Technical Notes

TECHNICAL NOTES are short manuscripts describing new developments or important results of a preliminary nature. These Notes cannot exceed 6 manuscript pages and 3 figures; a page of text may be substituted for a figure and vice versa. After informal review by the editors, they may be published within a few months of the date of receipt. Style requirements are the same as for regular contributions (see inside back cover).

Approximate Factorization with Source Terms

T. I.-P. Shih*
Carnegie Mellon University, Pittsburgh, Pennsylvania 15213
and
W. J. Chyu†
NASA Ames Research Center, Moffett Field, California 94035

Introduction

SOME of the most efficient numerical methods available for generating solutions to compressible flow problems utilize a technique known as approximate factorization in their construction. Examples of such methods include alternating direction implicit (ADI)[1,2] and lower-upper (LU)[3] methods. Previously, these methods were used primarily to study nonreacting flows in which the turbulence can be modeled by algebraic turbulence models such as the Baldwin-Lomax model. More recently, these methods have been used to study chemically reacting flows[4] and flows in which the turbulence needs to be modeled by more sophisticated approaches, such as the two-equation k-ϵ model.[5,6] For these latter problems, the governing equations contain source terms (i.e., nondifferentiated terms), whereas the previous problems do not.

When there are source terms in the governing equations, it is not clear in which factor the source terms should be placed after approximate factorization. Several different methods can be used to handle this situation. The objective of this Note is to examine these different methodologies in terms of approximate factorization error and to make a recommendation as to which method is the best.

Methodologies for Handling Source Terms

To illustrate the different methods for handling source terms in conjunction with approximate factorization, consider the Euler equation with an arbitrary source term:

$$\frac{\partial u}{\partial t} + \frac{\partial E}{\partial x} + \frac{\partial F}{\partial y} + \frac{\partial G}{\partial z} + S = 0 \tag{1}$$

Received Sept. 21, 1990; accepted for publication Oct. 12, 1990.

*Associate Professor, Department of Mechanical Engineering. Member AIAA.
†Research Scientist, Applied Aerodynamics Branch. Member AIAA.

In Eq. (1), u is a vector of conserved variables; E, F, and G represent fluxes of mass, momentum, and energy in the x, y, and z directions, respectively; and $S = S(u)$ is the source term.

Suppose that Eq. (1) is supplied with appropriate boundary and initial conditions, and that the domain is discretized so that the time-step size Δt and grid spacings Δx, Δy, and Δz are all constants. By using the Euler implicit time-differencing formula, Eq. (1) can be written as

$$[I + \Delta t(D_x A + D_y B + D_z C + H)]\Delta u_{ijk}^{n+1} = RHS \tag{2}$$

where $u_{ijk}^n = u(n\Delta t, i\Delta x, j\Delta y, k\Delta z)$; n, i, j, and k are non-negative integers; $\Delta u^{n+1} = u^{n+1} - u^n$; $D_x = \partial/\partial x$; $D_y = \partial/\partial y$; $D_z = \partial/\partial z$; I is a unit matrix; $A = (\partial E/\partial u)^n$; $B = (\partial F/\partial u)^n$; $C = (\partial G/\partial u)^n$; $H = (\partial S/\partial u)^n$; and

$$RHS = -\Delta t\left(\frac{\partial E}{\partial x} + \frac{\partial F}{\partial y} + \frac{\partial G}{\partial z} + S\right)_{ijk}^n \tag{3}$$

Equation (2) can be approximately factored in several different ways. Here, we shall illustrate the method for handling source terms by factoring according to eigenvalues as is done in the LU method. By using the LU method (i.e., flux-vector splitting followed by upwind differencing), Eq. (2) becomes

$$[I + \Delta t(L + U + H)]\Delta u_{ijk}^{n+1} = RHS \tag{4}$$

where

$$L = \delta_x^b A^+ + \delta_y^b B^+ + \delta_z^b C^+ \tag{5}$$

$$U = \delta_x^f A^- + \delta_y^f B^- + \delta_z^f C^- \tag{6}$$

In Eqs. (5) and (6), δ^b and δ^f are backward and forward difference operators, respectively; $W = W^+ + W^-$, where W is A, B, or C; and $W^\pm = 0.5[W \pm \rho(W)]$, where $\rho(W)$ is the spectral radius of W. Other ways of defining $W^\pm$ are possible; see Refs. 7 and 8.

