RELAXATION METHODS FOR HYPERBOLIC CONSERVATION LAWS

BRAM VAN LEER* AND WILLIAM A. MULDER**

Abstract. A class of unconditionally stable integration schemes for finding steady solutions to the time-dependent Euler equations is discussed. The schemes are time-accurate for small time-steps and turn into a relaxation method for large time-steps; the spatial discretization is upwind biased. Examples of these switched evolution/relaxation (SER) schemes, and a numerical comparison, are presented. Some attention is given to the standard ADI and AF methods, which do not belong to the SER family.

1. Introduction. Numerical methods for solving problems of steady compressible flow often are based on the time-dependent flow equations. In marching toward the steady state it is useful to distinguish two phases [1].

In the first or searching phase the numerical solution follows a path to the steady state with reasonable time-accuracy, a safeguard against selecting an unphysical solution. In the second or converging phase the numerical solution feels the attraction of the steady state and converges to it rapidly.

It is possible to go through both phases with a single numerical method. The paragon is Euler's method "backward"; for the equation

$$u_t = R(u) \ , \tag{1}$$

with R(u) some nonlinear finite-difference expression, it reads

$$\Delta_t u^n = \tau^n R^{n+1} \ ; \tag{2}$$

here $\tau^n \equiv t^{n+1} - t^n$, $\Delta_t u^n \equiv u^{n+1} - u^n$. If R(u) is continuously differentiable, with

$$K(u) \equiv dR(u)/du \ , \tag{3}$$

*University of Technology, Delft, The Netherlands
**University Observatory, Leiden, The Netherlands

linearization of (2) is possible:

$$(I-\tau^n K^n)\ \Delta_t u^n = \tau^n R^n. \tag{4}$$

In the present paper we study a class of methods that includes (4), namely,

$$(I-\tau^n M^n)\ \Delta_t u^n = \tau^n R^n\ , \tag{5}$$

where M(u) is a linear operator providing numerical stability for arbitrary values of τ.

For small τ scheme (5) is time-accurate, with truncation error in $\Delta_t u^n$ of the order $O(\tau^2)$ or better; for large τ it reduces to the relaxation scheme

$$-M^n \Delta_t u^n = R^n\ . \tag{6}$$

This is Newton's method if M = K, in which case the right-hand side of (6), the residual, converges quadratically to zero. While this type of convergence is highly desirable, achieving it usually involves a lot of work. The matrix K may be costly to form and too costly to invert by direct methods. The classic way out, represented by scheme (5), is to replace K by a simpler matrix M, usually one with a narrower bandwidth. The utmost simplification is to replace K by a scalar.

A convenient choice of the time-step is

$$\tau_n = \varepsilon ||u^n|| / ||R^n||\ , \tag{7}$$

where ε is a number determining the temporal accuracy. If the residual is large, as in the searching phase, (7) ensures that

$$||\Delta_t u^n|| / ||u^n|| \sim \varepsilon\ , \tag{8}$$

that is: the relative change of the numerical solution per time-step is of the order of ε. (In [2] the same accuracy constraint was enforced by a posteriori checking, reducing τ_n in case of violation, and recomputing $\Delta_t u^n$). Upon entering the converging phase the residual drops sharply and τ rises correspondingly, switching the scheme from (5) to (6).

Schemes of the form (5) that allow switching through (7) will be called Switched Evolution/Relaxation (SER) schemes. The aim of this paper is to indicate matrices M suitable for SER schemes approximating hyperbolic systems of conservation laws (§2), compare these in a numerical experiment (§3) and make recommendations about their use (§4).

Before getting to the main subject (§2.3 ff) we discuss two schemes that do not belong to the SER category but are very popular in aerodynamics, namely, Alternating Direction Implicit (ADI) and Approximate Factorization (AF) schemes.

2. Relaxation methods for hyperbolic equations.

2.1 General considerations. When solving flow problems in one dimension using a three-point discretization we have no excuse for not choosing Newton's method as the relaxation method. For example, the Jacobian K associated with a first-order-accurate upwind-difference approximation of the Euler equations is block-tridiagonal, and the main block-diagonal of -K is semi-dominant within a margin $0(\Delta x)$. Thus, the matrix $I-\tau K$ in (4) is block-diagonally dominant for all but the largest values of τ, and its block-LU decomposition requires no pivoting [1].

For one-dimensional upwind-biased schemes of a higher order of accuracy the formation and direct solution of the linear system (4) is more cumbersome but, possibly, worthwhile; for two-dimensional schemes of even the first order of accuracy the direct solution of (4) is out of the question. In this paper we choose to abandon (4) in favor of (5).

Nevertheless, it is fully legitimate to use a relaxation scheme of the type (4) in solving the more complete system of equations (5). The relaxation scheme may serve as a preconditioner, in conjugate-gradient methods, or as a "smoother," in multigrid methods; the latter possibility is briefly discussed in §2.5.

All classic iterative methods exploit that R is the sum of two one-dimensional expressions,

$$R(u) = R_x(u) + R_y(u) \ , \tag{10}$$

each with its own Jacobian

$$K(u) = K_x(u) + K_y(u). \tag{11}$$

This suggests relaxation schemes with

$$M(u) = M_x(u) + M_y(u) \ ; \tag{12}$$

we shall restrict ourselves to these from §2.3 onward. First we discuss two different ways to benefit from (10), not leading to SER schemes.

2.2 Alternating Direction Implicit and Approximate Factorization schemes. The ADI scheme commonly used for relaxation is a sequence of two different time-steps, each unstable by itself:

$$(I-\tau^n M_x^n)\ \Delta_t u^n = \tau^n R^n \ , \tag{13.1}$$

$$(I-\tau^{n+1}M_y^{n+1})\ \Delta_t u^{n+1} = \tau^{n+1}R^{n+1}\ , \tag{13.2}$$

$$\tau^{n+1} = \tau^n\ . \tag{13.3}$$

The original intention of ADI [3] was to combine unconditional stability with second-order accuracy in time through the choice $M_x = \frac{1}{2} K_x$, $M_y = \frac{1}{2} K_y$.

