UCLA
COMPUTATIONAL AND APPLIED MATHEMATICS

# Iterative Gardient-Newton Type Methods for Steady Shock Computations

Bjorn Engquist
Quyen Quang Huynh

November 1988

CAM Report 88-40

Department of Mathematics
University of California, Los Angeles
Los Angeles, CA. 90024-1555

# Iterative Gradient-Newton Type Methods for Steady Shock Computations

Bjorn Engquist[1]

Department of Mathematics
University of California
Los Angeles, CA 90024, U.S.A.

and

Quyen Quang Huynh

Department of Computer Sciences
Uppsala University, Sturegatan 4B 2TR
75223 Uppsala, Sweden

A class of modified Newton's methods are applied to difference approximations of the two-dimensional steady Burgers' equation and the transonic small disturbance equation. The solutions have sharp gradients which correspond to boundary layers and shock waves in fluid dynamics. The nonlinear terms in the differential equations are approximated by modern shock capturing schemes. The regularity of the coefficients is analyzed theoretically and its effect on the convergence on the Newton's method is studied numerically. Computational results from different types of gradient iterative methods and different types of preconditioners are presented. These methods are applied to the linear systems of the Newton iteration. The relative residuals in the Newton iterations are controlled such that a superlinear rate of convergence is preserved.

**1. Introduction.** We shall consider numerical solutions of nonlinear hyperbolic conservation laws. In two space dimensions the equation or system of equations have the form,

$$u_t + f(u)_x + g(u)_y = 0, \tag{1.1}$$

with appropriate initial and boundary conditions. These equations are used as mathematical models in many applications. In gas dynamics, for example, the unknown vector valued function $u(x, y, t)$ has four components representing density, momentum (2 components) and energy.

Even with smooth initial values the solution of (1.1) generically develops discontinuities. These discontinuities or shocks cause both theoretical and computational difficulties.

[1]Research supported by NSF-grant No. DMS88-11863, ONR-grant No. N00014-86-K-01691, AFOSR-grant No. AFOSR-87-0341 and NASA Consortium No. NCA2-372.

Weak solution must be considered and uniqueness might be a problem. For scalar equations there are unique solutions if extra constraints (entropy conditions) are added [Smoller, 83]. The solution is e.g. given as the limit of vanishing viscosity solutions,

$$\lim_{\varepsilon \to 0} \int \int |u^\varepsilon(x,y,t) - u(x,y,t)| dxdy = 0,$$
$$u_t^\varepsilon + \delta(u^\varepsilon)_x + g(u^\varepsilon)_y = \varepsilon \Delta u^\varepsilon. \tag{1.2}$$

The numerical approximations of (1.1) must work well when $u$ is smooth but also at discontinuities of $u$. Traditionally numerical schemes mimicked the equation (1.2). The standard schemes contain such added artificial viscosity, [Rizzi, Engquist, 87]. During the last fifteen years new classes of so called high resolution schemes have been developed. See e.g. the survey [Colella, Woodward, 84]. These algorithms are based more directly on the properties of (1.1) at discontinuities. The result is often approximations with sharp shocks without numerical oscillations. These algorithms are nonlinear and change structure depending on the solution. This adaptivity is a source of difficulties for direct steady state computations. The purpose of this paper is to study a few questions in the coupling of high resolution schemes with modern algebraic methods for steady state computations.

A three point difference approximation of (1.1) has the form,

$$\begin{aligned} u_{i,j}^{u+1} &= u_{i,j}^u - \frac{\Delta t}{\Delta x}(f(u_{i+1,j}^u, u_{i,j}^u) - f(u_{i,j}^u, u_{i-1,j}^u)) \\ &\quad - \frac{\Delta t}{\Delta y}(g(u_{i,j+1}^u, u_{i,j}^u) - g(u_{i,j}^u, u_{i,j-1}^u)), \ t_n = n\Delta t. \\ u_{i,j}^u &\sim u(x_i, y_j, t_n), \ x_i = i\Delta x, y_j = j\Delta y. \end{aligned} \tag{1.3}$$

The functions $f(\ ,\ )$ and $g(\ ,\ )$ are called numerical flux functions and are related to $f$ and $g$ in (1.1) via the consistency relation,

$$f(u,u) = fu, \quad g(u,u) = g(u). \tag{1.4}$$

It is common to compute with (1.3) for large time in order to approximate the steady solution of (1.1).

We shall here consider direct approximations of the steady equation,

$$f(u)_x + g(u)_y = 0, \tag{1.5}$$

with boundary conditions. When direct approximations are feasible they are usually much faster than time evolution techniques. There is no time index $n$ in the equations and the following system of algebraic equations has to be solved

$$F(U) = 0, \tag{1.6}$$

where $F$ and $U$ are vectors with components,

$$F_{i,j} = \frac{1}{\Delta x}(f(u_{i+1,j}, u_{i,j}) - f(u_{i,j}, u_{i-1,j})) + \frac{1}{\Delta y}(g(u_{i,j+1}, u_{ij}) - g(u_{i,j}, u_{i,j-1})),$$
$$U = (u_{ij}).$$

Some of the equations in $F$ should also contain the boundary conditions.

We mentioned earlier the adaptive feature of the high resolution schemes. This means that $f(\ ,\ )$ and $g(\ ,\ )$ depends strongly on the solution and often not in a smooth way. Thus $F$ is not a smooth function of $U$ which causes trouble when solving (1.6).

