

A METHOD OF LINES APPROACH TO THE NUMERICAL SOLUTION
 OF CONSERVATION LAWS

by

James M. Hyman
 Theoretical Division
 Los Alamos Scientific Laboratory
 Los Alamos, NM 87545

I. SUMMARY

The numerical solution of conservation laws is a highly complicated and problem-dependent process. The solution may contain dynamic interactions between shock waves, rarefaction waves and contact discontinuities. A method developed for a particular test problem may or may not work for another with stronger (or weaker) shocks and contact discontinuities. Methods which work well in one space dimension may or may not be easily extended to two or three dimensions.

The analysis in this report is based on solving systems of conservation laws in one space dimension using techniques that generalize to higher spatial dimensions. The approach is based on the philosophy of the method of lines (MOL).⁴ In the method of lines, the space and time discretizations of a partial differential equation (PDE) are decoupled and analyzed independently. First a method is selected to discretize the differential equation in space and incorporate the boundary conditions. The spectrum of this discrete operator is then used as a guide to choose an appropriate method to integrate the equations through time.

The dissipative effects of a numerical method are crucial to constructing reliable methods for conservation laws. This is particularly true when the solution is discontinuous as in a shock wave or contact discontinuity.

Choosing an accurate method to accomplish each of these tasks, space and time discretization and incorporating artificial dissipation in the numerical solution, determines the success of the calculation. In Sections III through VI we will consider each choice independently and combine them in Section VII to develop a class of particularly good explicit finite difference methods. In Section VIII, we present some numerical examples to illustrate the properties of the different methods and analyze their results in a summary, Section IX.

Before developing a good method for any system of PDEs one must have a basic understanding of the equations being solved. We now give a brief review of the general theory of conservation laws.

II. CONSERVATION LAWS

General Theory

A vector quantity W of length N is conserved as it evolves under the flow of a conservation law if the amount of each substance W_j , $j=1,2,\dots,N$, contained in any fixed volume V is due entirely to the flux $F_j(W)$ across the boundary ∂V of V . These conservation laws can be expressed in integral form as

$$\frac{d}{dt} \int_V W_j dx = - \int_{\partial V} F_j(W) \cdot n dS \quad (2.1)$$

where n denotes the outward normal to the boundary.

Moving the time derivative under the integral sign and applying the divergence theorem Eq. (2.1) can be rewritten as

$$\int_V \left[\frac{\partial}{\partial t} W_j + \text{div } F_j(W) \right] dx = 0. \quad (2.2)$$

By letting the volume V shrink to a point we obtain the system of PDEs

$$\frac{\partial}{\partial t} W_j + \text{div } F_j(W) = 0, \quad j = 1, 2, \dots, N \quad (2.3)$$

at every point where W and F are differentiable.

In one space dimension Eq. (2.3) can be written in vector form as

$$W_t + F(W)_x = 0 \quad (2.4)$$

or

$$W_t + G(W)W_x = 0 \quad (2.5)$$

where G is the N by N matrix gradient of F with respect to W .

Equation (2.5) is a first-order quasi-linear system of PDEs. This system is hyperbolic and well-posed if the eigenvalues of G are distinct and real. These eigenvalues, called the characteristic velocities, are the local signal speeds at which sharp disturbances propagate.

It is well known that a system of nonlinear conservation laws may fail to have a continuous solution after a finite time. Since conservation laws are derived from integral relations (2.1) these generalized solutions may still be admitted as long as they are measurable and bounded. There are instances, as in the Euler equations of gas dynamics, that there may be many different generalized solutions satisfying Eq. (2.1) with the same initial data. Within this set only one of these solutions has any physical significance. An important consideration in construction a numerical method is to build a mechanism, such as artificial dissipation, into the difference scheme that will automatically choose the physically relevant solution.

The physically relevant solution must satisfy the differential equation (2.4) in smooth regions and fulfill two additional constitutive relations across any discontinuities in the flow. The first constitutive relation, called the Rankine-Hugoniot jump conditions, states that the discontinuity must propagate with speed s satisfying the jump conditions

$$s[W_j] = [F_j(W)], \quad j = 1, 2, \dots, N.$$

Here $[]$ denotes the jump of the quantity in brackets across the discontinuity. These jump conditions are satisfied by the numerical solution if the equations are solved in divergence form (Eq. (2.4)) and the flux function is differenced with centered differences.

The second constitutive relation, called the entropy condition, states that entropy must increase across the shock discontinuity. This condition is satisfied by the limiting solution of the viscous equations as the viscosity is decreased to zero. Numerical methods for solving discontinuous solutions of Eq. (2.4) have a small artificial viscosity that helps select the physically relevant solution.

The methods developed in this paper will be described in terms of the Euler equations of gas dynamics. However, most of the techniques and results are equally valid for other hyperbolic systems. For a further discussion of the mathematical theory of general hyperbolic systems of conservation laws we refer the reader to Lax.^{6,7}

Euler Equations

The one-dimensional Eulerian equations of gas dynamics can be written in divergence form as

$$\dot{W}_t + F(W)_x = 0, \quad (2.6)$$

$$W = \begin{pmatrix} \rho \\ m \\ E \end{pmatrix}, F(W) = uW + \begin{pmatrix} 0 \\ p \\ pu \end{pmatrix},$$

where ρ = mass density, u = velocity, $m = \rho u$ = momentum, $E = \rho(I + \frac{1}{2}u^2)$ = total energy per unit volume, I = internal energy and p = pressure.

Equation (2.6) is hyperbolic if pressure is an increasing function of density at constant entropy. This is the case if we assume the equation of state to be that of a polytropic gas, i.e. $p = (\gamma - 1)\rho\epsilon$. The parameter γ is a constant greater than one and equal to the ratio of the specific heats of the gas. For this equation of state we have

$$\frac{dp}{d\rho} = \frac{\gamma p}{\rho} = c^2 > 0$$

at constant entropy. The quantity c is called the local sound speed of the gas and is related to the characteristic velocities u , $u + c$ and $u - c$ of Eq. (2.6).

III. SPACE DISCRETIZATION

To solve Eq. (2.6) numerically we must first choose an appropriate approximation of the spatial derivatives. The guiding principle in choosing a spatial approximation is that the discrete model should retain as closely as possible all the crucial properties of the original differential equation. Equations (2.6) reflect principles of conservation of mass, momentum, and energy which are the basis for the mathematical theory of fluid dynamics. These properties should be preserved in the difference formulation. This is best accomplished if the equations are integrated and differenced in divergence form using centered finite differences.