The related, more modern AF scheme [4] is a nesting of similar steps, without intermediate updating:

$$(I-\tau^n M_x^n)(I-\tau^n M_y^n)\ \Delta_t u^n = \alpha\tau^n R^n\ ; \tag{14}$$

here α is a parameter which, for the present purpose, may take the value 1 or 2. With $\alpha = 1$ and $M_x = K_x$, $M_y = K_y$, scheme (14) obviously approximates the backward Euler scheme (4) within a margin $O(\tau^3)$:

$$(I-\tau^n K^n+(\tau^n)^2 K_x^n K_y^n)\Delta_t u^n = \tau^n R^n\ . \tag{15}$$

It is equally obvious that this is not an SER scheme, since

$$\lim_{\tau^n\to\infty} \Delta_t u^n = 0\ , \tag{16}$$

regardless of the magnitude of R^n.

With $\alpha = 2$ the one-step scheme (14) advances as far in time as the two-step scheme (13). Inserting $M_x = K_x$, $M_y = K_y$ in both (13) and (14) and assuming that R(u) is linear in u, e.g. $R_x = K_x u$, $R_y = K_y u$, we may rewrite these schemes as

$$\text{(ADI)}\quad u^{n+2} = (I-\tau^n K_y)^{-1}(I+\tau^n K_x)(I-\tau^n K_x)^{-1}(I+\tau^n K_y)u^n\ , \tag{16.1}$$

$$\text{(AF,}\alpha\text{=2)}\quad u^{n+1} = (I-\tau^n K_y)^{-1}(I-\tau^n K_x)^{-1}(I+\tau^n K_x)(I+\tau^n K_y)u^n\ , \tag{16.2}$$

showing their identity. Eq. (16) also reveals why ADI and AF are such effective relaxation methods for discretized parabolic equations like the scalar diffusion equation. In this case the eigenvalues of K_x and K_y are negative real, and the corresponding eigenvectors can be efficiently removed from the solution by cycling the value of $1/\tau$ through the spectrum of eigenvalues; see Wachspress [5].

For discretized hyperbolic equations the eigenvalues of K_x and K_y are complex or purely imaginary; therefore, time-step cycling is useless if done with real values of τ. One must resort to complex arithmetics, a fact that was well understood by the authors of [3] (J. Douglas, private communication, 1983) but seems to have fallen into oblivion. A correct implementation of time-step cycling for the Euler equations is due to Liu and Lomax [6].

The eigenvalues of K_x and K_y, whether complex or real, usually are obtained from a Fourier analysis, i.e. under the assumption that the eigenvectors are harmonics. When this assumption breaks down, e.g. for strongly variable coefficients and/or strongly nonuniform grids, time-step cycling becomes hard to implement and ADI and AF methods lose their appeal.

Abarbanel, Dwoyer and Gottlieb [7] have succeeded, for the linear diffusion equation, in reducing the sensitivity of AF to the choice of the time-step, by adding to the right-hand side of (15) a term that roughly balances the factorization error on the left-hand side:

$$(I-\tau^n K+(\tau^n)^2 K_x K_y)\ \Delta_t u^n = \tau^n (I+\gamma\tau^n K_x K_y) R^n \ . \tag{17}$$

For the parameter γ they derive an optimum value. One observes, however, that for large τ_n this scheme reduces to

$$K_x K_y \Delta_t u^n = \gamma\ K_x K_y R^n \ , \tag{18.1}$$

which is a hard way to implement the explicit scheme with time-step γ:

$$\Delta_t u^n = \gamma\ R^n . \tag{18.2}$$

An SER scheme of the form (5), with $M^n = -I/\gamma$, would achieve the same.

We conclude that ADI and AF do not have special properties that makes these schemes advantageous in finding steady solutions to hyperbolic equations.

2.3 SER schemes of first-order accuracy. It is surprising how well the classic relaxation schemes for the discretized diffusion equation, i.e. Jacobi, Gauss-Siedel and line relaxation, are suited for discretized hyperbolic equations. Block versions of these methods may be applied, in all possible combinations, to any first-order upwind discretization of the Euler equations, for an arbitrary number of space dimensions. The more powerful combinations also apply to second-order upwind discretizations.

To fix our thoughts, let us examine the possibilities in two dimensions, with flux splitting providing the upwind logic, as in [1]. The Euler equations in a Cartesian frame read

$$u_t = - (f(u))_x - (g(u))_y \ ; \tag{19}$$

here u is the state vector of conserved quantities and the vectors f(u), g(u) contain their fluxes in the x-, y-direction. The fluxes can be split in forward fluxes $f^+(u)$, $g^+(u)$ and backward fluxes $f^-(u)$, $g^-(u)$ that are continuously differentiable [8].

Now define a computational grid of adjacent rectangular volumes; the volume centered on (x_i,y_j), measuring Δx_i by Δy_j, is referred to by a subscript ij. The operators causing a shift over one volume in the forward x-, y-direction are called T_x, T_y. In this notation the first-order flux-split upwind discretization of the right-hand side of (19) becomes

$$R^n \equiv - (f^+_{ij} - f^+_{i-1\ j} + f^-_{i+1\ j} - f^-_{ij})^n/\Delta x_i$$
$$- (g^+_{ij} - g^+_{i\ j-1} + g^-_{i\ j+1} - g^-_{ij})^n/\Delta y_i \ . \tag{20.1}$$

We may write the full backward-Euler scheme as

$$L^n_{ij}\Delta_t u^n_{ij} \equiv (\frac{I}{\tau^n} - K^n_{ij})\ \Delta_t u^n_{ij} = R^n_{ij} \ , \tag{20.2}$$

with

$$K^n_{ij} = - (A^x_{ij}+A^y_{ij}-B^x_{ij}T_x^{-1}-B^y_{ij}T_y^{-1}-C^x_{ij}T_x-C^y_{ij}T_y)^n \ , \tag{20.3}$$

$$A^x_{ij} = (\frac{df^+}{du} - \frac{df^-}{du})_{ij}/\Delta x_i \geq 0 \ , \tag{21.1}$$

$$A^y_{ij} = (\frac{dg^+}{du} - \frac{dg^-}{du})_{ij}/\Delta y_j \geq 0 \ , \tag{21.2}$$

$$B^x_{ij} = (\frac{df^+}{du})_{i-1\ j}/\Delta x_i \geq 0 \ , \tag{21.3}$$

$$B^y_{ij} = (\frac{dg^+}{du})_{i\ j-1}/\Delta y_j \geq 0 \ , \tag{21.4}$$

$$C^x_{ij} = - (\frac{df^-}{du})_{i+1\ j}/\Delta x_i \geq 0 \ , \tag{21.5}$$

$$C^y_{ij} = - (\frac{dg^-}{du})_{i\ j+1}/\Delta y_j \geq 0 \ . \tag{21.6}$$

Owing to the upwind differencing in (20.1), the main-diagonal blocks of $-K^n$ are comfortably large: if the matrix elements

(21.1)-(21.6) vary smoothly with i and j, $-K^n$ is semi-dominant within a margin $0(\Delta x)$. The matrix L^n will actually be block-diagonally dominant for a bias I/τ^n that is sufficiently large, but still only $0(\Delta x)$. Nevertheless, we shall not attempt to solve (20.2) directly by Gaussian elimination.

