NASA-TM-111237

JOURNAL OF COMPUTATIONAL PHYSICS 119, 295–324 (1995)

The Method of Space–Time Conservation Element and Solution Element—A New Approach for Solving the Navier–Stokes and Euler Equations

SIN-CHUNG CHANG

NASA Lewis Research Center, Cleveland, Ohio 44135

Received December 13, 1993; revised January 3, 1995

A new numerical framework for solving conservation laws is being developed. This new framework differs substantially in both concept and methodology from the well-established methods, i.e., finite difference, finite volume, finite element, and spectral methods. It is conceptually simple and designed to overcome several key limitations of the above traditional methods. A two-level scheme for solving the convection-diffusion equation

$$\partial u/\partial t + a\,\partial u/\partial x - \mu\,\partial^2 u/\partial x^2 = 0 \quad (\mu \geq 0)$$

is constructed and used to illuminate major differences between the present method and those mentioned above. This *explicit* scheme, referred to as the a-μ scheme, has two *independent* marching variables u_j^n and $(u_x)_j^n$ which are the numerical analogues of u and $\partial u/\partial x$ at (j, n), respectively. The a-μ scheme has the unusual property that its stability is limited only by the *CFL* condition, i.e., it is independent of μ. Also it can be shown that the amplification factors of the a-μ scheme are identical to those of the Leapfrog scheme if $\mu = 0$, and to those of the DuFort–Frankel scheme if $a = 0$. These coincidences are unexpected because the a-μ scheme and the above classical schemes are derived from completely different perspectives, and the a-μ scheme *does not* reduce to the above classical schemes in the limiting cases. The a-μ scheme is extended to solve the 1D time-dependent Navier–Stokes equations of a perfect gas. Stability of this *explicit* solver also is limited only by the CFL condition. In spite of the fact that it does not use (i) *any* techniques related to the high-resolution upwind methods, and (ii) *any* ad hoc parameter, the current *Navier–Stokes* solver is capable of generating highly accurate shock tube solutions. Particularly, for high-Reynolds-number flows, shock discontinuities can be resolved within one mesh interval. The inviscid ($\mu = 0$) a-μ scheme is reversible in time. It also is neutrally stable, i.e., free from numerical dissipation. Such a scheme generally cannot be extended to solve the Euler equations. Thus, the inviscid version is modified. Stability of this modified scheme, referred to as the a-ε scheme, is limited by the CFL condition and $0 \leq \varepsilon \leq 1$, where ε is a special parameter that controls numerical dissipation. Moreover, if $\varepsilon = 0$, the amplification factors of the a-ε scheme are identical to those of the Leapfrog scheme, which has no numerical dissipation. On the other hand, if $\varepsilon = 1$, the two amplification factors of the a-ε scheme become the same function of the Courant number and the phase angle. Unexpectedly, this function also is the amplification factor of the highly diffusive Lax scheme. Note that, because the Lax scheme is very diffusive and it uses a mesh that is staggered in time, a two-level scheme using such a mesh is often associated with a highly diffusive scheme. The a-ε scheme, which also uses a mesh staggering in time, demonstrates that it can also be a scheme with no numerical dissipation. The Euler extension of the a-ε scheme has stability conditions similar to those of the a-ε scheme itself. It has the unusual property that numerical dissipation at all mesh points can be controlled by a set of local parameters. Moreover, it is capable of generating accurate shock tube solutions with the CFL number ranging from close to 1 to 0.022

1. INTRODUCTION

The method of space-time conservation element and solution element [1–11] is a new numerical framework for solving conservation laws. This new approach differs substantially in both concept and methodology from the well-established methods, i.e., finite difference, finite volume, finite element, and spectral methods [12–16]. It is conceived and designed to overcome several key limitations of the above traditional methods. Thus, we shall begin this paper with a discussion of several considerations that motivate the current development:

(a) A set of physical conservation laws is a collection of statements of *flux conservation in space-time*. Mathematically, these laws are represented by a set of integral equations. The differential form of these laws is obtained from the integral form with the assumption that the *physical solution is smooth*. For a physical solution in a region of rapid change (e.g., a boundary layer), this smoothness assumption is difficult to realize by a numerical approximation that can use only a limited number of discrete variables. This difficulty becomes even worse in the presence of discontinuities (e.g., shocks). Thus, a method designed to obtain numerical solutions to the differential form without enforcing flux conservation is at a fundamental disadvantage in modeling physical phenomena with high-gradient regions. Particularly, it may not be used to solve flow problems involving shocks. Contrarily, a numerical solution obtained from a method that also enforces flux-conservation locally (i.e., down to a computational cell) and globally (i.e., over the entire computational domain) will always retain the basic physical reality of flux conservation even in a region

0021-9991/95 $12.00

involving discontinuities. For this reason, the enforcement of *both local and global flux conservation in space and time* is a tenet in the current development. To meet this requirement, first we define a set of *solution elements* which are subdomains in the space-time computation domain. Within each solution element, any physical flux vector is then approximated in terms of some simple smooth functions. In the last step, we divide the computational domain into *conservation elements* and demand that any flux be conserved over any space-time region that is the union of any combination of these elements. Note that a solution element generally is not a conservation element and vice versa.

Among the traditional methods, finite difference, finite element, and spectral methods are designed to solve the differential form of the conservation laws. Note that the set of integral equations usually solved in a finite-element scheme is equivalent to the differential form of the conservation laws assuming certain smoothness conditions. However, these integral equations generally are different from the integral equations representing the conservation laws. Even if they are cast into a conservative form, the resulting flux-conservation conditions generally do not represent the physical conservation laws.

The finite volume method is the only traditional method designed to enforce flux conservation. A finite-volume scheme may enforce flux conservation in space only, or in both space and time. As a preliminary to this enforcement, a flux must be assigned at any interface separating two neighboring conservation cells. In a typical finite-volume scheme, it is evaluated by extrapolating or interpolating the mesh values at the neighboring cells. This evaluation generally requires an ad hoc choice of a special flux model among many models available [17–19]. Generally numerical results obtained are dependent on which model one chooses. Also this process of interpolation and extrapolation generally is time consuming and has some undesirable side effects which will be discussed shortly.

Contrarily, by defining conservation elements wisely and considering the the spatial derivatives of dynamic variables as independent variables, current flux evaluation at an interface is carried out without interpolation or extrapolation. It is an integral part of the solution procedure.

(b) Space and time traditionally are treated separately in the time marching schemes. Generally one obtains a system of ordinary differential equations with time being the independent variable after a spatial discretization. As an example, elements in the finite element method usually are used for spatial discretization. These elements are domains in space only.

Because flux conservation is fundamentally a property in *space-time*, space and time are unified and treated on the same footing in the present method. Thus, conservation elements and solution elements used in the time-dependent version of the present method are domains in space-time. The significance of this unified approach cannot be overemphasized. As will be shown, it makes it easier for a numerical analogue to share the same space-time symmetry of the physical laws.

(c) In a finite-difference scheme, derivatives at mesh points are expressed in terms of mesh values of dependent variables by using finite-difference approximations. The accuracy of these approximations, especially those of higher-order accuracy, generally is excellent as long as dependent variables vary slowly across a mesh interval. However, it may not be adequate if these variables vary too rapidly. Thus, in a high-gradient region, e.g., a boundary layer, accuracy may demand the use of an extremely fine mesh. In turn, a prohibitively high computing cost may result.

The present method avoids the above pitfall by (i) expressing the numerical solution within a solution element as an expansion in terms of certain base functions, and (ii) considering the expansion coefficients as *the independent numerical variables to be solved for simultaneously*. For simplicity, Taylor's expansions will be used in the present paper. For this special case, the expansion coefficients are interpreted as the numerical analogues of the derivatives. Note that (i) van Leer [20] also has attempted to improve accuracy by introducing two independent numerical variables for each independent physical variable, and (ii) the current solution procedure has no resemblance with those used in compact difference schemes.

(d) The numerical variables used in a spectral method, i.e., the expansion coefficients, are global parameters pertaining to the entire computational domain. As a result, a spectral method generally (i) lacks local flexibility and thus may be applied only to problems with simple geometry, and (ii) is hindered by the fact that it must deal with a full matrix that is difficult to invert.

By design, only local parameters will be used in the present method. Moreover, solution elements and conservation elements are defined such that the set of discrete variables in any one of the numerical equations to be solved generally is associated with only two neighboring solution elements. The exception to this general rule occurs only in the situation in which numerical dissipation is introduced deliberately. Even in this special case, only the discrete variables associated with a few *immediately* neighboring solution elements will enter any equation to be solved. Thus, a scheme developed using the present method generally has the simplest stencil and one needs only to deal with a very sparse matrix if the scheme is implicit. Moreover, the maximum number of solution elements involved in a numerical equation of the current discretization framework is independent of the order of accuracy of a particular scheme. The order of accuracy can be raised by using a Taylor's expansion of higher order as the approximated solution within a solution element. Contrarily, the order of accuracy of a classical finite-difference scheme generally can be increased only by using variables of more mesh points in each of its equations. Usually, a side effect of this practice is an increase in numerical dissipation, a subject to be discussed shortly. Also it may be difficult to implement a high-order finite-difference scheme near a boundary because there are no *real* mesh points outside the boundary. The above discussions also point to another im-

portant advantage of the present method, i.e., *the specification of initial/boundary conditions generally is simpler, more flexible, and more accurate than that associated with a traditional method.* It is simpler because a smaller stencil is used [6]. Furthermore, it is more flexible and accurate because the spatial derivatives of dynamical variables, which are considered as independent numerical variables in the present method, can now be specified directly. Note that the current emphasis in reducing the size of the stencil is also consistent with a fundamental physical reality, i.e., in the absence of body force, *direct* physical interaction occurs only among the *immediate* neighbors

(e) With a few exceptions, numerical dissipation generally appears in a numerical solution of a time-marching problem. In other words, the numerical solution dissipates faster than the corresponding physical solution. For a nearly inviscid problem, e.g., flow with a high Reynolds number, this could be very serious because numerical dissipation may overwhelm physical dissipation and cause a complete distortion of solutions. One may argue that numerical dissipation can be reduced by increasing the order of accuracy of the scheme used. However, because the order of accuracy of a scheme is generally determined with the aid of Taylor's expansion, and the latter is valid only for a smooth solution, it has meaning only for a smooth solution. Thus the use of a scheme of higher-order accuracy may not reduce numerical dissipation associated with high-frequency Fourier components of a numerical solution. This is the reason that the Leapfrog scheme, which is free from numerical dissipation, can outperform schemes with higher-order accuracy in solving some wave equations [21].

In a study of finite-difference analogues of a simple convection equation [2], it was shown that a numerical analogue will be free from numerical dissipation if it does not violate certain space-time invariant properties of the convection equation. In other words, numerical dissipation may be considered as a result of *symmetry-breaking* by the numerical scheme. Because of its intrinsic nature of space-time unity, the current framework is an excellent vehicle for constructing a numerical analogue that shares the same space-time invariant properties with the physical equation.

It is recognized that a certain amount of numerical dissipation may be needed to prevent large dispersive errors [22] that are often caused by the presence of high-frequency disturbances (such as round-off errors). Therefore, in the present paper we shall construct a model scheme for a simple convection equation in which its numerical dissipation is controlled by a single adjustable parameter. The numerical dissipation is shut off when this parameter is set to zero. Furthermore, an Euler solver will be constructed such that *its numerical dissipation at all mesh points can be controlled by a set of local parameters.*

(f) High-resolution upwind methods [16] form a special class of the finite volume method. In these methods, the flux at an interface separating two neighboring conservation cells is also evaluated using a process of interpolation and extrapolation. This process generally is heavily dependent on characteristics-based techniques. For the 1D time-dependent case, the characteristics are curves in space-time, and the coefficient matrix associated with the Euler equations [23] also can be diagonalized easily. As a result, these techniques are easy to apply. However, for multidimensional cases, the characteristics are 2D or 3D surfaces in space-time [24]. Moreover, the coefficient matrices cannot be diagonalized simultaneously by the same matrix [23]. Because of the above complexites, application of these techniques to multidimensional problems is much more diffucult. Furthermore, high-resolution methods generally require the use of ad hoc parameters, e.g., flux-limiters and/or slope-limiters, and other ad hoc techniques. These ad hoc techniques may lead to numerical dissipation which varies from one place to another and from one Fourier component to another. In other words, numerical solutions may suffer annihilation of sharply different degrees at different locations and different frequencies [5, 25]. Also, these techniques generally are also difficult to apply in a space of higher dimension.

Although only the 1D time-marching schemes are constructed in the present paper, the current framework is developed to solve multidimensional problems. In order that 1D schemes can be extended to become multidimensional schemes in a straightforward manner, simplicity and generality weigh heavily in the development of the present method. Thus, we do not use characteristics-based techniques, and also try to avoid using ad hoc techniques. Note that, except the Navier–Stokes solver, other 1D schemes described in the present paper have been extended to become their 2D counterparts [7, 8] (the extension of the Navier–Stokes solver will be dealt with in a separate paper). Also, because of the similarity in their design, each of the 2D schemes described in [7, 8] shares with its 1D version virtually the same fundamental characteristics. Furthermore, it is shown in [7] that a 2D Euler time-marching solver, which uses a uniform stationary mesh, is capable of generating highly accurate solutions for a 2D shock reflection problem used by Helen Yee and others [26]. Specifically, *both the incident and the reflected shocks can be resolved by a single data point without the presence of numerical oscillations near the discontinuity.*

In addition to being difficult to apply in a space of higher dimension, the concept of characteristics generally is also not applicable to the Navier-Stokes equations, which is non-hyperbolic in nature. Therefore, the decision not to use characteristics-based techniques also makes it easier for the present framework to solve the Navier–Stokes equations.

This completes the discussion of the motivation for the current development. In summary, the development is guided by the following requirements: (i) to enforce both local and global flux consevation in space and time with flux evaluation at an interface being an integral part of the solution procedure and requiring no interpolation or extrapolation; (ii) space and time

are unified and treated on the same footing; (iii) mesh values of dependent variables and their derivatives are considered as independent variables to be solved for simultaneously; (iv) to use only local discrete variables; (v) solution elements and conservation elements be defined such that the simplest stencil will result; (vi) to minimize numerical dissipation, a numerical analogue should be constructed, as much as possible, to be compatible with the space-time invariant properties of the corresponding physical equations; and (vii) to exclude the use of the characteristics-based techniques, and to avoid the use of ad hoc techniques as much as possible. It is the purpose of this paper and its follow-ups [6–8] to show that the above requirements can be met with a simple unified numerical framework.

For any reader who is interested in getting an advance idea on how simple the present method can be, he is referred to the computer program listed at the end of the present paper. It is a shock-tube-problem solver constructed using the present method. The simplicity of the solver is easily appreciated by a comparison of the listed program and a typical program associated with high-resolution upwind methods *Not only is the listed program much smaller in size (it is self-contained and the main loop contains only 33 lines), but it contains no Fortran statements such as "if," "amax," and "amin" which are used so often in the programs implementing high-resolution methods.* The absence of the above Fortran statements in the listed program results from the effort in avoiding the use of the ad hoc techniques in the development of the present method. In spite of its simplicity, it will be shown in Section 7 that the present solver is capable of generating highly accurate shock tube solutions.

2. THE a-μ SCHEME

In this section, we consider a dimensionless form of the 1D convection–diffusion equation, i.e.,

$$\frac{\partial u}{\partial t} + a\frac{\partial u}{\partial x} - \mu\frac{\partial^2 u}{\partial x^2} = 0, \tag{2.1}$$

where the convection velocity a, and the viscosity coefficient μ (≥ 0) are constants. Let $x_1 = x$, and $x_2 = t$ be considered as the coordinates of a two-dimensional Euclidean space E_2. By using Gauss' divergence theorem in the space-time E_2, it can be shown that Eq. (2.1) is the differential form of the integral conservation law

$$\oint_{S(V)} \mathbf{h}\cdot d\mathbf{s} = 0. \tag{2.2}$$

As depicted in Fig. 1, here (i) $S(V)$ is the boundary of an arbitrary space-time region V in E_2, (ii) $\mathbf{h} = (au - \mu\partial u/\partial x, u)$ is a current density vector in E_2, and (iii) $d\mathbf{s} = d\sigma\,\mathbf{n}$ with $d\sigma$ and $\mathbf{n}$, respectively, being the area and the outward unit normal of a surface element on $S(V)$. Note that (i) $\mathbf{h}\cdot d\mathbf{s}$ is the *space-time* flux of $\mathbf{h}$ leaving the region V through the surface element $d\mathbf{s}$, and (ii) all mathematical operations can be carried out as though E_2 were an ordinary two-dimensional Euclidean space.

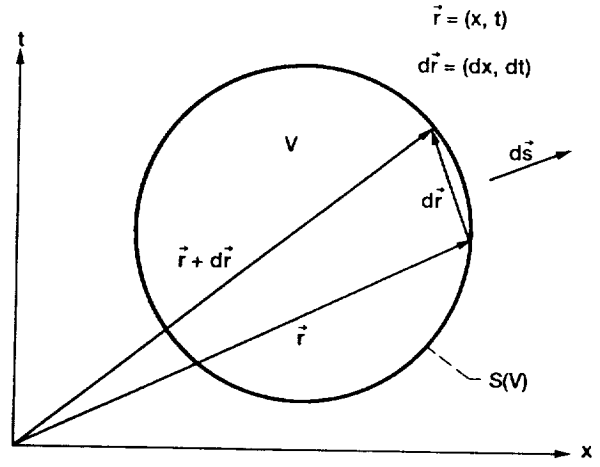

FIG. 1. A surface element $d\mathbf{s}$ and a line segment $d\mathbf{r}$ on the boundary $S(V)$ of a volume V in a space-time E_2.

At this juncture, note that the conservation law given in Eq. (2.2) is formulated in a form in which space and time are unified and treated on the same footing. *This unity of space and time is also a tenet in the following numerical development. It is a key characteristic that distinguishes the present method from most of the traditional methods.*

Let Ω denote the set of mesh points (j, n) in E_2 (dots in Fig. 2(a)), where $n = 0, \pm\frac{1}{2}, \pm 1, \pm\frac{3}{2}, \pm 2, \pm\frac{5}{2}, \ldots$, and, for each n, $j = n \pm \frac{1}{2}, n \pm \frac{3}{2}, n \pm \frac{5}{2}, \ldots$. There is a solution element (SE) associated with each $(j, n) \in \Omega$. Let the solution element SE(j, n) be the *interior* of the *space-time* region bounded by a dashed curve depicted in Fig. 2(b). It includes a horizontal line segment, a vertical line segment, and their immediate neighborhood. For the following discussions, the exact size of this neighborhood does not matter.

