Flux Vector Splitting and Approximate Newton Methods

Dennis C. Jespersen† and Thomas H. Pulliam†

NASA Ames Research Center
Moffett Field, CA 94035

Abstract. Some iterative methods for the solution of the steady Euler equations can be analyzed by considering them as modifications of modified Newton's method. The analysis leaves the equations in block coupled form, as opposed to the usual technique of diagonalizing the equations and examining the representative scalar equation. The analysis technique is applied to various flux splittings of the Euler equations. We highlight here the use of numerical tools and Fourier techniques to examine the stability through eigensystem analysis.

The quasi-one-dimensional Euler equations are employed; we give analysis and numerical results for flow through nozzles with shocks. In particular, we examine the convergence of various modified Newton methods obtained from implicit schemes for flux splitting.

§1. Introduction.

This paper presents a point of view and an analysis technique for some iterative methods for the steady Euler equations. The basic approach employed here is to view an iterative scheme as Newton's method or a modified Newton's method. Newton's method solves a nonlinear equation of the form \( F(Q) = 0 \), where \( F \) can be either an analytic operator or some numerical operator, such as a finite difference operator. The solution \( Q \) is obtained by solving a sequence of linear problems of the form

\[
DG(Q^{(n)})(Q^{(n+1)} - Q^{(n)}) = -F(Q^{(n)})
\]

where \( DG(Q^{(n)}) = \partial F / \partial Q \), the Jacobian of \( F \) with respect to \( Q \), evaluated at a given \( Q^{(n)} \). It is well known that Newton's method converges quadratically in the vicinity of the solution. By a modified Newton's method we mean an iteration of the form (1.1) where \( DG(Q^{(n)}) \neq \partial F / \partial Q \); such an algorithm may be used for reasons of efficiency (if the exact Jacobian of \( F \) is costly to compute) or of practicality (if it is impracticable to compute the exact Jacobian). A modified Newton's method will not necessarily be convergent in the vicinity of the solution; we shall investigate here various modified Newton methods which can arise from differencing schemes for the Euler equations.

The Euler equations are a system of nonlinear partial differential equations describing mass, momentum, and energy balance for inviscid compressible flow. We will consider flux vector splitting as the basic spatial differencing technique. Flux vector splitting is based on the partition of a flux vector into groups which have certain specified properties. For instance, the Euler equations fluxes can be split into two groups, the first group having a flux Jacobian with all positive eigenvalues and the second group having a flux Jacobian with all negative eigenvalues. With this splitting the spatial derivatives can be approximated by type dependent differences, producing an efficient algorithm without the need for artificial viscosity.

The Steger-Warming version of flux vector splitting employs such a plus-minus splitting. An implicit algorithm for the integration of the flux-split equations has the drawback of requiring the computation of a rather complicated flux Jacobian matrix at each grid point, causing increased expense. A natural modification of the iterative scheme replaces the complicated matrix by a simpler and seemingly more natural choice. This modification of the iterative method may harm the convergence properties of the method. For example, Steger and Warming (Ref. 1) found that use of the modified method resulted in a restrictive time step limitation. This finding was one of the original motivations for this work. Our analysis here sheds some light on this restriction.

We will also look at a flux vector splitting based on a velocity-sound speed split. One reason for considering such splittings is that the velocity term can be type-dependent differenced while the sound speed term can be central differenced, and no artificial viscosity is needed. This form of splitting is also the basis for the parabolised Navier-Stokes method of Schiff and Steger, where the velocity-sound speed splitting is

†Research Scientist, Computational Fluid Dynamics Branch.

This paper is declared a work of the U.S. Government and therefore is in the public domain.
used in the stability analysis of their method.

Each numerical scheme we consider will be cast in the form of a Newton-type method and analyzed as such. Variations such as changing the type or order of the differencing schemes employed on either side of Eq. (1.1) or modifying the Jacobian $DG$ to reflect different flux vector approximations will be examined.