Phase and Damping Errors

The derivative of the flux function F determines the phase velocities of the solution and hence the shock speeds. Therefore, the errors in approximating F and its derivative should be made as small as possible. These errors can be divided into two classes; phase or dispersion errors and damping or dissipation errors.

Centered finite difference approximations of F_x preserve many of the conservation properties of Eq. (2.6). The second-, fourth- and sixth-order finite difference approximations to F_x can be written as

$$F_x = (F_{i+1} - F_{i-1}) / (2\Delta x) + O(\Delta x^2), \quad (3.1a)$$

$$F_x = (-F_{i+2} + 8F_{i+1} - 8F_{i-1} + F_{i-2}) / (12\Delta x) + O(\Delta x^4), \quad (3.1b)$$

and

$$F_x = (F_{i+3} - 9F_{i+2} + 45F_{i+1} - 45F_{i-1} + 9F_{i-2} - F_{i-3}) / (60\Delta x) + O(\Delta x^6). \quad (3.1c)$$

Here F_i refers to the value of $F(W(x_i, t))$ and

$\Delta x = x_{i+1} - x_i$ is the mesh spacing.

The errors in a finite difference approximation can be computed for numerical approximations of traveling wave solutions to Eq. (2.6). Consider the solution to Eq. (2.6) with periodic boundary conditions on the unit interval and constant initial pressure and velocity v . The solution is a traveling wave and satisfies the simple linear hyperbolic convective equation

$$\rho_t + v\rho_x = 0, \quad (3.2)$$

with the initial conditions $\rho(x, 0) = g(x)$, and the solution $\rho(x, t) = g(x - vt)$.

When the initial conditions consist of a single frequency, $g(x) = a \sin(kx) + b \cos(kx)$, then the phase error introduced by the finite difference approximation of ρ_x will be the same using second-, fourth- or sixth-order differences if the number of mesh points in the calculations satisfy

$$M_2 \cong 0.36 M_4^2 \cong 0.12 M_6^3. \quad (3.3)$$

Here M_j is the number of mesh points when using j -th order finite differences.⁴

The table below compares the number of points per wave length necessary to obtain a given phase error e in the solution to (3.2) at time t using second-, fourth- and sixth-order centered differences.

2nd order M_2	4th order M_4	6th order M_6	Accuracy $e/(vkt)$
4	4	3	2.6
8	5	4	0.65
16	7	5	0.16
32	10	7	0.04
64	14	8	0.01
128	19	10	0.0025
256	27	13	0.0006

Table I. Points per wavelength for second-, fourth- and sixth-order differences to have the same accuracy.

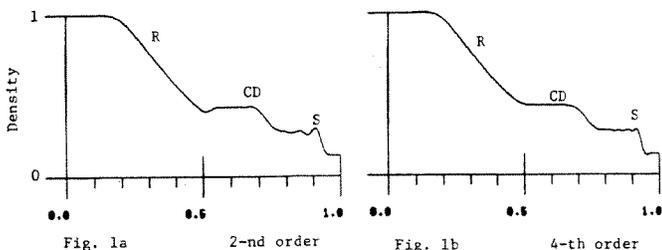
In a calculation where the solution contains many different frequencies, the high modes (2-5 points per wavelength) are approximated equally poorly with all the methods. The middle modes (6-16 points per wavelength) are computed much more accurately with the fourth and sixth-order differences than with the second order method. The sixth-order differences are more accurate for the lower modes than either second- or fourth-order differences.

The relationship of the accuracies of the different methods compared to the number of points per wavelength is even more impressive in higher dimensions. In two space dimensions the numbers in Table I should be squared; in three dimensions cubed.

The next step is to determine if this linear analysis is applicable to nonlinear equations with shocks and contact discontinuities. Figure 1 displays the solution to an initial value problem for Eq. (2.6). The problem was solved twice, the only difference being that the spatial differences were changed from second- to fourth-order.

The example is the shock tube Riemann problem described completely in Section VIII. The shock and contact discontinuity in Figs. 1a and 1b are much sharper when the higher order differences are used. Also, the post shock oscillations are reduced even though the same artificial dissipation was used in each

calculation. The high order differences are able to resolve the discontinuities better and require less artificial dissipation to eliminate post shock oscillations.



IV. BOUNDARY CONDITIONS

Before calculating the solution to any differential equation one should determine if the boundary conditions are consistent with a well posed problem. A numerical method cannot be expected to generate reasonable results for a problem which does not have a well defined reasonable solution. The importance of proper boundary conditions cannot be overstressed, the boundary conditions exert one of the strongest influence on the behavior of the solution. Also, the errors introduced into the calculation from improper boundary conditions persist even as the mesh spacing tends to zero.

A common error in prescribing boundary conditions for conservation laws is to over or under specify the number of boundary conditions. Overspecification usually results in nonsmooth solutions with mesh oscillations near the boundary. Underspecification does not insure the solution is unique and the numerical solution may tend to wander around in steady state calculations. In either case the results of the calculation are not accurate and one should be skeptical of even the qualitative behavior of the solution.

Once it has been determined that the differential equations and boundary conditions are well posed, special care must be taken to preserve this in the difference approximation. This can best be done by enforcing constituent relationships on the difference equations such that the discrete equations are consistent with as many relationships that can be derived from the boundary conditions and differential equation as possible.

Fictitious Points

To define the solution at fictitious points, the boundary conditions are differentiated with respect to time and combined with Eq. (2.6) to obtain differential constraints for the extrapolation formulas. This technique will be shown for reflecting boundary conditions to the Euler equations.

The reflecting boundary conditions for a thermally insulated wall for Eq. (2.6) at $x = x_0$ are

$$u(x_0, t) = 0, \quad I_x(x_0, t) = 0. \quad (4.1)$$

The thermally insulated boundary condition, $I_x = 0$, is obtained from the limit of the viscous dissipative equations as the viscosity and heat dissipation tend to zero. This condition is necessary to prevent a boundary layer in the difference approximation of inviscid calculations due to the presence of artificial dissipation.