In the following section we shall study the regularity of $F$ and show that very sharp shock resolution and $f \in C^1$ is mutually exclusive. We have to relax the sharpness of the discontinuities in the numerical solution in order to have an algebraic system which is suitable for numerical methods.

Some natural algebraic methods are outlined in section 3. The linear system resulting from a damped inexact Newton method is approximated by preconditioned gradient type methods.

In section 4 we shall apply these methods to two hyperbolic conservation laws in two space dimensions. The Burgers' equation

$$(\frac{1}{2}u^2)_x + (\frac{1}{2}u^2)_y = \epsilon \Delta u \tag{1.7}$$

and the transonic small disturbance equation,

$$(K\phi_x - \frac{1}{2}(\gamma+1)\phi_x^2)_x + \phi_{yy} = 0 \tag{1.8}$$

are studied. In (1.7) we have a conservation law plus added viscosity. As in equation (1.2) we are interested in small values of $\epsilon$.

**2. Regularity of Numerical Fluxes.** Consider Newton's method applied to the system (1.1): $F(U) = 0$. In the Kantorovich convergence theorem, the mapping $F$ is assumed to be continuously differentiable.

Let us consider a three point formula for the one dimensional Burgers' equation. The nonlinear term $f(u)_x = 0.5(u^2)_x$ is discretized as follows

$$[f(u)]_x = 1/\Delta x[f(u_i, u_{i+1}) = f(u_{i-1}, u_i)] \tag{2.1}$$

where $f(\ ,\ )$ represents the numerical flux.

We shall now present various numerical fluxes for the Burgers' equation and compare their accuracy and smoothness properties. The shock speed is given by $u_S = 0.5(u_\ell + u_r)$, $f_+(u) = 0.5\max(u,0)^2$, $f_-(u) = 0.5\min(u,0)^2$.

*Godunov flux* (first order) [Godunov, 59]

$f_G(u_\ell, u_r) = \max[f_+(u_\ell), f_-(u_r)]$

If $u_S = 0$ i.e. $u_1 = -u_r$, then $\partial f_G/\partial u_1 = u_\ell$ or 0

*Roe flux* (first order) [Roe, 85]

$f_R(u_\ell, u_r) = 0.5[f(u_\ell) + f(u_r)] - 0.5|0.5(u_\ell + u_r)|(u_r - u_\ell)$

If $u_S = 0$ then $\partial f_R/\partial u_\ell = u_\ell$ or 0

*E-O flux* (first order) [Engquist, Osher 80]

$f_{E-O}(u_\ell, u_r) = f_+(u_\ell) + f_-(u_r)$

If $u_S = 0$ then $\partial f_{E-O}/\partial u_\ell = u_1$

*TVD flux* (second order, with van Leer limiter) [van Leer, 74]

$$\begin{aligned} f_{\mathrm{TVD}}(u_{i-1}, u_i, u_{i+1}, u_{i+2}) = & f_{E-O}(u_i, u_{i+1}) = 0.5\Psi(R^-_{i+1}) \\ & (f_{E-O}(u_i, u_{i+1}) - f(u_i)) + 0.5\Psi(R^+_{i+1}) \\ & (f(u_{i+1}) - f_{E-O}(u_i, u_{i+1})) \end{aligned}$$

where

$$\begin{aligned} R^+_i &= (f(u_i) - f_{E-O)}(u_{i-1}, u_i))/(f(u_{i+1}) - f_{E-O}(u_i, u_{i+1})) \\ R^-_i &= (f_{E-O}(u_i, u_{i+1}) - f(u_i))/(f_{E-O}(u_{i-1}, u_i) - f(u_{i-1})) \end{aligned}$$

$\Psi_{VL}(R) = (|R| + R)/(1 + |R|)$

Clearly the first derivatives of $f_G$ and $f_R$ exhibit jumps while $f_{E-O}$ is a $C^1$ function. The flux limiter $\Psi_{VL}(R)$ is differentiable.

It was shown in [Engquist, Osher, 80], that the E-O scheme admits a discrete representation of a steady shock with two interior states. Next, we shall show that for a three points scheme admitting a steady shock profile with at most one interior state, its numerical flux cannot be a $C^1$ function. Thus the sharpest possible steady discrete shock profile has two interior states for schemes with $C^1$ numerical fluxes. The Godunov and Roe schemes have discrete shock profiles with one interior state but their corresponding numerical fluxes are not $C^1$.

Consider a scalar hyperbolic conservation law

$$u_t + [f(u)]_x = 0 \tag{2.2}$$

with $f'' > 0$ and $f'(0) = 0$.

Eq. (2.2) is approximated by a three points scheme in conservation form,

$$u_i^{n+1} = u_i^n + \lambda[f(u_i^n, u_{i+1}^n) - f(u_{i-1}^n, u_i^n)], \tag{2.3}$$

where $\lambda = \Delta t/h$.

The regularity property of numerical flux is given in the following theorem.