The analysis technique employed here will differ from standard analysis of differencing schemes applied to the Euler equations in that the equations will be treated as a system rather than diagonalized and reduced to uncoupled scalar equations. We will study the stability of algorithms around a steady solution using both Fourier analysis and "total matrix analysis". By "total matrix analysis" we mean the explicit formation and numerical solution of a matrix eigenvalue problem whose eigenvalues determine the stability of the iteration process. The solution of the eigenvalue problem is carried out using standard software from the IMSL library. Both techniques lead to a generalized eigenvalue problem which is solved numerically; it will be this examination of the eigensystem of the schemes which will form the basis of our analysis. The Fourier technique enables one to make predictions concerning the stability of the algorithm; these predictions provide good guidelines. The comparison with the results from the total matrix analysis will show the efficacy and usefulness of the Fourier technique.

We would like to highlight here the use of numerical techniques to analyze nonlinear systems. When one is willing to use numerical techniques in the analysis, one need not make any of the standard simplifying assumptions such as constant coefficients, periodic boundary conditions, or a scalar equation. The total matrix analysis makes none of these assumptions. The equations are retained as coupled systems and analyzed as such. We are willing to employ the computer to solely our eigenvalue problems and use the results in the development of our algorithms.

The steady Euler equations for quasi-one-dimensional flow in a nozzle are considered here. We will look at a typical implicit algorithm in "delta" form. Results are given for steady flows with shocks.

### §2. Iterative Schemes.

The steady Euler equations for flow in a quasi-one-dimensional nozzle may be written

\[ \frac{\partial E}{\partial z} + H = 0, \]

where $E = E(Q)$, $H = H(Q)$, and $Q = (\rho, m, e)^T$.

The flow variables are $\rho$, the density, $m = \rho u$ with $u$ the flow velocity, and $e$, the total energy per unit volume. The flux $E$ and source term $H$ are

\[ E = a(x) \cdot (\rho u, \rho u^2 + p, (e + p)u)^T, \]

\[ H = a'(x) \cdot (0, -\rho, 0)^T, \]

where the pressure $p$ is given by

\[ p = (\gamma - 1)(e + \frac{1}{2}\rho u^2) \]

and $a(x)$ is the cross-sectional area for nozzle flow.

A general formulation of flux vector splitting is as follows. We write $E(\delta z) = E_1(Q) + E_2(Q)$ and, with $A(Q) := E_1/\partial Q$, write $A(Q) = A_1(Q) + A_2(Q).$ (The notation $x := y$ will mean that $x$ is defined to be $y.$) We will give examples of such splittings later. Note that one obvious choice for $A_1$ and $A_2$ is $A_1 = \partial E_1/\partial Q$ and $A_2 = \partial E_2/\partial Q,$ but we will consider other possibilities as well. Assume we have some space grid with typical mesh length $\Delta x$, and let $h$ denote a time step. (For the Fourier analysis technique to be described below, the mesh is thought of as locally uniform, while for the total matrix analysis technique there is no restriction on the mesh. We write, for simplicity's sake, the formulas as if the mesh were uniform.) Let $\delta$ (with various subscripts) denote a spatial differencing operator (for example, a central differencing operator would be $\delta u_j = (u_{j+1} - u_{j-1})/2\Delta x$). Then the general iterative algorithm we will consider has the form (given an initial guess $Q^0$)

\[ M(Q^n)\Delta Q^n = P(Q^n) \]

\[ Q^{n+1} = Q^n + \Delta Q^n, \]

where

\[ M(Q) := I + h(\delta_1, \delta_2, \delta_3) + h \frac{\partial H}{\partial Q} \]

\[ P(Q) := -h(\delta_1, E_1(Q) + \delta_2, E_2(Q)) - h H(Q). \]

This general form is based on an implicit approximation in "delta" form, using the backward Euler method in time, of the time-dependent equation $Q_t + E_z + H = 0.$