To incorporate these boundary conditions into our numerical solution when using fourth-order centered differences we will introduce two fictitious points at $x_{-1} = x_0 - \Delta x$ and $x_{-2} = x_0 - 2\Delta x$ outside the region of integration. At these points we need an approximation to ρ , ρu , and E to preferably fourth-

order.

Combining Eqs. (2.6), and (4.1) at $x = x_0$ we have

$$0 = -(\rho u)_t = (\rho u^2 + p)_x = p_x = (\gamma - 1)I\rho_x,$$

$$E_x = [\rho(I + \frac{1}{2}u^2)]_x = 0,$$

and

$$(\rho u)_{xx} = -(\rho_t)_x = -(\rho_x)_t = 0.$$

Since these equations are valid for all time and $\gamma \neq 1$ we have

$$\rho_x = m = m_{xx} = E_x = 0 \quad (4.2)$$

as auxiliary boundary conditions at $x = x_0$ consistent with the original problem. The nonphysical solution at the fictitious points outside the region of integration needs to be chosen such that a finite difference approximation of Eqs. (4.2) is satisfied at the boundary.

When we replace the derivatives in these auxiliary boundary conditions by the standard centered finite differences we see that Eqs. (4.2) are satisfied if and only if

$$\rho_{-i} = \rho_i, \quad m_{-i} = -m_i, \quad E_{-i} = E_i \quad (4.3)$$

for $i = 1$ or 2 .

No Fictitious Points

There is not always a simple extrapolation formula such as Eq. (4.3) to extend the solution to the fictitious points. For these problems it is often better to use uncentered differences near the boundary. This method will be described for the linear hyperbolic system of M equations

$$W_t = H(x)W_x \quad (4.4)$$

with the boundary conditions

$$SW_0 = b(t), \quad x = x_0. \quad (4.5)$$

Difficulties arise in defining the solution at the boundary when $0 < \text{Rank}(S) < \text{Rank}(H) = M$ and there does not exist a unique solution W_0 of (4.5). If $\text{Rank}(S) = 0$ then all the characteristics are outgoing and using either uncentered differences at the points near the boundary or straight forward extrapolation to the fictitious points gives accurate results. When $\text{Rank}(S) = M$ then all the characteristics are entering the boundary and all the components of the solution can be solved for on the boundary. Uncentered spatial differences can then be used at the points near the boundary will result in an accurate approximation of the boundary conditions.

When $\text{Rank}(S)$ is greater than zero but less than M then by differentiating Eq. (4.5) with respect to time and replacing W_t from Eq. (4.4) we have

$$SH(x)W_x = b'(t), \quad x = x_0. \quad (4.6)$$

Approximating W_x by second-order one-sided differences gives us

$$SH_0 W_0 = [SH_0(4W_1 - W_2) - 2\Delta x b'(t)]/3 \quad (4.7)$$

where $H_0 = H(x_0)$. Equation (4.7) gives us additional information about the boundary conditions that is consistent with both the original boundary conditions (4.5) and the differential Eq. (4.4). If we still do not have enough boundary conditions to solve for W_0 uniquely then we can continue by differentiating (4.6) with respect to time and using Eq. (4.4) again.

It is often the case that H_0 is nonlinear and the above procedure must be iterated. Usually one or two

iterations are sufficient for a stable accurate boundary approximation.

Once W_0 has been found we can use uncentered finite differences to approximate the spatial derivatives at the mesh point nearest the boundary or we can extrapolate the solution to fictitious points outside the region of integration. This extrapolation can be done by replacing the derivatives in Eqs. (4.6) with second-order centered differences and solving for W_{-1} .

Imbedded Regions

There are many initial boundary value problems where it is essential to introduce artificial boundaries to reduce the computing time and storage of a calculation. These problems are usually posed in a domain much larger than the subregion where the solution is of interest. The subregion is blocked off and imbedded in the original problem by creating artificial boundaries. The boundary conditions at the artificial boundary are chosen such that the solution on the full domain would automatically satisfy these internal boundary conditions if the full problem were solved. The goal, of course, is to approximate the original problem as closely as possible on the reduced domains.

Consider the initial boundary value problem for the Euler equations on the half line $[0, \infty)$, with reflecting boundary conditions at $x = 0$. Suppose we are interested only in the behavior of the solution in the interval $[0, 1]$ and wish to restrict the domain of our computation to a neighborhood of this region. First we map the interval $[0, \infty)$ into $[0, b]$ with a map such as

$$y = \begin{cases} x & 0 \leq x \leq 1 \\ b + (1-b)/x & 1 < x < \infty \end{cases} \quad (4.8)$$

In this new coordinate system Eq. (2.6) transforms to

$$W_t + s(y) F_y = 0, \quad y \in [0, b]$$

where

$$s(y) = \begin{cases} 1 & 0 \leq y \leq 1 \\ (b-y)/(b-1)^2 & 1 < y < b \end{cases} \quad (4.9)$$

The solution to (4.9) is identical to the solution of our original problem. Therefore, the transformed system has the correct number of signals entering and leaving through the artificial break point at $x = 1$.

In this transformed system a wave slows down in the region $(1, b)$ and approaches zero speed as x nears b . This causes a wave train to squeeze up, with the lower frequencies being pushed into higher ones as in Fig. 2a. These high frequencies cannot be computed accurately and it is best to add some dissipation to damp them out as they approach the transformed boundary b . This damping should be chosen such that the signals propagating into the region of interest $[0, 1]$ depend in some sense on an average of the solution outside this region, i.e. $(1, b)$. A possible form for the dissipation is

$$W_t + s(y) F_y = (\Delta y d(y) W_y)_y \quad (4.10)$$

where

$$d(y) = \begin{cases} 0 & 0 \leq y \leq 1 \\ \delta[(y-1)/(b-1)]^2 & 1 < y \leq b \end{cases}$$

The graph in Fig. 2b shows the functional form of the two coefficients. Notice that the equation is

unchanged in the interval $[0, 1]$ and becomes parabolic in the interval $(1, b]$. In fact at $y = b$ the equation reduces to a simple diffusion equation.

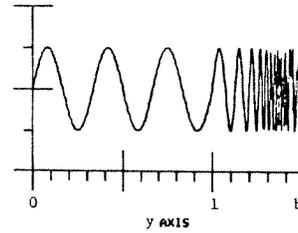


Fig. 2a Graph of $\sin(4\pi x(y))$

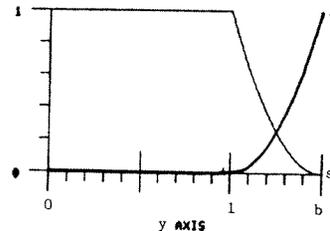


Fig. 2b Form of d and s

Boundary conditions must be given for all the variables at $y = b$ for the problem to be well-posed. The boundary condition for steady flow at infinity ($W_y = 0$) gave the best results in a series of test problems.