THEOREM. *Let the scalar hyperbolic conservation law* (2.2) *be approximated by a consistent three points scheme in conservation form. Assume that this scheme admits the following discrete representation of a steady shock* $u_\ell > 0; f(u_\ell) = f(u_r)$, *and for any* $u_m, u_\ell \geq u_m \geq u_r$

$$\begin{aligned} u_i &= u_\ell, \quad i \leq -1 \\ u_0 &= u_m, \\ u_i &= -u_\ell, \quad i \geq 1 \end{aligned}$$

*Then the numerical flux* $f(u_\ell, u_r)$ *cannot be a* $C^1$ *function.*

*Proof.* If (2.3) is applied to the point $i = -1$, then in terms of the states $u_1$ and $u_0$, we have

$$u_{-1}^{n+1} = u_{-1}^n + \lambda[f(u_1^n, u_0^n) - f(u_1^n, u_1^n)] \tag{2.4}$$

Since there are multiple discrete representations of the same steady shock profile with one interior state $u_0$, $f(u_1, u_0)$ depends only on $u_1$ i.e.

$$\begin{gathered} f(u_0, u_r) - f(u_r), \ u_r < 0 \\ u_1 > u_0 > u_r \end{gathered} \tag{2.5}$$

Hence from (2.4) and (2.5)

$$f(u, v) = f(u), \ u > 0, \ u > v > \underline{u} \text{ with } f(\underline{u}) = f(u) \tag{2.6}$$

$$f(u, v) = f(v), \ v < 0, \ \underline{v} > u > v \text{ with } f(\underline{v}) = f(u) \tag{2.7}$$

From (2.6) and (2.7), it follows that $f(u, v)$ cannot be a $C^1$ function since $\nabla f(\ ,\ )$ is discontinuous at $(u, v) = (u, \underline{u})$ or equivalently $(u, v) = (\underline{v}, v)$.

**3. Algebraic Methods.** Consider the inexact Newton's method (IN) applied to (1.6),

$$U^{k+1} = U^k + S^k, \tag{3.1a}$$

$$J(U^k)S^k = -F(U^k) + r^k. \tag{3.1b}$$

The method is called inexact if there is an error ($r^k \neq 0$) in the solution of the linear system (3.1b). The error is controlled by a sequence $\{g_k\}$ such that $||r^k||/||f(U^k)|| \leq q_k$. The Jacobian matrix of $F$ is denoted by $J$. Modifications in order to improve the global convergence properties can be done at the updating step (3.1a) as follows

$$U^{k+1} = U^k + \alpha_k S^k$$

The method is then called damped inexact Newton (DIN). The idea of a global method is to make sure that each step decreases the value of some norm of $F : R^n \rightarrow R^n$. If we choose the $l_2$ norm $||f(U)||$, solving the system of nonlinear equations $F(U) = 0$ is equivalent to minimizing $g = 1/2F(U)^T\ F(U)$.

Naturally one wants to choose a direction $S$ such that in this direction

$$g(U^k + \alpha_k S^k) < g(U^k) \quad \text{for some } 0 < \alpha_k \leq 1$$

It is easy to show that the vector $S^k$ is a descent direction $\nabla\ g(U^k)^T S < 0$, if $||r^k||$ is small enough.

Our damped inexact Newton method is based on an algorithm in [Dembo, Steihaug, 83]. We have incorporated a very simple backtrack technique instead of a more complicated quadratic or cubic backtrack which is described in [Dennis, Schnabel, 83]. Starting with $\alpha_k = 1$, $\alpha_k$ is reduced by a factor $\frac{1}{2}$ until a descent condition is satisfied.

From (3.1b) we see that at each Newton step, a large linear system of the form

$$AU = b \tag{3.2}$$

needs to be solved. Several iterative gradient methods have been proposed recently to solve (3.2), where $A$ is nonsymmetric and possesses a positive definite symmetric part. In our examples, we have found the truncated GCR method called ORTHOMIN(i) to be particularly attractive in term of computational effort and storage [Vinsome, 76]. This method, is a modification of the GCR method where only the last i direction vectors need to be saved. It is worth pointing out that the truncated version of the GMRES method proposed in [Saad, Schultz, 86], does require the positive definiteness of the symmetric part of $A$ although its full version does not.

A survey and comparison of generalized gradient methods for nonsymmetric problems, is given by in [Saad, Schultz, 85]. We choose to adopt here the minimal residual (MR) method and the ORTHOMIN(1) method because of their simplicity. In particular, the MR method which is identical to ORTHOMIN(0) is a simple two-steps algorithm.

For symmetric problems, the convergence of iterative gradient methods can be accelerated by reducing the condition number of $A$. It is also well known that the rate of convergence depends on the clustering of eigenvalues into groups. However a similar theory does not exist in general for nonsymmetric problems. It is therefore necessary to conduct extensive numerical experiments for nonsymmetric matrices.

Preconditioning techniques transform the original matrix into a matrix with better properties. If $C$ is a preconditioning matrix, instead of solving $AU = b$, we solve $AC^{-1}CU = b$.

All the preconditioners discussed in this paper were first constructed for symmetric matrices with $C = LL^T$. We generalize them to nonsymmetric cases by choosing $C = LU$ such that $\mathrm{diag}(U) = I$. We also make sure that $C(= LU)$ is symmetric when $A$ is symmetric.