For any $(x, t) \in$ SE(j, n), $u(x, t)$, and $\mathbf{h}(x, t)$, respectively, are approximated by $u^*(x, t; j, n)$ and $\mathbf{h}^*(x, t; j, n)$ which we shall define shortly. Let

$$u^*(x, t; j, n) = u_j^n + (u_x)_j^n(x - x_j) + (u_t)_j^n(t - t^n), \tag{2.3}$$

where (i) u_j^n, $(u_x)_j^n$, and $(u_t)_j^n$ are constants in SE(j, n), and (ii) (x_j, t^n) are the coordinates of the mesh point (j, n). Note that

$$u^*(x_j, t^n; j, n) = u_j^n, \quad \frac{\partial u^*(x, t; j, n)}{\partial x} = (u_x)_j^n, \quad \frac{\partial u^*(x, t; j, n)}{\partial t} = (u_t)_j^n, \tag{2.4}$$

Moreover, if we identify u_j^n, $(u_x)_j^n$, and $(u_t)_j^n$, respectively, with

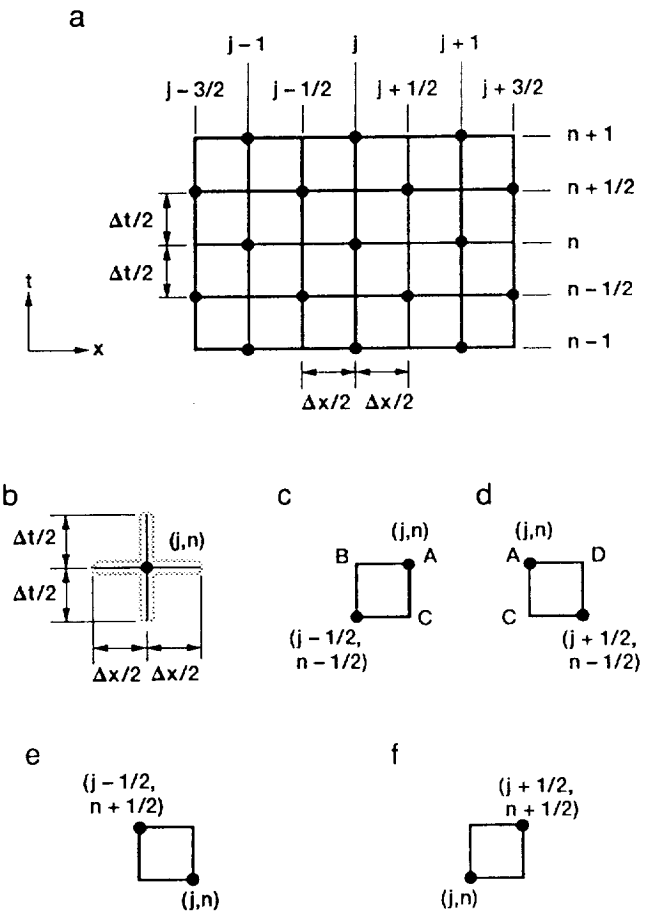

FIG. 2. The SEs and CEs of type I: (a) The relative positions of SEs and CEs; (b) SE(j, n); (c) CE (j, n); (d) CE$_+(j, n)$; (e) CE$_-(j - \frac{1}{2}, n + \frac{1}{2})$; (f) CE $(j + \frac{1}{2}, n + \frac{1}{2})$.

the values of u, $\partial u/\partial x$, and $\partial u/\partial t$ at (x_j, t^n), the expression on the right side of Eq. (2.3) becomes the first-order Taylor's expansion of $u(x, t)$ at (x_j, t^n). As a result of these considerations, u_j^n, $(u_x)_j^n$, and $(u_t)_j^n$ will be considered as the numerical analogues of the values of u, $\partial u/\partial x$, and $\partial u/\partial t$ at (x_j, t^n), respectively.

We shall require that $u = u^*(x, t; j, n)$ satisfy Eq. (2.1) within SE(j, n). As a result of Eq. (2.4), this implies that

$$(u_t)_j^n = -a(u_x)_j^n. \tag{2.5}$$

Because Eq. (2.3) is a first-order Taylor's expansion, the diffusion term in Eq. (2.1) has no counterpart in Eq. (2.5). As a result, the diffusion term has no impact on how $u^*(x, t; j, n)$ varies with time *within* SE(j, n). However, as will be shown shortly, through its role in the numerical analogue of Eq. (2.2), it does influence time-dependence of numerical solutions. Note that, for a higher-order scheme, how $u^*(x, t; j, n)$ varies with time within SE(j, n) will be influenced by the presence of the diffusion term. Combining Eqs. (2.3) and (2.5), one has

$$u^*(x, t; j, n) = u_j^n + (u_x)_j^n[(x - x_j) - a(t - t^n)], \quad (x, t) \in \text{SE}(j, n). \tag{2.6}$$

Because $\mathbf{h} = (au - \mu\partial u/\partial x, u)$, we define

$$\mathbf{h}^*(x, t; j, n) = (au^*(x, t; j, n) - \mu\partial u^*(x, t; j, n)/\partial x, u^*(x, t; j, n)). \tag{2.7}$$

Let E_2 be divided into nonoverlapping rectangular regions (see Fig. 2(a)) referred to as conservation elements (CEs). As depicted in Figs. 2(c) and 2(d), the CE with its top-right (top-left) vertex being the mesh point $(j, n) \in \Omega$ is denoted by CE (j, n) (CE$_+(j, n)$). Obviously the boundary of CE (j, n) (CE$_+(j, n)$), excluding two isolated points B and C (C and D), is formed by the subsets of SE(j, n) and SE$(j - \frac{1}{2}, n - \frac{1}{2})$ (SE$(j + \frac{1}{2}, n - \frac{1}{2})$). The current approximation of Eq. (2.2) is

$$F_+(j, n) \overset{\text{def}}{=} \oint_{S(\text{CE}_+(j,n))} \mathbf{h}^* \cdot d\mathbf{s} = 0 \tag{2.8}$$

for all $(j, n) \in \Omega$. In other words, the total flux leaving the boundary of any conservation element is zero. Note that the flux at any interface separating two neighboring CEs is calculated using the information from a single SE. As an example, the interface AC depicted in Figs. 2(c) and 2(d) is a subset of SE(j, n). Thus the flux at this interface is calculated using the information associated with SE(j, n). Also note that an SE is the *interior* of a space-time region. Thus the vertices B, C, and D, strictly speaking, do not belong to any SE. As a result, $\mathbf{h}^*$ is not defined at these points. However, contributions to the above integral from these isolated points are zero no matter what values of $\mathbf{h}^*$ are assigned to them. For this reason, one may simply exclude them from the above surface integration.

Because the surface integration across any interface separating two neighboring CEs is evaluated using the information from a single SE, obviously the local conservation condition Eq. (2.8) will lead to a global conservation relation, i.e., *the total flux leaving the boundary of any space-time region that is the union of any combination of CEs will also vanish.*

Because each $S(\text{CE}_+(j, n))$ is a simple closed curve in E_2 (see Fig. 1), the surface integration in Eq. (2.8) can be converted into a line integration. Let

$$\mathbf{g}^* \overset{\text{def}}{=} (-u^*, au^* - \mu\partial u^*/\partial x), \quad d\mathbf{r} \overset{\text{def}}{=} (dx, dt). \tag{2.9}$$

Thus, $d\mathbf{r}$ is normal to $d\mathbf{s}$ and points in the tangential direction of the line segment joining the two points (x,t) and $(x + dx, t + dt)$. Because $d\mathbf{s} = \pm(dt, -dx)$ [1, p.14], we have

$$\mathbf{h}^* \cdot d\mathbf{s} = \pm \mathbf{g}^* \cdot d\mathbf{r}, \tag{2.10}$$

where the upper (lower) sign should be chosen if the 90° rotation from $d\mathbf{s}$ to $d\mathbf{r}$ is in the counterclockwise (clockwise) direction. By combining Eqs. (2.8) and (2.10), one concludes that

$$F_-(j, n) = \oint_{S(\mathrm{CE}_-(j,n))}^{c.c.} \mathbf{g}^* \cdot d\mathbf{r}. \tag{2.11}$$

Note that the notation *c.c.* indicates that the line integration should be carried out in the counterclockwise direction. Substituting Eq. (2.6) into Eq. (2.11), and using the fact that the boundary of a CE is formed by the subsets of two SEs, one has

$$\begin{aligned} &\frac{4}{(\Delta x)^2} F_-(j, n) \\ &= \pm \left(\frac{1}{2}\right)[(1 - \nu^2 + \xi)(u_x)_j^n + (1 - \nu^2 - \xi)(u_x)_{j-1/2}^{n-1/2}] \\ &\quad + \frac{2(1 \mp \nu)}{\Delta x}(u_j^n - u_{j-1/2}^{n-1/2}), \end{aligned} \tag{2.12}$$

where

$$\nu \overset{\text{def}}{=} \frac{a\Delta t}{\Delta x}, \quad \xi \overset{\text{def}}{=} \frac{4\mu\Delta t}{(\Delta x)^2}. \tag{2.13}$$

Note that (i) the parameter ν is the Courant number, and (ii) a more efficient method of flux evaluation will be presented later in this section.

With the aid of Eqs. (2.8) and (2.12), u_j^n and $(u_x)_j^n$ can be solved in terms of $u_{j-1/2}^{n-1/2}$ and $(u_x)_{j-1/2}^{n-1/2}$ if $1 - \nu^2 + \xi \neq 0$; i.e., for all SE(j, n),

$$\begin{aligned} \mathbf{q}(j, n) = {} & Q_+\mathbf{q}(j - \tfrac{1}{2}, n - \tfrac{1}{2}) \\ & + Q_-\mathbf{q}(j + \tfrac{1}{2}, n - \tfrac{1}{2}) \quad (1 - \nu^2 + \xi \neq 0). \end{aligned} \tag{2.14}$$

Here

$$\mathbf{q}(j, n) \overset{\text{def}}{=} \begin{pmatrix} u_j^n \\ (\Delta x/4)(u_x)_j^n \end{pmatrix} \tag{2.15}$$

for all $(j, n) \in \Omega$, and

$$Q_+ \overset{\text{def}}{=} \left(\frac{1}{2}\right)\begin{pmatrix} 1 + \nu & 1 - \nu^2 - \xi \\ \dfrac{-(1 - \nu^2)}{1 - \nu^2 + \xi} & \dfrac{-(1 - \nu)(1 - \nu^2 - \xi)}{1 - \nu^2 + \xi} \end{pmatrix} \tag{2.16}$$

and

$$Q_- \overset{\text{def}}{=} \left(\frac{1}{2}\right)\begin{pmatrix} 1 - \nu & -(1 - \nu^2 - \xi) \\ \dfrac{1 - \nu^2}{1 - \nu^2 + \xi} & \dfrac{-(1 + \nu)(1 - \nu^2 - \xi)}{1 - \nu^2 + \xi} \end{pmatrix}. \tag{2.17}$$

Because numerical variables at a higher time level can be evaluated in terms of those at a lower time level by using Eq. (2.14), it defines a marching scheme. Furthermore, because this scheme models Eq. (2.1) which is characterized by two parameters a and μ, hereafter it will be referred to as the a-μ scheme.

As a preliminary for future developments, we apply Eq. (2.14) successively and obtain

$$\begin{aligned} \mathbf{q}(j, n + 1) = {} & (Q_+)^2\mathbf{q}(j - 1, n) \\ & + (Q_+Q_- + Q_-Q_+)\mathbf{q}(j, n) \\ & + (Q_-)^2\mathbf{q}(j + 1, n) \quad (1 - \nu^2 + \xi \neq 0). \end{aligned} \tag{2.18}$$

A result of Eq. (2.18) is

$$\mathbf{q}(j, n + 1) \to \mathbf{q}(j, n) \quad \text{as } \Delta t \to 0, \tag{2.19}$$

if a, μ, and Δx are held constant. The proof follows from the fact that

$$\begin{aligned} &(Q_+)^2 \to 0, \quad (Q_+Q_- + Q_-Q_+) \to 1, \\ &(Q_-)^2 \to 0 \quad \text{as } \Delta t \to 0, \end{aligned} \tag{2.20}$$

if a, μ, and Δx are held constant.

Alternatively, Eq. (2.19) can be proved using the fact that the total flux of $\mathbf{h}^*$ leaving the boundary of any space-time region that is the union of any combination of CEs vanishes. Consider the union of $\mathrm{CE}_+(j, n + 1)$ and $\mathrm{CE}_-(j + \frac{1}{2}, n + \frac{1}{2})$ (see Fig. 2). This union is a rectangle with the vertices $(j + \frac{1}{2}, n + 1)$, $(j, n + 1)$, (j, n) and $(j + \frac{1}{2}, n)$. The flux leaving this rectangle through its two vertical edges approaches zero as $\Delta t \to 0$. Because the total flux leaving its boundary vanishes, one concludes that the total flux leaving its two horizontal edges also approaches zero as $\Delta t \to 0$. In other words, the flux entering the rectangle through the lower horizontal edge approaches that leaving through the upper horizontal edge as $\Delta t \to 0$. Because these two fluxes are evaluated using $\mathbf{q}(j, n)$ and $\mathbf{q}(j, n + 1)$, respectively, the above limiting condition implies a limiting relation between $\mathbf{q}(j, n)$ and $\mathbf{q}(j, n + 1)$. Similarly, by considering the union of $\mathrm{CE}_-(j, n + 1)$ and $\mathrm{CE}_+(j - \frac{1}{2}, n + \frac{1}{2})$, one obtains another limiting relation for $\mathbf{q}(j, n)$ and $\mathbf{q}(j, n + 1)$. Equation (2.19) is a result of the above two limiting relations.

The a-μ scheme has several nontraditional features. They are summarized in the following remarks:

(a) Space and time are unified and treated on the same footing in the construction of the a-μ scheme.

(b) The expansion coefficients u_j^n and $(u_x)_j^n$ in Eq. (2.6) are treated as independent variables; i.e., $(u_x)_j^n$ is not expressed in terms of u_j^n's by using a finite-difference approximation.

(c) As a result of Eq. (2.12), each of the conservation conditions $F_\pm(j, n)$ involves only numerical variables associated with two neighboring SEs. This fact remains true for a scheme of higher-order accuracy in which Eq. (2.3) is replaced by a Taylor's expansion of higher-order. The contrast with the finite difference method and its physical significance were dicussed in Section 1.

(d) The a-μ scheme has the simplest stencil, i.e., a triangle with a vertex at the upper time level and the other two vertices at the lower time level. Equation (2.14), which relates numerical variables at these vertices, was derived using the flux conservation conditions $F_-(j, n) = 0$. Because the flux at an interface separating two neighboring CEs is evaluated using information of a single SE, no interpolation or extrapolation is required. Moreover, accuracy of flux evaluation is enhanced by requiring that $u = u^*(x, t; j, n)$ satisfy Eq. (2.1) within SE(j, n). This makes the use of characteristics-based techniques less necessary.

(e) The a-μ scheme uses a mesh that is staggered in time. As will be explained in Appendix A, for a two-level scheme using such a mesh, e.g., the Lax scheme [12, p.97], generally the numerical variable at $(j, n + 1)$ does not approach that at (j, n) as $\Delta t \to 0$, if a, μ, and Δx are held constant. This is a key reason why the Lax scheme is very diffusive when the Courant number ν is small. According to Eq. (2.19), the a-μ scheme is an exception to the above general rule.

(f) Equation (2.1) can be solved numerically using the Leapfrog/DuFort–Frankel scheme [12, p.161]. This scheme is reduced to the Leapfrog scheme [12, p.100] if diffusion is absent (i.e., $\mu = 0$), and to the DuFort–Frankel scheme [12, p.114] if convection is absent (i.e., $a = 0$). It is well known that a solution of any of the above schemes is formed by two decoupled solutions with each being associated with a mesh that is also staggered in time. Traditionally the von Neumann stability analysis for the above schemes is performed without taking into account this decoupled nature [12]. In Appendix A, it is performed separately for each decoupled solution using the mesh depicted in Fig. 2(a). It is shown that the amplification factors of the Leapfrog/DuFort–Frankel scheme are

$$A_\pm = \left\{ \frac{1}{1+\xi} [\xi \cos(\theta/2) - i\nu \sin(\theta/2) \pm \sqrt{[\xi \cos(\theta/2) - i\nu \sin(\theta/2)]^2 + 1 - \xi^2}] \right\}^2. \tag{2.21}$$

Here θ, $-\pi < \theta \le \pi$ [1, p.30], is the phase angle variation per Δx. Note that, in the present paper, the amplification factors are defined to be those between the time levels n and $n + 1$, i.e., they are the amplification factors of the solution after two marching steps. The reason behind this definition is that the mesh points at the time levels n and $n + 1$ are not staggered. Let $1 - \nu^2 \neq 0$. Then the amplification factors $G_\pm^{(1)}$ of the current a-μ scheme (see Eq. (6.9)) are identical to those given by Eq. (2.21) except that the parameter ξ should be replaced by $\hat{\xi} \overset{\text{def}}{=} \xi/(1 - \nu^2)$. Because (i) $\hat{\xi} = \xi = 0$ if $\mu = 0$, and (ii) $\nu = 0$ and thus $\hat{\xi} = \xi$, if $a = 0$, one concludes that $G_\pm^{(1)}$ are completely identical to those of the Leapfrog scheme if $\mu = 0$, and to those of the DuFort–Frankel scheme if $a = 0$. These coincidences are unexpected because the a-μ scheme and the above classical schemes are derived from completely different perspectives. Moreover, the a-μ scheme is a two-level scheme with two variables u_j^n and $(u_x)_j^n$ associated with the mesh point (j, n), while the above classical schemes are three-level schemes with a single variable u_j^n associated with the same point.

Because the amplification factors of the inviscid a-μ scheme (i.e., the a-μ scheme with $\mu = 0$) are identical to those of the Leapfrog scheme, the former, as in the case of the latter, is neutrally stable (i.e., free of numerical dissipation) if $\nu^2 < 1$. Note that the case with $\mu = 0$ and $\nu^2 = 1$ is ruled out by the assumption $1 - \nu^2 + \xi \neq 0$ of Eq. (2.14). Similarly, the pure-diffusion a-μ scheme (i.e., the a-μ scheme with $a = 0$), as in the case of the DuFort–Frankel scheme, is unconditionally stable. Furthermore, it is proved in Section 6 that the stability of the general a-μ scheme, as in the case of the Leapfrog/DuFort–Frankel scheme, *is independent of μ, and restricted only by the CFL condition*, i.e., $\nu^2 \le 1$. The a-μ scheme is the only *two-level explicit* scheme known to the author to possesss the above properties. Also it will be shown later that the same stability condition is retained by a natural 1D time-dependent Navier–Stokes extension of the a-μ scheme.

Because stability of the a-μ scheme is restricted only by the CFL condition, the stability bound for Δt is proportional to Δx. In contrast, the stability condition of a typical classical explicit scheme generally is more restrictive than the CFL condition. For a small mesh Reynolds number, the stability bound for Δt is approximately proportional to $(\Delta x)^2$ for the MacCormack scheme [12, p.102].

Because a neutrally stable numerical analogue of the pure convection equation

$$\frac{\partial u}{\partial t} + a \frac{\partial u}{\partial x} = 0 \tag{2.22}$$

usually becomes unstable when it is applied to a nonlinear inviscid generalization of Eq. (2.22), the inviscid a-μ scheme will be modified in Section 3 such that it can be extended to model the Euler equations. In this new version, numerical dissipation is introduced in a way that allows its magnitude to be adjusted by a special parameter.

(g) The conservation relations for CE$_-(j - \frac{1}{2}, n + \frac{1}{2})$ and CE $(j + \frac{1}{2}, n + \frac{1}{2})$ (see Figs. 2(e) and 2(f)) are

$$F_-(j - \tfrac{1}{2}, n + \tfrac{1}{2}) = 0, \quad F_-(j + \tfrac{1}{2}, n + \tfrac{1}{2}) = 0, \tag{2.23}$$

respectively. Combining Eqs. (2.12) and (2.23) and assuming $1 - \nu^2 - \xi \neq 0$, one has

$$\mathbf{q}(j, n) = \hat{Q}_+ \cdot \mathbf{q}(j + \tfrac{1}{2}, n + \tfrac{1}{2}) + \hat{Q}_- \mathbf{q}(j - \tfrac{1}{2}, n + \tfrac{1}{2}) \quad (1 - \nu^2 - \xi \neq 0). \tag{2.24}$$

Here

$$\hat{Q}_+ \overset{\text{def}}{=} \left(\frac{1}{2}\right) \begin{pmatrix} 1 + \nu & -(1 - \nu^2 + \xi) \\ \dfrac{1 - \nu^2}{1 - \nu^2 - \xi} & \dfrac{-(1 - \nu)(1 - \nu^2 + \xi)}{1 - \nu^2 - \xi} \end{pmatrix}, \tag{2.25}$$

and

$$\hat{Q}_- \overset{\text{def}}{=} \left(\frac{1}{2}\right) \begin{pmatrix} 1 - \nu & 1 - \nu^2 + \xi \\ \dfrac{-(1 - \nu^2)}{1 - \nu^2 - \xi} & \dfrac{-(1 + \nu)(1 - \nu^2 + \xi)}{1 - \nu^2 - \xi} \end{pmatrix}. \tag{2.26}$$

Equation (2.24) defines a backward marching scheme, i.e., the numerical variables at the time level n are determined in terms of those at the time level $(n + \frac{1}{2})$. Recall that both the forward marching scheme Equation (2.14) and the backward marching scheme Eq. (2.24) are derived using the same set of conservation relations. As a matter of fact, Eqs. (2.14) and (2.24) are equivalent if $(1 - \nu^2)^2 \neq (\xi)^2$ is assumed. For the above reason, the a-μ scheme may be referred to as a *two-way marching* scheme. For the case $\mu > 0$, it will be proved in Section 6 that the a-μ scheme cannot be stable for both the forward and the backward marching directions, except for the singular case $\nu^2 = 1$ which is also on the threshold of instability. Thus, for all practical purposes the viscous a-μ scheme is irreversible in time. On the other hand, it is neutrally stable for both the forward and backward marching directions, and thus is reversible in time, if $\mu = 0$, and $\nu^2 < 1$. Again, the a-μ scheme is the only two-level explicit two-way marching scheme known to the author.