If the source term H in Eq. (4) is zero, then LU factoring of Eq. (4) readily gives

$$(I + \Delta t L)(I + \Delta t U)\Delta u_{ijk}^{n+1} = RHS \tag{7}$$

Now, the question is if the source term H is nonzero, then where should it be placed? Several alternatives are possible; three of them are as follows:

Method 1:

$$[I + \Delta t(L + H)][I + \Delta t U]\Delta u_{ijk}^{n+1} = RHS \tag{8}$$

Method 2:

$$(I + \Delta tH)(I + \Delta tL)(I + \Delta tU)\Delta u_{ijk}^{n+1} = RHS \tag{9}$$

Method 3:

$$(N + \Delta tL)N^{-1}(N + \Delta tU)\Delta u_{ijk}^{n+1} = RHS$$

$$N = I + \Delta tH \tag{10}$$

Note that method 3 is implemented in the following sequence: $(N + \Delta tL)\Delta u_{ijk}^{*} = RHS$ and $(N + \Delta tL)\Delta u_{ijk}^{n+1} = N\,\Delta u_{ijk}^{*}$.

Most investigators use method 1 given by Eq. (8)[4-6] because of either simplicity or computational efficiency. If the source term H is stiff, then using method 2 given by Eq. (9) may be even more efficient. This is because the factor associated with the source term can be called many times (e.g., m times by using a time step of $\Delta t/m$) before a single call is made to the other two factors. This approach would be similar to the time-split method of MacCormack.[9] The ideas utilized in method 3 given by Eq. (10) has been used to construct diagonalized LU schemes[4,5]; however, these same investigators did not use the concepts in method 3 to treat source terms, presumably on the grounds of efficiency.

Here, we note that efficiency is not the only criterion for evaluating numerical methods. Accuracy and robustness are others. In the next section, we examine the errors introduced by these three methods for handling source terms.

Error Analysis

The three methods given by Eqs. (8-10) for handling source terms produce different approximate factorization errors. This error needs to be small if transient solutions are of interest. This error also affects the stability and, hence, the robustness of the algorithm. In the following, we compare the three methods for handling source terms on the basis of this error.

The approximate factorization error in method 1 given by Eq. (8) is

$$E_{af1} = \text{Eq. (8)} - \text{Eq. (4)} = \Delta t^2(LU + HU)\Delta u_{ijk}^{n+1}$$

By taking the second norm of this equation, we can readily show that the relative error due to the approximate factorization is bounded by

$$\epsilon_{af1} = \|E_{af1}\|/\|\Delta u_{ijk}^{n+1}\| \le \Delta t^2(\|LU\| + \|HU\|) \tag{11}$$

Similarly, bounds on the relative errors created by methods 2 and 3 given by Eqs. (9) and (10) are given by

$$\epsilon_{af2} = \|E_{af2}\|/\|\Delta u_{ijk}^{n+1}\| \le \Delta t^2[\|LU\| + \|H(L + U)\|] + \Delta t^3\|LUH\| \tag{12}$$

$$\epsilon_{af3} = \|E_{af3}\|/\|\Delta u_{ijk}^{n+1}\| \le \Delta t^2\|N^{-1}\|^*\|LU\| \tag{13}$$

By comparing Eq. (11) with Eq. (12), it can be seen that $\epsilon_{af1} \le \epsilon_{af2}$. Thus, treating source terms according to method 1 given by Eq. (8) produces less approximate factorization errors than those produced by method 2 given by Eq. (9). However, for both methods 1 and 2, Eqs. (11) and (12) indicate that the upper bounds on approximate factorization errors are directly proportional to $\|H\|$. Since $\|H\|$ can be quite high, this implies that the approximate factorization errors produced can be quite significant when either one of these methods is used.