The classic relaxation schemes for the linear system (20.2) may be thought to arise from one particular way of approximating K^n: in (20.3) one or more shift operations are replaced by scalar multiplications. This is common practice in Fourier analysis, where, for a single mode with spatial frequencies ξ and η, we have

$$T_x^{\pm 1} = e^{\pm i\xi} I \ , \quad T_y^{\pm 1} = e^{\pm i\eta} I \ , \tag{22.1}$$

or

$$\Delta_t u^n_{i\pm 1\ j} = e^{\pm i\xi} \Delta_t u^n_{ij} \ , \tag{22.2}$$

$$\Delta_t u^n_{ij\pm 1} = e^{\pm i\eta} \Delta_t u^n_{ij} \ . \tag{22.3}$$

However, we wish to avoid complex arithmetics and complete dependence on Fourier analysis; see §2.2. In replacing $T_x^{\pm 1}$, $T_y^{\pm 1}$ we shall restrict ourselves to real-valued multipliers $s_x^{\pm}$, $s_y^{\pm}$:

$$T_x^{\pm 1} = s_x^{\pm} I \ , \quad T_y^{\pm 1} = s_y^{\pm} I \ , \tag{23.1}$$

or

$$\Delta_t u^n_{i\pm 1\ j} = s_x^{\pm} \Delta_t u^n_{ij} \ , \tag{23.2}$$

$$\Delta_t u^n_{ij\pm 1} = s_y^{\pm} \Delta_t u^n_{ij} \ . \tag{23.3}$$

There are three important values: s = -1, 0 and 1. The approximation of $T^{\pm 1}$ by $s^{\pm} = -1$ is exact for a saw-tooth component in $\Delta_t u^n$. It is seen from (20.3) and (21) that this choice makes the main-diagonal blocks more strongly dominant, leading to underrelaxation for long waves. The longest waves would be best represented by s = 1, but this value makes scheme (20) unstable. The stability condition on s, as shown in the Appendix, is

$$s \leq 0 \ . \tag{24}$$

The value s = 0 will be considered the standard value. It is computationally attractive: an off-diagonal block multiplied by zero need not be computed at all, and does not influence the main diagonal.

In simplifying the right-hand side of (20.3), it is useful to distinguish the five cases listed below.

(i) All four shift operators replaced.
The block version of point/Jacobi relaxation. If the volumes with i+j even and i+j odd are updated alternatingly, checkerboard relaxation results.

(ii) Three shift operators replaced.
A combination of Jacobi and Gauss-Seidel relaxation. If, for instance, T_x^{-1} is the one operator that is not replaced, the Gauss-Seidel process requires sweeps in the forward x-direction.

(iii) Two shift operators replaced that work along different coordinate axes.
Full Gauss-Seidel relaxation. If, for instance, T_x^{-1} and T_y^{-1} are retained, one must make sweeps in the forward x-direction and the forward y-direction.

(iv) Two shift operators replaced that work along the same coordinate axis.
Line/Jacobi relaxation. If, for instance, T_y and T_y^{-1} are retained, the line relaxation is in the y direction; updating the volumes with i even and i odd alternatingly yields zebra relaxation.

(v) One shift operator replaced.
Line/Gauss-Seidel relaxation. If, for instance, T_x is the one operator that is replaced, the line relaxation is in the y-direction and the Gauss-Seidel sweep goes in the forward x-direction.

It is generally recommended in the cases (ii)-(v) to cycle through all shift operators when replacing one or more. Gauss-Seidel relaxation should not be used in the direction of a cyclic coordinate, as this causes a long-wave closure error which is hard to remove. Pattern relaxation (checkerboard, zebra) does not have this drawback.

A nonlinear variant of the Gauss-Seidel scheme results when the fluxes at t^{n+1}, needed in the original backward Euler scheme (2), are not all approximated linearly,

$$(f^{\pm})^{n+1} = (f^{\pm})^n + (\frac{df^{\pm}}{du})^n \Delta_t u^n \ , \tag{25.1}$$

but are actually updated,

$$(f^{\pm})^{n+1} = f^{\pm}(u^n+\Delta_t u^n) \ , \tag{25.2}$$

in any volume where $\Delta_t u^n$ is already known. Implemented this way, the scheme no longer includes the off-diagonal blocks of K^n.

A very radical simplification, one that significantly reduces the relaxation power of the schemes, is achieved in replacing any of the blocks of K^n replaced by its spectral radius. Particularly well-known is the simplified point/Jacobi scheme with $s_x^+ = s_y^\pm = 0$, that is,

$$(I/\tau^n+\rho_{ij}^n I)\ \Delta_t u_{ij}^n = R_{ij}^n \ , \tag{26.1}$$

or

$$\Delta_t u_{ij}^n = \frac{\tau^n}{1+\tau^n\rho_{ij}^n} R_{ij}^n \ . \tag{26.2}$$

where ρ_{ij}^n is the spectral radius of $(A^x+A^y)_{ij}^n$. The factor multiplying R_{ij}^n may be interpreted as a locally adjusted time-step value, which for $\tau^n \to \infty$ approaches the local stability limit $(\rho_{ij}^n)^{-1}$.

2.4 SER schemes of second-order accuracy. The upwind spatial discretization of second-order accuracy, described and tested by Mulder and Van Leer [1] for the one-dimensional flow equations, applies in a straightforward way to the two-dimensional equations. In a Cartesian grid the two dimensions completely decouple, owing to property (10).