The incomplete factorization method which was first proposed in [Dupont, Kendall, Rachford, 68] for self-adjoint elliptic difference equations will be described here for five points schemes approximating the linear advection-diffusion equation, (3.3), which can be seen as a linearization of the Burgers' equation:

$$\epsilon \nabla^2 \phi - u\phi_x - v\phi_y = 0 \tag{3.3}$$

Eq. (3.3) is written in finite difference form as

$$(A\phi)_{ij} = s_{ij}\phi_{i,j-1} + w_{ij}\phi_{i-1,j} + c_{ij}\phi_{i,j} + e_{ij}\phi_{i+1,j} + n_{ij}\phi_{i,j+1} = 0.$$

The resultant matrix $A$ is sparse and nonsymmetric, and has five diagonals. It is possible to approximate $A$ in the form, $C = A + R$, where $C$ is the product $LU$ and $R$ is the defect matrix. $L$ and $U$ are defined to be respectively the lower and upper triangular matrices with no more than three entries per row,

$$(L\phi)_{ij} = v_{ij}\phi_{ij} + t_{ij}\phi_{i-1,j} + g_{ij}\phi_{i,j-1} \tag{3.4}$$

$$(U\phi)_{ij} = \phi_{ij} + f_{ij}\phi_{i,j+1} + k_{ij}\phi_{i+1,j} \tag{3.5}$$

The product $LU$ has seven diagonals

$$\begin{aligned}(LU\phi)_{ij} = {} & s_{ij}\phi_{i,j-1} + w_{ij}\phi_{i-1,j} + d_{ij}\phi_{i,j} + e_{ij}\phi_{i+1,j} \\ & + n_{ij}\phi_{i,j+1} + y_{ij}\phi_{i-1,j+1} + z_{ij}\phi_{i+1,j-1} = 0.\end{aligned} \tag{3.6}$$

The new points $y_{ij}$ and $z_{ij}$ involved in the product $LU$ are those corresponding to $\phi_{i-1,j+1}$ and $\phi_{i+1,j-1}$ respectively.

We choose to equate the non-zero elements of $A$ which are off the main diagonal with the corresponding elements of $LU$. We shall make one assumption which is row-sum$(A) \geq 0$. If we impose, row-sum$(A)$ = row-sum$(C)$, we can solve uniquely for the five elements $v, t, g, f$ and $k$ in term of $w, e, s, n$ and $c$. They are given recursively by the following formulas

$$
\begin{aligned}
&W - \textit{Preconditioner} \quad ([\text{Wong, 78}]) \\
&t_{ij} = w_{ij}, \quad g_{ij} = s_{ij} \\
&v_{ij} = c_{ij} - g_{ij} f_{i,j-1} - t_{ij} k_{i-1,j} - t_{ij} f_{i-1,j} - g_{ij} k_{i,j-1} \\
&f_{ij} = n_{ij}/v_{ij}, \quad k_{ij} = e_{ij}/v_{ij}
\end{aligned}
$$

Here we adopt the convection that the elements $t, g, v, f$ and $k$ are set to zero if they cannot be computed by the above algorithm. It is easy to see that the row-sum of $R$ is zero by construction.

DKR-*Preconditioner* [Dupont, Kendall, Rachford, 68] is the same as $W$-preconditioner, except for the formula for $v$,

$$v_{ij} = (1 + \alpha h^2) c_{ij} - g_{ij} f_{i,j-1} - t_{ij} k_{i-1,j} - t_{ij} f_{i-1,j} - g_{ij} k_{i,j-1}.$$

We denote the preconditioning in [Meijerink, Van Der Vorst, 77] by the MV preconditioning. It forces equality of the elements of the preconditioning matrix $C$ and the matrix $A$ on the diagonals in positions defined by the non-zero diagonals of the matrix $A$. The MV preconditioner is constructed by the following formulas,

MV-*Preconditioner* is the same as $W$-preconditioner, except for another formula for $v$

$$v_{ij} = c_{ij} - g_{ij} f_{i,j-1} - t_{ij} k_{i-1,j}.$$

All the preconditioning techniques described above can be applied to symmetric and nonsymmetric matrices. As an example of a preconditioned gradient method, we show the preconditioned ORTHOMIN(1) method.

$X^0$ given, let $r^0 = b - AX^0$
Solve $CZ^0 = r^0$ and set $p^0 = Z^0$
For $k = 0$ step 1 until convergence *do*
$\alpha_{k+1} = (r^k, Ap^k)/(Ap^k, Ap^k)$
$X^{k+1} = X^k + \alpha_{k+1} p^k$, $r^{k+1} = r^k - \alpha_{k+1} Ap^k$
Solve $CZ^{k+1} = r^{k+1}$
$p^{k+1} == Z^{k+1} + \beta_k p^k$, $\beta_k = -(AZ^{k+1}, Ap^k)/(Ap^k, Ap^k)$

**4. Two Nonlinear Examples.** The first example is the *Burgers's equation* (1.7) in the square $0 \leq x, y \leq 1$. We shall use central differences to discretize the second

derivatives terms and the shock capturing schemes to approximate the nonlinear terms. We obtain an equation $F(U) = 0$ where the unknowns $u_{ij}$ are ordered as the components of the vector $U$.

Let us first verify directly the regularity of the E-O differencing

$$[f(u)]_x \approx 1/\Delta x[\Delta_+ \ f_-(u_i) + \Delta_- f_+(u_i)].$$

The following estimates are valid for the Burgers' equation:

$$|f_+(a) - f_\pm(b) - f'_\pm(b)(a-b)| \leq 0.5\ (a-b)^2 \quad \forall\ a, b \tag{4.1}$$

$$|f'_\pm(a) - f'_\pm(b)| \leq |a-b| \quad \forall\ a, b \tag{4.2}$$

Using (4.1) and (4.2), we can show that there exists constants $C_1$ and $C_2$ which depend on the grid size $h$ such that:

$$||F(U) - F(V) - J(V)[U-V]|| \leq C_1||U-V||^2 \quad \forall\ U, V \in R^n \tag{4.3}$$

$$||J(U) - J(V)|| \leq C_2||U-V|| \quad \forall\ U, V \in R^n \tag{4.4}$$

Therefore the mapping $F$ is twice Frechet differentiable, which guarantees convergence of the Newton's method.