In order to compare methods 1 and 2 with method 3, we assume that the source term H is positive definite. When this is the case, $\|N^{-1}\| = \rho(N^{-1}) = \rho[I + \Delta tH)^{-1}]$, but $\rho(N^{-1}) = 1/\{\text{minimum}|\lambda_N|\} = 1/\{\text{minimum}|1 + \Delta t\lambda_H|\}$, where λ_N and λ_H are the eigenvalues of N and H, respectively. If H is positive definite, then the smallest eigenvalue of H is bounded by zero so that

$$\|N^{-1}\| \le 1 \tag{14}$$

With Eq. (14), Eq. (13) becomes

$$\epsilon_{af3} = \|E_{af3}\|/\|\Delta u_{ijk}^{n+1}\| \le \Delta t^2\|LU\| \tag{15}$$

With Eq. (15) derived, we can compare the approximate factorization error produced by method 3 with those produced by methods 1 and 2. By comparing Eq. (15) with Eqs. (11) and (12), it can be seen that, of the three methods presented for treating source terms, method 3 produces the least amount of approximate factorization error. In addition, it is important to note that Eq. (13) and (14) indicates that, for method 3, the source term reduces approximate factorization error instead of increasing it as do the other methods.

This comparison assumed H to be positive definite. If H is not positive definite, then

$$\|N^{-1}\| = \|(I + \Delta tH)^{-1}\| = \|I - \Delta tH + (\Delta tH)^2 - \cdots\| \approx \|I - \Delta tH\|$$

Substitution of this equation into Eq. (13) gives

$$\epsilon_{af3} \le \Delta t^2\|I - \Delta tH\|^*\|LU\| \le \Delta t^2\|LU\| + \Delta t^3\|H\|^*\|LU\| \tag{16}$$

Comparing Eq. (16) with Eqs. (11) and (12) again shows that method 3 produces less approximate factorization errors than those produced by methods 1 and 2.

Summary

This Note discussed and evaluated three methods for treating source terms in schemes which utilize approximate factorization. It was noted that methods 1 and 2 can lead to more efficient algorithms. This is especially true if factors that do not contain source terms can be diagonalized. However, in terms of approximate factorization error, method 3 was shown to be the best because it produces the least amount of this error when compared to the other methods presented. Accordingly, method 3 may be preferred when the norms of source terms are large and transient solutions are of interest.

References

[1] Beam, R. M., and Warming, R. F., "An Implicit Factored Scheme for the Compressible Navier-Stokes Equations," *AIAA Journal,* Vol. 16, No. 4, 1978, pp. 393–402.

[2] Briley, W. R., and McDonald, H., "On the Structure and Use of Linearized Block Implicit Schemes," *Journal of Computational Physics,* Vol. 34, No. 1, 1980, pp. 54–73.

[3] Jameson, A., and Yoon, S., "Lower-Upper Implicit Schemes with Multiple Grids for the Euler Equations," *AIAA Journal,* Vol. 26, No. 9, 1988, pp. 1025–1026.

[4] Shuen, J. S., and Yoon, S., "Numerical Study of Chemically Reacting Flows Using a Lower-Upper Symmetric Successive Overrelaxation Scheme," *AIAA Journal,* Vol. 27, No. 12, 1989, pp. 1752–1760.

[5] Yokoda, J., "A Diagonally Inverted LU Implicit Multigrid Scheme for the 3-D Navier-Stokes Equations and a Two Equation Model of Turbulence," AIAA Paper 89-0467, Jan. 1989.

[6] Nichols, R. H., "A Two-Equation Model for Compressible Flows," AIAA Paper 90-0494, Jan. 1990.

[7] Steger, J. L., and Warming, R. F., "Flux Vector Splitting of the Inviscid Gasdynamics Equations with Application to Finite-Difference Methods," *Journal of Computational Physics,* Vol. 40, No. 2, 1981, pp. 263–293.

[8] Van Leer, B., "Flux-Vector Splitting for the Euler Equations," *Lecture Notes in Physics,* Vol. 170, 1982, pp. 507–512.

[9] MacCormack, R. W., "Numerical Solution of the Interaction of a Shock Wave with a Laminar Boundary Layer," *Lecture Notes in Physics,* Vol. 8, 1971, pp. 151–163.