The relaxation schemes are essentially those of §2.3, but with the Jacobians in (20) evaluated at $(i\pm\frac{1}{2},j)$, $(i,j\pm\frac{1}{2})$, rather than (i,j), $(i\pm1,j)$, $(i,j\pm1)$; for details see [1]. Except for this adjustment, the second-order terms in the scheme are not accounted for in M_x^n or M_y^n. The second-order SER schemes therefore deviate more strongly from the full backward Euler scheme than the corresponding first-order schemes.

The stability analysis in the Appendix reveals that Jacobi relaxation is no longer stable, pattern relaxation is stable with sufficiently strong underrelaxation, whereas Gauss-Seidel relaxation is stable only when forward and backward sweeps are alternated (symmetric Gauss-Seidel). In demanding computations, e.g. on strongly non-uniform grids, with strong oblique shocks in the solution, without a preferential flow direction, it is recommended to use the best possible relaxation methods, i.e. symmetric line/Gauss-Seidel and symmetric point/Gauss-Seidel.

2.5 Multigrid relaxation. Any of the SER schemes of §2.3 or §2.4 may be used as the "smoother" in a multigrid cycle, with the symmetric Gauss-Seidel scheme as the first choice. The nonlinear version is suited for use in a "full-approximation-storage scheme" (FAS; for a review of multigrid concepts see [9]), while the linearized version is appropriate for a "correction scheme." The latter combination was successfully applied to the test problem of §3 by Mulder [10]. The FAS scheme for the Euler equations implemented by Jameson [11] does not include an SER scheme; relaxation is provided by a four-stage Runge-Kutta method with local time-step values.

3. A numerical comparison. The SER schemes of §2.3 and §2.4, and also the ADI and AF schemes of §2.2, were used to compute the steady transonic flow through a straight channel with a circular bump on the lower wall. The inflow Mach number was 0.85, the thickness of the bump was 4.2% of the chord length. The steady flow exhibits a shock almost choking the channel.

In marching toward the steady state, the isenthalpic Euler equations were used. At the walls reflection conditions were imposed with the help of mirror-image zones; the arc was described according to small-disturbance theory (thickness ignored, flow angle prescribed). Total pressure and cross-flow velocity (=0) were given at the inlet, static pressure at the outlet. The details of the equations and the boundary conditions can be found in [10], where the same problem was used for a multigrid experiment; for numerical solutions to this problem by other authors see [15].

Figure 1 shows the distributions of the pressure coefficient on both walls, obtained from solutions with first-order (a) and second-order (b) spatial accuracy, on a uniform grid of 32×16 zones.

In Table 1 and Table 2 are listed some of the many data gathered on the convergence speed achieved by the various schemes. The convergence process was monitored by the quantity RES defined by

$$RES^n = \max_{k,i,j} \left(\frac{|R_k^n|}{|u_k^n| + h_k^n} \right)_{ij} , \tag{27.1}$$

where k=1,2,3 indicates the different conservation laws and h is a bias vector preventing division by zero. The time-step for SER schemes was chosen according to

$$\tau^n = \varepsilon / RES^n , \tag{27.2}$$

which is similar to Eq. (7).

As a rule, the number of iterations needed to reduce the residuals by a factor of 10^{-10} is smaller when a scheme deviates less from

Experiment number	Kind of relaxation	Order of approximation	One cycle involves	Number of iterations per cycle	Number of iterations till convergence	Cpu-time spent (minutes)
1	line/GS	1	line(x),GS(+y), line(x),GS(-y); line(y),GS(+x), line(y),GS(-x)	2	123	1.2
2	line/GS	1	line(x),GS(+y), line(y),GS(+x)	1	174	1.7
3	line/GS	1	line(x),GS(+y), line(x),GS(-y)	1	204	1.7
4	line/GS	1	line(y),GS(+x), line(y),GS(-x)	1	92	0.83
5	line/GS	1	line(x),GS(+y)	½	243	2.1
6	line/GS	1	line(y),GS(+x)	½	162	1.4
7	line/GS	2	see nr. 1	2	183	2.4
8	line/GS	2	see nr. 3	1	221	2.6
9	line/GS	2	see nr. 4	1	313	3.8
10	zebra	1	line(x),pat(y)	½	243	2.0
11	zebra	1	line(y),pat(x)	½	162	1.4
12	zebra	2	line(y), pat(x;s=-0.25)	½	796	9.3
13	zebra	2	line(y), pat(x;s=-1.0)	½	575	6.8
14	line/Jacobi	1	line(x),Jac(y)	½	466	3.5
15	line/Jacobi	1	line(y),Jac(x)	½	294	2.2
16	line/Jacobi	1	line(x),Jac(y), line(y),Jac(x)	1	234	1.8
17	ADI	1	line(x),line(y)	1	486	3.7
18	AF(α=1)	1	line(x),line(y)	1	452	2.3

Table 1. Data on the convergence speed achieved by the line-relaxation schemes (SER and other) in solving the transonic flow problem of §3, on a grid of 32 x 16 zones. Problem parameters: M_∞ = 0.85, arc thickness = 4.2% of chord length. The iteration count is based on a unit including two line relaxations (regardless of their direction), making comparisons more or less meaningful. The cpu-time is given in minutes on an Amdahl V7B computer. The value of ε in the time-step formula (27.2) was 1.0 for all SER schemes. With ADI and AF the time-step, based on ε = 0.5, was frozen at the start, fixing the free-stream Courant number at a value of 4.13 for both experiments. Under-relaxation was applied only in the second-order zebra scheme (s = -0.25, -1.0), for the sake of stability.

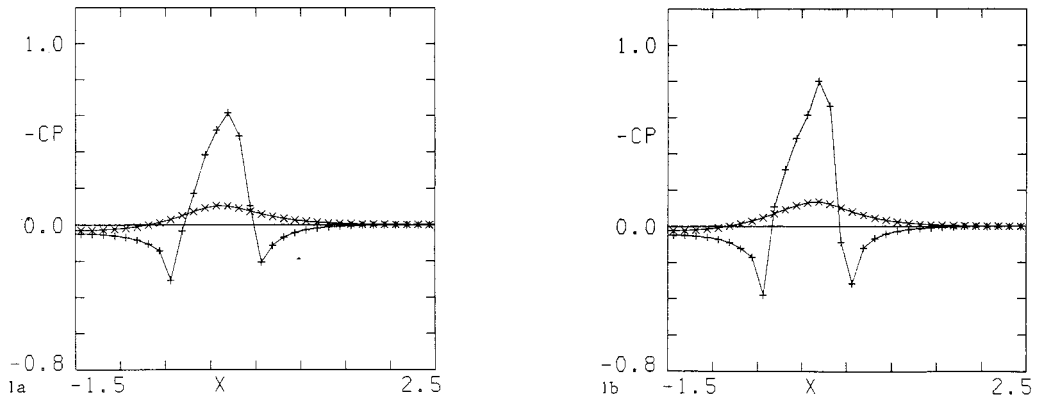


Figure 1. (a) Pressure coefficient on lower (+) and upper (x) wall for Mach 0.85 flow in a channel with a circular arc of 4.2% thickness on the lower wall, as computed from a solution on a 32×16 grid with first-order accuracy. Boundary conditions on the arc according to thin-airfoil theory. (b) As before, but computed with second-order accuracy.