The spatial derivative operators are written in a general form in Eq. (2.3). The subscripts $i$ and $j$ stand for left and right, respectively, and reflect the possibility of allowing the spatial differencing operators to be of different orders of accuracy on the two sides of the equation. This may be desirable for reasons of computational efficiency. The subscripts 1 and 2 are associated with the splitting of the fluxes into two distinct parts and give us the flexibility of employing different differencing operators for each term. In the plus-minus flux split algorithm, backward differencing is used for the "plus" terms and forward differencing for the "minus" terms. (See Warming and Beam for more on implicit algorithms in computational fluid dynamics.)
The natural choice for \( A_1 \) and \( A_2 \) is \( A_1 = \partial E_1 / \partial Q \)
and \( A_2 = \partial E_2 / \partial Q \), but the choice of \( A_1 \) and \( A_2 \) has no effect on the steady solution. Thus one is free to use other \( A_1 \) and \( A_2 \) in order to, for example, attain greater computational efficiency.

If we let \( h \to \infty \) and if we have \( A_1 = \partial E_1 / \partial Q \) and \( A_2 = \partial E_2 / \partial Q \), then algorithm (2.2) becomes precisely Newton’s method for the solution of the nonlinear discrete system of equations \( P(Q) = 0 \). Thus in this case (i.e., \( A_1 = \partial E_1 / \partial Q \) and \( A_2 = \partial E_2 / \partial Q \)) iteration (2.2) is stable for all sufficiently large time steps and all \( Q_0 \) sufficiently close to a solution \( \hat{Q}^* \). One of our goals in this paper is to analyze the time step limitation (if any) imposed by a different (perhaps more convenient) choice of \( A_1 \) and \( A_2 \). We will study the stability of algorithm (2.2) near a solution \( Q^* \). Thus we will be using the term “stability” in a sense different from the usual computational fluid dynamics sense. In this paper, to say iteration (2.2) is stable near a solution \( \hat{Q}^* \) of \( P(Q^*) = 0 \) is to say that there is some norm and some \( \varepsilon > 0 \) such that if \( \| Q_0 - Q^* \| < \varepsilon \), then \( \| Q^n - Q^* \| \to 0 \) as \( n \to \infty \). We will not be considering the behavior of (2.2) for arbitrary initial data \( Q_0 \), we will only be looking at perturbations about a solution \( Q^* \).

The iteration (2.2) can be written as a fixed point iteration

\[
Q^{n+1} = G(Q^n),
\]

where \( G(Q^*) = Q^* + M(Q^*)^{-1}P(Q^*) \).

The stability of this fixed-point iteration near a solution \( Q^* \) of \( Q^* = G(Q^*) \) is determined by the eigenvalues of the Jacobian matrix \( D(G(Q^*)) \). A necessary and sufficient condition for stability of the iteration near \( Q^* \) is that all the eigenvalues of the Jacobian are inside the unit circle. Using the definition

\[
DG(Q^*)(Q) := \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} [G(Q^* + \varepsilon Q) - G(Q^*)]
\]

and the fact that \( P(Q^*) = 0 \), we easily compute

\[
DG(Q^*) = I + M(Q^*)^{-1} \frac{\partial P}{\partial Q}(Q^*).
\]

Thus, we want to study the eigenvalue problem \( DG(Q^*)u = \lambda u \) (where \( DG \) is given by (2.6)), which is equivalent to the generalized eigenvalue problem

\[
Lv = \lambda Kv,
\]

where

\[
L = M(Q^*) + \frac{\partial P}{\partial Q}(Q^*) = I + h (\delta_{1,1} A_1 - \delta_{1,1} (\partial E_1 / \partial Q) + \delta_{2,1} A_2 - \delta_{2,1} (\partial E_2 / \partial Q))
\]

\[
K = M(Q^*).
\]

We will study this generalized eigenvalue problem by two methods. The first method is a Fourier method. The second is the total matrix analysis, where we construct the matrices \( L \) and \( K \) and then solve the problem (2.8) numerically using standard software.