By imbedding the equation in the subregion into a well-posed problem in a slightly larger domain the difficulty of maintaining the correct number of boundary conditions at the artificial boundary was solved automatically. Furthermore, the information entering the region at this boundary depends on some global average properties of the solution outside the subregion.

Characteristic Form

In problems where the solution is sensitive to the approximation of the boundary conditions it may be more stable to transform the boundary conditions or the equation into characteristic form. The extrapolation formulas are then derived to extrapolate the outgoing characteristic variables to the fictitious points.

At a subsonic inflow boundary, $|u| < c$, the boundary conditions should be of the form

$$u = \alpha_1(u-c) + \beta_1(t) \quad (4.11)$$

and

$$u + c = \alpha_2(u-c) + \beta_2(t), \quad (4.12)$$

where α_i and β_i are functions of t alone. At subsonic outflow boundary the boundary conditions should be of the form

$$u - c = \alpha_3 u + \alpha_4(u+c) + \beta_3(t). \quad (4.13)$$

When the flow is supersonic, $|u| > c$, at the boundary the either three or no boundary conditions are given, depending on whether it is an inflow or outflow boundary.

Characteristic variables are also important when no amount of algebra seems to yield enough relationships to uniquely define all the solution variables at the fictitious points. When this happens one is forced to extrapolate on some of the variables without any boundary relationships to guide the extrapolation. It is usually best to extrapolate on outgoing characteristic variables and use their values at the fictitious points to provide the extra needed information.

Differential Form

Whatever extrapolation formula is used there may be some inherent truncation error in the extrapolated solution at the fictitious points. Some of these truncation errors can be eliminated by changing the differential form of the equation at the boundary.

For example, the reflecting boundary conditions (4.1) and (4.2) can be incorporated in the Euler equations at the boundary to give

$$\begin{pmatrix} \rho \\ m \\ E \end{pmatrix}_t + \begin{pmatrix} m \\ 0 \\ pu \end{pmatrix}_x = 0 \quad (4.14)$$

at the boundary. By differencing these equations, rather than Eq. (2.6), at the boundary we have prevented some of the possible truncation errors inherent in the extrapolation formula, from creeping into our calculation.

Notice that the modified Eq. (4.14) has been kept in divergence form. This is particularly important to maintain conservation when shocks are reflected at the wall.

Using the modified differential form of the equations is especially important when there is a removable singularity at the boundary. These terms should be replaced by their equivalent nonsingular form obtained using L'Hôpital's rule.

V. ARTIFICIAL DISSIPATION

Artificial dissipation or artificial viscosity is a special form of truncation error either inherent to a finite difference approximation or resulting from explicitly adding an additional term to the equation.

This dissipation is the leading truncation error in the numerical approximation and is chosen on the basis of the expected form of the solution.

The purpose of the artificial term is to remove many of the numerical difficulties by dissipating or damping out the high frequencies of the solution. This approach does in some sense mock up the effects of the viscous and dissipative terms discarded in the derivation of the Euler equations in that it primarily dissipates the high wave numbers, but it has little to do with true heat dissipation or viscosity.

There are six primary reasons for including artificial dissipation in the numerical approximation. They are:

1. To achieve proper entropy production across shock fronts.
2. To smooth out nonphysical discontinuities in the flow.
3. To solve the problem of the energy cascade when computing only a finite number of modes.
4. To compensate for spatial interpolation errors, such as the Gibb's phenomenon, near discontinuities in the solution.
5. To counteract the dispersion error in the numerical scheme.
6. To stabilize certain time differencing methods.

The form of a good artificial dissipation term tailored for a particular problem will depend on which of these points are most important. It is therefore essential to designing a numerical method to have a basic understanding of each of them. In this section we will review each reason for adding artificial dissipation and suggest a form which works for a large class of problems.

Entropy Production

The most common reason given for adding artificial dissipation is so that one can calculate shock waves. Entropy increases across a shock front, but Eq. (2.6) has no mechanism for the increase. We must add a term to the equation which will allow entropy to increase by the proper amount. The term should be in conservation form to maintain the Rankine-Hugoniot jump conditions and therefore give the correct shock speed.

Nonphysical Discontinuities

Another desired effect of the artificial dissipation is to smooth out nonphysical discontinuities in the flow. That is, it would be advantageous if the artificial dissipation were formulated in such a way that physical shocks are stable and nonphysical sudden compression shocks are unstable. These nonphysical

discontinuities often occur in the initial conditions and can be smoothed out by using more artificial dissipation in the first few time steps than later in the calculation.

Energy Cascade

Typically, in Eq. (2.6) energy enters the system at low wave numbers and cascades upward through the high wave numbers where it is eventually dissipated by molecular viscosity and enters the system as heat (Kolmogoroff hypothesis). In numerical calculations the energy spectrum is limited by the number of mesh points. When there is no artificial dissipation in the system the energy cascade backs up at the higher frequencies and shows up in the calculation as high frequency noise or trash. Some of this energy is aliased or reflected back into the lower wave numbers. This closed loop energy cascade can destroy the accuracy in all wave numbers during even moderately short computations.

Gibbs Phenomenon

Artificial dissipation can help compensate for some of the errors introduced by approximating F and F_x with an interpolant whose values agree with F only at a discrete set of points.

The errors in the interpolant are most severe near discontinuities in the function being approximated. At these points the continuity conditions used to derive the interpolant break down and oscillations appear in the calculation.

These oscillations can destroy the accuracy of the calculation by creating nonlinear instabilities or introducing nonphysical features in the flow such as negative mass or pressure. The oscillations may generate new artifacts into the calculation such that the numerical calculation is stable but converges to the wrong solution [Harten et. al. (1976)]. In reacting flows these overshoots can trigger a chemical reaction and lead to meaningless results. Adding artificial dissipation to the numerical approximation damps the high frequencies and helps reduce superfluous oscillations in the solution.