Exper-iment number	Kind of relaxation	Order of approx. imation	One cycle involves	Number of iterations per cycle	Number of iterations till convergence	Cpu-time spent (minutes)
19	point/GS	1	GS(+x,+y); GS(-x,-y)	2	557	2.6
20	point/GS	1	GS(+x,+y); GS(+x,-y)	2	597	2.8
21	point/GS	1	GS(+x,+y); GS(-x,+y)	2	606	2.6
22	point/GS	1	GS(+x,+y)	1	713	3.0
23	point/GS	2	see nr. 19	2	575	4.0
24	point/Jacobi	1	Jac(x,y)	1	1367	4.4

Table 2. Convergence data for the point-relaxation schemes (SER only).

the backward Euler scheme. This rule applies, too, when comparing schemes of the first order of accuracy to those of the second order.

Among the schemes bearing the same name there is a considerable spread in performance which is hard to explain in detail, especially because it appears to be very sensitive to the precise implementation of the boundary conditions. Line/Gauss-Seidel relaxation, for instance, seems to solve the present test problem most efficiently with the line in the y-direction and symmetric sweeps in the x-direction. This result was obtained, however, by ignoring at the solid walls the contributions from the mirror-image zones to the relaxation matrix. If these contributions are included, as they should be in the full backward Euler method, the relaxation process does not even converge, in flagrant disagreement with our rule-of-thumb. Obviously, an analysis of the interference of numerical boundary conditions with relaxation schemes is due; meanwhile we chose to use the less complete linearization for all line relaxations in the y-direction.

Zebra relaxation, while second best for the first-order scheme, slows down considerably when used for the second-order scheme; this is due to the strong underrelaxation needed for stability or for convergence.

The performance of the ADI and AF schemes is found to depend critically on the value of the time-step; when regarded as relaxation methods these cannot be called robust. Attempts to optimize the choice of the time-step were not uniformly successful and therefore did not make the schemes more robust. In the present experiments a fixed time-step was used, determined by RES^0 and $\varepsilon = 0.5$. The choice $\varepsilon = 1.0$ leads to divergence for ADI and to non-convergence for AF.

When comparing the number of iterations till convergence to the cpu-time spent, one must realize that the block-elements of the matrix L^n used in the relaxation schemes were not computed at every time level. Their values, and the value of τ^n, were updated only when RES^n dropped below some control level and remained frozen until the next lower level was reached. For these levels the following sequence of fractions of RES^0 were used: 10^{-1}, 3×10^{-2}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-6}. Freezing has no significant effect on the relaxation process, except for AF, where it may change slow convergence into non-convergence (observed for $\varepsilon = 1.0$).

Along with the blocks of L^n, all blocks derived from these were frozen, i.e. the inverses of the main diagonal blocks needed for point relaxation, or the block elements of all line-wise LU-decompositions needed for line relaxation. This strategy leads to large

savings on cpu-time; its weakness lies in the prerequisite that lots of storage space be available. In practice one will have to trade the upper limit to storage space for a lower limit to cpu-time; the trade-off is highly computer-dependent. Consider, for example, point/Jacobi, checkerboard and symmetric nonlinear point/Gauss-Seidel relaxation, each requiring the storage of only one block: the inverse of the main-diagonal block. While the Gauss-Seidel scheme offers the strongest relaxation per iteration, it may finish last when implemented on a supercomputer, since its update step does not vectorize.

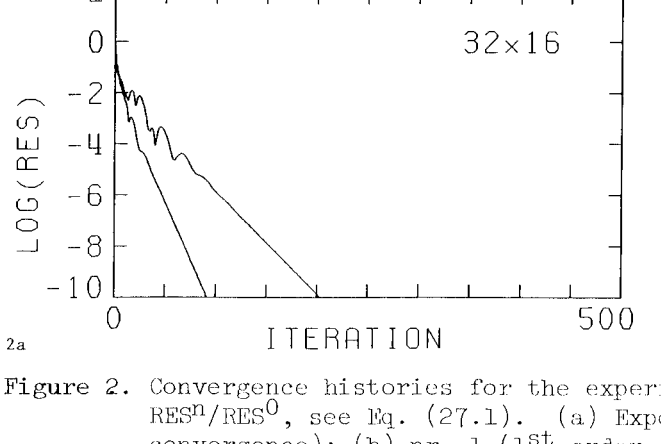
Figures 2 through 7 show the convergence histories of some of the experiments compiled in Tables 1 and 2; RES stands for RES^n/RES^0. In most cases the residual norm does not decrease monotonically; various periodic and quasi-periodic fluctuations can be discerned. These correspond to the sequencing of sweep directions (as in Figure 6) or line directions (as in Figure 2b) and to the bouncing of disturbances from wall to wall (as in Figure 4).

4. Conclusions and recommendations. In the preceding sections it has been demonstrated that Switched Evolution/Relaxation schemes incorporating a classic relaxation method are robust means to compute steady discontinuous solutions of hyperbolic systems of conservation laws such as the Euler equations. It is crucial that the spatial discretization be upwind biased. If storage space is not restricted, the most complex relaxation methods are also the most efficient, owing to the possibility of keeping the coefficient blocks frozen during many iteration steps. Alternating-Direction Implicit and Approximate-Factorization methods are less efficient than equally complex SER methods, and not at all robust.