Let $U^*$ be the solution of (4.1). In order to ensure convergence, the initial guess $U^0$ ought to be chosen so that [Dennis, Schnabel, 83]

$$\begin{aligned} &||U^0 - U^*||1/2\alpha\beta, \\ &||J(U^*)^{-1}|| \leq \alpha, ||J(U^0) - J(U^*)|| \leq \beta||U^0 - U^*||. \end{aligned} \tag{4.5}$$

The relation (4.5) tells us that the radius of convergence is inversely proportional to the product $\alpha\beta$. The size of $\alpha$ depends strongly on the structure of the solution. If there is a boundary layer solution (out Type A below) the eigenvalues of $J(U^*)$ are well bounded away from zero. In the one dimensional case $\sigma(J) \leq -C/\varepsilon$, [Kreiss, Kreiss, 86]. For shock solutions (our Type B below) the radius of convergence is much smaller. There are eigenvalues of $J$ of the order $e^{-1/\varepsilon}$.

The 2-D Burgers's equation (4.1) was solved on a square with two sets of boundary conditions (Figs. 1 and 2).

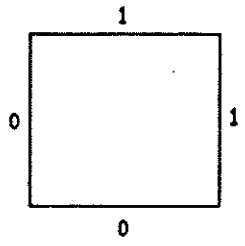


Fig. 1. Type A.

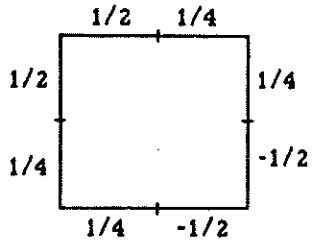


Fig. 2. Type B.

Note that in the case of type B boundary condition (Fig. 2), there is a jump in the middle of each side of the square. These jumps will indeed trigger switching mechanism of various upwind schemes.

The computations were done with a $31 \times 31$ grid if not otherwise noted. The iterative procedure stops when the norm of the residue $||R||$ is less than $10^{-5}$. At each Newton step, the linear system is solved iteratively by a minimal residue method (MR) and the convergence criteria are fixed at 20 MR iterations or inner residue $||r|| < 10^{-5}$. In table 1 are listed the numbers of Newton iterations for the Godunov, Roe and E-O schemes. The number of iterations are the same for these three schemes which are identical for the type A boundary conditions. With the type A boundary conditions there are no switchings involved because the boundary layers are at $x = 1$ and $y = 1$. The rapid convergence is indicated by the computational results. Note that the total number of MR iterations decreases as $\epsilon$ becomes smaller. This is due to the fact that the Jacobian matrix reduces to a triangular matrix for $\epsilon = 0$ and the LU factorization of the Jacobian matrix is exact.

In table 2 are reported the number of Newton iterations for the type B boundary conditions. The linear system at each Newton step is preconditioned by the DKR, W and MV techniques. the MV preconditioner proves to be the most robust while the inner iteration together with the DKR and W preconditioners fails to converge for $\epsilon \leq 10^{-2}$. In table 3, as predicted the Newton iteration together with the E-O scheme converges for all values of $\epsilon$ and the number of iterations does not vary greatly with respect to $\epsilon$.

1. Number of iterations , ( . ) = total number of MR iterations,

type A boundary conditions , DKR ($\alpha$ = 1) preconditioner

| $\varepsilon$ | $10^{-1}$ | $10^{-2}$ | $10^{-3}$ | $10^{-4}$ | $10^{-5}$ | $10^{-6}$ |
|---|---|---|---|---|---|---|
| | 5 (95) | 6 (79) | 9 (57) | 10 (37) | 10 (27) | 10 (16) |

2. Number of iterations , ( . ) = total number of MR iterations,

type B boundary conditions , E-O scheme

* no convergence

| $\varepsilon$ | $10^{-1}$ | $10^{-2}$ | $10^{-3}$ | $10^{-4}$ | $10^{-5}$ | $10^{-6}$ |
|---|---|---|---|---|---|---|
| DKR $\alpha = 1$ | 5 (96) | * | * | * | | |
| W | 6 (115) | * | * | * | | |
| MV | 13 (225) | 8 (142) | 9 (121) | 12 (147) | 15 (200) | 16 (214) |

3. Number of Newton iterations , ( . ) = total number of MR iterations,

type B boundary conditions , MV preconditioner

* no convergence

| $\varepsilon$ | $10^{-1}$ | $10^{-2}$ | $10^{-3}$ | $10^{-4}$ | $10^{-5}$ | $10^{-6}$ |
|---|---|---|---|---|---|---|
| Godunov | 10 (182) | 8 (146) | 10 (119) | 20 (217) | 30 (368) | * |
| Roe | 12 (215) | 7 (140) | 14 (200) | * | * | * |
| E-O | 13 (225) | 8 (142) | 9 (121) | 12 (147) | 15 (200) | 16 (214) |

4. Number of Newton iterations , ( . ) = total number of MR iterations

E-O schemes , type B boundary conditions , MV preconditioner

| $\varepsilon$ | $10^{-1}$ | $10^{-2}$ | $10^{-3}$ | $10^{-4}$ | $10^{-5}$ | $10^{-6}$ |
|---|---|---|---|---|---|---|
| h = 1/14 | 4 (77) | 5 (63) | 8 (54) | 9 (55) | 11 (54) | 11 (56) |
| h = 1/30 | 13 (225) | 8 (142) | 9 (121) | 12 (147) | 15 (200) | 16 (214) |
| h = 1/44 | 10 (400) | 7 (280) | 9 (287) | 14 (353) | 17 (419) | 22 (567) |

Tables 1-4

Table 4 reports the number of Newton iterations for three different grid sizes. We observe that the total work is still very modest with respect to the number of unknowns. The TVD scheme did not converge for $\varepsilon \leq 10^{-3}$. For $\varepsilon = 10^{-2}$, 12 Newton iterations were needed for convergence.