Dispersion Error

Dispersion errors come from the inexactness in both the time and space differencing methods. The dispersion errors due to the different modes of the solution traveling at different and incorrect velocities can accumulate and destroy the accuracy of the computation. This is particularly true for the higher modes even in calculations of flows which should have only smooth solutions. Increasing the accuracy in both the time and space differencing methods will reduce the dispersion in the low and middle frequencies, but not the high modes. It may be best to damp these out by some form of artificial dissipation.

Stabilization of Time Integration Methods

The ability of artificial dissipation to stabilize what may otherwise be an unstable time differencing method for Eq. (2.6) lies in the fact that it shifts the spectrum of the spatial operator such that the solution to the modified equation is mathematically and numerically more stable. This is especially true for such standard methods as forward Euler and improved Euler as will be seen in Section VI.

Differential Form

For many problems the artificial dissipation inherent to the time integration method is sufficient to compensate for the energy cascade problem and also the entropy production in weak shocks. For strong shocks it is necessary to add significantly more dissipation. The extra dissipation can be added by

explicitly adding a dissipative truncation error to Eq. (2.6). This is done in all the shock calculations in Section VII. The modified equation for these calculations can be written as

$$W_t + F_x = (\Delta x \frac{dW_x}{dx})_x, \quad (5.1)$$

where

$$d = \delta \left(\frac{\Delta t}{\Delta x} \right)^{\ell-1} \lambda_{\max} \left| \lambda_{\max} \right|^{\ell-1} = \delta \left(\frac{\Delta t}{\Delta x} \right)^{\ell-1} (|u|+c)^{\ell} \quad (5.2)$$

and λ_{\max} is the largest characteristic velocity of the system and δ is called the artificial dissipation coefficient. Numerical experiments indicate that choosing the parameter $\ell = 1$ works best for most flows.

It may seem strange at first to use a first-order artificial dissipation term with a fourth- or sixth-order approximation of the derivatives of the flux function. In calculations with strong shock waves there is not always a one-to-one correlation between the formal order of accuracy of a difference scheme and the true accuracy of the calculation. The most reliable and accurate artificial dissipation terms known happen to be of low order and we are stuck with them until better ones can be developed.

VI. TIME DISCRETIZATION

In choosing the "best" numerical method to integrate the Euler equations through time one has to consider the accuracy, stability, storage requirements, computational complexity and the relative cost of the different methods. These factors are dependent on each other and tradeoffs must be made as to which criteria are more important for a particular problem.

Spectral Analysis

Both the phase and damping errors depend on the spectrum of the differential equation and the time step size. The time step can be varied during the calculation to reduce the numerical integration errors, but the spectrum of the differential equations is determined by the spatial difference operator. A good integration method depends on how accurately it can integrate a particular set of equations. For this reason the spectrum of the spatial difference operator is the most important guide in selecting an efficient numerical method to integrate through time. The spectrum can be determined by analyzing the linearized continuous time - discrete space approximation of the partial differential equation.

Equation (2.6) is solved after adding artificial dissipation and therefore we must analyze a system of the form (5.1). Most of the essential properties of this system are also found in the simple prototype equation

$$\rho_t + \rho_x = (\Delta x \delta \rho_x)_x. \quad (6.1)$$

A semi-discrete approximation of (6.1) results when the spatial derivatives are approximated by finite differences on a mesh of N points. This system can be written in the form of ordinary differential equations (ODEs) $y' = Ay + \delta \Delta x By = Cy = f(y)$.

The vector y is an array of the approximate solution at the mesh points and the prime denotes the derivative of y with respect to time.

When second-order centered differences are used the eigenvalues of A are imaginary and the eigenvalues of B negative real. The eigenvalues of C are complex and lie on the ellipses graphed in Fig. 3 for values of $\delta = 0.0, 0.2, \dots, 1.0$.

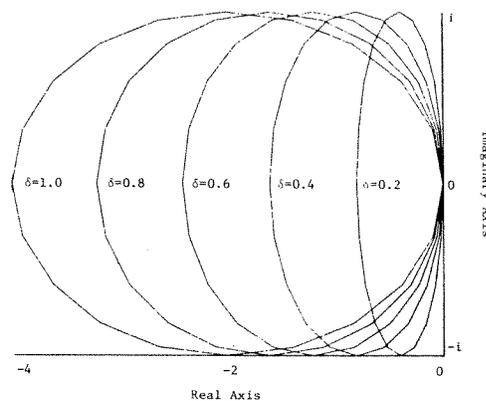


Fig. 3. The eigenvalues of $\Delta x C$

We shall first analyze Eq. (6.2) when there is no artificial dissipation (i.e. $\delta=0$) and include the effects of the dissipation as a perturbation on this equation. When $\delta=0$ Eq. (6.2) is dispersive since the eigenvalues of A lie on the imaginary axis. These eigenvalues, λ , are equal to $i\alpha u$, $i\alpha(u+c)$ and $i\alpha(u-c)$, where α depends upon the spatial order of approximation. When second- or fourth-order centered differences are used and the boundary conditions are periodic on the unit interval the corresponding α 's are

$$\alpha_2 = (\sin(2\pi j \Delta x)) / \Delta x, \quad (6.3)$$

and

$$\alpha_4 = (8 \sin(2\pi j \Delta x) - \sin(4\pi j \Delta x)) / 6\Delta x, \quad (6.4)$$

for

$$j = -N/2, -N/2 + 1, \dots, N/2 \text{ and } \Delta x = 1/N.$$

To facilitate studying the properties of different time integration methods we use a well known result from ordinary differential equations. The isolation theorem,⁸ states that the stability and accuracy of a numerical integration method for Eq. (6.2) is determined entirely by how it approximates the decoupled diagonalized system

$$y' = \lambda_i y, \quad (6.6)$$

with the solution $y_i(t) = y_i(0)e^{\lambda_i t}$ where the λ_i are the eigenvalues of C .

Numerical Methods

Equation (6.6) (hence 6.2) is a multirate system of equations since some ODE components change on vastly different time scales than others. These systems can have accuracy and stability restrictions that can make standard explicit integration methods inefficient.