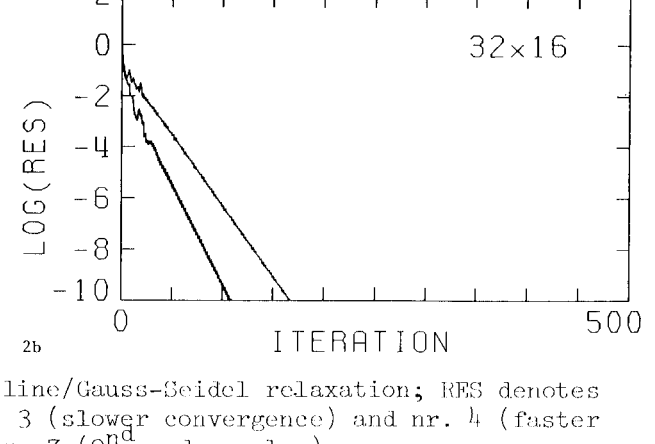
It is not surprising that other advocates of upwind differencing, independently or through interaction, have come to the same conclusions. Chakravarthy [12] has applied the point/Gauss-Seidel scheme to a variety of aerodynamic problems, with remarkable success; Dadone and Napolitano [13] recently turned from using AF to using SER schemes.

All SER schemes allow of underrelaxation, which improves short wave damping. It turns out that overrelaxation, a standard routine for the iterative solution of second-order elliptic equations, does not work for first-order equations (see the Appendix).

We recommend the further development of and experimentation with SER schemes requiring only one block evaluation and inversion. Neither the complexity nor the storage requirements of such schemes are extravagant, so that their application, even to three-dimensional flow problems, is within the capacity of today's computers. This view differs somewhat from Jameson's [11], who puts more emphasis on the storage aspect and therefore excludes any blocks from his



2a



2b

Figure 2. Convergence histories for the experiments with line/Gauss-Seidel relaxation; RES denotes RES^n/RES^0, see Eq. (27.1). (a) Experiments nr. 3 (slower convergence) and nr. 4 (faster convergence); (b) nr. 1 (1^{st} order, fast) and nr. 7 (2^{nd} order, slow).

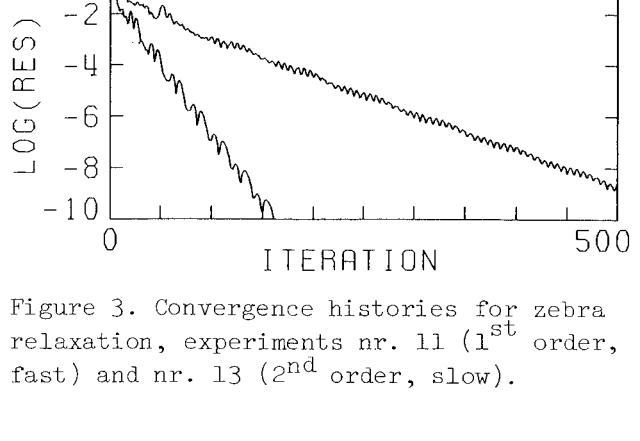


Figure 3. Convergence histories for zebra relaxation, experiments nr. 11 (1^{st} order, fast) and nr. 13 (2^{nd} order, slow).

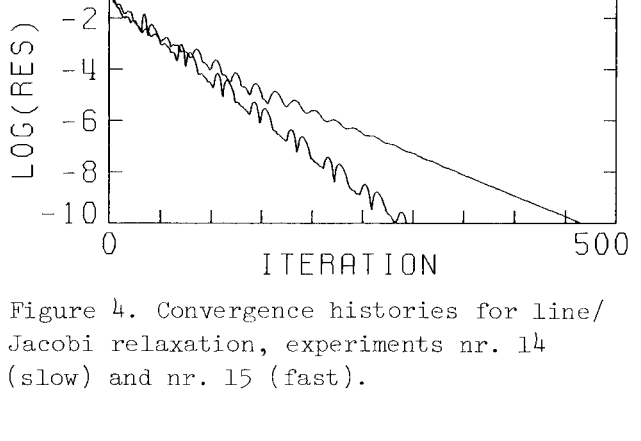
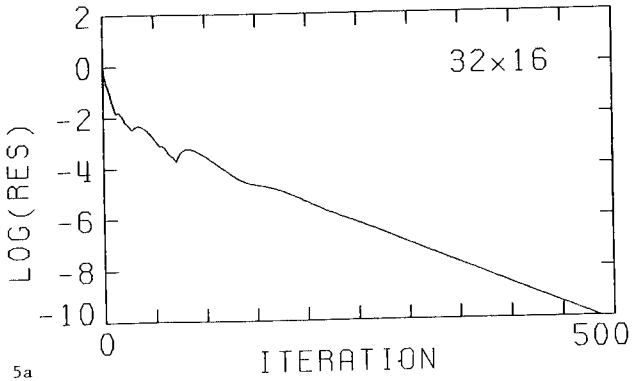
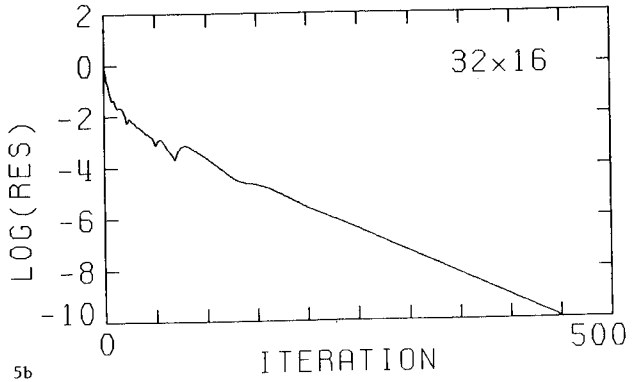


Figure 4. Convergence histories for line/Jacobi relaxation, experiments nr. 14 (slow) and nr. 15 (fast).



5a



5b

Figure 5. Convergence histories for non-SER relaxation. (a) Experiment nr. 17 (ADI); (b) nr. 18 (AF).

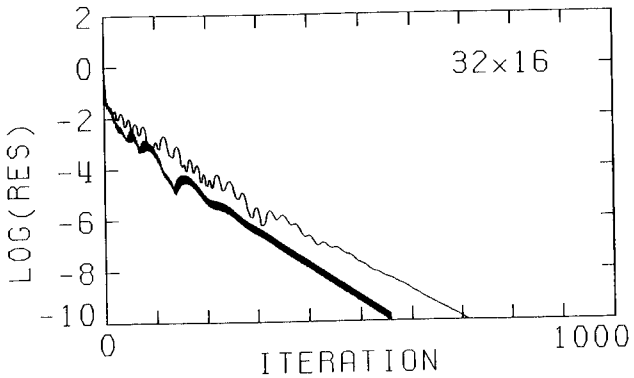


Figure 6. Convergence histories for point/ Gauss-Seidel relaxation, experiments nr. 19 (fast) and nr. 22 (slow).