To illustrate the efficiency of different preconditioners on the eigenvalues of the matrix A in (3.2) we shall consider the linearization (3.3) of the Burgers' equation. A linear upwind scheme [Fiadeiro, Veronis, 77] with the same character as our earlier algorithms for the nonlinear equations is used in the discretization. The velocities are given by $u = 1 - x,\ v = y$ and the boundary conditions are

$$\phi(0,y) = \phi(x,0) = 0 \quad \phi(1,y) = \phi(x,1) = 1 \tag{4.6}$$

We choose a matrix $A$ of order 225 and compute the eigenvalues of $A$ and $C^{-1}A$. The matrix $C$ is constructed by the three preconditioning algorithms described above. Figs. 3 to 6 show plots of eigenvalues in increasing order of magnitude for $\varepsilon = 1$. All eigenvalues computed are real and negative. In terms of the ratio $r = \lambda_{\max}/\lambda_{\min}$ for $\varepsilon = 1$, the W and DKR preconditioners perform better than the MV preconditioner as predicted from symmetric cases.

The coefficient matrix $A$ resulting from a certain ordering of the finite difference equations, becomes a triangular matrix as $\varepsilon$ tends to zero. Moreover if $A$ is a triangular matrix, the DKR, W and MV are all exact factorizations i.e. $A = C = LU$. Hence the matrix $C^{-1}A$ is "better conditioned" for smaller $\varepsilon$.

Our final examples is the *transonic small disturbance equation* (TSD) which is written in conservative form

$$[K\phi_x - 1/2(\gamma+1)\phi_x^2]_x + \phi_{yy} = 0, \tag{4.7}$$

or

$$-[f(u)]_x + v_y = 0, \tag{4.8}$$

where

$$f(u) = 1/2(\gamma+1)\phi_x^2 - K\phi_x,$$

$$u = \phi_x, \quad v = \phi_y, \quad K = (1-(M_\infty)^2)/(\delta^{2/3}M_\infty)$$

Then the E-O approximation of Equation (5.2) gives us ($u_i = \Delta_-^x \phi_{ij}/h$)

$$-1/h[\Delta_+^x f_-(u_i) + \Delta_-^x f_+(u_i)] + 1/h^2 \Delta_+^y \Delta_-^y \phi_{ij} = 0 \tag{4.9}$$

Let $f'(u) = (\gamma+1)u - K$ and $\underline{u} = K/(\gamma+1)$.

Note that $\underline{u}$ is determined from $f'(\underline{u}) = 0$. Then

$$f'_+(u) = f'(u) \text{ if } u \geq \underline{u}; \quad f'_+(u) = 0 \text{ if } u < \underline{u}$$

and

$$f'_-(u) = f'(u) \text{ if } u \leq \underline{u}; \quad f'_-(u) = 0 \text{ if } u > \underline{u}$$

Similar to the Burgers' equation, we can show that the numerical fluxes are twice Frechet differentiable and this property guarantees convergence of the Newton iteration.

The Jacobian matrix $J$ with respect to the linearized E-O schemes possesses a structure of 6 diagonals. The matrices $L$, $U$ and $C$ have the same diagonal structures as in (3.4), (3.5) and (3.6) respectively. In the derivations of the DKR and MV preconditioners, the leftmost diagonal of $J$ is ignored. For the W-preconditioner, the entries of $L$ and $U$ are functions of all the elements of $J$ with rowsum$(C)$ = rowsum$(J)$.

For the TSD equation, ORTHOMIN(1) was used to solve the Newton equation. The computations were done for a parabolic arc airfoil with a thickness ratio $\delta = 6\%$. Unless mentioned explicitly otherwise, all computations were done on a grid $51 \times 30$. The iterative process stops when $||R||_2 < 10^{-5}$. The E-O scheme gave the smallest number of iterations and the displayed results are with this scheme. Godunov's and Roe's schemes did also give converging results. Table 5 shows performance of the Newton's method with respect to the W, DKR and MV preconditioners which are applied to the inner iterations. For $M_\infty = 0.895$ and $M_\infty = 0.916$, the ORTHOMIN(1) method with the W and DKR preconditioners does not converge. As in the case of the Burgers' equation, the MV preconditioner proves to be the most robust.

Next the forcing sequence $\{q^k\}$ will be invoked to control how accurately the Newton equation should be solved. In all the computational results presented here we use the initial $q^0 = 0.1$. Recall that $q^k$ is given by

$$q^{k+1} = c||r^k||/||F(\Phi^k)||, \quad 0 < c < 1.$$

Numerical results are presented in table 6 for different values of $c$. Clearly the convergence rate is superlinear as predicted by the theory. A well balanced adjustment of the parameter $c$ helps to minimize both the number of outer iterations and the total number of inner iterations. Convergence histories of the damped inexact and the inexact Newton's methods are shown in figure 7 for $M_\infty = 0.895$. Oscillations of the IN method are more pronounced as the Mach number goes up. The DIN method is more robust but both methods converged in most cases.