We now describe a new class of numerical methods, called iterative multistep (IMS) methods⁵ that overcome some of the difficulties in solving multirate systems. An IMS method sequence begins with an explicit predictor of order k , such as forward Euler ($k=1$)

$$y_{n+1}^{(1)} = y_n + \Delta t f_n, \quad (6.6)$$

and a corrector of order $k+1$, in this case the improved Euler method

$$y_{n+1}^{(2)} = y_n + \frac{1}{2} \Delta t (f_{n+1}^{(1)} + f_n). \quad (6.7a)$$

Here $n+1$ refers to time t_{n+1} and the superscript is the iteration index, $f_{n+1}^{(1)} = f(y_{n+1}^{(1)})$. After the corrector cycle, additional iterations are based on the simple recurrence relation

$$y_{n+1}^{(i)} = y_{n+1}^{(i-1)} + c_i \Delta t [f_{n+1}^{(i-1)} - f_{n+1}^{(i-2)}] \quad (6.7b)$$

for $i = 3, 4, \dots$. The constants c_i depend on the iteration count and the predictor-corrector method used

to start the process. The c_i are chosen to increase the order of accuracy of the method for linear autonomous systems and each iteration. When Eqs. (6.7) are used to start the iteration the constants c_i have the simple explicit formula $c_i = 1/i$, $i = 3, 4, \dots$. This method is called the iterated Runge-Kutta method since the stability region after the i -th iteration is equivalent to the stability region of an i -th order Runge-Kutta method. These regions are symmetric about the real axis and are shown in Fig. 4 below.

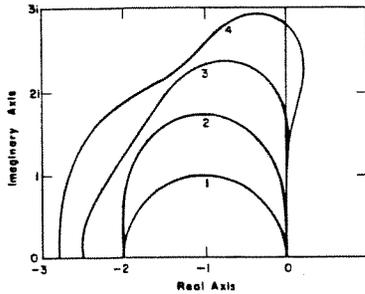


Fig. 4. Stability regions for the iterated Runge-Kutta method for $i=1, 2, 3$ and 4.

The method is stable if Δt is chosen small enough that $\lambda \Delta t$ lies within its stability region. For Eq. (6.2) this must hold for all the eigenvalues of C . The stability regions increase on each iteration and the approximations will converge to the exact solution when solving linear autonomous systems such as Eq. (6.2).

Another IMS method which has excellent stability and accuracy properties for the ODEs arising from the Euler equations is the iterated leap-frog method. The second-order leap-frog predictor is given by

$$y_{n+1}^{(1)} = (1-r^2)y_n + r^2y_{n-1} + \Delta t(1+r)f_n, \quad (6.8a)$$

where

$$r = \left[\frac{(t_n - t_{n-1})}{(t_{n+1} - t_n)} \right]^{-1/2}$$

and the third-order leap-frog corrector is

$$y_{n+1}^{(2)} = [(2-r)(1+r)^2y_n + r^3y_{n-1} + \Delta t(1+r)^2f_n + \Delta t(1+r)f_{n+1}^{(1)}] / (2+3r). \quad (6.8b)$$

The IMS coefficients for this method are $c_i = 3/10, 7/30, 4/21, 451/2800, 314/2255, 1153/9420,$ and $126/1153$ for $i=3, 4, \dots, 9$ when $r=1$. The c_i are functions of r and are not known for general r at this time.

The iterated leap-frog method has stability regions that are particularly good along the imaginary axis, as seen in Fig. 5 below.

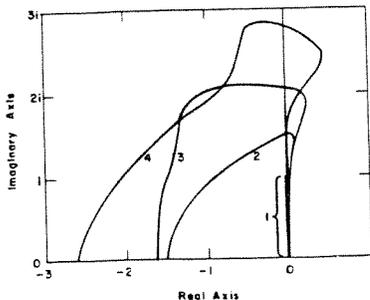


Fig. 5. Stability regions for the iterated leap-frog method when $i = 1, 2, 3$ and 4.

The leap-frog predictor is unstable for systems of equations with eigenvalues having a nonzero real part. Therefore, when artificial dissipation is added or the boundary conditions shift the spectrum of the discretized equation the leap-frog method cannot be used without the corrector cycle. The first corrector

application extends the bound on the maximum time step by 50%, increases the method to third-order and is stable in smooth regions of the solution with or without any spatial artificial dissipation. Another difficulty with using the leap-frog predictor is a unique type of error due to time and space mesh decoupling. The odd and even points of a mesh are only weakly coupled when integrating conservation laws and errors with frequency $= 2\Delta x$ can degrade the accuracy of the solution with high frequency noise. The corrector cycle couples the mesh points among the three time levels and prevents this weak instability.

When integrating nonlinear equations, the IMS methods reduce to the order of the predictor-corrector or Runge-Kutta starting method. The stability regions still expand with extra iterations but the order of accuracy remains the same. The coefficients for the IMS methods can also be chosen to increase the stability by a maximum amount on each iteration while retaining the order of accuracy of the starting method.

Stability

For most numerical methods it is the largest eigenvalue λ_{\max} of the linearized equations that determines the stability condition. When we make this assumption stability restrictions on Δt can be derived using Figs. 4 and 5. When second-order centered differences are used in space and the leap-frog predictor is used in time, then if $\delta=0$ the stability condition requires $\Delta t |\lambda_{\max}| \leq 1$ or (using Eq. (6.3))

$$\Delta t \max(|u|+c) \left| \frac{\sin(2\pi j \Delta x)}{\Delta x} \right| = \frac{\Delta t}{\Delta x} \max(|u|+c) \leq 1.$$

This is the usual Courant-Friedrichs-Lewy stability condition for explicit methods when solving the Euler equations. If fourth-order centered differences are used in space and the leap-frog predictor-corrector method in time, the corresponding stability condition is

$$\Delta t \frac{8}{6\Delta x} \max(|u| + c) \leq 1.5.$$

Notice in Fig. 4 that some integration schemes such as forward Euler are unconditionally unstable for all $\Delta t > 0$ when the spectrum of the discretized system lies on the imaginary axis. It is well known that forward Euler is the heart of many standard methods to solve Eq. (2.6) and in fact is not always unconditionally unstable. This is because of the addition of artificial dissipation shifts the eigenvalues of the linearized system to the left so they have a negative real part as seen in Fig. 3.

We caution the reader that this stability analysis is linear and is not necessarily valid for highly nonlinear phenomena such as shockwaves. In practice to prevent nonlinear instabilities, it is necessary to restrict the time step such that a shock will not move more than one mesh point per time step.