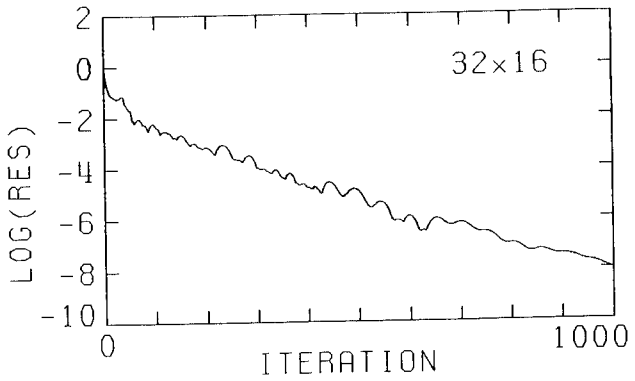


Figure 7. Convergence history for point/ Jacobi relaxation, experiment nr. 24.

relaxation schemes. A modification of his technique proposed by Turkel [14] indeed introduces a coefficient block, for the sake of efficiency.

Appendix: Stability of SER schemes. A necessary condition for the stability of the schemes of §2.3 and §2.4 is that plane waves moving in a coordinate direction shall not be amplified. Any such wave can be described by the scalar linear convection equation

$$q_t + aq_x = 0 \tag{A1.1}$$

starting from the discrete initial-value distribution

$$q_k^0 = q_0^0 e^{i\xi k}. \tag{A1.2}$$

In practice it appears that, in the special case of the Euler equations, this condition is also sufficient; the proof remains to be given. The present stability analysis is still based on the initial-value problem (A1). Pattern relaxation requires a more elaborate analysis; we shall only state some results.

In the case of first-order upwind differencing, the one-dimensional SER scheme including line relaxation is identical to the backward Euler scheme, hence is stable. With Gauss-Seidel relaxation it still is identical to the backward Euler scheme if the sweep direction is the same as the wave direction. A wave travelling against the sweep direction is accounted for in the relaxation matrix only by a main-diagonal element, just as in Jacobi relaxation. It therefore suffices to study Jacobi relaxation; introducing the Courant number

$$\sigma = \frac{a\tau}{\Delta\xi} \geq 0 , \tag{A2}$$

the SER scheme for Eq. (A1.1) reads

$$\{1+\sigma(1-s)\}\Delta_t q_k^n = -\sigma(1-T^{-1})q_k^n. \tag{A3}$$

As explained in §2.3, a scalar s has taken the place of the shift operator T^{-1}. The amplification factor $g(\xi,\sigma,s)$ of this scheme follows upon inserting (A1.2):

$$\{1+\sigma(1-s)\}(g-1) = -\sigma(1-e^{-i\xi}) , \tag{A4.1}$$

or

$$g = 1 - \frac{\sigma}{1+\sigma(1-s)}(1-e^{-i\xi}) . \tag{A4.2}$$

This corresponds to the forward Euler scheme with effective Courant number $\sigma/\{1+\sigma(1-s)\}$, which for first-order upwind differencing is stable up to a Courant-number value of 1. It follows that

$$s \leq \frac{1}{\sigma} \tag{A5.1}$$

for stability; if we insist on stability for arbitrarily large σ we find

$$s \leq 0 \ , \tag{A5.2}$$

which is Eq. (24). This condition also suffices to stabilize pattern relaxation, which is closer in performance to Gauss/Seidel than to Jacobi relaxation.

Overrelaxation, meant to damp efficiently the longest waves ($\xi \sim 0$, $T = I$), cannot be achieved through the approximation (23.1). In approximating $T^{\pm 1}$, two diagonals must be involved:

$$T^{\pm 1} = sI + (1-s)T^{\mp 1} \ . \tag{A6}$$

The scheme for adverse waves, e.g. with T^{-1} expressed in I and T during a backward sweep, becomes

$$\{I+\sigma(1-s)(I-T)\}\Delta_t q_k^n = -\sigma(I-T^{-1})q_k^n \ , \tag{A7}$$

with amplification factor

$$g = 1 - \frac{1-e^{-i\xi}}{1/\sigma+(1-s)(1-e^{i\xi})} \tag{A8}$$

For small α this reduces to

$$g = \frac{i\xi+O(\xi^2)}{1/\sigma-(1-s)i\xi+O(\xi^2)} \ , \tag{A9}$$

and if we choose σ much larger than the largest spatial frequency that can be represented on a finite grid, e.g. $1/\sigma = O(\xi^2)$, we finally get

$$g = 1 + \frac{1}{1-s} + O(\xi) \ . \tag{A10}$$

For $s = 2$, g is of the order of ξ, which is precisely the aim of overrelaxation. From (A6) we see that this choice means to replace backward differencing by forward differencing and vice versa, i.e.

$$I - T^{-1} = T - I \ . \tag{A11}$$

Unfortunately, $s = 2$ leads to violent growth of the shortest waves for all but the smallest values of σ; e.g., $\xi = \pi$ and $\sigma = \frac{1}{2}$ make the denominator in (A9) vanish while the numerator remains finite.

We may therefore conclude that overrelaxation with first-order upwind SER schemes is not possible.

It is tempting to investigate if underrelaxation may be improved through the use of two diagonals. If we approximate T^{-1} according to

$$T^{-1} = sI + (1+s)\, T^{+1} \quad , \tag{A12}$$

which is satisfied by the shortest waves ($\alpha = \pi$, $T = -1$), the scheme for adverse waves becomes

$$[I+\sigma\{(1-s)I-(1+s)T\}]\Delta_t q_k^n = -\sigma(I-T^{-1})q_k^n \quad , \tag{A13}$$

with amplification factor

$$g = 1 - \frac{1-e^{-i\xi}}{1/\sigma+\{(1-s)-(1+s)e^{i\xi}\}} \tag{A14}$$

Inserting $\xi = \pi-\phi$, with ϕ small, we get

$$g = \frac{1/\sigma-i(2+s)\phi+O(\phi^2)}{1/\sigma+2-i(1+s)\phi+O(\phi^2)} \quad , \tag{A15}$$

for large enough σ, i.e. $1/\sigma = O(\phi^2)$, this leads to

$$g = -i(1+\tfrac{1}{2}s)\phi + O(\phi^2) \quad , \tag{A16}$$

indicating extra strong damping if

$$s = -2, \tag{A17}$$

Inserting (A17) into (A12) yields

$$I + T^{-1} = -(I + T) \quad , \tag{A18}$$

which, for the shortest waves, is as good an extrapolation as (A11) is for the longest waves.