(h) Several invariant properties of Eq. (2.1) with respect to space and time are discussed in [2]. In the same paper, these properties are also defined for the numerical analogues of Eq. (2.1). It is also shown that the neutral stability of several finite-difference analogues of Eq. (2.22) can be established by using their invariant properties with respect to space-time inversion. Because solutions of Eq. (2.22) do not dissipate with time, it is not surprising that solutions of a numerical analogue also will not dissipate with time, i.e., the scheme is neutrally stable, if it shares with Eq. (2.22) some space-time invariant properties. It will be shown in a future paper that the a-μ scheme shares with Eq. (2.1) the same space-time invariant properties. Also note that these invariant properties are closely linked with the other properties discussed in (a), (e), (f), and (g).

This completes the discussion on nontraditional features of the a-μ scheme. In the following, it will be shown that this scheme can also be constructed from a completely different perspective. As a part of this construction, SEs and CEs of different types will be used and discussed.

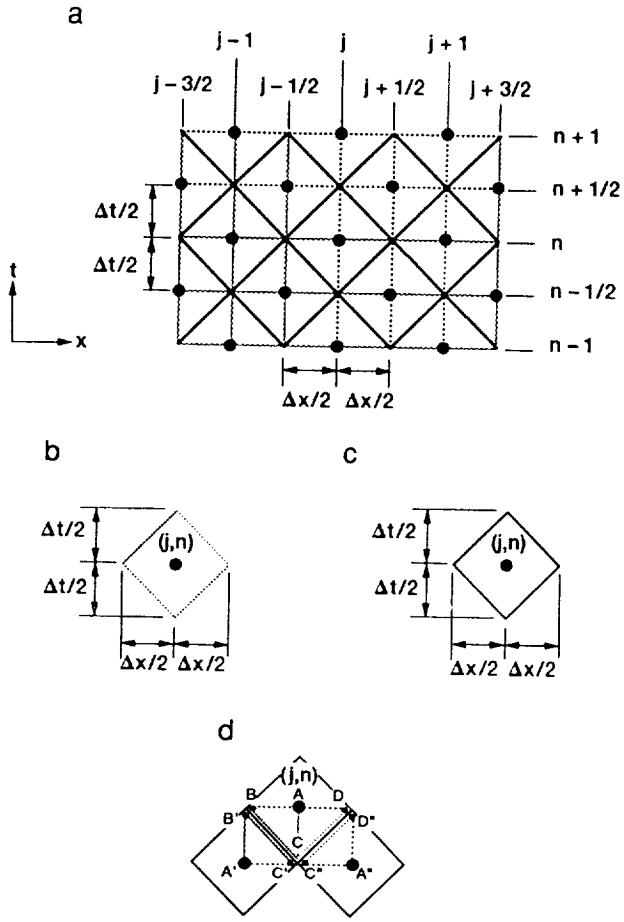

FIG. 3. The SEs and CEs of type II: (a) The relative positions of SEs and CEs; (b) SE(j, n); (c) CE(j, n); (d) Three neighboring CEs.

In the new construction, the locations of mesh points (dots in Fig. 3(a)) are identical to those used in the original construction. However, SE(j, n) is defined to be the *interior* of a rhombus centered at (j, n) (see Fig. 3(b)). CE(j, n) is the union of SE(j, n) and its boundary. Readers are warned *not* to confuse the sides of the rhombus with the characteristics of Eq. (2.22). Any one of these sides is simply a line segment joining two points of intersection (not marked by dots) of horizontal and vertical mesh lines. For any $(x, t) \in$ SE(j, n), $u(x, t)$ and $\mathbf{h}(x, t)$, respectively, again are approximated by $u^*(x, t; j, n)$ and $\mathbf{h}^*(x, t; j, n)$ which are defined by Eqs. (2.3) and (2.7), respectively. However, Eq. (2.5) will be derived from a consideration of flux conservation.

Let Eq. (2.2) be approximated by

$$\oint_{S(V^*)} \mathbf{h}^* \cdot d\mathbf{s} = 0, \tag{2.27}$$

where V^* is *the union of any combination of CEs*. Because an

SE is the interior of a CE, $\mathbf{h}^*$ is not defined on $S(V^*)$, the boundary of V^*. As a result, the above surface integration is to be carried out over a surface that is in the interior of V^* and immediately adjacent to $S(V^*)$. A necessary condition of Eq. (2.27) is that, for all $(j, n) \in \Omega$,

$$\oint_{S(\mathrm{CE}(j,n))} \mathbf{h}^* \cdot d\mathbf{s} = 0; \tag{2.28}$$

i.e., the total flux leaving any conservation element is zero.

Note that the center of a current SE no longer sits on an interface separating two CEs. It coincides with the center of a CE. Thus $\mathbf{h}^*$ at one side of an interface is evaluated using information from one SE, while that at the other side is evaluated using information from another SE. As an example, $\mathbf{h}^*$ at BC and $B'C'$ depicted in Fig. 3(d), respectively, are evaluated using information from $\mathrm{SE}(j, n)$ and $\mathrm{SE}(j - \frac{1}{2}, n - \frac{1}{2})$. Another necessary condition for Eq. (2.27) *is the equality between the fluxes entering and leaving any interface*. This can be seen by applying Eq. (2.27) separately to two neighboring CEs, and then to their union. Obviously the local flux conservation relations at all interfacs, and within all CEs (i.e., Eq. (2.28)) are equivalent to the global conservation relation Eq. (2.27). The equations representing the above conservation conditions are the numerical equations to be solved. Note that, in the current construction, a flux is not preassigned at an interface using an interpolation or extrapolation of information from both sides of this interface. The present method of interface flux evaluation obviously is different from that used in the finite volume method which was discussed in Section 1.

By using Eqs. (2.3) and (2.7), one concludes that, for any $(x, t) \in \mathrm{SE}(j, n)$, the divergence of $\mathbf{h}^*$ in E_2 is

$$\begin{aligned} \nabla \cdot \mathbf{h}^* &\stackrel{\text{def}}{=} \frac{\partial [a u^*(x, t; j, n) - \mu \partial u^*(x, t; j, n)/\partial x]}{\partial x} \\ &\quad + \frac{\partial u^*(x, t; j, n)}{\partial t} \\ &= a(u_x)_j^n + (u_t)_j^n. \end{aligned} \tag{2.29}$$

Because $(u_x)_j^n$ and $(u_t)_j^n$ are constants within an SE, Eq. (2.29) implies that $\nabla \cdot \mathbf{h}^*$ is also a constant. Thus Eq. (2.28) coupled with Gauss' divergence theorem implies that, within any SE,

$$\nabla \cdot \mathbf{h}^* = 0. \tag{2.30}$$

Equation (2.5) is a direct result of Eqs. (2.29) and (2.30).

Note that Eq. (2.30) follows from Eq. (2.28) because $u^*(x, t; j, n)$ defined in Eq. (2.3) is a first-order Taylor's expansion For a higher-order expansion, the condition that Eq. (2.30) being valid uniformly within an SE is stronger than Eq. (2.28) For the general case, the stronger condition should be imposed. Because Eq. (2.30) is the numerical analogue of Eq. (2.1), the imposition of the stronger condition ensures that, *within an SE, the numerical solution uniformly satisfies the differential form of the conservation law Eq.* (2.2).

With the aid of Gauss' divergence theorem, Eq. (2.30) implies that the surface integration of $\mathbf{h}^*$ over any closed surface located within any SE vanishes. As a result,

$$\oint_{S(\triangle ABC)} \mathbf{h}^* \cdot d\mathbf{s} = 0, \quad \oint_{S(\triangle A'B'C')} \mathbf{h}^* \cdot d\mathbf{s} = 0, \tag{2.31}$$

where the triangles $\triangle ABC$ and $\triangle A'B'C'$ are those depicted in Fig. 3(d). Because the net flux of $\mathbf{h}^*$ entering an interface from both sides vanishes, the sum of the flux leaving $\mathrm{CE}(j, n)$ through BC and that leaving $\mathrm{CE}(j - \frac{1}{2}, n - \frac{1}{2})$ through $B'C'$ vanishes. Thus, Eq. (2.31) implies that $F_\cdot(j, n) = 0$, where $F_\cdot(j, n)$ is defined in Eq. (2.11). Similarly, it can be shown that $F_\cdot(j, n) = 0$.

Assuming Eqs. (2.3) and (2.7), it has been shown that both Eqs. (2.5) and (2.8) can be derived using Eq. (2.27). Conversely, Eq. (2.27) also follows from Eqs. (2.5) and (2.8). Obviously both the forward marching scheme Eq. (2.14) and the backward marching scheme Eq. (2.22) can also be obtained by assuming Eqs. (2.3), (2.7), and (2.27).

Note that the equivalence between Eq. (2.27) and the pair of equations Eqs. (2.5) and (2.8) hinges on the fact that $\nabla \cdot \mathbf{h}^* = 0$ within an SE of either type I or type II. As will be shown immediately, this condition can be used to simplify evaluation of the flux across a simple curve that *lies entirely within an SE of either type*.

According to the top expression given in Eq. (2.29), $\nabla \cdot \mathbf{h}^* = 0$ implies that there exists a function $\psi^*(x, t; j, n)$ such that

$$\frac{\partial \psi(x, t; j, n)}{\partial t} = a u^*(x, t; j, n) - \mu \frac{\partial u^*(x, t; j, n)}{\partial x} \tag{2.32}$$

and

$$-\frac{\partial \psi(x, t; j, n)}{\partial t} = u^*(x, t; j, n) \tag{2.33}$$

for any $(x, t) \in \mathrm{SE}(j, n)$. Substituting Eq. (2.6) into Eqs. (2.32) and (2.33), one concludes that, up to an arbitrary constant,

$$\begin{aligned} \psi(x, t; j, n) = -\frac{(u_x)_j^n}{2} \{ &[(x - x_j) - a(t - t^n)]^2 \\ &+ 2\mu(t - t^n) \} - u_j^n [(x - x_j) - a(t - t^n)]. \end{aligned} \tag{2.34}$$

Moreover, with the aid of Eq. (2.9), Eqs. (2.32) and (2.33) imply that

$$\mathbf{g}^* \cdot d\mathbf{r} = d\psi. \tag{2.35}$$

Let $(x, t) \in \mathrm{SE}(j, n)$ and $(x', t') \in \mathrm{SE}(j, n)$. Let Γ be a simple

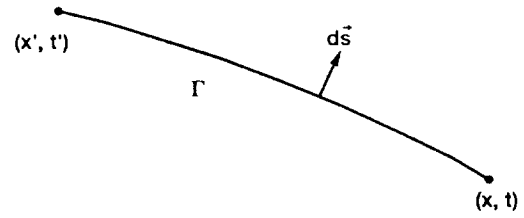

FIG. 4. A simple curve Γ joining (x, t) and (x', t').

curve joining (x, t) and (x', t'), and *lying entirely within* SE(j, n) (see Fig. 4). Then Eqs. (2.10) and (2.35) imply that

$$\int_\Gamma \mathbf{h}^* \cdot d\mathbf{s} = \psi(x', t'; j, n) - \psi(x, t; j, n). \tag{2.36}$$

Here we assume that $d\mathbf{s}$ points to the right of Γ if one moves forward from (x, t) to (x', t') (see Fig. 4). Equation (2.36) states that the flux of $\mathbf{h}^*$ across the curve Γ is given by the difference in the values of ψ at its two end-points. For this reason, $\psi(x, t; j, n)$ will be referred to as the potential function associated with SE(j, n). Obviously, Eq. (2.12) can be obtained using Eq. (2.36).

Note that a generalized a-μ scheme with a moving mesh was constructed in [1]. This scheme is reduced to the present a-μ scheme when the mesh becomes stationary. In [1], the generalized scheme is subjected to a thorough theoretical and numerical analysis on stability, dissipation, dispersion, consistency, truncation error, and accuracy. It is shown that it has many advantages over the MacCormack and the Leapfrog/Dufort–Frankel schemes. Particularly, by using a new discrete Fourier error analysis, it is shown that the generalized scheme is more accurate than the Leapfrog/DuFort–Frankel scheme by one order (in a sence defined in [1]) in both initial-value specification and the main marching scheme. Other key results of [1] are summarized in the following remarks:

(a) For the generalized scheme, (i) stability and accuracy can be improved, and (ii) dissipation and dispersion can be reduced, if the space-time mesh is allowed to evolve with the physical variables such that the local convective motion of physical variables relative to the moving mesh is kept to a minimum.

(b) For a numerical analogue of Eq. (2.22) that has both principal and spurious amplification factors, a numerical solution with periodic boundary conditions is the sum of a principal solution and a spurious solution [1, p.32]. Only the principal solution contributes to the accuracy of the scheme. Note that (i) the behaviors of the principal and the spurious solutions as functions of time are determined by the principal and spurious amplification factors, respectively; (ii) both two amplification factors of the present inviscid a-μ scheme are of unit magnitude; and (iii) given an accurate initial-value specification, the spurious solution at $t = 0$ generally is very small compared with the principal solution. As a result, the spurious solution of the present a-μ scheme generally is negligible. Furthermore, for the inviscid a-μ scheme, it is shown that [1, pp. 36-37] (i) the principal solution has no dispersion if $\nu = 0$ or in the limit of $\nu^2 \to 1$; and (ii) each Fourier component of the principal solution has a convection velocity not more than a and not less than $(2/\pi)a$ for all phase angles and $\nu^2 < 1$. In other words, *the dispersion associated with the inviscid a-μ scheme is small compared with that associated with a typical finite-difference scheme.*

In conclusion, a model scheme has been constructed from two different perspectives using SEs and CEs of different types. Using either perspective, one can say that *a numerical solution generated using the current framework satisfies* (i) *the differential form of the conservation law uniformly within an SE, and* (ii) *the integral form over any region that is the union of any combination of CEs.* The second perspective that used the SEs and CEs of type II depicted in Fig. 3 was used in the initial development of the present method [1]. In addition, it also was adopted to develop several new solvers for the 2D steady, incompressible Navier–Stokes equations [4, 9–11]. However, in these new solvers, the CEs and SEs depicted in Fig. 3 are replaced by CEs and SEs of rectangular shape in the 2D *spatial* computational domain. It was shown that, for a laminar channel flow with $\mathrm{Re}_L = 100$, an accurate solution can be obtained by using as few as six SEs across the channel.

Because (i) the first perspective is easier to use in constructing explicit schemes, and (ii) the schemes to be discussed in the present paper are exclusively explicit, the first perspective will be adopted in the present paper hereafter.

3. THE a-ε SCHEME

The inviscid a-μ scheme is neutrally stable and reversible in time. It is well known that a neutrally stable numerical analogue of Eq. (2.22) generally becomes unstable when it is extended to model the Euler equations. It is also obvious that a scheme that is reversible in time cannot model a physical problem that is irreversible in time, e.g., an inviscid flow problem involving shocks. In this section, we assume $\mu = 0$ and attempt to modify the inviscid a-μ scheme such that it can be extended to model the Euler equations.

The current path of development is almost identical to that given in Section 2. We continue to assume Eqs. (2.3)–(2.7), and use SEs of type I depicted in Fig. 2. In addition to $\mu = 0$, the *only* other modification is the replacement of the assumption $F_-(j, n) = 0$ by

$$F_-(j, n) = \pm \frac{\varepsilon(1 - \nu^2)(\Delta x)^2}{4} (du_x)_j^n, \tag{3.1}$$

where ε is a parameter independent of numerical variables, and

$$(du_s)_j^n \overset{\text{def}}{=} (\tfrac{1}{2})[(u_s)_{j+1/2}^{n-1/2} + (u_s)_{j-1/2}^{n-1/2}] - (u_{j+1/2}^{n-1/2} - u_{j-1/2}^{n-1/2})/\Delta x. \quad (3.2)$$

In other words, we add two terms of the same magnitude but with opposite signs, respectively, to the right sides of the original conservation conditions $F_+(j, n) = 0$ and $F_-(j, n) = 0$. The beauty of this modification will be fully explained later in this section. For now it suffices to say that this modification injects a higher-order *finite-difference* error into the inviscid a-μ scheme. It breaks the space-time symmetry of the latter. In turn, numerical dissipation is introduced as a result of this symmetry breaking. Because the magnitude of the terms added in this modification is controlled by ε, numerical dissipation is controlled by ε in the modified scheme just as physical dissipation is controlled by μ in the a-μ scheme. Note that, as a result of Eq. (3.1) and the assumption $\mu = 0$, the modified scheme is characterized by two parameters a and ε. Thus, hereafter it will be referred to as the a-ε scheme. Also note that, because there is no upwind bias in the a-ε scheme, *upwind bias is not the source of numerical dissipation*. Additional remarks on Eqs. (3.1) and (3.2) are:

(a) By definition, $F_-(j, n)$ and $F_+(j, n)$ represent total fluxes leaving $\text{CE}_-(j, n)$ and $\text{CE}_+(j, n)$, respectively (see Figs. 2(c) and 2(d)). Because $F_-(j, n) \neq 0$ if $\varepsilon \neq 0$, $\text{CE}_-(j, n)$ and $\text{CE}_-(j, n)$ generally are no longer conservation elements in the a-ε scheme.

(b) Let $\text{CE}(j, n)$ be the union of $\text{CE}_+(j, n)$ and $\text{CE}_-(j, n)$ (see Fig. 5(b)). Note that *this definition of* $\text{CE}(j, n)$ *differs from that given in Section* 2 *and depicted in Fig.* 3(c). Let

$$F(j, n) \overset{\text{def}}{=} \oint_{S(\text{CE}(j,n))} \mathbf{h}^* \cdot d\mathbf{s}. \quad (3.3)$$

Because the net flux entering the interface separating $\text{CE}_+(j, n)$ and $\text{CE}_-(j, n)$ is zero. $F(j, n)$ is the sum of $F_+(j, n)$ and $F_-(j, n)$. With the aid of Eq. (3.1), we have

$$F(j, n) = F_+(j, n) + F_-(j, n) = 0; \quad (3.4)$$

i.e., the total flux leaving $\text{CE}(j, n)$ vanishes. As a result, $\text{CE}(j, n)$ *is a conservation element in the* a-ε *scheme*. Note that Eq. (3.4) leads to a global conservation relation in the form of Eq. (2.27), where V^* is the union of any combination of these new CEs.

(c) Because $\xi = 0$ if $\mu = 0$, Eq. (3.4) coupled with Eq. (2.12) implies that

$$u_j^n = \left(\frac{1}{2}\right)[(1 + \nu)u_{j-1/2}^{n-1/2} + (1 - \nu)u_{j+1/2}^{n-1/2}] + \frac{\Delta x(1 - \nu^2)}{8}[(u_s)_{j-1/2}^{n-1/2} - (u_s)_{j+1/2}^{n-1/2}]. \quad (3.5)$$

Thus, u_j^n is independent of ε.

