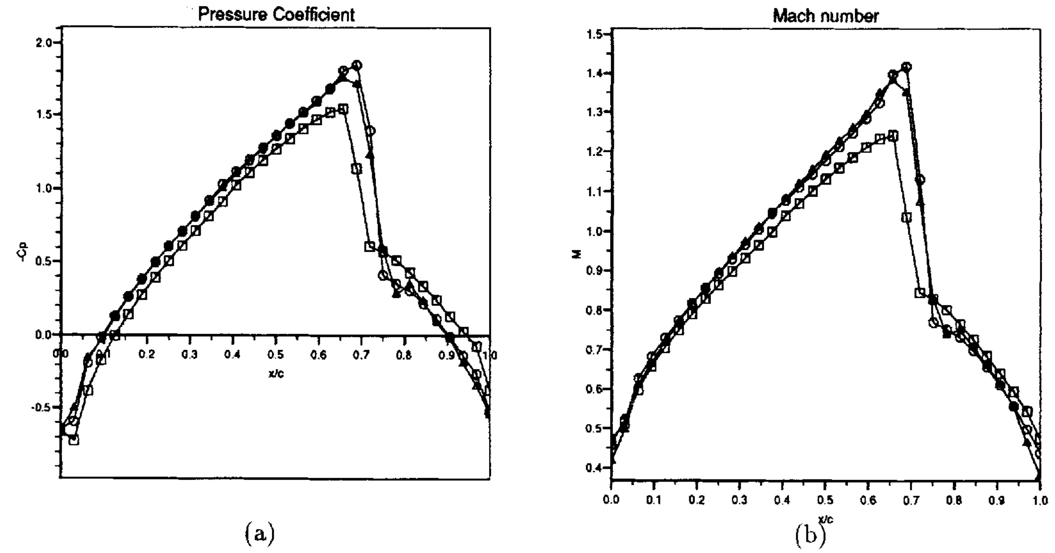


Figure 10. Pressure coefficient (a) and Mach number (b) for transonic channel flow: □ first-order scheme; ⊕ after one defect correction; △ second-order vertex-centred scheme

first-order scheme and after a single defect correction. Figure 8 shows a detailed comparison of the two solutions and a third independent solution, calculated on the same mesh, from a vertex-centred scheme,[3] which is designed to be second-order accurate. The defect correction iteration gives a clear improvement in the peaks of both Mach number and pressure coefficient around the leading edge of the forebody, which are characteristic of this geometry. However, the accuracy of the vertex-centred scheme is not matched and little further improvement was observed after subsequent iterations, as was the case in the linear-advection analogue (Figure 3). Table II indicates the cost in CPU time for these calculations and a clear saving is made by adopting the solution from a single defect correction iteration.

Figures 9 and 10 show the results for transonic channel flow at $M_{\infty} = 0{\cdot}675$, before and after the first defect correction iteration, and those from the vertex-centred scheme. This test problem appeared in Reference 20. The shock strength and resolution are significantly improved by defect correction, giving remarkable agreement with the vertex-centred method. However, oscillatory behaviour is beginning to develop, which becomes more apparent in subsequent iterations. This is due to the lack of any special treatment of the shock region in the higher-order operator. This agreement is not so surprising since, in the vicinity of shocks, the vertex-centred scheme is only first-order accurate, and, so, the defect correction iteration improves the accuracy of the lower-order operator by a small constant, as is the case in the linear-advection analogue (Figure 3).

6. CONCLUSIONS

The results for linear advection clearly demonstrate that second-order accuracy may be obtained in only one defect correction iteration on highly distorted meshes. Indeed, this property is obtained even in cases where the fixed point of the iteration is not itself second-order accurate. It has been found from the results reported here, and other experimental tests, that the finite termination property described above is crucially dependent on the consistency of the lower-order operator.

The finite termination property has been further exemplified by numerical results for Poisson's equation, which show that fourth-order accuracy can be achieved at the cost of only two inversions of a stable second-order operator.

For transonic Euler calculations, conservation is imperative for the correct location of shocks. Thus, the lower-order operator should be both conservative and consistent for the finite termination property to hold. The results indicate that improvements can be obtained after a single iteration despite the lack of consistency of the lower-order scheme, whose accuracy is derived from supraconvergence. However, judging from the results for linear advection, it seems unlikely that the finite termination property has been achieved in these Euler results.

The defect correction procedure adopted here was not found to be an efficient iterative technique for solving the higher-order problem to a given tolerance.

ACKNOWLEDGEMENTS

This work was supported by British Aerospace and the Science and Engineering Research Council. The authors express their thanks to the staff of the St. Giles Café for their hospitality.