VII NUMERICAL ALGORITHMS

The general flow of a MOL computer code has a well defined structure. The code must:

1. Define the initial conditions for the PDEs.
2. Incorporate the boundary conditions into the discrete system.
3. Define W_t , i.e. evaluate and difference the flux function and the add artificial dissipation.
4. Predict the solution and update the time ($t \rightarrow t + \Delta t$) or correct the solution (t is unchanged).
5. Repeat the cycle if the problem is unfinished (go to 2).

The numerical method is determined by the decisions made in steps 2, 3 and 4. In step 2 we recommend incorporating the boundary conditions into the discrete system by using fictitious points. This approach can be used with any of the procedures described in

Section IV. The extrapolation formula for the fictitious points allows more freedom to include information about the PDEs and the boundary conditions into the difference scheme than does using uncentered differences.

In Step 3 we recommend using second-order differences only in the initial debugging stages of the program and later switching to at least fourth-order. The higher order methods reduce the phase errors and the amount of artificial dissipation needed to stabilize the calculation. However, the extra work required to use sixth-order differences may not be justified unless this accuracy is balanced by also using a high order method in time.

The artificial dissipation term added in step 3 is crucial to the success of any numerical method in shock calculations. The dissipation in Eq. (5.1) can be approximated when $\ell=1$ by second-order differences at time t_n and $x=x_i$ as

$$(\Delta x d_w)_x \cong \phi_{i+\frac{1}{2}}^n - \phi_{i-\frac{1}{2}}^n \quad (7.1)$$

where

$$\phi_{i+\frac{1}{2}}^n = (d_{i+1}^n + d_i^n)(w_{i+1}^n - w_i^n)/(2\Delta x)$$

and

$$d_i^n = \delta_i (|u| + c)_i^n.$$

Choosing δ between zero and one works adequately for most problems.

When the location of a contact discontinuity may be tracked accurately then a switch may be used to reduce δ away from shocks. In one-dimensional calculations a shock can be detected by a discontinuous negative velocity gradient and the artificial dissipation coefficient can be increased accordingly. For example, a possible switch is to replace $\phi_{i+\frac{1}{2}}^n$ in (7.1) by $\alpha \phi_{i+\frac{1}{2}}^n$ where $\alpha=1$ if $u_{i+1}^n > u_i^n + (\Delta x/3)$ and, say, $\alpha=1/3$ otherwise.

A second type of switch changes the artificial dissipation coefficient in the predictor and corrector cycles. This is a particularly good switch for the iterated Runge-Kutta method. The forward Euler predictor cycle is less stable (Fig. 4) than the improved Euler corrector method and requires more artificial dissipation to be stable for hyperbolic equations. We, therefore, add a large amount of artificial dissipation in the predictor cycle. In the corrector cycle the artificial dissipation coefficient is reduced, set to zero or even reversed in sign to add antidiffusion and counteract the effects of the overly diffused predicted solution.

A third type switch is designed to prevent contact discontinuities from smearing in long-time or steady-state calculations by artificially compressing them. Harten³ has proposed modifying Eq. (2.6) by adding the derivative of an artificial compression function to the right hand side. This function is chosen such that a shock or contact discontinuity for Eq. (2.6) is a shock for the modified equation. That is the contact discontinuity is artificially compressed to reduce numerical smearing.

These improved artificial dissipation strategies all help reduce the numerical errors away from shocks, but no one method stands out as best for all problems. For this reason it is one of the most active areas in developing new methods for the Euler equations.

Of the ODE methods used in Step 4 the leap-frog predictor-corrector method has outperformed any other method we have tested. If the solution can be stored on more than two time levels then the high order Adams-Bashford-Moulton methods may be more competitive. (See Shampine and Gordon⁹).

VIII. NUMERICAL RESULTS

Riemann Problem

An initial value problem for the Euler equation is called a Riemann problem if the initial data consists of two constant states. The initial conditions chosen in this example were also used in a survey article by Sod (1978) to allow comparisons with other difference schemes. Initially a gas ($\gamma=1.4$) is separated in a shock tube by a diaphragm at $x=0.5$. The left and right states of the system are:

$$\begin{aligned} u(x,0) &= 0 & u(x,0) &= 0 \\ \rho(x,0) &= 1, 0 \leq x \leq \frac{1}{2}; & \rho(x,0) &= 0.125, \frac{1}{2} < x \leq 1 \\ p(x,0) &= 1 & p(x,0) &= 0.1 \end{aligned}$$

at $t=0$ the diaphragm is burst and by $t=0.25$ the gases have developed into a shock wave on the far right, a contact discontinuity near $x=0.7$ and a rarefaction wave to the left of $x=0.5$ as seen in Fig. 6.

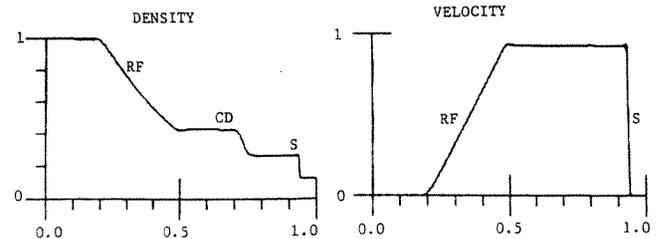


Fig. 6. Density and velocity at $t = 0.25$.

The problem was solved on a mesh of 100 points, with fourth-order spatial difference and an artificial dissipation term given by Eqns. (5.1) and (7.1) with $\ell=1$ and $\delta=0.5$. The iterated Runge-Kutta was used in time with one corrector cycle and $|\lambda_{\max}|\Delta t = 1$.

When the ends of the shock tube are approximated by a reflecting boundary condition, Eq. (4.1), the shock reflects from the right boundary and passes through the contact discontinuity by $t=0.5$ in Fig. 7, below.

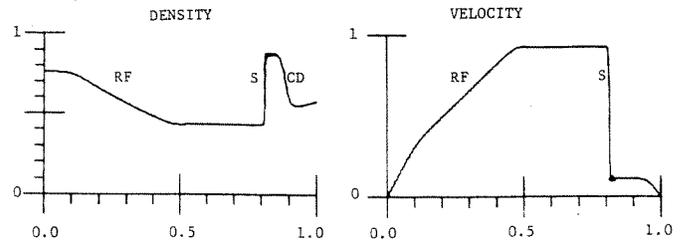


Fig. 7. Solution $t=0.5$ with reflecting boundary conditions.