(d) Because $(u_s)_{j\pm1/2}^{n-1/2}$ is a numerical analogue of $\partial u/\partial x$ at $(j \pm \frac{1}{2}, n - \frac{1}{2})$, the simple average

$$(\tfrac{1}{2})[(u_s)_{j+1/2}^{n-1/2} + (u_s)_{j-1/2}^{n-1/2}]$$

is a numerical analogue of $\partial u/\partial x$ at $(j, n - \frac{1}{2})$, the midpoint of a line segment joining $(j + \frac{1}{2}, n - \frac{1}{2})$ and $(j - \frac{1}{2}, n - \frac{1}{2})$ (see Fig. 2(a)). Note that $(j, n - \frac{1}{2}) \notin \Omega$ if $(j, n) \in \Omega$. Also note that

$$(u_{j+1/2}^{n-1/2} - u_{j-1/2}^{n-1/2})/\Delta x$$

is a central-difference analogue of $\partial u/\partial x$ at $(j, n - \frac{1}{2})$. Thus, $(du_s)_j^n$ represents the *difference of two numerical analogues of* $\partial u/\partial x$ *at the same mesh point* $(j, n - \frac{1}{2})$. By using Taylor's expansion at $(j, n - \frac{1}{2})$, it can be shown that $(du_s)_j^n = O[(\Delta x)^2]$, if $(u_s)_{j\pm1/2}^{n-1/2}$ are identified with $\partial u(x_{j\pm1/2}, t^{n-1/2})/\partial x$, respectively. Hereafter a quantity is denoted by $O[(\Delta x)^l]$ if there exists a constant $C > 0$ such that the absolute value of this quantity $\leq C\,|\Delta x|^l$ for all sufficiently small $|\Delta x|$. Note that we have constructed an expression of $O[(\Delta x)^2]$ without explicitly introducing the factor $(\Delta x)^2$. This natural construction leads to the simple stability conditions to be given in Eq. (3.14). It is possible only because there are two discrete variables u_j^n and $(u_s)_j^n$ associated with the mesh point (j, n).

(e) Equation (3.1) could have been written as $F_\pm(j, n) = \pm\varepsilon'(du_s)_j^n$ with $\varepsilon' = \varepsilon(1 - \nu^2)(\Delta x)^2/4$. However, this *simplified* expression would lead to much more complicated equations later.

This completes the discussion of Eqs. (3.1) and (3.2). Now, let $1 - \nu^2 \neq 0$. Then Eqs. (2.12), (3.1), and (3.2) can be used to obtain the current counterparts of Eqs. (2.14) and (2.18). They are

$$\mathbf{q}(j, n) = M_+\mathbf{q}(j - \tfrac{1}{2}, n - \tfrac{1}{2}) + M_-\mathbf{q}(j + \tfrac{1}{2}, n - \tfrac{1}{2}) \quad (1 - \nu^2 \neq 0) \quad (3.6)$$

and

$$\mathbf{q}(j, n + 1) = (M_+)^2\mathbf{q}(j - 1, n) + (M_+M_- + M_-M_+)\mathbf{q}(j, n) + (M_-)^2\mathbf{q}(j + 1, n) \quad (1 - \nu^2 \neq 0), \quad (3.7)$$

respectively. Here

$$M_+ \overset{\text{def}}{=} (\tfrac{1}{2})\begin{pmatrix} 1 + \nu & 1 - \nu^2 \\ \varepsilon - 1 & 2\varepsilon - 1 + \nu \end{pmatrix} \quad (3.8)$$

and

$$M \overset{\text{def}}{=} (\tfrac{1}{2}) \begin{pmatrix} 1-\nu & -(1-\nu^2) \\ 1-\varepsilon & 2\varepsilon - 1 - \nu \end{pmatrix} \tag{3.9}$$

Obviously, $M_\pm = Q_\pm$ if $\varepsilon = 0$ and $\xi = 0$. Furthermore, the limiting condition given in Eq. (2.19) is still valid *if we assume that* $\varepsilon = \varepsilon(\Delta t)$ *and* $\lim_{\Delta t \to 0} \varepsilon(\Delta t) = 0$. However, unlike the a-μ scheme, the a-ε scheme *is not a two-way marching scheme if* $\varepsilon \neq 0$.

Equation (3.6) represents a pair of equations. The first is Eq. (3.5). With the aid of Eqs. (2.5) and (2.13), the second equation can be expressed as

$$(u_t)_j^n = (u_{j+1/2}'^n - u_j'^n{}_{-1/2})/\Delta x + (2\varepsilon - 1)(du_t)_j^n. \tag{3.10}$$

Here

$$u_{j\pm 1/2}'^n \overset{\text{def}}{=} u_{j\pm 1/2}^{n-1/2} + (\Delta t/2)(u_t)_{j\pm 1/2}^{n-1/2}; \tag{3.11}$$

i.e., $u_{j\pm 1/2}'^n$ is a first-order Taylor's approximation of u at $(j \pm \frac{1}{2}, n)$. Thus, the expression on the right side of Eq. (3.10) is the sum of a central-difference approximation of $\partial u/\partial x$ at (j, n) and the extra term $(2\varepsilon - 1)(du_t)_j^n$. Because $(du_t)_j^n = O[(\Delta x)^2]$, *the presence of this extra term will not lower the order of accuracy of the entire sum as an approximation of* $\partial u/\partial x$ *at* (j, n). *Also note that this extra term vanishes when* $\varepsilon = \frac{1}{2}$ *while the term associated with* $(du_t)_j^n$ *in Eq.* (3.1) *vanishes when* $\varepsilon = 0$.

Next we shall study the influence of ε on the stability and numerical dissipation of the a-ε scheme. Let $G_\pm^{(2)}$ and $G^{(2)}$ be the principal and spurious amplification factors of the a-ε scheme, respectively. Then, it will be shown in Section 6 that

$$G_\pm^{(2)} = |\lambda_\pm(\varepsilon, \nu, \theta)|^2, \tag{3.12}$$

with

$$\lambda_\pm(\varepsilon, \nu, \theta) \overset{\text{def}}{=} \varepsilon \cos(\theta/2) - i\nu \sin(\theta/2) \pm \sqrt{(1-\varepsilon)[(1-\varepsilon)\cos^2(\theta/2) + (1-\nu^2)\sin^2(\theta/2)]}. \tag{3.13}$$

Also it will be proved that

$$0 \leq \varepsilon \leq 1 \quad \text{and} \quad \nu^2 < 1 \tag{3.14}$$

are necessary and sufficient conditions for the stability of the a-ε scheme. Thus, Eq. (3.14) will be assumed in the remainder of this section.

It was pointed out in Section 2 that the amplification factors of the Leapfrog scheme are identical to those of the inviscid a-μ scheme. Because the latter scheme is a special case of the a-ε scheme with $\varepsilon = 0$, $G_\pm^{(2)}$ become the amplification factors of the Leapfrog scheme when $\varepsilon = 0$. This fact can be reverified by comparing Eqs. (2.21), (3.12), and (3.13) with $\xi = 0$ and $\varepsilon = 0$.

Also, we have

$$\lambda_\pm(1, \nu, \theta) = \cos(\theta/2) - i\nu \sin(\theta/2). \tag{3.15}$$

Thus, $G_+^{(2)} = G_-^{(2)}$ when $\varepsilon = 1$. Moreover, it is shown in Appendix A that the *coalesced amplification factor is identical to that of the Lax scheme*. Note that, like the Leapfrog scheme, a solution of the Lax scheme is also composed of two decoupled solutions with each being associated with a mesh that is staggered in time. However, because the Lax scheme is a two-level scheme, it does not have a spurious amplification factor.

Thus, at one extreme, i.e., when $\varepsilon = 0$, $G_\pm^{(2)}$ become the amplification factors of the Leapfrog scheme, which *is free of numerical dissipation*. At another extreme, i.e., when $\varepsilon = 1$, $G_+^{(2)}$ and $G^{(2)}$ coalesce into one and it becomes the amplification factor of the Lax scheme, which is notorious for its large diffusive errors. From the above observations, one may infer the conclusion that will be established shortly, i.e., *the* a-ε *scheme becomes more diffusive as the value of* ε *increases. Note that, because the Lax scheme is very diffusive and uses a mesh that is staggered in time, a two-level scheme using such a mesh is usually associated with a highly diffusive scheme* [27]. The a-ε scheme demonstrates that it can also be a scheme with no diffusive error!

As a result of Eq. (3.14), the expression under the radical sign in Eq. (3.13) is nonnegative. Thus, it can be shown that

$$1 - |G_\pm^{(2)}| = \chi_\pm(\varepsilon, \nu, \theta) \overset{\text{def}}{=} \varepsilon\{(1-\nu^2)\sin^2(\theta/2) + 2\cos(\theta/2) \times [(1-\varepsilon)\cos(\theta/2) \mp \sqrt{(1-\varepsilon)[(1-\varepsilon)\cos^2(\theta/2) + (1-\nu^2)\sin^2(\theta/2)]}]\}. \tag{3.16}$$

Because solutions to the physical equation Eq. (2.22) do not dissipate with time, a numerical analogue to Eq. (2.22) is said to be free of numerical dissipation if its solutions also do not dissipate with time, i.e., its amplification factors are of unit magnitude. As a result, numerical dissipation of the a-ε scheme may be measured by $1 - |G_\pm^{(2)}|$, i.e., $\chi_\pm(\varepsilon, \nu, \theta)$. Obviously the a-ε scheme is free of numerical dissipation if $\varepsilon = 0$. Also, by using Eqs. (3.14) and (3.16), it is shown in Section 6 that, for all θ with $-\pi < \theta \leq \pi$, and all ε and ν satisfying Eq. (3.14), we have

$$0 \leq \chi_-(\varepsilon, \nu, \theta) + 4\varepsilon(1-\varepsilon)\cos^2(\theta/2) \leq \chi(\varepsilon, \nu, \theta) \leq \min\{1, 4\varepsilon\} \tag{3.17}$$

and

$$0 \leq \chi_+(\varepsilon, \nu, \theta) \leq \varepsilon(1-\nu^2)\sin^2(\theta/2). \tag{3.18}$$

The significance of Eqs. (3.17) and (3.18) is discussed in the following remarks:

(a) Note that (i) the behaviors of the principal and the spurious solutions of the a-ε scheme are determined by its principal and spurious amplification factors, respectively; and (ii) because $0 \le \varepsilon(1 - \varepsilon)$ if $0 \le \varepsilon \le 1$, Eq. (3.17) implies that $\chi_-(\varepsilon, \nu, \theta) \le \chi_+(\varepsilon, \nu, \theta)$. Thus, the spurious solution will not dissipate more slowly than the principal solution. Let ε be not too close to 0 or 1. Then Eq. (3.17) also implies that the Fourier components of the spurious solution with smaller $|\theta|$ i.e., longer wavelength, will dissipate much faster than those of the principal solution. In other words, the spurious solution will rapidly disappear from the long-wavelength components of a numerical solution. Note that $\chi_+(\frac{1}{2}, \nu, 0) = 1$. Thus, *the long-wavelength components of the spurious solution are annihilated almost completely in a single time step if* $\varepsilon = \frac{1}{2}$, *i.e., if the last term in Eq.* (3.10) *is dropped.*

(b) The upper bound of $\chi_+(\varepsilon, \nu, \theta)$ given in Eq. (3.18) is proportional to $\sin^2(\theta/2)$. As a result, the long-wavelength Fourier components in the principal solution are nearly free of numerical dissipation. On the other hand, short-wavelength components may decay rapidly.

(c) For a fixed ε, Eq. (3.18) implies that the principal solution is more diffusive for a smaller $|\nu|$. How to compensate this effect is a subject to be discussed in Section 7.

(d) Equations (3.17) and (3.18) imply that, for all ν with $\nu^2 < 1$ and all θ with $-\pi < \theta \le \pi$, we have

$$0 \le \chi_-(\varepsilon, \nu, \theta) \le \varepsilon, \quad 0 \le \chi_+(\varepsilon, \nu, \theta) \le \min\{1, 4\varepsilon\}, \tag{3.19}$$

which, according to Eq. (3.16), is equivalent to

$$1 - \varepsilon \le |G_-^{(2)}| \le 1, \quad 1 - \min\{1, 4\varepsilon\} \le |G^{(2)}| \le 1. \tag{3.20}$$

As a result, by choosing ε small enough, both $|G_-^{(2)}|$ and $|G^{(2)}|$ can be confined within an arbitrarily narrow range. As noted previously, the spurious part of a numerical solution generally is insignificantly small assuming a smooth initial condition. It does not contribute to accuracy and usually dissipates faster than the principal part. Thus, our primary concerns is how the principal part dissipates. From Eq.(3.20), one concludes that, for any ε with $0 < \varepsilon < 1$, $|G_+^{(2)}|$ will be bounded *uniformly* from below by a *positive* number $1 - \varepsilon$ for all ν with $\nu^2 < 1$ and all θ with $-\pi < \theta \le \pi$. By choosing an ε of proper magnitude, *one can suppress artificial numerical oscillations without causing large diffusive errors for any combination of* ν *and* θ. This fact contrasts sharply with what one expects from typical classical schemes which are usually very diffusive with respect to certain ν and θ, while not at all with respect to other ν and θ. As an example, we consider the Lax–Wendroff scheme [12, p.101]. Its amplification factor is of unit magnitude, for all θ at $\nu = 0$, or $\nu = 1$. On the other hand, the amplification factor = 0 if $\nu^2 = \frac{1}{2}$ and $\theta = \pi$.

In nonlinear flow solutions, e.g., shock-tube solutions to be discussed in Section 7, analogues of ν are dependent on local velocity components. Thus, they may vary from one location to another. Also, at some neighborhood, the Fourier spectrum of the local solution may have peaks spread over a wide range of θ. Thus, for a numerical analogue of Eq. (2.22), a large variation in numerical diffusivity with respect to θ and ν generally means that numerical solutions obtained using its nonlinear extensions will suffer annihilations of sharply different degrees at different locations and different θ. Such selective annihilations may cause large distortions of numerical solutions [25].

This completes the discussion of stability and numerical dissipation. Other key subjects, i.e., consistency and the truncation error, are discussed in Section 7 of [5].

In conclusion, the a-ε scheme has been constructed to solve Eq. (2.22). It has the unique property that numerical dissipation can be controlled by a parameter ε. *Because neither characteristics-based techniques nor knowledge about the upwind direction is used in the construction of the* a-ε *scheme, as will be shown in the next section, it can be easily extended to model the Euler equations.*

4. THE EULER SOLVER

We consider a dimensionless form of the 1D unsteady Euler equations of a perfect gas. Let ρ, v, p, and γ be the mass density, velocity, static pressure, and constant specific heat ratio, respectively. Let

$$u_1 = \rho, \quad u_2 = \rho v, \quad u_3 = p/(\gamma - 1) + (\tfrac{1}{2})\rho v^2, \tag{4.1}$$

$$f_1 = u_2, \tag{4.2}$$

$$f_2 = (\gamma - 1)u_3 + (\tfrac{1}{2})(3 - \gamma)(u_2)^2/u_1, \tag{4.3}$$

and

$$f_3 = \gamma u_2 u_3/u_1 - (\tfrac{1}{2})(\gamma - 1)(u_2)^3/(u_1)^2. \tag{4.4}$$

Then the Euler equations can be expressed as

$$\frac{\partial u_m}{\partial t} + \frac{\partial f_m}{\partial x} = 0, \quad m = 1, 2, 3. \tag{4.5}$$

The integral form of Eq. (4.5) in space-time E_2 is

$$\oint_{S(V)} \mathbf{h}_m \cdot d\mathbf{s} = 0, \quad m = 1, 2, 3, \tag{4.6}$$

where $\mathbf{h}_m = (f_m, u_m)$, $m = 1, 2, 3$, are the space-time mass, momentum, and energy current density vectors, respectively.

As a preliminary, let

$$f_{m,k} \stackrel{\text{def}}{=} \partial f_m/\partial u_k, \quad m, k = 1, 2, 3, \tag{4.7}$$

and let F be the matrix formed by $f_{m,k}$, $m, k = 1, 2, 3$. Let c be the sonic speed. Moreover, for any numbers $a_1, a_2, \ldots, a_n$, let $\text{diag}(a_1, a_2, \ldots, a_n)$ denote the diagonal matrix with $a_1, a_2, \ldots, a_n$ being the diagonal elements on the first, second, ..., and nth rows, respectively. Then there exists a 3×3 matrix G such that

$$G^{-1}FG = \text{diag}(v, v - c, v + c), \tag{4.8}$$

where G^{-1} is the inverse of G. Note that v, c, F, G, and G^{-1} are functions of u_m, $m = 1, 2, 3$. These functions are given explicitly in [5].

Consider SEs of type I depicted in Fig. 2. For any $(x, t) \in \text{SE}(j, n)$, $u_m(x, t)$, $f_m(x, t)$, and $\mathbf{h}_m(x, t)$ are approximated by $u_m^*(x, t; j, n)$, $f_m^*(x, t; j, n)$, and $\mathbf{h}_m^*(x, t; j, n)$, respectively. They will be defined shortly. Let

$$\begin{aligned} u_m^*(x, t; j, n) \overset{\text{def}}{=} & (u_m)_j^n + (u_{mx})_j^n(x - x_j) \\ & + (u_{mt})_j^n(t - t^n), \quad m = 1, 2, 3, \end{aligned} \tag{4.9}$$

where $(u_m)_j^n$, $(u_{mx})_j^n$, and $(u_{mt})_j^n$ are constants in $\text{SE}(j, n)$. Obviously, they can be considered as the numerical analogues of the values of u_m, $\partial u_m/\partial x$, and $\partial u_m/\partial t$ at (x_j, t^n), respectively.

Let $(f_m)_j^n$ and $(f_{m,k})_j^n$ denote the values of f_m and $f_{m,k}$, respectively, when u_m, $m = 1, 2, 3$, respectively, assume the values of $(u_m)_j^n$, $m = 1, 2, 3$. Let

$$(f_{mx})_j^n \overset{\text{def}}{=} \sum_{k=1}^{3} (f_{m,k})_j^n (u_{kx})_j^n, \quad m = 1, 2, 3, \tag{4.10}$$

and

$$(f_{mt})_j^n \overset{\text{def}}{=} \sum_{k=1}^{3} (f_{m,k})_j^n (u_{kt})_j^n, \quad m = 1, 2, 3. \tag{4.11}$$

Because

$$\frac{\partial f_m}{\partial x} = \sum_{k=1}^{3} f_{m,k} \frac{\partial u_k}{\partial x} \tag{4.12}$$

and

$$\frac{\partial f_m}{\partial t} = \sum_{k=1}^{3} f_{m,k} \frac{\partial u_k}{\partial t}, \tag{4.13}$$

$(f_{mx})_j^n$ and $(f_{mt})_j^n$ can be considered as the numerical analogues of the values of $\partial f_m/\partial x$ and $\partial f_m/\partial t$ at (x_j, t^n), respectively. As a result, we assume that

$$\begin{aligned} f_m^*(x, t; j, n) = & (f_m)_j^n + (f_{mx})_j^n(x - x_j) \\ & + (f_{mt})_j^n(t - t^n), \quad m = 1, 2, 3. \end{aligned} \tag{4.14}$$

Because $\mathbf{h}_m = (f_m, u_m)$, we also assume that

$$\begin{aligned} \mathbf{h}_m^*(x, t; j, n) &= (f_m^*(x, t; j, n), u_m^*(x, t; j, n)), \\ & m = 1, 2, 3. \end{aligned} \tag{4.15}$$

Note that, by their definitions: (i) $(f_m)_j^n$ and $(f_{m,k})_j^n$, $m = 1, 2, 3$, are functions of $(u_m)_j^n$, $m = 1, 2, 3$; (ii) $(f_{mx})_j^n$, $m = 1, 2, 3$, are functions of $(u_m)_j^n$ and $(u_{mx})_j^n$, $m = 1, 2, 3$; and (iii) $(f_{mt})_j^n$ are functions of $(u_m)_j^n$ and $(u_{mt})_j^n$, $m = 1, 2, 3$.