It is easily verified that scheme (A13) with (A17) is unconditionally stable.

With second-order upwind differencing the SER schemes including line relaxation or downwind Gauss-Seidel sweeping deviate from the backward Euler scheme, as the second-order terms are not treated implicitly. For Eq. (A1.1) these schemes read

$$\{I+\sigma(I-T^{-1})\}\Delta_t q_k^n = -\sigma(I-T^{-1})\{1+\tfrac{1}{4}(T-T^{-1})\}; \tag{A19}$$

their amplification factor is

$$g = \frac{1-\sigma(1-e^{-i\alpha})\ \frac{i}{2}\sin\xi}{1+\sigma(1-e^{-i\xi})} \tag{A20.1}$$

$$= \frac{1 + 2\sigma \sin^2\frac{\xi}{2}\cos^2\frac{\xi}{2} - i\sigma \sin\xi \sin^2\frac{\xi}{2}}{1 + 2\sigma \sin^2\frac{\xi}{2} + i\sigma \sin\xi}\ , \tag{A20.2}$$

with modulus not exceeding 1. For $\sigma \to \infty$ we have $g = -\frac{i}{2}\sin\xi$, hence $|g| \leq \frac{1}{2}$.

With Jacobi relaxation, or upwind Gauss-Seidel sweeping, the second-order SER scheme becomes unstable for any value of σ. Like scheme (A3) it is equivalent to the forward Euler scheme, used with an effective Courant number $\sigma/\{1+\sigma(1-s)\}$. The locus of the Fourier transform of the spatial-differencing operator in the complex plane now has fourth-order contact with the imaginary axis for $\alpha=0$, whereas the stability domain for the forward Euler scheme has only second-order contact. To match the stability domain with the spectral locus, a two-step technique would have to be used.

The instability is insuperable for pure Jacobi relaxation, but can be suppressed for pattern relaxation by taking negative values of s (down to $-\sqrt{2}$ if necessary), and for Gauss-Seidel relaxation by alternating between upwind and downwind sweeps. In the worst case, $\sigma \to \infty$, the amplification factors are

$$g_{\text{upwind}} = 1 - \frac{1}{1-s}(1-e^{-i\xi})(1+\frac{i}{2}\sin\xi)\ , \tag{A21}$$

$$g_{\text{downwind}} = -\frac{i}{2}\sin\xi, \tag{A22}$$

and the modulus of their product remains safely below 1, for $s \leq 0$.

Underrelaxation through the use of two diagonals works as well in the second-order case as in the first-order case. Inserting (A12) into scheme (A19) yields, for small $\phi = \pi-\xi$,

$$g = \frac{1/\sigma - i(3+s)\phi + 0(\phi^2)}{1/\sigma + 2 - i(1+s)\phi + 0(\phi^2)}\ , \tag{A23}$$

from which it is seen that, for sufficiently large σ, extra strong damping results when

$$s = -3\ . \tag{A24}$$

With this value of s the second-order scheme for the adverse waves is unconditionally stable.

It is worthwhile to mention that underrelaxation makes a scheme, whether first-order or second-order accurate, a good smoother for use in a multigrid strategy.

REFERENCES

(1) W. A. MULDER and B. VAN LEER, Experiments with implicit upwind methods for the Euler equations, J. Comp. Phys., 59 (1985), pp. 232-246.

(2) A. DADONE and M. NAPOLITANO, An implicit Lambda-Scheme, AIAA paper nr. 82-0972, AIAA-ASME 3rd Joint Thermophysics, Fluids, Plasma and Heat Transfer Conference, St. Louis, Mo., June 1982.

(3) J. DOUGLAS and J. GUNN, A general formulation of alternating direction methods, I, Numer. Math. 6 (1964), pp. 428-453.

(4) R. W. BEAM and R. F. WARMING, An implicit finite-difference algorithm for hyperbolic systems in conservation-law form, J. Comp. Phys. 22 (1976), pp. 87-110.

(5) E. WACHSPRESS, Iterative solution of elliptic systems, Prentice Hall, Englewood Cliffs, N.J., 1966.

(6) Y. LIU and H. LOMAX, Nonstationary relaxation methods for the Cauchy-Riemann and the one-dimensional Euler equations, AIAA paper 83-1901, AIAA 6th Computational Fluid Dynamics Conference, Danvers, Mass., July 1983.

(7) S. S. ABARBANEL, D.L. DWOYER and D. GOTTLIEB, Improving the convergence rates of parabolic ADI methods, ICASE Report 82-28 (1982).

(8) B. VAN LEER, Flux-vector splitting for the Euler equations, Lecture Notes in Physics 170 (1982), pp. 507-512.

(9) A. BRANDT, Guide to multigrid development, Lecture Notes in Mathematics 960 (1982), pp. 220-312.

(10) W. A. MULDER, Multigrid relaxation for the Euler equations, preprint (1984), Leiden University Observatory, submitted to J. Comp. Phys.

(11) A. JAMESON, Numerical solution of the Euler equations for compressible inviscid fluids, this volume. pp.

(12) S. R. CHAKRAVARTHY and S. OSHER, Higher-resolution applications of the Osher upwind scheme for the Euler equations, AIAA paper 83-1943, AIAA 6th Computational Fluid Dynamics Conference, Danvers, Mass., July 1983.

(13) M. NAPOLITANO and A. DADONE, Three-dimensional implicit lambda-method, NASA Contractor Report 172264 (ICASE), October 1983.

(14) E. TURKEL, Acceleration to a steady state for the Euler equations, this volume, pp.

(15) A. RIZZI and H. VIVIAND, eds., Numerical methods for the computation of inviscid transonic flows with shock waves, Notes on Numerical Fluid Mechanics, Vol. 3, Vieweg, Braunschweig, 1981.