Two versions of flux vector splitting are considered. The first is the Steger–Warming version. In this version the fluxes \( E_1 \) and \( E_2 \) are defined as follows. Write

\[
A = \partial E / \partial Q = XAX^{-1}
\]

where the columns of \( X \) are the eigenvectors of \( A \) and \( A \) is a diagonal matrix whose diagonal entries are the eigenvalues of \( A \). Then write

\[
\lambda = \lambda^+ + \lambda^-
\]

where the eigenvalues are split via

\[
\lambda = \lambda^+ + \lambda^- = \text{max}(\lambda, 0) + \text{min}(\lambda, 0).
\]

Define

\[
A^+ := XA^+ X^{-1} \quad \text{and} \quad A^- := XA^- X^{-1}.
\]

Finally, define

\[
E^+ := E^+ \quad \text{and} \quad E^- := E^-.
\]

Then

\[
E_1 := E^+ \quad \text{and} \quad E_2 := E^-.
\]

We will consider for \( A_1 \) and \( A_2 \) the choice \( A_1 := \partial E_1 / \partial Q, A_2 := \partial E_2 / \partial Q \) (i.e., the true Jacobians, which are computationally expensive) and also the choice \( A_1 := A^+, A_2 := A^- \) (computationally less expensive).

The second version of flux vector splitting we will look at is a velocity–sound speed splitting defined as follows. With notation as above, we have \( A = \text{diag}(u, u + c, u - c) \). Write

\[
A = A_1 + A_2 = \text{diag}(u, u + c, u - c).
\]

Define

\[
A_1 := XA_1 X^{-1} \quad \text{and} \quad A_2 := XA_2 X^{-1}.
\]

Define

\[
E_1 := A_2 Q \quad \text{and} \quad E_2 := A_2.
\]

Then we take \( E_1 := E_1 \) and \( E_2 := E_2 \). For \( A_1 \) we can take either \( A_1 := \partial E_1 / \partial Q \) (the true Jacobian) or \( A_1 := A_1 \), and similarly for \( A_2 \).
§3. Fourier Analysis. 

The stability of algorithm (2.2) near a solution $\mathbf{x}^*$ is determined by the eigenvalues of the generalized eigenvalue problem (2.7). If all the eigenvalues are inside the unit circle the iteration is stable near $\mathbf{Q}^*$, otherwise the iteration is unstable. A Fourier analysis is a quick and easy way to study the eigenvalues of (2.7). For the Fourier analysis, assume constant space coefficients and assume a Fourier mode for $v_j$, $v_j = e^{i j \theta}$ where $y_j$, independent of $j$, is a vector with three components. (In a rigorous analysis $\theta$ would be $k \pi \Delta y$ for some $k$; our analysis is approximate, and we will be less precise about the relation of $\theta$ and $A \mathbf{x}$. We think of $|\theta|$ less than $\pi/2$ as corresponding to low space frequencies and $|\theta|$ between $\pi/2$ and $\pi$ as corresponding to high space frequencies.) Inserting $v_j = e^{i j \theta}$ into (2.7), applying the difference operators, and dividing out a common factor of $e^{-i j \theta}$, gives us a 3-by-3 eigenproblem

$$\mathbf{L} y = \lambda \mathbf{K} y$$  \hspace{1cm} (3.1a)

where

$$\mathbf{C} = I + h (\hat{B}_1 \mathbf{e}(\theta) A_1 - \hat{B}_{1,r} (\theta) \partial E_1 / \partial Q) + \hat{B}_2 \mathbf{e}(\theta) A_2 - \hat{B}_{2,r} (\theta) \partial E_2 / \partial Q),$$

$$\mathbf{K} = I + h (\hat{B}_1 \mathbf{e}(\theta) A_1 (Q) + \hat{B}_2 \mathbf{e}(\theta) A_2 (Q)) + h \hat{H} \mathbf{H}.$$  \hspace{1cm} (3.1b)

In (3.1), $\hat{B}(\theta)$ is the symbol of the difference operator $\mathbf{B}$, i.e., $\hat{B}(\theta) e^{-i j \theta}$ times the operator $\mathbf{B}$ applied to the mesh function $e^{i j \theta}$. For example, if $\mathbf{B}(w_j) = (w_j + 1 - w_{j+1})/2 \Delta y$, then $\hat{B}(\theta) = i \sin(\theta)/\Delta y$.