Moreover, we assume that, for any $(x, t) \in \text{SE}(j, n)$, $u_m = u_m^*(x, t; j, n)$ and $f_m = f_m^*(x, t; j, n)$ satisfy Eq. (4.5); i.e.,

$$\frac{\partial u_m^*(x, t; j, n)}{\partial t} + \frac{\partial f_m^*(x, t; j, n)}{\partial x} = 0. \tag{4.16}$$

According to Eqs. (4.9) and (4.14), Eq. (4.16) is equivalent to

$$(u_{mt})_j^n = -(f_{mx})_j^n. \tag{4.17}$$

Because $(f_{mx})_j^n$ are functions of $(u_m)_j^n$ and $(u_{mx})_j^n$, Eq. (4.17) implies that $(u_{mt})_j^n$ are also functions of $(u_m)_j^n$ *and* $(u_{mx})_j^n$. From this result and the facts stated following Eq. (4.15), one concludes that *the only independent discrete variables needed to be solved in the current marching scheme are* $(u_m)_j^n$ *and* $(u_{mx})_j^n$.

From Eq. (4.16), one concludes that the generalization of the potential function $\psi(x, t; j, n)$ introduced in Section 2 to the current solver are $\psi_m(x, t; j, n)$, $m = 1, 2, 3$, which satisfy

$$\frac{\partial \psi_m(x, t; j, n)}{\partial t} = f_m^*(x, t; j, n) \tag{4.18}$$

and

$$-\frac{\partial \psi_m(x, t; j, n)}{\partial x} = u_m^*(x, t; j, n). \tag{4.19}$$

Substituting Eqs. (4.9) and (4.14) into Eqs. (4.18) and (4.19), and using Eq. (4.17), one concludes that, up to an arbitrary constant,

$$\begin{aligned} \psi_m(x, t; j, n) = & (f_m)_j^n(t - t^n) - (u_m)_j^n(x - x_j) \\ & + (\tfrac{1}{2})(f_{mt})_j^n(t - t^n)^2 - (\tfrac{1}{2})(u_{mx})_j^n(x - x_j)^2 \\ & + (f_{mx})_j^n(x - x_j)(t - t^n). \end{aligned} \tag{4.20}$$

By using an argument similar to that leading to Eq. (2.36), one concludes that

$$\int_\Gamma \mathbf{h}_m^* \cdot d\mathbf{s} = \psi_m(x', t'; j, n) - \psi_m(x, t; j, n). \tag{4.21}$$

Here Γ is a simple curve joining (x, t) and (x', t'), and *lying*

entirely within SE(j, n). We also assume that $d\mathbf{s}$ points to the right of Γ if one moves forward from (x, t) to (x', t').

As in the a-ε scheme, we assume that the flux of $\mathbf{h}_m^*$ is conserved over CE(j, n), i.e.,

$$\oint_{S(\mathrm{CE}(j,n))} \mathbf{h}_m^* \cdot d\mathbf{s} = 0. \tag{4.22}$$

Combining Eqs. (4.21) and (4.22), one has

$$\begin{aligned} &\psi_m(x_j - \Delta x/2, t^n; j, n) - \psi_m(x_j + \Delta x/2, t^n; j, n) \\ &\quad + \psi_m(x_{j-1/2} + \Delta x/2, t^{n-1/2}; j - \tfrac{1}{2}, n - \tfrac{1}{2}) \\ &\quad - \psi_m(x_{j-1/2}, t^{n-1/2} + \Delta t/2; j - \tfrac{1}{2}, n - \tfrac{1}{2}) \\ &\quad + \psi_m(x_{j+1/2}, t^{n-1/2} + \Delta t/2; j + \tfrac{1}{2}, n - \tfrac{1}{2}) \\ &\quad - \psi_m(x_{j+1/2} - \Delta x/2, t^{n-1/2}; j + \tfrac{1}{2}, n - \tfrac{1}{2}) = 0. \end{aligned} \tag{4.23}$$

Substitution of Eq. (4.20) into Eq. (4.23) yields

$$(u_m)_j^n = \tfrac{1}{2}[(u_m)_{j-1/2}^{n-1/2} + (u_m)_{j+1/2}^{n-1/2} + (s_m)_{j-1/2}^{n-1/2} - (s_m)_{j+1/2}^{n-1/2}], \tag{4.24}$$

where, for all $(j, n) \in \Omega$,

$$\begin{aligned} (s_m)_j^n \overset{\text{def}}{=} \frac{\Delta x}{4}(u_{mx})_j^n + \frac{\Delta t}{\Delta x}(f_m)_j^n \\ + \frac{(\Delta t)^2}{4\Delta x}(f_{mt})_j^n, \quad m = 1, 2, 3. \end{aligned} \tag{4.25}$$

Equation (4.24) forms the first half of the current marching scheme. The second half which solves $(u_{mx})_j^n$ will come from a generalization of Eq. (3.10).

For all $(j, n) \in \Omega$, let

$$\begin{aligned} (du_{mx})_j^n \overset{\text{def}}{=} \tfrac{1}{2}[(u_{mx})_{j+1/2}^{n-1/2} + (u_{mx})_{j-1/2}^{n-1/2}] \\ - [(u_m)_{j+1/2}^{n-1/2} - (u_m)_{j-1/2}^{n-1/2}]/\Delta x \end{aligned} \tag{4.26}$$

and

$$(u'_m)_{j+1/2}^n \overset{\text{def}}{=} (u_m)_{j+1/2}^{n-1/2} + (\Delta t/2)(u_{mt})_{j+1/2}^{n-1/2}, \tag{4.27}$$

for $m = 1, 2, 3$. Because Eqs. (4.26) and (4.27) are the generalizations of Eqs. (3.2) and (3.11), respectively, a natural generalization of Eq. (3.10) is

$$\begin{aligned} (u_{mx})_j^n = [(u'_m)_{j+1/2}^n - (u'_m)_{j-1/2}^n]/\Delta x \\ + (2\varepsilon - 1)(du_{mx})_j^n, \quad m = 1, 2, 3, \end{aligned} \tag{4.28}$$

where ε is a parameter independent of numerical variables. Note that the last term in Eq. (4.28) vanishes if $\varepsilon = \frac{1}{2}$. The marching scheme presented in [3] is formed by Eqs. (4.24) and (4.28) with $\varepsilon = \frac{1}{2}$.

To construct a larger class of generalizations to Eq. (3.10), for all $(j, n) \in \Omega$, let

$$(\hat{u}_m)_j^n \overset{\text{def}}{=} \tfrac{1}{2}[(u_m)_{j+1/2}^{n-1/2} + (u_m)_{j-1/2}^{n-1/2}], \quad m = 1, 2, 3. \tag{4.29}$$

Let $(\hat{\varepsilon}_m)_j^n$, $m = 1, 2, 3$, be parameters that can be functions of $(\hat{u}_m)_j^n$, $m = 1, 2, 3$. There can be many choices of these functions. Let $(\hat{g}_{mk})_j^n$ be the value of the (m, k)-element of the matrix G when u_m, $m = 1, 2, 3$, respectively, assume the values of $(\hat{u}_m)_j^n$, $m = 1, 2, 3$. Similarly, let $(\hat{g}_{mk}^{-1})_j^n$ be the value of the (m, k)-element of the matrix G^{-1} when u_m, $m = 1, 2, 3$, respectively, assume the values of $(\hat{u}_m)_j^n$, $m = 1, 2, 3$. Let

$$(\hat{e}_{mk})_j^n \overset{\text{def}}{=} \sum_{l=1}^{3} (\hat{g}_{ml})_j^n (\hat{\varepsilon}_l)_j^n (\hat{g}_{lk}^{-1})_j^n, \quad m, k = 1, 2, 3. \tag{4.30}$$

Then Eq. (3.10) can be generalized as

$$\begin{aligned} (u_{mx})_j^n = [(u'_m)_{j-1/2}^n - (u'_m)_{j-1/2}^n]/\Delta x \\ + \sum_{k=1}^{3} [2(\hat{e}_{mk})_j^n - \delta_{mk}](du_{kx})_j^n, \end{aligned} \tag{4.31}$$

where $m = 1, 2, 3$, and δ_{mk} is the kronecker-delta symbol.

Consider the special case in which, for all $(j, n) \in \Omega$, $(\hat{\varepsilon}_1)_j^n = (\hat{\varepsilon}_2)_j^n = (\hat{\varepsilon}_3)_j^n$. Let $(\hat{\varepsilon}_m)_j^n = (\hat{\varepsilon})_j^n$, $m = 1, 2, 3$. Then $(\hat{\varepsilon}_{mk})_j^n = (\hat{\varepsilon})_j^n \delta_{mk}$, and thus Eq. (4.31) is reduced to

$$\begin{aligned} (u_{mx})_j^n = [(u'_m)_{j+1/2}^n - (u'_m)_{j-1/2}^n]/\Delta x \\ + [2(\hat{\varepsilon})_j^n - 1](du_{mx})_j^n, \quad m = 1, 2, 3, \end{aligned} \tag{4.32}$$

Note that Eq. (4.32) reduces to Eq. (4.28) if $(\hat{\varepsilon})_j^n = \varepsilon$ for all $(j, n) \in \Omega$.

Recall that both v and c are functions of u_m, $m = 1, 2, 3$. For all SE(j, n), let $\hat{v}_j^n$ and $\hat{c}_j^n$, respectively, denote the values of v and c when u_m, $m = 1, 2, 3$, respectively, assume the values of $(\hat{u}_m)_j^n$, $m = 1, 2, 3$. It is shown in [5] that the marching scheme formed by Eqs. (4.24) and (4.31) is stable if, for all $(j, n) \in \Omega$,

$$(\hat{\nu}_{\max})_j^n < 1, \quad 0 \le (\hat{\varepsilon}_m)_j^n \le 1, \quad m = 1, 2, 3, \tag{4.33}$$

where

$$(\hat{\nu}_{\max})_j^n \overset{\text{def}}{=} (|\hat{v}_j^n| + |\hat{c}_j^n|)\frac{\Delta t}{\Delta x}. \tag{4.34}$$

We conclude this section by introducing some possible modifications to the above solver. Note that $(u'_m)_{j-1/2}^n$, by its definition,

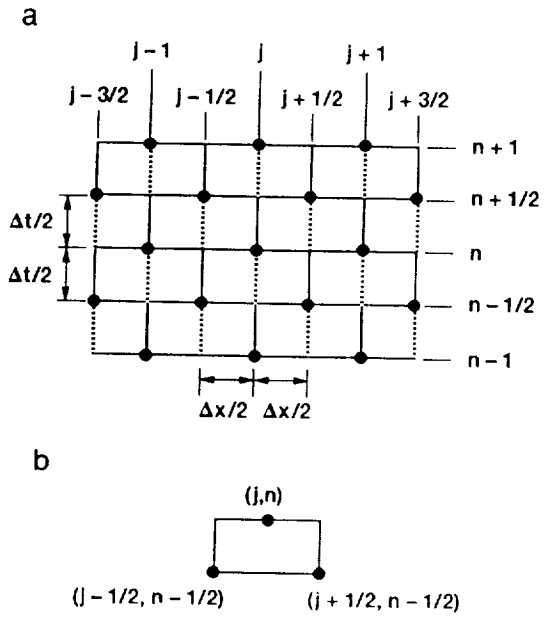

FIG. 5. The mesh and CEs of the a-ε scheme. (a) The relative positions of CEs and mesh points. (b) CE(j, n).

represents a finite-difference approximation of u_m at $(j \pm \frac{1}{2}, n)$. As a result,

$$(u'^{c}_{mx})^n_j \overset{\text{def}}{=} [(u'_m)^n_{j+1/2} - (u'_m)^n_{j-1/2}]/\Delta x, \quad m = 1, 2, 3, \tag{4.35}$$

respectively, are the central-difference approximations for $\partial u_m/\partial x$, $m = 1, 2, 3$, at (j, n). Note that $(u'^{c}_{mx})^n_j$ is the first term on the right side of each of Eqs. (4.28), (4.31), and (4.32). The above central-difference approximation is valid as long as no discontinuity of u_m (or its derivatives) occurs between $(j - \frac{1}{2}, n)$ and $(j + \frac{1}{2}, n)$ (see Fig. 5). In the following discussion, we develop alternates which are valid even in the presence of discontinuity.

Let

$$(u_{mx\pm})^n_j \overset{\text{def}}{=} \pm \frac{(u'_m)^n_{j\mp 1/2} - (u_m)^n_j}{\Delta x/2}, \quad m = 1, 2, 3, \tag{4.36}$$

where $(u_m)^n_j$ can be obtained from Eq. (4.24). Because $(u'_m)^n_{j-1/2}$, $(u_m)^n_j$, and $(u'_m)^n_{j+1/2}$, are the numerical analogues of u_m at $(j - \frac{1}{2}, n)$, (j, n) and $(j + \frac{1}{2}, n)$, respectively, $(u_{mx-})^n_j$ and $(u_{mx+})^n_j$ are two numerical analogues of the value of $\partial u_m/\partial x$ at (j, n) with one being evaluated from the left and another from the right. Note that

$$(u'^{c}_{mx})^n_j = \tfrac{1}{2}[(u_{mx-})^n_j + (u_{mx+})^n_j]. \tag{4.37}$$

In case a discontinuity occurs between (j, n) and $(j + \frac{1}{2}, n)$ but not between (j, n) and $(j - \frac{1}{2}, n)$, one would expect that $|(u_{mx+})^n_j| \gg |(u_{mx-})^n_j|$. Moreover, because (j, n) and $(j - \frac{1}{2}, n)$ are on the same side of the discontinuity while (j, n) and $(j + \frac{1}{2}, n)$ are on the opposite sides, $(u_{mx})^n_j$ should be a weighted average of $(u_{mx+})^n_j$ and $(u_{mx-})^n_j$ biased toward the one with the smaller magnitude.

As a result of the above considerations, $(u'^{c}_{mx})^n_j$ can be replaced by

$$(u'^{w_\alpha}_{mx})^n_j \overset{\text{def}}{=} W_\alpha((u_{mx-})^n_j, (u_{mx+})^n_j; \alpha), \quad m = 1, 2, 3, \tag{4.38}$$

Here α is an adjustable constant and the function W_α is defined by (i) $W_\alpha(0, 0, \alpha) = 0$ and (ii)

$$W_\alpha(x_+, x_-; \alpha) = \frac{|x_+|^\alpha x_+ + |x_-|^\alpha x_-}{|x_+|^\alpha + |x_-|^\alpha}, \quad (|x_+| + |x_-| > 0), \tag{4.39}$$

where x_+ and x_- are any two real variables. Note that $W_\alpha(x_+, x_+; \alpha) = (x_+ + x_+)/2$; i.e., $(u'^{w_\alpha}_{mx})^n_j = (u'^{c}_{mx})^n_j$, if $\alpha = 0$ or $|x_-| = |x_+|$. Also the expression on the right side of Eq. (4.39) represents a weighted average of x_+ and x_- with the weight factors $|x_+|^\alpha/(|x_+|^\alpha + |x_-|^\alpha)$ and $|x_-|^\alpha/(|x_+|^\alpha + |x_-|^\alpha)$. *For $\alpha > 0$, this average is biased toward the one among x_+ and x_- with the smaller magnitude. For the same value of $|x_+|$ and $|x_-|$, the bias increases as α increases.* Thus, we should always choose $\alpha \geq 0$.

Note that the special weighted averages $W_\alpha(x_+, x_+; 1)$ and $W_\alpha(x_-, x_+; 2)$ are used in the slope-limiters proposed by van Leer [28] and van Albada [29], respectively.

The above modification, i.e., $(u'^{c}_{mx})^n_j$ replaced by $(u'^{w_\alpha}_{mx})^n_j$, is first given in [3]. It is shown in [3] and also Section 7 of the current paper that it is an efficient tool to suppress overshoots and/or numerical oscillations near a discontinuity. Moreover, *because $(u_{mx\pm})^n_j$ are constructed using only the data associated with the mesh points $(j - \frac{1}{2}, n - \frac{1}{2})$ and $(j + \frac{1}{2}, n - \frac{1}{2})$, the effect of this modification is highly local; i.e., it generally will not cause the smearing of shock discontinuities.*

However, there may be a price to pay for the above modification. Because a fractional power is costly to evaluate, so is $W_\alpha(x_+, x_+; \alpha)$ if α is not an integer. Moreover, because the bias of this weighted average increases with α, a situation may arise such that the use of an α with $|\alpha| < 1$ may be desirable. To obtain a computationally efficient weighted average of arbitrary small bias, let

$$W(x_-, x_+; \alpha, \beta) \overset{\text{def}}{=} (1 - \beta)W_\alpha(x_+, x_-; 0) + \beta W_\alpha(x_-, x_+; \alpha), \tag{4.40}$$

where $\beta \geq 0$ is an adjustable weight factor, and α generally is an integer. Because $W_\alpha(x_-, x_-; 0)$ is the simple average of

x_- and x_+, Eq. (4.40) defines *a linear weighted average of this simple average and the nonlinear weighted average defined in Eq.* (4.39). Obviously, $W(x_-, x_+; \alpha, \beta) = (\frac{1}{2})(x_- + x_+)$ if $x_- = x_+$. Furthermore, because

$$W_o(x_-, x_+; -\alpha) = \frac{|x_-|^\alpha x_+ + |x_+|^\alpha x_-}{|x_+|^\alpha + |x_-|^\alpha}, \quad (|x_+| + |x_-| > 0), \tag{4.41}$$

alternatively, $W(x_-, x_+; \alpha, \beta)$ can also be expressed as

$$W(x_-, x_+; \alpha, \beta) = \left(\frac{1+\beta}{2}\right) W_o(x_-, x_-; \alpha) + \left(\frac{1-\beta}{2}\right) W_o(x_+, x_+; -\alpha). \tag{4.42}$$

The application of the more general modification, i.e., $(u^c_{m\lambda})^n_j$ is replaced by

$$(u^w_{m\lambda})^n_j \stackrel{\text{def}}{=} W((u_{m\lambda-})^n_j, (u_{m\lambda+})^n_j; \alpha, \beta), \quad m = 1, 2, 3, \tag{4.43}$$

will be demonstrated in Section 7.

Finally, note that $W(x_-, x_+; \alpha, \beta)$ can be further generalized by a linear weighted average of several $W_o(x_-, x_+; \alpha)$ with different values of α.

5. THE NAVIER–STOKES SOLVER

We consider a dimensionless form of the 1D unsteady Navier–Stokes equations of a perfect gas [12, pp.191–193]. (Note: the expressions on the right sides of the last three equations in Eq. (5-47) of [12] have incorrect signs in the earlier versions. The conduction heat-flux vector should be proportional to the negative of the gradient of temperature.) These equations are extensions of the Euler equations defined in Section 4. Thus, unless specified otherwise, the symbols, definitions, and equations given there will be used in this section.

Let Re_L and Pr denote the Reynolds number and Prandtl number, respectively. They are assumed to be nonnegative constants. Let

$$\tilde{f}_1 \stackrel{\text{def}}{=} 0, \tag{5.1}$$

$$\tilde{f}_2 \stackrel{\text{def}}{=} \frac{4}{3\,\mathrm{Re}_L}\frac{u_2}{u_1}, \tag{5.2}$$

and

$$\tilde{f}_3 \stackrel{\text{def}}{=} \frac{2}{3\,\mathrm{Re}_L}\left(\frac{u_2}{u_1}\right)^2 + \frac{\gamma}{\mathrm{Re}_L\,\mathrm{Pr}}\left[\frac{u_3}{u_1} - \frac{(u_2)^2}{2(u_1)^2}\right]. \tag{5.3}$$

Then, the Navier–Stokes euations can be expressed as

$$\frac{\partial u_m}{\partial t} + \frac{\partial f_m}{\partial x} - \frac{\partial^2 \tilde{f}_m}{\partial x^2} = 0, \quad m = 1, 2, 3, \tag{5.4}$$

The integral form of Eq. (5.4) in space-time E_2 is Eq. (4.6) with

$$\mathbf{h}_m \stackrel{\text{def}}{=} (f_m - \partial \tilde{f}_m/\partial x, u_m), \quad m = 1, 2, 3, \tag{5.5}$$

As a preliminary, let

$$\tilde{f}_{m,k} \stackrel{\text{def}}{=} \partial \tilde{f}_m/\partial u_k, \quad m, k = 1, 2, 3, \tag{5.6}$$

and

$$\tau_1 \stackrel{\text{def}}{=} \frac{4}{3\,\mathrm{Re}_L}, \quad \tau_2 \stackrel{\text{def}}{=} \frac{\gamma}{\mathrm{Re}_L\,\mathrm{Pr}}, \quad \tau_3 \stackrel{\text{def}}{=} \tau_2 - \tau_1. \tag{5.7}$$

Let $\tilde{F}$ denote the 3×3 matrix formed by $\tilde{f}_{m,k}$, $m, k = 1, 2, 3$. Then Eqs. (5.1)–(5.3) imply that

$$\tilde{F} = \begin{pmatrix} 0 & 0 & 0 \\ -\dfrac{\tau_1 u_2}{(u_1)^2} & \dfrac{\tau_1}{u_1} & 0 \\ \tau_3\dfrac{(u_2)^2}{(u_1)^3} - \tau_2\dfrac{u_3}{(u_1)^2} & -\dfrac{\tau_3 u_2}{(u_1)^2} & \dfrac{\tau_2}{u_1} \end{pmatrix}. \tag{5.8}$$

Again we consider SEs of type I depicted in Fig. 2. For any $(x, t) \in \mathrm{SE}(j, n)$, $u_m(x, t)$, $f_m(x, t)$, $\tilde{f}_m(x, t)$, and $\mathbf{h}_m(x, t)$, respectively, are approximated by $u^*_m(x, t; j, n)$, $f^*_m(x, t; j, n)$, $\tilde{f}^*_m(x, t; j, n)$, and $\mathbf{h}^*_m(x, t; j, n)$; $u^*_m(x, t; j, n)$ and $f^*_m(x, t; j, n)$, respectively, are defined in Eqs. (4.9) and (4.14); $\tilde{f}^*_m(x, t; j, n)$ and $\mathbf{h}^*_m(x, t; j, n)$ will be defined immediately.