This problem was also solved with a nonreflecting boundary condition at $x=1$ using the mapping technique given by Eq. (4.10). Three fictitious points were included between $x=1$ and $x=b=1.03$ in Fig. 8 below.

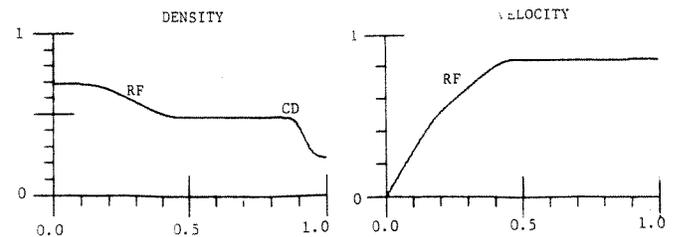


Fig. 8. Solution at $t=0.5$ with an artificial boundary at $x=1$.

Notice the shock has passed through the boundary with almost no reflections.

Generalized Euler Equations

The general one-dimensional form of the Euler equation is

$$(AW)_t + [AF(W)]_x = p \begin{pmatrix} 0 \\ A \\ -A_t \end{pmatrix} \quad (8.1)$$

where W and F are as in (2.6) and $A(x,t)$ is an area term depending on the geometry and dimensionality of the problem. For example, $A=1$ for slab symmetry, $A=x$ for cylindrical symmetry and $A=x^2$ for spherical symmetry. In this problem $A(x,t)$ is the cross-sectional area of an imploding cylindrical duct.

The cylinder is 100 cm long and collapses at a constant velocity of 1 cm per unit time from a radius of 1 cm to 0.25 cm. The collapse progresses up the cylinder behind a hinge from $x=0$ at $t=0$ to $x=93.8$ at $T=7$. A shock forms in the gas and is maintained at the hinge location by exponentially accelerating the velocity of the hinge.

This action causes the collapsing cylinder to act as a velocity accelerator. That is, the imploding wall pushes the gas and accelerates it up the cylinder. The velocity of the gas is over 20 times the velocity of the collapsing cylinder walls by the time it has reached the end of the cylinder. A more detailed description of this problem can be found in Ref. 1.

Initially the system is at rest and $\rho=0.15$, $p=1.3$ and $\gamma=5/3$. The hinge is advanced according to

$$h(t) = 78.93 \beta (\exp(\beta t) - 1)$$

where $\beta = (\gamma-1)/(\gamma+1) = 1/4$. A cross-section of the cylinder, the gas velocity and maximum sound speed ($|u|+c$) are shown at time $t=7$ in Fig. 9.

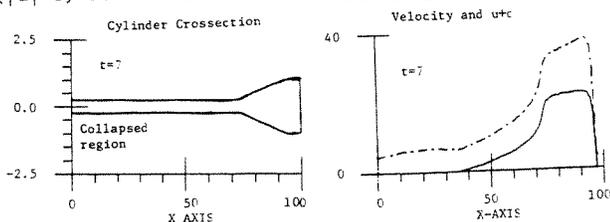


Fig. 9. The imploding cylinder, the gas velocity (—) and maximum characteristic velocity $u+c$ (---) at $t=7$.

This solution was calculated with $\Delta x=1$, $\delta=0.4$, fourth-order centered difference and the leap-frog predictor-corrector method in time with $(|u|+c)\Delta t=0.75$. The numerical method encountered no difficulties in solving Eq. (8.1) and the shock developed in strength and position as predicted in Ref. 1.

IX. SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

In this paper we have followed a MOL approach to constructing accurate and robust numerical methods for hyperbolic PDEs derived from conservation laws. The approach has proved to be straight forward and has lead to some excellent new methods for solving the Euler equations. It is our belief that a similar approach may yield some equally useful methods in other "problem areas" of numerical analysis such as the Navier-Stokes equations, reacting or combusting flows and nonlinear diffusion equations.

Acknowledgements

The author would like to thank Dr. Joe Dendy, Dr. Don Durack and Dr. Burton Wendroff for a number of helpful discussions and suggestions in developing the methods presented in this paper.

REFERENCES

1. Colgate, S., Hyman, J. M. and Durack, D. (1977), Shock Exponentiation in Axial Jets, Los Alamos Scientific Laboratory preprint.
2. Harten, A., Hyman, J. M., Lax, P. D., and Keyfitz, B. (1976), "On Finite-Difference Approximations and Entropy Conditions for Shocks," Comm. on Pure and Applied Math., Vol. XXIX, pp. 297-322.
3. Harten, A. (1978), "The Artificial Compression Method for Computation of Shocks and Contact Discontinuities: II. Self-Adjusting Hybrid Schemes," Math. of Comp., Vol. 32, 142, pp. 363-389.
4. Hyman, J. M. (1976), "The Method of Lines Solution of Partial Differential Equations," NYU report C00-3077-139.
5. Hyman, J. M. (1979), Explicit A-Stable Iterative Methods for the Solution of Differential Equations, in preparation.
6. Lax, P. D. (1973), Hyperbolic Systems of Conservation Laws and the Mathematical Theory of Shock waves, SIAM, Phil., Pa.
7. Lax, P. D. (1978), The Numerical Solution of the Equations of Fluid Dynamics, Lectures on Combustion Theory, NYU report C00-3077-153, pp. 1-60.
8. Lomax, H. (1967), An Operational Unification of Finite Difference Methods for the Numerical Integration of ODEs, NASA TR R-262.
9. Shampine, L. F. and Gordon, M. K. (1975), Computer Solution of Ordinary Differential Equations - The Initial Value Problem, W. H. Freeman and Company, San Francisco, CA.
10. Sod, G. A. (1978), "A Survey of Several Finite Difference Methods for Systems of Nonlinear Hyperbolic Conservation Laws," J. Comp. Phys., Vol. 27, 1, pp. 1-31.

ADVANCES IN COMPUTER METHODS FOR PARTIAL DIFFERENTIAL EQUATIONS - III

Proceedings of the Third IMACS International Symposium
on Computer Methods for Partial Differential Equations
held at Lehigh University - Bethlehem, Pennsylvania, U.S.A.
June 20-22, 1979

Edited by

R. VICHNEVETSKY
Rutgers University
New Brunswick, N.J. 08903 (USA)

and

R.S. STEPLEMAN
RCA Laboratories
Princeton, N.J. 08540 (USA)

Published by IMACS
1979