The matrices $A_1$, $A_2$, $\partial E_1 / \partial Q$, $\partial E_2 / \partial Q$, and $\partial H / \partial Q$ are evaluated at $\mathbf{x}^*$. The mesh size $\Delta y$ and time step $h$ enter as parameters; the effect on stability of varying these parameters can thus be studied. In particular, upper limits on the CFL number (defined as $(|u| + c) h / \Delta y$) for stability can be predicted. We will try to show (by examples) that if for some $\theta$ there is an eigenvalue of (3.1a) which has modulus greater than 1, then the iteration (2.2) is unstable near $\mathbf{x}^*$, and also that an instability predicted by (3.1a) agrees closely with instability determined by computing all the eigenvalues of the total matrix problem (2.7). Note that for $\theta = 0$ the problem (3.1a) has $\lambda = 1$ an eigenvalue (since by consistency $\hat{B}_1 (0) = \hat{B}_2 (0) = 0$; this also uses the fact that $\partial H / \partial Q$ has its first and last rows identically zero). However, we argue that the Fourier mode $v_j = e^{i j \theta}$ for $\theta = 0$ is to be eliminated from consideration because it is a constant mode and is not allowed by realistic boundary conditions. Experience has shown that an instability frequently appears at a high frequency mode (often with $\theta = \pi$, corresponding to a spatial pattern $+1, -1, +1, -1, \ldots$).

This analysis is like the von Neumann analysis for initial value problems. Notice, however, that we can let $h \to \infty$ in (2.2), obtaining a Newton–like algorithm; the stability of this algorithm near a solution $Q^*$ can then be studied by Fourier analysis. In this case there is no time step or CFL number any more, so the analysis is different from the usual von Neumann analysis. We mention again that we are not considering arbitrary initial values, so also in this case the analysis differs from the von Neumann analysis.

The actual mechanics of the analysis are as follows. Fix a set of flow parameters, some $A \mathbf{x}$, and a CFL number. Pick a convenient integer $N$ (10 or 20, say) for the number of Fourier modes, and let $\theta_n = \pi n / N$ for $1 \leq n \leq N$. For each $\theta_n$ we form the matrices $\mathbf{L}$ and $\mathbf{K}$ of (3.1b) and solve the generalized eigenvalue problem (3.1a). This gives three eigenvalues $\lambda_k(\theta_n)$, $k = 1, 2, 3$. The amplification factor $|\lambda_n|$ for $\theta_n$ is taken to be $\max_{1 \leq k \leq 3} |\lambda_k(\theta_n)|$. Then the maximum over $n$ of the $|\lambda_n|$ is the overall amplification factor; if this is greater than 1 we predict instability. We also compute an “average” amplification factor via

$$\sqrt{\frac{1}{N} \sum_{n=1}^{N} |\lambda_n|^2 / N}.$$  \hspace{1cm} (3.2)

For stable cases it may be possible to predict an optimal or near optimal CFL number which would minimize this average amplification factor. This prediction of optimal CFL number via local Fourier analysis may be relevant for implicit methods which use local time stepping as an acceleration procedure; it may give a good basis on which to pick the local CFL number. Finally, the amplification factor of the $\theta = \pi$ mode, the highest space frequency mode, is computed as $\max_1 \lambda_k(\theta_n)$. This mode is often the mode that goes unstable first, and as such it is of interest to compute its amplification factor.

§4. Total Matrix Analysis.

The total matrix analysis employs an actual computational code which solves the quasi-one-dimensional Euler equations for flow through nozzles. The code can handle arbitrary nozzle shapes, variable mesh spacing and can compute flows with shocks. The implicit delta form of the equations is solved, see (2.2) and (2.3). The fluxes, flux Jacobians, and differencing operators are chosen to agree with corresponding cases from the Fourier analysis and are defined as above.

The strategy is to form the matrices $L$ and $K$ of (2.8) and compute all the eigenvalues of the generalized eigenvalue problem (2.7). The base solution $\mathbf{x}^*$ used to form the matrices is chosen as the converged solution for a specific geometry and set of flow conditions.