Both $\tilde{f}_m$ and $\tilde{f}_{m,k}$ are functions of u_m, $m = 1, 2, 3$. Let $(\tilde{f}_m)^n_j$ and $(\tilde{f}_{m,k})^n_j$, respectively, denote the values of $\tilde{f}_m$ and $\tilde{f}_{m,k}$ when u_m, $m = 1, 2, 3$, respectively, assume the values of $(u_m)^n_j$, $m = 1, 2, 3$. Let

$$(\tilde{f}_{m\lambda})^n_j \stackrel{\text{def}}{=} \sum_{k=1}^{3} (\tilde{f}_{m,k})^n_j (u_{k\lambda})^n_j, \quad m = 1, 2, 3, \tag{5.9}$$

and

$$(\tilde{f}_{mt})^n_j \stackrel{\text{def}}{=} \sum_{k=1}^{3} (\tilde{f}_{m,k})^n_j (u_{kt})^n_j, \quad m = 1, 2, 3. \tag{5.10}$$

Using an argument similar to that leading to Eq. (4.14), we assume that

$$\tilde{f}^*_m(x, t; j, n) = (\tilde{f}_m)^n_j + (\tilde{f}_{m\lambda})^n_j (x - x_j) + (\tilde{f}_{mt})^n_j (t - t^n), \quad m = 1, 2, 3. \tag{5.11}$$

As a result of Eq. (5.5), we also assume that

$$\mathbf{h}_m^*(x, t; j, n) = \left(f_m^*(x, t; j, n) - \frac{\partial \tilde{f}_m^*(x, t; j, n)}{\partial x}, u_m^*(x, t; j, n)\right), \quad m = 1, 2, 3, \tag{5.12}$$

Also, we assume that, for any $(x, t) \in \mathrm{SE}(j, n)$, $u_m = u_m^*(x, t; j, n)$, $f_m = f_m^*(x, t; j, n)$, and $\tilde{f}_m = \tilde{f}_m^*(x, t; j, n)$ satisfy Eq. (5.4), i.e.,

$$\frac{\partial u_m^*(x, t; j, n)}{\partial t} + \frac{\partial}{\partial x}\left[f_m^*(x, t; j, n) - \frac{\partial \tilde{f}_m^*(x, t; j, n)}{\partial x}\right] = 0. \tag{5.13}$$

The above condition again leads to Eq. (4.17). Thus, the diffusion term in Eq. (5.4) has no impact on how $u_m^*(x, t; j, n)$ varies with time *within* $\mathrm{SE}(j, n)$. This same fact was observed in Section 2. The reason behind it and its significance were also discussed there. As a result of Eq.(4.17), and other definitions given earlier in this section, one can conclude that the only independent discrete variables needed to be solved in the current solver, as in the Euler solver described in Section 4, are also $(u_m)_j^n$ and $(u_{mx})_j^n$.

A comparison between Eqs. (4.16) and (5.13) reveals that, for the current solver, Eqs. (4.18) and (4.19) should be replaced by

$$\frac{\partial \psi_m(x, t; j, n)}{\partial t} = f_m^*(x, t; j, n) - \frac{\partial \tilde{f}_m^*(x, t; j, n)}{\partial x} \tag{5.14}$$

and

$$-\frac{\partial \psi_m(x, t; j, n)}{\partial x} = u_m^*(x, t; j, n), \tag{5.15}$$

respectively. Note that Eqs. (5.15) and (4.19) are identical. According to Eq. (5.11), the second term on the right side of Eq. (5.14) is simply the constant $-(\tilde{f}_{mx})_j^n$. Thus, for the current solver, Eq. (4.20) should be replaced by

$$\begin{aligned}\psi_m(x, t; j, n) = {} & (\acute{f}_m)_j^n(t - t^n) - (u_m)_j^n(x - x_j) \\ & + (\tfrac{1}{2})(f_{mt})_j^n(t - t^n)^2 - (\tfrac{1}{2})(u_{mx})_j^n(x - x_j)^2 \\ & + (f_{mx})_j^n(x - x_j)(t - t^n),\end{aligned} \tag{5.16}$$

where

$$(\acute{f}_m)_j^n \overset{\text{def}}{=} (f_m)_j^n - (\tilde{f}_{mx})_j^n. \tag{5.17}$$

The only difference between Eqs. (4.20) and (5.16) is that $(f_m)_j^n$ in Eq. (4.20) is replaced by $(\acute{f}_m)_j^n$ in Eq. (5.16). Obviously, Eq. (4.21) is still valid for the current solver. Because $\psi_m(x, t; j, n)$ is independent of $(\tilde{f}_m)_j^n$ and $(\tilde{f}_{mt})_j^n$, Eq. (4.21) implies that the last two parameters are irrelevant in flux evaluation. Moreover, because the current solver will be constructed using only flux-balance conditions, these parameters are also irrelevant in the following construction.

For all $(j, n) \in \Omega$, we assume that

$$\oint_{\partial \mathrm{SE}(j,n)} \mathbf{h}_m^* \cdot d\mathbf{s} = 0. \tag{5.18}$$

With the aid of Eqs.(5.16) and (4.21), Eq. (5.18) implies that, for all $(j, n) \in \Omega$,

$$\begin{aligned}(u_m)_j^n - (u_m)_{j\pm 1/2}^{n-1/2} & \pm \frac{\Delta x}{4}[(u_{mx})_{j\cdot 1/2}^{n-1/2} + (u_{mx})_j^n] \\ & \pm \frac{\Delta t}{\Delta x}[(\acute{f}_m)_{j\cdot 1/2}^{n-1/2} - (\acute{f}_m)_j^n] \\ & \pm \frac{(\Delta t)^2}{4\Delta x}[(f_{mt})_{j\cdot 1/2}^{n\cdot 1/2} + (f_{mt})_j^n] = 0.\end{aligned} \tag{5.19}$$

Adding the two equations given in Eq. (5.19) results in

$$(u_m)_j^n = \tfrac{1}{2}[(u_m)_{j\cdot 1/2}^{n-1/2} + (u_m)_{j\cdot 1/2}^{n\cdot 1/2} + (\acute{s}_m)_{j\cdot 1/2}^{n\cdot 1/2} - (\acute{s}_m)_{j\cdot 1/2}^{n\cdot 1/2}], \tag{5.20}$$

where, for all $(j, n) \in \Omega$,

$$\begin{aligned}(\acute{s}_m)_j^n \overset{\text{def}}{=} {} & \frac{\Delta x}{4}(u_{mx})_j^n + \frac{\Delta t}{\Delta x}(\acute{f}_m)_j^n \\ & + \frac{(\Delta t)^2}{4\Delta x}(f_{mt})_j^n, \quad m = 1, 2, 3.\end{aligned} \tag{5.21}$$

Equations (5.20) and (5.21) are the current counterparts of Eqs. (4.24) and (4.25), respectively. By using Eq. (5.20), $(u_m)_j^n$ can be solved explicitly in terms of discrete variables at the next lower time level.

By substraction of the two equations given in Eq. (5.19) and using Eq. (5.17), one has

$$\begin{aligned}\frac{\Delta x}{4}(u_{mx})_j^n & + \frac{(\Delta t)^2}{4\Delta x}(f_{mt})_j^n \\ & + \frac{\Delta t}{\Delta x}(\tilde{f}_{mx})_j^n = (b_m)_j^n, \quad m = 1, 2, 3,\end{aligned} \tag{5.22}$$

where, for all $(j, n) \in \Omega$, and $m = 1, 2, 3$,

$$\begin{aligned}(b_m)_j^n \overset{\text{def}}{=} {} & \frac{\Delta t}{\Delta x}(f_m)_j^n + \tfrac{1}{2}[(u_m)_{j+1/2}^{n-1/2} \\ & - (u_m)_{j\ 1/2}^{n\ 1/2} - (\acute{s}_m)_{j+1/2}^{n\ 1/2} - (\acute{s}_m)_{j\ 1/2}^{n\ 1/2}].\end{aligned} \tag{5.23}$$

Note that $(f_m)_j^n$, $m = 1, 2, 3$, are functions of $(u_m)_j^n$, $m = 1, 2, 3$, and the latter can be evaluated by using Eq. (5.21). Thus, $(b_m)_j^n$, $m = 1, 2, 3$, can also be evaluated in terms of the variables at the $(n - \frac{1}{2})$th time level.

To proceed, note that Eqs. (4.10), (4.11) and (4.17) imply that

$$(f_{mt})_j^n = -\sum_{k=1}^{3}\sum_{l=1}^{3}(f_{m,l})_j^n(f_{l,k})_j^n(u_{kx})_j^n. \tag{5.24}$$

Moreover, for all $(j, n) \in \Omega$, let

$$(u_{mx}^{+})_j^n \stackrel{\text{def}}{=} \frac{\Delta x}{4}(u_{mx})_j^n, \qquad m = 1, 2, 3, \tag{5.25}$$

$$(f_{m,k}^{+})_j^n \stackrel{\text{def}}{=} \frac{\Delta t}{\Delta x}(f_{m,k})_j^n, \qquad m, k = 1, 2, 3, \tag{5.26}$$

$$(\tilde{f}_{m,k}^{-})_j^n \stackrel{\text{def}}{=} \frac{4\Delta t}{(\Delta x)^2}(\tilde{f}_{m,k})_j^n, \quad m, k = 1, 2, 3, \tag{5.27}$$

and

$$(a_{mk})_j^n \stackrel{\text{def}}{=} \delta_{mk} + (\tilde{f}_{m,k}^{-})_j^n - \sum_{l=1}^{3}(f_{m,l}^{+})_j^n(f_{l,k}^{+})_j^n, \quad m, k = 1, 2, 3. \tag{5.28}$$

With the aid of Eqs. (5.9) and (5.24)–(5.27), Eq. (5.22) can be reexpressed as

$$\sum_{k=1}^{3}(a_{mk})_j^n(u_{kx}^{+})_j^n = (b_m)_j^n, \quad m = 1, 2, 3. \tag{5.29}$$

Because $(f_{m,k}^{-})_j^n$ and $(\tilde{f}_{m,k}^{+})_j^n$, $m, k = 1, 2, 3$, are all functions of $(u_m)_j^n$, $m = 1, 2, 3$, so are $(a_{mk})_j^n$ $m, k = 1, 2, 3$. Thus, $(a_{mk})_j^n$ can also be evaluated in terms of the variables at the $(n - \frac{1}{2})$th time level. It follows that, for each $(j, n) \in \Omega$, Eq. (5.29) represents a system of three linear equations for three unknowns $(u_{mx}^{-})_j^n$, $m = 1, 2, 3$. These unknowns (and thus $(u_{mx})_j^n$, $m = 1, 2, 3$, through Eq. (5.25)) can be solved easily by a matrix inversion. Equations (5.20) and (5.29) form the current marching scheme.

6. STABILITY ANALYSIS

The stability of the a-μ and a-ε schemes will be studied using the von Neumann analysis. For all $(j, n) \in \Omega$, let

$$\mathbf{q}(j, n) = \mathbf{q}^*(n, \theta)e^{ij\theta} \quad (i \stackrel{\text{def}}{=} \sqrt{-1}, -\pi < \theta \leq \pi), \tag{6.1}$$

where $\mathbf{q}^*(n, \theta)$ is a 2×1 column matrix. Substituting Eq. (6.1) into Eq. (2.18), one obtains

$$\mathbf{q}^*(n + 1, \theta) = [Q(\nu, \xi, \theta)]^2\mathbf{q}^*(n, \theta), \tag{6.2}$$

where

$$Q(\nu, \xi, \theta) \stackrel{\text{def}}{=} e^{-i\theta/2}Q_{+} + e^{i\theta/2}Q_{-}. \tag{6.3}$$

According to Eq. (6.2), the amplification matrix is the square of the matrix $Q(\nu, \xi, \theta)$. Substituting Eqs. (2.16) and (2.17) into Eq. (6.3), one has

$$Q(\nu, \xi, \theta) = \begin{pmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{pmatrix}$$

where

$$\begin{aligned} q_{11} &= \cos(\theta/2) - i\nu\sin(\theta/2) \\ q_{12} &= -i(1 - \nu^2 - \xi)\sin(\theta/2) \\ q_{21} &= \frac{i(1 - \nu^2)\sin(\theta/2)}{1 - \nu^2 + \xi} \\ q_{22} &= -\frac{1 - \nu^2 - \xi}{1 - \nu^2 + \xi}[\cos(\theta/2) + i\nu\sin(\theta/2)]. \end{aligned} \tag{6.4}$$

Let

$$\eta(\nu, \xi, \theta) \stackrel{\text{def}}{=} \xi\cos(\theta/2) - i\nu(1 - \nu^2)\sin(\theta/2). \tag{6.5}$$

Then the eigenvalues of $Q(\nu, \xi, \theta)$ are

$$\sigma_{\pm}(\nu, \xi, \theta) \stackrel{\text{def}}{=} \frac{\eta(\nu, \xi, \theta) \pm \sqrt{[\eta(\nu, \xi, \theta)]^2 + (1 - \nu^2)^2 - \xi^2}}{1 - \nu^2 + \xi}. \tag{6.6}$$

Thus the amplification factors $G_{+}^{(1)}$ and $G_{-}^{(1)}$ of the a-μ scheme are given by

$$G_{\pm}^{(1)} = [\sigma_{\pm}(\nu, \xi, \theta)]^2. \tag{6.7}$$

Note that

$$G_{+}^{(1)} \to 1, \quad G_{-}^{(1)} \to \left(\frac{1 - \nu^2 - \xi}{1 - \nu^2 + \xi}\right)^2 \quad \text{as } \theta \to 0 \tag{6.8}$$

if $1 - \nu^2 \geq 0$. Because the amplification factor of a plane-wave solution to Eq. (2.1) approaches 1 as $\theta \to 0$, $G_{+}^{(1)}$ and $G_{-}^{(1)}$ are referred to as the principal and the spurious amplification factors, respectively. Moreover, Eqs. (6.5)–(6.7) imply that

$$G_{\pm}^{(1)} = \left\{ \frac{1}{1+\hat{\xi}} [\hat{\xi}\cos(\theta/2) - i\nu\sin(\theta/2) \pm \sqrt{[\hat{\xi}\cos(\theta/2) - i\nu\sin(\theta/2)]^2 + 1 - \hat{\xi}^2}] \right\}^2. \tag{6.9}$$

if $1 - \nu^2 \neq 0$, and $\hat{\xi} \stackrel{\text{def}}{=} \xi/(1-\nu^2)$. Similarity between Eqs. (6.9) and (2.21) was noted in Section 2.

In [1], the stability of the a-μ scheme is studied using a rigorous discrete Fourier analysis. The von Neumann stability analysis can be considered as a limiting case of the discreate Fourier analysis. By using Eqs. (4.29) and (4.30) in [1], one can infer that the a-μ scheme is stable if and only if, for all θ with $-\pi < \theta \leq \pi$,

$$\max\{|G_{+}^{(1)}|, |G_{-}^{(1)}|\} \leq 1 \quad \text{if } Q(\nu, \xi, \theta) \text{ is nondefective} \tag{6.10}$$

and

$$|G_{-}^{(1)}| < 1 \quad \text{if } Q(\nu, \xi, \theta) \text{ is defective.} \tag{6.11}$$

Note that $G_{-}^{(1)} = G^{(1)}$ if $Q(\nu, \xi, \theta)$ is defective [30, p. 353]. Assuming $\xi \geq 0$ and $1 - \nu^2 + \xi \neq 0$ (the latter is a basic assumption of Eq. (2.14)), it is proved in [1] that the current scheme is stable if and only if $\nu^2 \leq 1$.

Let $(1-\nu^2)^2 \neq \xi^2$ such that both Eqs. (2.14) and (2.24) are valid. Combining Eqs. (6.5)–(6.7), one has

$$G_{-}^{(1)} G^{(1)} = \left(\frac{1 - \nu^2 - \xi}{1 - \nu^2 + \xi} \right)^2. \tag{6.12}$$

Because the amplification factors of the backward-marching scheme are $(G_{+}^{(1)})^{-1}$ and $(G_{-}^{(1)})^{-1}$, stability of both Eqs. (2.14) and (2.24) requires that $|G_{+}^{(1)}| = |G_{-}^{(1)}| = 1$. According to Eq. (6.12), the last condition cannot be met if $\mu > 0$ and $\nu^2 \neq 1$. This result was used in a discussion given in Section 2.

Next we study the stability of the a-ε scheme. By substituting Eq. (6.1) into Eq. (3.7), one has

$$\mathbf{q}^*(n+1, \theta) = [M(\varepsilon, \nu, \theta)]^2 \mathbf{q}(n\theta), \tag{6.13}$$

where

$$M(\varepsilon, \nu, \theta) \stackrel{\text{def}}{=} e^{-i\theta/2} M_1 + e^{i\theta/2} M \ . \tag{6.14}$$

According to Eq. (6.13), the amplification matrix of the a-ε scheme is the square of the matrix $M(\varepsilon, \nu, \theta)$. Substituting Eqs. (3.8) and (3.9) into Eq. (6.14), one has

$$M(\varepsilon, \nu, \theta) = \begin{pmatrix} \cos(\theta/2) - i\nu\sin(\theta/2) & -i(1-\nu^2)\sin(\theta/2) \\ i(1-\varepsilon)\sin(\theta/2) & (2\varepsilon - 1)\cos(\theta/2) - i\nu\sin(\theta/2) \end{pmatrix}. \tag{6.15}$$

The eigenvalues $\lambda_{\pm}(\varepsilon, \nu, \theta)$ of $M(\varepsilon, \nu, \theta)$ were given in Eq. (3.13). The principal amplification factor $G_{+}^{(2)}$ and the spurious amplification factor $G_{-}^{(2)}$ of the a-ε scheme were given in Eq. (3.12). Note that

$$G_{+}^{(2)} \to 1, \quad G_{-}^{2} \to 2\varepsilon - 1 \quad \text{as } \theta \to 0 \tag{6.16}$$

if Eq. (3.14) is assumed. Moreover, from Eqs. (6.10) and (6.11), one infers that the a-ε scheme is stable if and only if, for all θ with $-\pi < \theta \leq \pi$,

$$\max\{|G_{+}^{(2)}|, |G_{-}^{(2)}|\} \leq 1 \quad \text{if } M(\varepsilon, \nu, \theta) \text{ is nondefective} \tag{6.17}$$

and

$$|G_{-}^{(2)}| < 1 \quad \text{if } M(\varepsilon, \nu, \theta) \text{ is defective.} \tag{6.18}$$

Equation (3.13) implies that

$$|\lambda_{+}(\varepsilon, \nu, 0)|\,|\lambda_{-}(\varepsilon, \nu, 0)| = |2\varepsilon - 1|. \tag{6.19}$$

By using Eqs. (3.12) and (6.17)–(6.19), one concludes that stability requires that $|2\varepsilon - 1| \leq 1$, i.e., $0 \leq \varepsilon \leq 1$. Thus the first part of Eq. (3.14) is necessary for stability. Equation (3.13) also implies that

$$\lambda_{+}(\varepsilon, \nu, \pi) = -i\nu \pm \sqrt{(1-\varepsilon)(1-\nu^2)}. \tag{6.20}$$

Thus,

$$\max\{|\lambda_{+}(\varepsilon, \nu, \pi)|, |\lambda_{-}(\varepsilon, \nu, \pi)|\} > 1 \quad \text{if } \nu^2 > 1; \varepsilon \leq 1. \tag{6.21}$$

The first part of Eq. (3.14) coupled with Eqs. (6.17), (6.18), and (6.21) implies that $\nu^2 \leq 1$ is necessary for stability. Because the case $\nu^2 = 1$ is ruled out by the basic assumption $1 - \nu^2 \neq 0$ of Eq. (3.6), the second part of Eq. (3.14) is also necessary for stability. The proof that Eq. (3.14) is also sufficient for stability will be given later in this section.