Finally, we shall compare the performance of the DIN method and an SLOR method. Our SLOR method is implemented as follows:

$$B(\phi^k)\delta\phi^k = -L(\phi^k) \tag{4.10}$$

$$\phi^{k+1} = \phi^k + \delta\phi^k \tag{4.11}$$

$$L(\phi) = [k - (\gamma+1)\phi_x]\phi_{xx} + \phi_{yy} = 0 \tag{4.12}$$

$$s\delta\phi_{i,j-1} + c\delta\phi_{ij} + n\delta\phi_{i,j+1} = -\omega L(\phi_{ij}) - \omega w\delta\phi_{i-1,j} - z\delta\phi_{i-2,j} \tag{4.13}$$

$$\phi_{ij}^{k+1} = \phi_{ij}^k + \delta\phi_{ij}^k \tag{4.14}$$

In (4.13), the correction $\delta\phi_{ij}$ is solved on each successive vertical line $i$. The coefficients $s, c, n, w$ and $z$ are the entries of the Jacobian matrix $J$. $\omega$ is the relaxation parameter. Table 7 gives the number of iterations and computing time of the DIN method and SLOR method. The relaxation parameter $\omega$ in the SLOR method is assumed to have optimal values which are determined by numerical experiments. The convergence properties of the Newton's method is clearly superior.

Number of Newton iterations , E-O schemes

* no convergence

| $M_\infty$ | 0.839 subsonic | 0.872 transonic | 0.895 transonic | 0.916 transonic |
|---|---|---|---|---|
| W | 4 | 6 | * | * |
| DKR $\alpha = 1$ | 4 | 6 | * | * |
| MV | 4 | 6 | 7 | 9 |

Table 5

Number of Newton iterations , ( . ) = total number of ORTHOMIN (1) iterations

| $M_\infty$ | 0.839 subsonic | 0.872 transonic | 0.895 transonic | 0.916 transonic |
|---|---|---|---|---|
| $\epsilon = 1$ | 5 (68) | 7 (110) | 11 (117) | 17 (237) |
| $\epsilon = 0.5$ | 4 (62) | 6 (154) | 7 (152) | 10 (336) |

Table 6

Number of iterations , ( . ) = CPU time , E-O scheme

| $M_\infty$ | 0.839 subsonic | 0.872 transonic | 0.895 transonic | 0.916 transonic |
|---|---|---|---|---|
| SLOR | 142 (1:22.78) | 294 (2:40.23) | 408 (3:38.12) | 564 (4:58.26) |
| DIN | 5 (29.11) | 7 (34.86) | 12 (47.04) | 19 (1:02.60) |