The maximum of the moduli of the eigenvalues is computed as a function of CFL number (here CFL number is defined as $\max_1 |\omega_j| + c_j h / (\Delta y)$ where $j$ refers to the mesh points). The maximum eigenvalue computed this way and an $L_2$ average over all the eigenvalues will be compared with the maximum, average, and “$\pi$” $\max$ eigenvalues computed via Fourier analysis.
§5. Numerical Results.

In this section we describe the results of some numerical experiments. A large number of experiments have been run, and representative cases have been chosen. We will graph the modulus of the eigenvalues as a function of the CFL number. Three curves are plotted from the Fourier analysis: the maximum eigenvalue \( \max_k \max_n |\lambda_k(\theta_n)| \), the \( L_2 \) average \( \sqrt{\sum_n |\lambda_n|^2 / N} \), and the maximum eigenvalue at \( \pi \), \( \max_k |\lambda_k(\pi)| \). The results from the total matrix analysis of the generalized eigenvalue problem (2.7) will be plotted as two sets of points, first the modulus of the maximum eigenvalue and second the \( L_2 \)-average of the eigenvalues.

In Figure 1 we show typical results from an analysis of the Steger–Warming flux splitting. In this case a straight nozzle is used and the exact solution \( Q^* \) is constant. For the total matrix analysis Dirichlet boundary conditions were used at the inflow and outflow boundaries. There were 13 interior mesh points, so \( L \) and \( K \) were 39-by-39 matrices. The Mach number \( (u/c) \) was taken to be 0.5. For the Fourier analysis the Mach number was 0.5 and density was 1.0.

Figure 1a shows the case of second-order upwind differencing on both sides of the equation, with the choice \( A_1 = \partial E_1/\partial Q \) and \( A_2 = \partial E_2/\partial Q \) (the correct Jacobians). Twenty Fourier modes were used in the Fourier analysis. In this case the maximum eigenvalue from the Fourier analysis is always less than 1, leading to a prediction of stability for all CFL numbers. Also, the maximum eigenvalue \( \to 0 \) as \( \text{CFL} \to \infty \); this is a reflection of the fact that for infinite time step this method becomes Newton’s method. The points plotted as circles show the maximum eigenvalue from total matrix analysis. The points follow the trend of the curve of maximum eigenvalues from the Fourier analysis. The points plotted as squares show the \( L_2 \)-average of the eigenvalues from computation using the matrices \( L \) and \( K \); these points agree well with the \( L_2 \)-average eigenvalue curve from the Fourier analysis.

In Figure 1b we have the case of second-order differencing on both sides of the equation, with \( A_1 := A_1^+ \neq \partial E_1/\partial Q \) and \( A_2 := A_2^- \neq \partial E_2/\partial Q \) (i.e., not the exact Jacobians). In this case we notice that the Fourier analysis predicts instability for CFL number greater than 2, while the total matrix analysis predicts instability for CFL number greater than 3. The curve from the Fourier analysis gives an excellent prediction of the shape of the curve from the total matrix analysis as well as a good (and, in this case, safe) estimate of the upper stability limit on CFL number. The Fourier analysis also predicts that the *12 = \( \pi \) mode is the first to go unstable.

Figure 1c gives the results of analysis of the method which uses the exact Jacobian matrices but which uses different orders of differencing on the two sides of the equation, here first order spatial differencing on the implicit operator and second order spatial differencing on the explicit operator. This choice is used in two-dimensional problems for reasons of computational efficiency, e.g. in Buning and Steger. In this case the Fourier analysis predicts the stability of the iteration for all CFL numbers, with the maximum eigenvalue being minimized at a CFL number of 10. The total matrix analysis follows this prediction quite closely, with maximum eigenvalue always less than 1, and maximum eigenvalue minimized at a CFL number of 9.