To prove Eqs. (3.17) and (3.18), note that Eq. (3.16) implies that

$$\chi_{\pm}(\varepsilon, \nu, \theta) = \varepsilon(\chi' \mp \chi''), \tag{6.22}$$

where

$$\chi' \stackrel{\text{def}}{=} (1-\nu^2)\sin^2(\theta/2) + 2(1-\varepsilon)\cos^2(\theta/2) \tag{6.23}$$

and

$$\chi'' \stackrel{\text{def}}{=} 2\cos(\theta/2) \times \sqrt{(1-\varepsilon)[(1-\varepsilon)\cos^2(\theta/2) + (1-\nu^2)\sin^2(\theta/2)]}. \tag{6.24}$$

With the aid of Eq. (3.14) and $-\pi < \theta \leq \pi$, , Eqs. (6.23) and (6.24) imply that

$$\chi' = \chi'' = 0 \quad \text{if } \varepsilon = 1;\ \theta = 0, \tag{6.25}$$

$$\chi' \begin{cases} =0, & \text{if } \varepsilon = 1;\ \theta = 0; \\ >0, & \text{if } \varepsilon \neq 1;\ \text{or } \theta \neq 0, \end{cases} \tag{6.26}$$

$$\chi'' \geq 2(1-\varepsilon)\cos^2(\theta/2) \geq 0, \tag{6.27}$$

$$\chi' - \chi'' \leq (1-\nu^2)\sin^2(\theta/2), \tag{6.28}$$

$$(\chi' - \chi'')(\chi' + \chi'') = (\chi')^2 - (\chi'')^2 = (1-\nu^2)^2\sin^4(\theta/2). \tag{6.29}$$

For the case $\varepsilon = 1$ and $\theta = 0$, Eqs. (3.17) and (3.18) follow immediately from Eqs. (6.22) and (6.25). Thus, in the following proof of Eqs. (3.17) and (3.18), we assume that

$$\varepsilon \neq 1 \quad \text{or} \quad \theta \neq 0. \tag{6.30}$$

Combining Eqs. (6.26), (6.27), and (6.30), one concludes that

$$\chi' + \chi'' > 0. \tag{6.31}$$

Equations (6.29) and (6.31) imply that

$$\chi' - \chi'' \geq 0. \tag{6.32}$$

Equation (3.18) now follows from Eqs. (3.14), (6.22), (6.28), and (6.32). The validity of the first inequality sign in Eq. (3.17) follows from Eq. (3.18) and the fact that $\varepsilon(1-\varepsilon) \geq 0$ if $0 \leq \varepsilon \leq 1$. The validity of the second inequality sign follows from the fact that

$$\chi(\varepsilon, \nu, \theta) - \chi_*(\varepsilon, \nu, \theta) = 2\varepsilon\chi'' \geq 4\varepsilon(1-\varepsilon)\cos^2(\theta/2). \tag{6.33}$$

Equation (6.33) is a simple result of Eqs. (6.22) and (6.27). To establish the validity of the last inequality sign in Eq. (3.17), note that

$$\begin{aligned} \chi(\varepsilon, \nu, \theta) &= \varepsilon(\chi' + \chi'') = \varepsilon[2\chi' - (\chi' - \chi'')] \leq 2\varepsilon\chi' \\ &= 2\varepsilon[(1-\nu^2)\sin^2(\theta/2) \\ &\quad + 2(1-\varepsilon)\cos^2(\theta/2)] \\ &\leq \max\{2\varepsilon(1-\nu^2), 4\varepsilon(1-\varepsilon)\} \leq 4\varepsilon, \end{aligned} \tag{6.34}$$

where Eqs. (6.22), (6.32), (6.23), and (3.14) have been used Moreover, because $|G^{(2)}| \geq 0$, Eq. (3.16) implies that

$$\chi(\varepsilon, \nu, \theta) \leq 1. \tag{6.35}$$

The validity of the last inequality sign in Eq. (3.17) now follows from Eqs. (6.34) and (6.35). Q.E.D

Next we shall prove that Eq. (3.14) is also sufficient for stability. Note that, as a result of Eqs. (3.17) and (3.18), $0 \leq \chi_*(\varepsilon, \nu, \theta)$, and thus $|G^{(2)}| \leq 1$, for all ε, ν, and θ satisfying Eq. (3.14) and $-\pi < \theta \leq \pi$. As a result, Eq. (6.17) is always satisfied. To complete the proof, we need only show that Eq.(6.18) is also satsfied. To proceed, note that $G_*^{(2)} = G^{(2)}$ if $M(\varepsilon, \nu, \theta)$ is defective. From Eqs. (3.12)–(3.14), one also concludes that $\varepsilon = 1$ is necessary if $G_*^{(2)} = G^{(2)}$. Moreover, Eq. (6.15) implies that $M(1, \nu, 0)$ is the identity matrix. Thus, one concludes that $\varepsilon = 1$ and $\theta \neq 0$ are necessary if $M(\varepsilon, \nu, \theta)$ is defective. Because (i)

$$G^{(2)} = [\cos(\theta/2) - i\nu\sin(\theta/2)]^2 \quad \text{if } \varepsilon = 1 \tag{6.36}$$

and (ii)

$$|[\cos(\theta/2) - i\nu\sin(\theta/2)]^2| < 1 \quad \text{if } \nu^2 < 1;\ \theta \neq 0, \tag{6.37}$$

one arrives at the conclusion that Eq. (6.18) is also satisfied. Q.E.D

7. NUMERICAL RESULTS

In [1], numerical solutions of Eq. (2.1) generated by the MacCormack [12, p.102], the Leapfrog/DuFort–Frankel, and the a-μ schemes are compared with the corresponding analytical solutions for different values of physical coefficients, mesh parameters and total marching times. These comparisons show that the a-μ scheme is far superior to the Leapfrog/DuFort–Frankel scheme in accuracy and has a substantial advantage over the MacCormack scheme in both accuracy and stability.

In this section, accuracy of both the Euler and the Navier–Stokes solvers will be evaluated numerically using a shock tube problem suggested by Sod [31]. Because the a-ε scheme may be considered as a special case of the Euler solver, no separate numerical evaluation for the a-ε scheme will be given.

Let the specific heat ratio $\gamma = 1.4$. At $t = 0$, let (i) $(\rho, v, p) = (1, 0, 1)$, i.e., $(u_1, u_2, u_3) = (1, 0, 2.5)$ if $x < 0$, and (ii) $(\rho, v, p) = (0.125, 0, 0.1)$, i.e., $(u_1, u_2, u_3) = (0.125,$

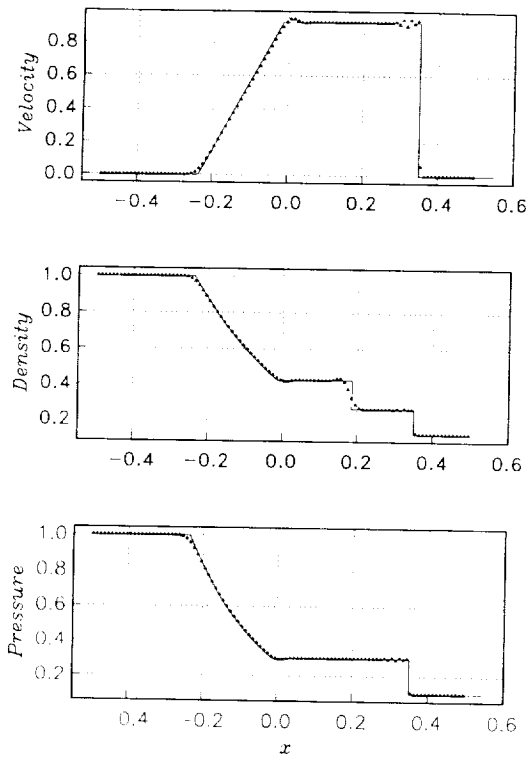

FIG. 6. The Euler solution ($\varepsilon = 0.5$, $\alpha = 0$, $\Delta t = 0.004$, CFL $\doteq 0.88$).

0, 0.25) if $x > 0$. For all $(j, n) \in \Omega$, let $x_j = j\Delta x$, and $t^n = n\Delta t$. Then (i)

$$((u_1)_j^0, (u_2)_j^0, (u_3)_j^0) = \begin{cases} (1, 0, 2.5) & \text{if } j = -\frac{1}{2}, -\frac{3}{2}, \ldots; \\ (0.125, 0, 0.25), & \text{if } j = \frac{1}{2}, \frac{3}{2}, \ldots, \end{cases} \quad (7.1)$$

and (ii) $(u_m)_j^0 = 0$, $j = \pm\frac{1}{2}, \pm\frac{3}{2}, \ldots$, for $m = 1, 2, 3$. Hereafter, we assume that $n \geq 0$.

The above initial conditions coupled with several equations given in Sections 4 and 5 imply that, for both the Euler and the Navier–Stokes solvers, $(u_m)_j^n$ is a constant and $(u_m)_j^n = 0$ in two separate regions that are defined by $j \leq -(n + \frac{1}{2})$ and $j \geq (n + \frac{1}{2})$, respectively. Thus, one needs to evaluate the above variables only if $|j| < (n + \frac{1}{2})$.

Without exception, $\Delta x = 0.01$ is assumed in this section. Also, all numerical results will be compared with the exact weak solution at $t = 0.2$. Because, at $t = 0.2$, the effect of the initial discontinuity at $t = 0$ is far from reaching the spatial regions defined by $x > 0.5$ and $x < -0.5$, respectively, numerical computations, unless specified otherwise, will be simplified by assuming that, for all n with $t^n \leq 0.2$, (i)

$$((u_1)_j^n, (u_2)_j^n, (u_3)_j^n) = \begin{cases} (1, 0, 2.5) & \text{if } x_j < -0.5; \\ (0.125, 0, 0.25) & \text{if } x_j > 0.5, \end{cases} \quad (7.2)$$

and (ii) $(u_m)_j^n = 0$ if $|x_j| > 0.5$. Because $\Delta x = 0.01$, the above assumptions imply that the computation domain can be limited to $|j| \leq 50$.

In the initial evaluation, we consider the Euler marching scheme defined by Eqs. (4.24) and (4.28). Numerical results (triangles) obtained assuming $\Delta t = 0.004$ and $\varepsilon = \frac{1}{2}$ are compared with the exact solutions (solid lines) in Fig. 6. Because each marching step advances the solution from t to $t + \Delta t/2$, these results at $t = 0.2$ are obtained after 100 steps. Also it can be estimated that CFL $\doteq 0.88$, where CFL is defined to be the maximum value of $(|v| + |c|)\Delta t/\Delta x$. Thus the numerical calculation is carried out within the stability limits given by Eq. (4.33). Note that the agreements between the numerical results and the exact solutions are excellent. Particularly, the shock discontinuity is resolved almost within one mesh interval, and the contact discontinuity is resolved in four mesh intervals. Also, there are only slight numerical overshoots and/or oscillations near these discontinuities.

According to the discussions given in Sections 3, 4, and 6, the Euler solver behaves like the Leapfrog scheme, if $\varepsilon = 0$, and like the Lax scheme, if $\varepsilon = 1$. The former is free from numerical dissipation while the latter is highly diffusive. *The current scheme with $\varepsilon = \frac{1}{2}$ can be considered as a scheme midway between the above two celebrated schemes.*

Moreover, the last term on the right side of Eq. (4.28) vanishes if $\varepsilon = \frac{1}{2}$. The remaining term is simply a central-difference approximation for $(u_{mx})_j^n$.

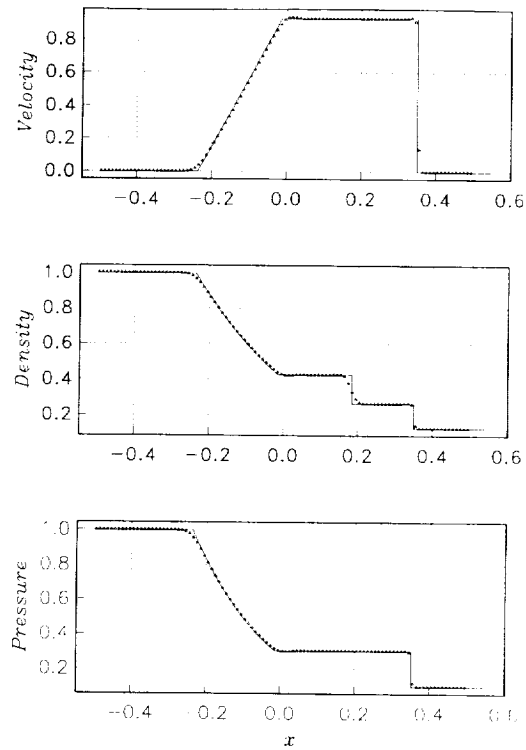

FIG. 7. The Euler solution ($\varepsilon = 0.5$, $\alpha = 1$, $\Delta t = 0.004$, CFL $\doteq 0.88$).

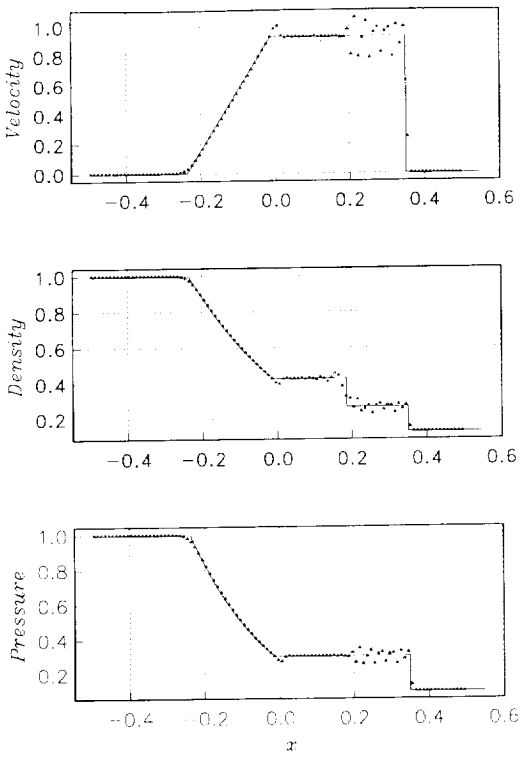

FIG. 8. The Euler solution ($\varepsilon = 0.1$, $\alpha = 0$, $\Delta t = 0.004$, CFL $\doteq 0.88$).

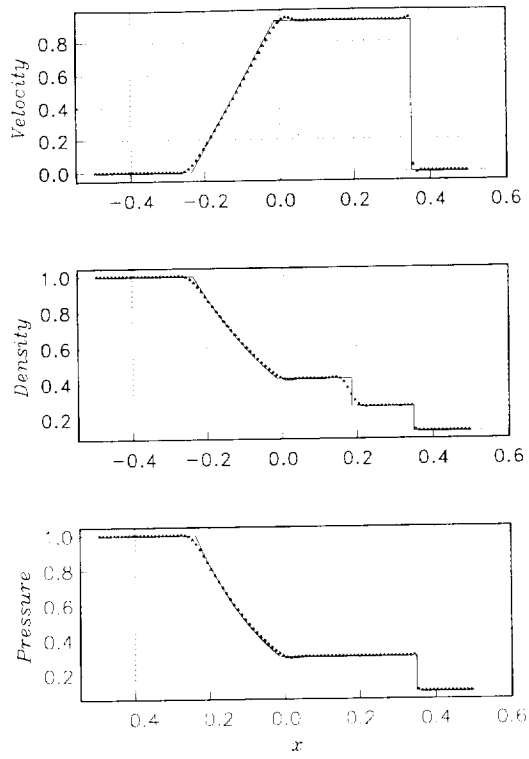

FIG. 9. The Euler solution ($\varepsilon = 0.7$, $\alpha = 0$, $\Delta t = 0.004$, CFL $\doteq 0.88$).

Let Eq. (4.28) be modified with $(u^c_{mx})^n_j$ being replaced by $(u^{wo}_{mx})^n_j$ (see Eqs. (4.35) and (4.38)). Again assuming that $\Delta t = 0.004$ and $\varepsilon = \frac{1}{2}$, the numerical results obtained with $\alpha = 1$ are given in Fig. 7. The results obtained with $\alpha = 2$, and $\alpha = 3$ are almost identical to those shown in Fig. 7 [5]. The effectiveness of the above modification as a tool to surpress numerical wiggles near discontinuities is apparent. It was explained in Section 4 why this modification does not cause the smearing of shock discontinuities. Furthermore, the modification has no discernable effect on the smooth part of the solution. Because $(u^{wo}_{mx})^n_j = (u^c_{mx})^n_j$ if $\alpha = 0$, in the following discussion, it should be understood that the above modification is turned off if $\alpha = 0$.

Note that the results shown in Figs. 6 and 7 can be generated using the sample program listed at the end of the present paper. It is coded assuming $\varepsilon = 0.5$ and without imposing the conditions given in Eq. (7.2). The parameter α is represented by *ia* in the code.

Let $\alpha = 0$ and $\Delta t = 0.004$. The numerical results obtained with $\varepsilon = 0.1$, and $\varepsilon = 0.7$, respectively, are given in Figs. 8 and 9. Note that the case with $\varepsilon = 0.5$ are given in Fig. 6. For $\varepsilon = 0.1$, because the scheme has very small numerical dissipation, pronounced wiggles appear in large regions near discontinuities. However, because of the same reason, *the smooth part of the solution is highly accurate*. The results shown in Figs. 6, 8, and 9, and other results obtained with $\varepsilon = 0.3$ and $\varepsilon = 0.9$ [5] are consistent with the theoretical

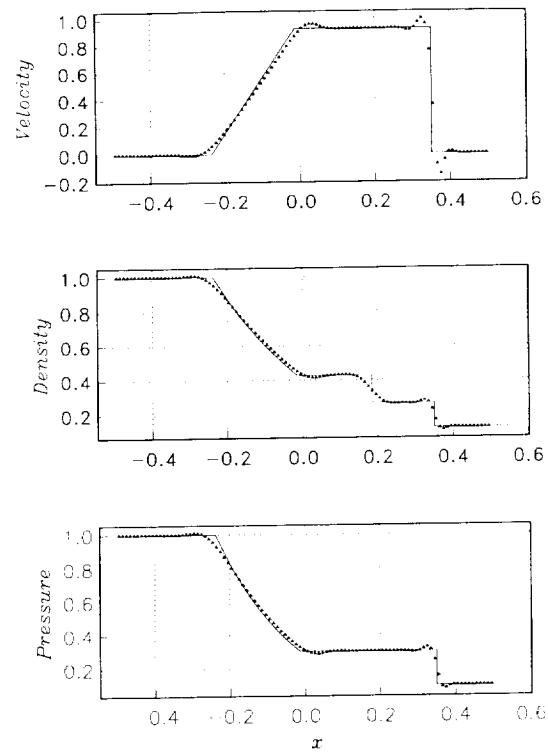

FIG. 10. The Euler solution ($\varepsilon = 0.5$, $\alpha = 0$, $\Delta t = 0.0004$, CFL $\doteq 0.088$).