Table 7

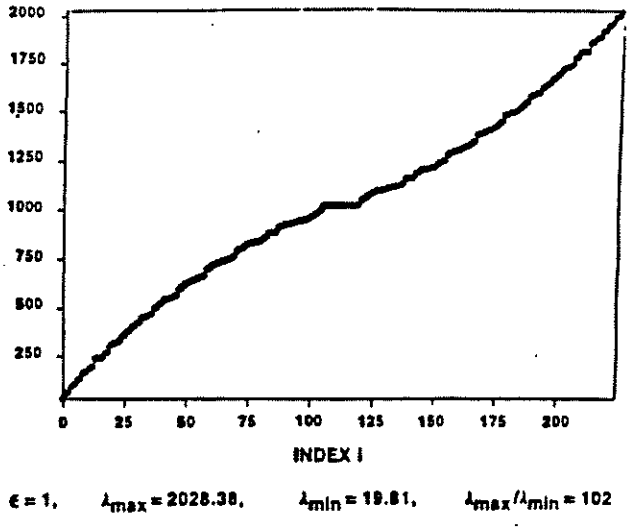


Figure 3. $\lambda(A)$

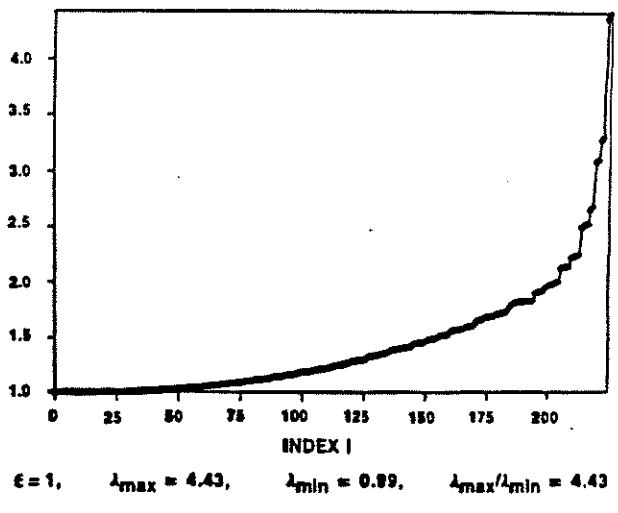


Figure 4. $\lambda(C^{-1}A)$, w preconditioner

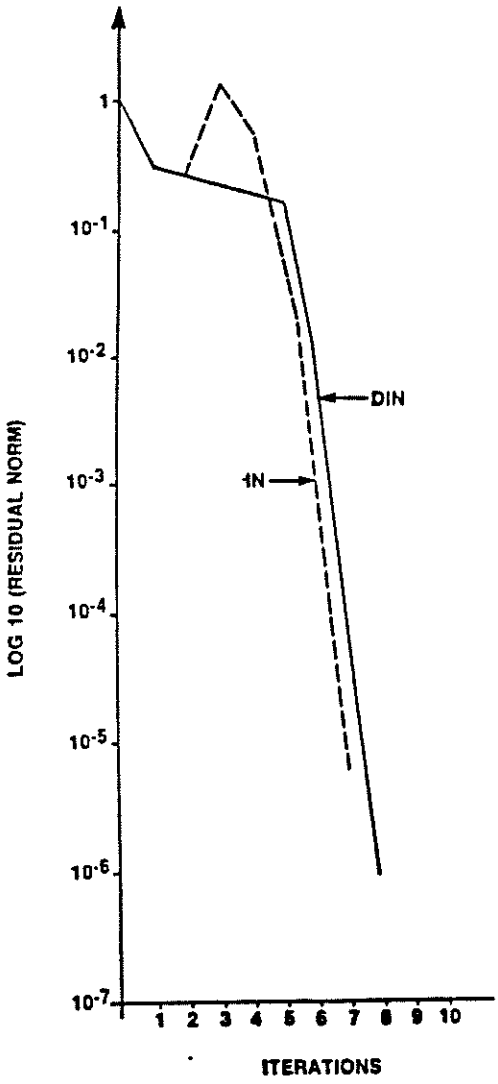


Figure 7.

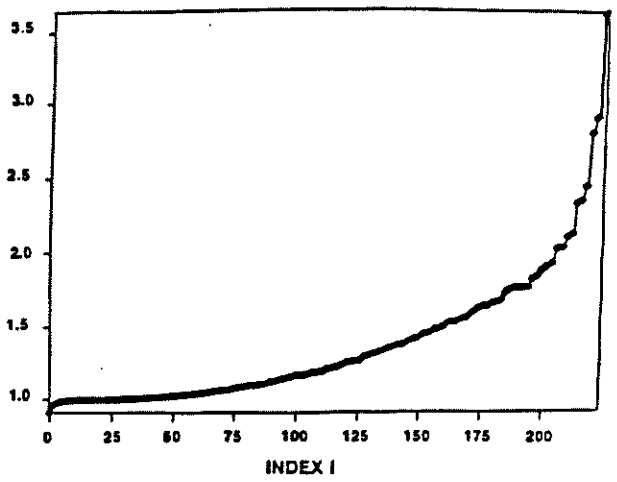


Figure 5. $\lambda(C^{-1}A)$, DKR preconditioner

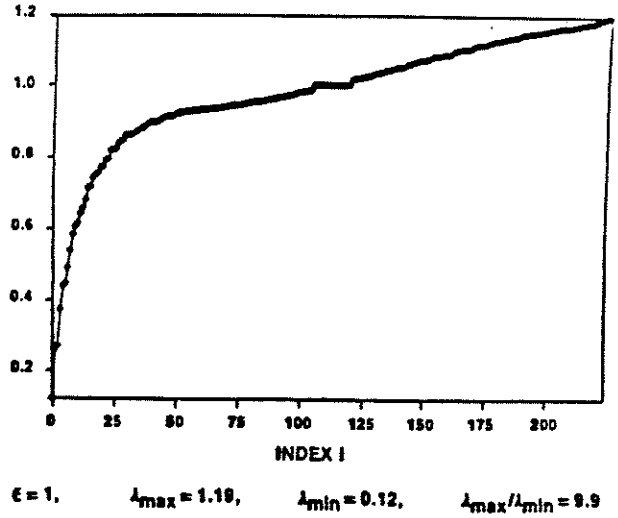


Figure 6. $\lambda(C^{-1}A)$, MV preconditioner

## References


P. Colella and P.R. Woodward, J. Comp. Phys. **54**, 174 (1984)

R.S. Dembo and T. Steihaug, *Mathematical Programming*, **26**, 190 (1983).

J.E. Dennis, Jr. and R.B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations* (Prentice-Hall, New Jersey, 1983).

T. Dupont, R. Kendall, and H.H. Rachford, Jr., *SIAM J. Numer. Anal.* **5**, 559 (1968).

B. Engquist and S. Osher, *Math. Comp.* **34**, 45 (1980).

M.E. Fiadeiro G. Veronis, *Tellus* **29**, 512 (1977).

S.K. Godunov, *Mat. Sbornik* **47** (in Russian), 271 (1959).

G. Kreiss and H.O. Kreiss, "Convergence ot steady state of solutions of Burgers' equation", in *Advances in Numerical and Applied Mathematics*, edited by J.C. South, Jr. and M.Y. Hussaini, (ICASE Report No. 86-18, 1986), p. 1.

J.A. Meijerink and H.A. Van der Vorst, *Math. Comp.* **31**, 148 (1977).

A. Rizzi and B. Engquist, J. Comp. Phys. 72, **1** (1987).

P.L. Roe, "Some contributions to the Modelling of Discontinuous Flows", in *Lectures in Applied Mathematics* (Amer. Math. Soc., Providence, RI, 1985), Vol. 22.

Y. Saad and M.H. Schultz, *Math. Comp.* **44**, 417 (1985).

Y. Saad and M.H. Schultz, *SIAM J. Sci. Statist. Comput.*, **7**, 856 (1986).

J. Smoller, "Shock wave and reaction-diffusion equations", Springer-Verlag, New York, 1983.

B. van Leer, *J. Comput. Phys.* **14**, 361 (1974).

P.K. W.Vinsome, "Orthomin, an iterative method for solving sparse sets of simultaneous linear equations", in *Proc. Fourth Symposium on Reservoir Simulation* (Society of Petroleum Engineers of AIME, 1976), p. 149.

Y.S. Wong, Ph.D. thesis, Oxford University, England, 1978.