In Figure 2 we show results from an analysis of the velocity–sound speed splitting. Again a straight nozzle and constant \( Q^* \) were used for the total matrix analysis. For Figure 2a the true Jacobians are used for \( A_1 \) and \( A_2 \). The results again show the stability of this method and demonstrate that as CFL \( \to \infty \) (so the algorithm approaches Newton’s method), the maximum eigenvalue \( \to 0 \). In Figure 2b we show the result when the incorrect Jacobians are used for \( A_1 \) and \( A_2 \). In this case the maximum eigenvalue computed via the Fourier analysis is greater than 1 for a CFL number greater than about 12, while the maximum eigenvalue from the total matrix analysis remains less than 1 until CFL \( > 150 \) (approximately). The curves of maximum eigenvalue have the same shape and the Fourier analysis prediction gives a safe upper bound on the CFL number for stability. (We do not claim that the Fourier analysis always gives a safe upper limit on the CFL number, but it did in this case.) The results from analysis of the velocity–sound speed splitting are always quite similar to the results from the plus–minus splitting, so in the following we will only show results from experiments with the plus–minus splitting.

We will next show some results from experiments using nonconstant \( Q^* \). We consider flow in an expanding nozzle, with the exact solution having a shock at about \( x = 4.8 \). Graphs of the density, momentum, and energy for the numerical solution \( Q^* \) are given in Figure 3. The flow is supersonic ahead of and subsonic behind the shock; the inlet Mach number is 1.26. The mesh used for computing \( K \) and \( L \) for the total matrix analysis was clustered in the vicinity of the shock. The nozzle shape was that used by Yee.

Figure 4 shows results of the total matrix analysis with the matrices \( K \) and \( L \) computed using Dirichlet boundary conditions, with the boundary values from the exact solution. In Figure 4a we have the situation with second-order differencing on both sides of the equation and the correct Jacobians used for \( A_1 \) and \( A_2 \). The Fourier analysis predicts unconditional stability, with the maximum eigenvalue tending to 0 as CFL \( \to \infty \), and this prediction is borne out by the total matrix analysis. Figure 4b shows the situa-
tion with second-order differencing on both sides of the equation and the incorrect Jacobians used for \( A_1 \) and \( A_2 \). In this case the Fourier analysis predicts instability for CFL number greater than about 2, while the total matrix analysis predicts instability for CFL number greater than 10. The Fourier analysis predicts the correct trend of the maximum eigenvalue curve and gives a reasonable quantitative estimate of the upper limit for stability. This is rather remarkable, since the Fourier analysis assumes constant mesh size and is based on flow conditions at only one point (a Mach number of 0.5 which is consistent with the flow conditions downstream of the shock), while in the total matrix analysis the mesh is nonuniform and the flow condition \( Q^* \) varies quite a bit over the nozzle.

In Figure 4c we show the result of the analysis when the correct Jacobians are used and the differencing is first-order on the left, second-order on the right. The Fourier analysis predicts unconditional stability, with maximum eigenvalue minimized at about CFL = 5, while the total matrix analysis shows unconditional stability with maximum eigenvalue \( Q^* \) minimized at about CFL = 11.

In Figure 5 we have the same situation as in Figure 4, except now the matrices \( K \) and \( L \) are computed using a space extrapolation boundary condition, where momentum and energy are extrapolated and density is fixed at the subsonic outflow. These boundary conditions are imposed implicitly, and they have been shown to be well posed and unconditionally stable (when coupled with an unconditionally stable interior scheme) by Yee, Ref. [6]. Also see this reference for a more detailed description of the diverging nozzle flow and the implicit method.

Figure 53 shows the stability of the algorithm when the correct Jacobians and second-order differencing on both sides are used. The results here are very similar to those in Figure 4a. Figure 5b gives the situation when the incorrect Jacobians are used on the left, hand side. Again the curve from the Fourier analysis gives a conservative prediction of the instability bound, with predicted stability limit of about CFL = 2, while the total matrix analysis shows instability for CFL greater than about 10.