REFERENCES

1. A. Brandt, 'Multi-level adaptive solutions to boundary value problems', *Math. Comput.*, **31,** 333–390 (1977).
2. P. I. Crumpton and G. J. Shaw, 'Cell vertex finite volume discretizations in three dimensions', *Int. j. numer. methods fluids*, **14,** 505–527 (1992).
3. P. I. Crumpton and G. J. Shaw, 'A vertex centred finite volume method with shock detection', *Oxford University computing Lab. Report no. 92/5* (1992).
4. J. D. Denton, 'An improved time marching method for turbo machinery calculations', in *Proc. IMA Conf. on Num. Meth. in Aero. Fluid Dynamics*, Academic Press, London, 1982, pp. 19–35.
5. E. Dick, 'A flux-difference splitting method for steady Euler equations', *J. Comput. Phys.*, **76,** 19–32 (1988).
6. L. Fox, *Numerical Solution of Ordinary and Partial Differential Equations*, Pergamon, New York, 1962.
7. W. Hackbusch, *Multi-grid Methods and Applications*, Springer, Berlin, 1985.
8. M. G. Hall, 'Cell-vertex multigrid schemes for solution of the Euler equations', in K. W. Morton and M. J. Baines (eds.), *Proc. Conf. on Numerical Methods for Fluid Dynamics*, University of Reading, Oxford University Press, Oxford, 1985, pp. 303–345.
9. M. G. Hall, 'A vertex centroid scheme for improved finite volume solution of the Navier–Stokes equations', *AIAA Paper 91-1540*, 1991.
10. P. W. Hemker, 'A note on defect correction processes with an approximate inverse of deficient rank', *J. Comput. Appl. Math.*, **8,** 137–140 (1982).
11. P. W. Hemker, 'Mixed defect correction iteration for the accurate solution of a singular perturbation problem', *Computing* (Suppl.), **5,** 123–145 (1984).
12. P. W. Hemker and B. Koren, 'Defect correction and non-linear multi-grid for steady Euler equations', *Report NM-R9007*, Centre for Mathematics and Computer Science, Amsterdam, 1990.
13. P. W. Hemker and S. P. Spekreijse, 'Multiple grid and Osher's scheme for the efficient solution of the steady Euler equations', *Appl. Numer. Math.*, **2,** 475–493 (1986).
14. M. R. Hestenes and E. Stiefel, 'Methods of conjugate gradients for solving linear systems', *J. Res. Nat. Bureau Standards*, **49,** 409–436 (1952).
15. B. Koren, 'Defect correction and multigrid for an efficient and accurate computation of airfoil flows', *J. Comput. Phys.*, **77,** 183–206 (1988).
16. P. Lax and B. Wendroff, 'Systems of conservation laws', *Commun. Pure Appl. Math.*, **XIII,** 217–237 (1960).
17. B. van Leer, 'Flux vector splitting for the Euler equations', *Lecture Notes in Physics* **170,** 507–512 (1982).
18. J. A. Meijerink and H. A. van der Vorst, 'An iterative solution method for linear systems of which the coefficient matrix is a symmetric M-matrix', *Math. Comput.*, **31,** 148–162 (1977).
19. K. W. Morton and M. F. Paisley, 'A finite volume scheme with shock fitting for the steady Euler equations', *J. Comput. Phys.*, **80,** 168–203 (1989).
20. R. H. Ni, 'A multiple grid method for solving the Euler equations', *AIAA J.*, **20,** 1565–1571 (1982).
21. V. Pereyra, 'On improving an approximate solution of a functional equation by deferred corrections', *Numer. Math.*, **8,** 376–391 (1966).
22. V. Pereyra, 'Highly accurate numerical solution of casilinear elliptic boundary value problems in *n* dimensions', *Math. Comput.*, **24,** (no. 112) 771–783 (1970).

23. G. Shaw and P. Wesseling, 'Multi-grid solution of the compressible Navier–Stokes equations on a vector computer, in F. G. Zhuang and Y. L. Zhu (eds.), *Proc. of the Tenth International Conference on Numerical Methods in Fluid Dynamics*, Lecture Notes in Physics, Vol. 264, 1986, p. 567.
24. E. Süli, 'The accuracy of finite volume methods on distorted partitions', in J. R. Whiteman (ed.), *The Mathematics of Finite Elements and Applications VII*, Academic Press, New York, 1991, pp. 253–260.
25. P. Wesseling, *An Introduction to Multigrid Methods*, Wiley, Chichester, 1992.