In Figure 6 we give results of the analysis for a smooth subsonic exact solution in a converging-diverging nozzle. Figure 7 shows a plot of the solution \( Q^* \); the inlet Mach number was 0.2. A uniform mesh and Dirichlet boundary conditions were used. Figure 6a shows the by-now-familiar result when the correct Jacobians are used, while the results when the incorrect Jacobians are used are given in Figure 6b. Once again the Fourier analysis predicts the correct trend and gives a good estimate of the upper CFL limit for stability.

In the preceding result the Fourier analysis was done using a Mach number of 0.2, which corresponds to the inflow condition for the total matrix analysis. We can examine the variation of the Fourier results by computing the eigenvalues for a range of Mach numbers corresponding to Mach numbers in the flow. This case was chosen for this study because the exact solution is smooth and the results will not be distorted by the presence of a shock or other flowfield peculiarities. Figure 8 shows the maximum eigenvalue from the Fourier analysis for a range of Mach number between 0.2 and 0.4 (0.4 was the maximum Mach number of the exact solution). (Note the the total matrix analysis solution is now the solid line and the symbols giving the results of the Fourier analysis are now not connected.) We see that the total maximum from the Fourier analysis varies between 1.6 and 3.1 but the overall prediction of a stability bound in terms of CFL number is consistent.


We have considered iterative methods for the steady-state Euler equations; these iterative methods can be thought of as modifications of Newton's method. The convergence behavior of these methods near a solution has been shown to be determined by a matrix eigenvalue problem. A Fourier method has been proposed to analyze these eigenvalue problems. Comparison of the Fourier method with a numerical total matrix analysis of the full eigenvalue problem shows good agreement even with shocks and clustered meshes.

We would like to emphasize that this analysis is essentially bound up with the analysis of systems of differential equations as opposed to scalar equations. The use of the "incorrect" Jacobian matrix is rare in the scalar case—it is a phenomenon that usually occurs only for systems of equations. The Fourier analysis is similar to the usual analysis for initial-value problems. The differences are that total matrix analysis of the generalized eigenvalue problem (2.7) gives a necessary and sufficient condition for stability of the iteration (2.2) near a solution \( Q^* \), and we make no claim that the Fourier analysis gives either a necessary or sufficient condition for convergence, only that the Fourier analysis gives good qualitative and quantitative predictions about the stability of these modified Newton algorithms.

One of the original motivations for this work was the experimental finding of Steger and Warming that use of the incorrect Jacobian matrices in the plus–minus algorithm led to an upper limit on the CFL number for stability. Our analysis for the one-dimensional case has shown rigorously the existence of such a CFL limit. Also, we have considered the use of first order differencing on the implicit operator and second order differencing on the explicit operator. We have shown this scheme is stable and has an optimum CFL number of about 10.
These techniques will be extended in the near future to the two-dimensional case, to help in studying the questions that arise not only from incorrect Jacobian matrices but also from the approximate factorization commonly employed in implicit algorithms in two dimensions.

References


Fig. 1. Results for flux splitting, straight nozzle, constant $Q^*$, fixed boundary conditions.

Fig. 2. Results for velocity–sound speed splitting, straight nozzle, constant $Q^*$, fixed boundary conditions.
4a. Second order differencing, exact Jacobians.

4b. Second order differencing, exact Jacobian on right, approximate Jacobian on left.

4c. First order differencing on left, second order differencing on right, exact Jacobians.

5a. Second order differencing, exact Jacobians.

5b. Second order differencing, exact Jacobian on right, approximate Jacobian on left.

Fig. 4. Results for flux splitting, diverging nozzle, fixed boundary conditions.

Fig. 5. Results for flux splitting, diverging nozzle, extrapolated boundary conditions.
6a. Second order differencing, exact Jacobians.

6b. Second order differencing, exact Jacobian on right, approximate Jacobian on left.

Fig. 6. Results for flux splitting, converging-diverging nozzle, subcritical solution, fixed boundary conditions.

Fig. 7. Numerical solution $Q^*$ for converging-diverging nozzle.

Fig. 8. Maximum eigenvalue from Fourier analysis for three different Mach numbers.

---

REPRODUCIBILITY OF THE ORIGINAL PAGE IS POOR