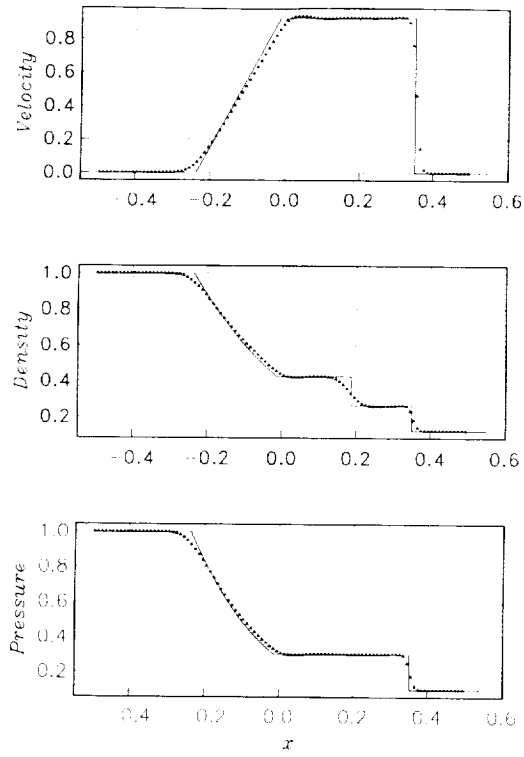

FIG. 11. The Euler solution ($\varepsilon = 0.5$, $\alpha = 1$, $\Delta t = 0.0004$, CFL $\doteq 0.088$).

prediction that the Euler solver becomes progressively diffusive as the value of ε increases from 0 to 1.

The above numerical results are all generated assuming $\Delta t = 0.004$. The numerical results shown in Figs. 10 and 11 are generated with $\Delta t = 0.0004$ (i.e., CFL $\doteq 0.088$). Note that now it takes 1000 marching steps to advance the solution to $t = 0.2$. Other defining conditions for these figures are identical to those for Figs. 6 and 7, respectively. A glance over Figs. 6, 7, 10, and 11 reveals that the current Euler solver is more diffusive at a smaller CFL. Note that, by considering the truncation error, it was shown in [5] that, for constant ε and Δx, the a-ε scheme becomes more diffusive as Δt decreases. A similar conclusion can also be reached by studying the amplification factors given in Eqs. (3.12) and (3.13). Because the Euler solver is a straightforward extension of the a-ε scheme, one would expect that the former also behaves similarly.

Also, as the value of CFL decreases, the diffusive effect of replacing $\alpha = 0$ with $\alpha = 1$ generally becomes more discernable. In other words, numerical dissipation introduced by replacing $\alpha = 0$ with $\alpha > 0$, is greater when CFL is small.

To modify the above Euler solver such that it can compensate for the observed effect of increasing numerical dissipation as Δt decreases, in the following discussions, we shall consider the more general marching scheme defined by Eqs. (4.24) and (4.32). The parameter $(\hat{\varepsilon})_j^n$ in Eq. (4.32) will be dependent on the mesh position (j, n) and the ratio $\Delta t/\Delta x$. Moreover, the term $(u_{mx}^c)_j^n$ (see Eq. (4.35)) in Eq. (4.32) will be replaced by $(u_{mx}^w)_j^n$, which is defined in Eq. (4.43). The weight factor β will also be dependent on (j, n) and $\Delta t/\Delta x$.

To proceed, let

$$\zeta(x) \overset{\text{def}}{=} x \exp(1 - x), \quad 0 \le x \le 1. \tag{7.3}$$

Because ζ is an increasing function within its domain, we have

$$\zeta(x) \le \zeta(1) = 1, \quad 0 \le x \le 1. \tag{7.4}$$

For all $(j, n) \in \Omega$, let

$$(\hat{\varepsilon})_j^n = b\zeta((\hat{\nu}_{\max})_j^n) \tag{7.5}$$

and

$$(u_{mx}^w)_j^n = W((u_{mx-})_j^n, (u_{mx+})_j^n; \alpha, \sqrt{(\hat{\nu}_{\max})_j^n}), \tag{7.6}$$

where $(\hat{\nu}_{\max})_j^n$ is defined in Eq. (4.34), and b and α are constants that do not vary from one mesh point to another. Because $(\hat{\varepsilon}_m)_j^n = (\hat{\varepsilon})_j^n$, $m = 1, 2, 3$, is assumed in Eq. (4.32), Eqs. (4.33), (7.4) and (7.5) require that (i) $(\hat{\nu}_{\max})_j^n$ be in the domain of $\zeta(x)$, and (ii) $0 \le b \le 1$.

Note that $(\hat{\nu}_{\max})_j^n$ is proportional to $\Delta t/\Delta x$. Thus, Eqs. (7.3) and (7.5) imply that $(\hat{\varepsilon})_j^n$ is an increasing function of $\Delta t/\Delta x$, i.e., it decreases as Δt decreases if other parameters are held constant. Because numerical dissipation decreases as $(\hat{\varepsilon})_j^n$ decreases, with other factors being equal, the replacement of a constant ε with $(\hat{\varepsilon})_j^n$ has an effect in reducing numerical dissipation as Δt decreases. This effect will compensate for the observed opposite effect on numerical dissipation as Δt decreases with ε, Δx, *and the total marching time is being held constant.*

Moreover, for a fixed α, $W(x_-, x_+; \alpha, \beta) \to (x_- + x_+)/2$ as $\beta \to 0$. This fact, coupled with Eq. (4.37), implies that the numerical dissipation introduced as a result of replacing $(u_{mx}^c)_j^n$ with $(u_{mx}^w)_j^n$ will decrease as β decreases. Because $(\hat{\nu}_{\max})_j^n$ is proportional to $\Delta t/\Delta x$, with other factors being equal, the replacement of $(u_{mx}^c)_j^n$ by $(u_{mx}^w)_j^n$ defined in Eq. (7.6), has an effect in reducing numerical dissipation as Δt decreases. This effect will compensate for the observed opposite effect on numerical dissipation as Δt decreases with α, β, Δx, and *the total marching time is held constant.* Note that $W_0(x_-, x_+; \alpha)$ is a special case of $W(x_-, x_+; \alpha, \beta)$ with $\beta = 1$.

Assuming that $\alpha = 1$ and $b = 0.5$, the numerical results shown in Figs. 12, 13, and 14 are generated with $\Delta t = 0.004$ (CFL $\doteq 0.88$), $\Delta t = 0.0004$ (CFL $\doteq 0.088$), and $\Delta t = 0.0001$ (CFL $\doteq 0.022$), respectively. Note that the results shown in Fig. 12 are almost identical to those shown in Fig. 7 which were generated assuming the same conditions but using a simpler marching scheme. However, the results shown in Fig. 13 are far less diffusive than their counterparts shown in Fig. 11. One can conclude from this comparison and the results shown in

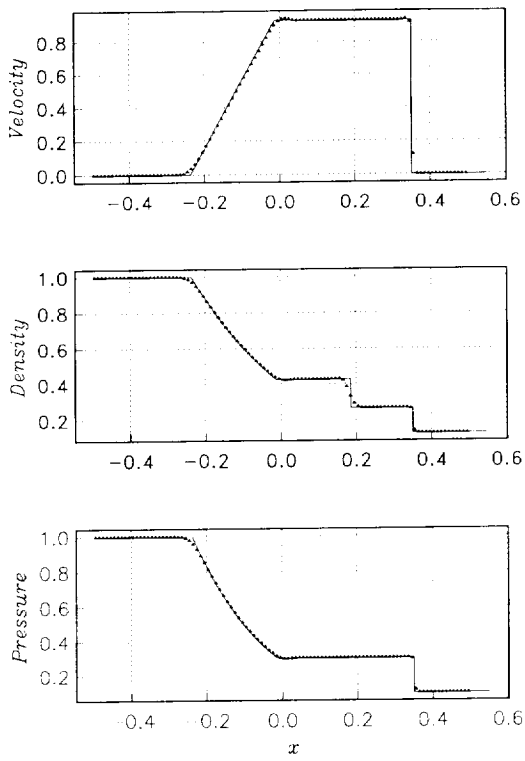

FIG. 12. The Euler solution ($b = 0.5$, $\alpha = 1$, $\Delta t = 0.004$, CFL $\doteq 0.88$).

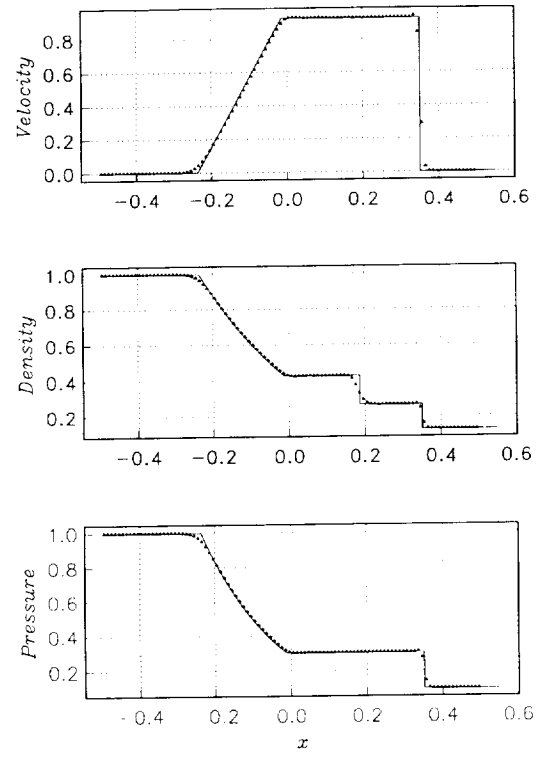

FIG. 13. The Euler solution ($b = 0.5$, $\alpha = 1$, $\Delta t = 0.0004$, CFL $\doteq 0.088$).

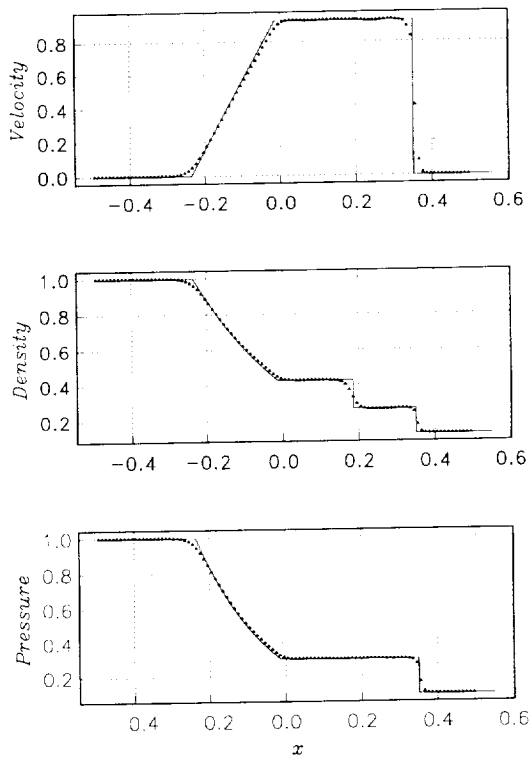

FIG. 14. The Euler solution ($b = 0.5$, $\alpha = 1$, $\Delta t = 0.0001$, CFL $\doteq 0.022$).

Fig. 14 that the current modified Euler solver is capable of generating accurate numerical solutions even for the case with a very small CFL.

In the above modified Euler scheme, $(\hat{\varepsilon})_j^n$ and β are expressed as two special functions of $(\hat{\nu}_{\max})_j^n$, respectively. They are only two among many possible choices. The investigation of other choices is a subject to be studied in the future.

The most general marching scheme presented in Section 4 is that defined by Eqs. (4.24) and (4.31). It requires several matrix multiplications at each mesh points and, therefore, is much more costly. Thus, its use is difficult to justify unless a substantial gain in accuracy can be made. How this most general marching scheme can be applied wisely is left for a future study.

This completes the numerical study of the Euler solver. We conclude this section with a numerical evaluation of the Navier–Stokes marching scheme defined by Eqs. (5.20) and (5.29). Again the initial conditions defined in Eq. (7.1) are assumed, and the numerical solutions are compared with *the exact weak solution of the Euler equations* at $t = 0.2$. The numerical results shown in Figs. 15–17 are generated assuming $\Delta t = 0.004$, $\Delta x = 0.01$, $\gamma = 1.4$, and Pr $= 0.72$. The value of the Prandtl number used here is that for air at standard conditions. The values of the Re_L for these figures are 2000, 6000, and 10,000, respectively.

From the results shown in these figures, one concludes that, for a high-Reynolds-number flow, the shock can be resolved

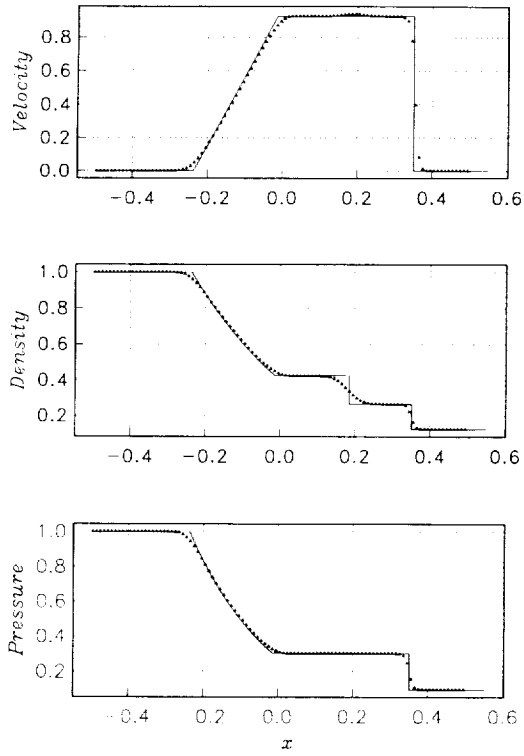

FIG. 15. The Navier–Stokes solution ($Re_L = 2000$).

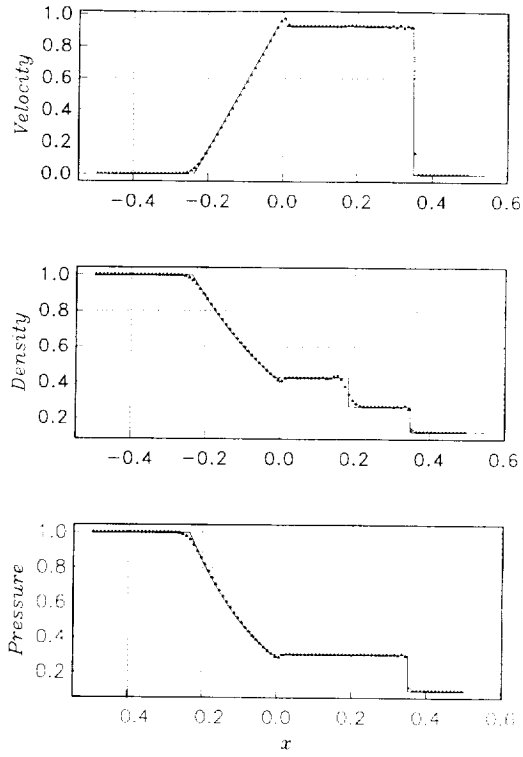

FIG. 17. The Navier–Stokes solution ($Re_L = 10000$).

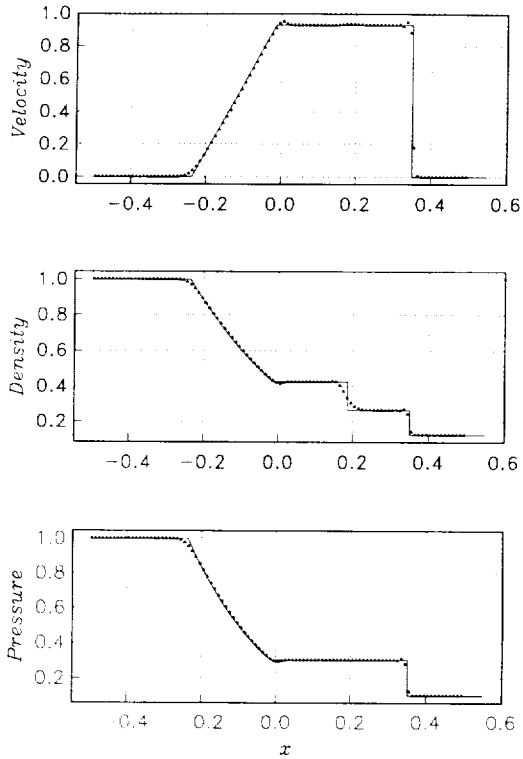

FIG. 16. The Navier–Stokes solution ($Re_L = 6000$).

within one mesh interval by the current Navier–Stokes solver. Also the contact discontinuity can be resolved within a few mesh intervals. *Note that these results are obtained without using any ad hoc parameters or techniques.* Because the Reynolds number is inversely proportional to the physical viscosity, as expected, numerical overshoots and oscillations shown in these figures increase slightly as the values of the Reynolds number increase.

Furthermore, through repeated numerical experiments using different physical and mesh parameters, it is established that the current Navier–Stokes solver is stable if, for all $(j, n) \in \Omega$,

$$0 \le \mathrm{Re}_L, \quad 0 \le \mathrm{Pr}, \qquad (\hat{\nu}_{\max})_j^n < 1. \tag{7.7}$$

However, because a Navier–Stokes problem is fundamentally an initial-value/boundary-value problem, the current explicit marching scheme obviously cannot model such a problem unless the boundary effect is small, i.e., when the contribution of the viscous terms to Eqs. (5.20) and (5.29) is small compared to that of the convection terms. In general, this implies that the current scheme is applicable only to high-Reynolds-number flows. Note that the Leapfrog/Dufort–Frankel and the a-μ schemes [1] also encounter a similar limitation in modelling Eq. (2.1).

Finally, note that the current Navier–Stokes solver with $Re_L = \infty$ (i.e., the physical viscosity vanishes) and $Pr = 0$ can be considered as a nonlinear extension of the inviscid a-μ

scheme. Because the latter scheme is neutrally stable, generally one would expect that a nonlinear extension of such a scheme is unstable. However, it has been shown numerically that the current Navier–Stokes solver is stable even for the above limiting case as long as $(\hat{\nu}_{\max})_j^n < 1$ for all $(j, n) \in \Omega$.

8. CONCLUSIONS AND DISCUSSIONS

Several key limitations of the finite difference, finite volume, finite element, and spectral methods were discussed in Section 1. The method of space-time conservation element and solution element was conceived to overcome these limitations.

Using the a-μ scheme as an example, major differences between the present method and those mentioned above were explained in Section 2. This explicit scheme has the unusual property that its stability is limited only by the CFL condition, i.e., it is independent of μ. Also, it was shown that its amplification factors are identical to those of the Leapfrog scheme, if $\mu = 0$, and to those of the DuFort–Frankel scheme, if $a = 0$. These coincidences are rather unexpected because the a-μ scheme and the above classical schemes are derived from completely different perspectives, and the current scheme *does not* reduce to the above classical schemes in the limiting cases.

The inviscid a-μ scheme is reversible in time. Obviously the Euler extension of such a scheme cannot model a physical problem that is irreversible in time, e.g., an inviscid flow problem involving shocks. Thus, the inviscid version was modified in Section 3 to form the a-ε scheme. This new scheme has the unusual property that numerical dissipation is controlled by an adjustable parameter ε. As a matter of fact, for all wavelengths, numerical dissipation can be *uniformly* bounded from above by an arbitrary small number by choosing a small enough ε. Stability of the a-ε scheme is limited by the CFL condition and $0 \leq \varepsilon \leq 1$. Moreover, if $\varepsilon = 0$, the amplification factors of the a-ε scheme are identical to those of the Leapfrog scheme, which has no numerical dissipation. On the other hand, if $\varepsilon = 1$, they unexpectedly become identical to each other and to the amplification factor of the highly diffusive Lax scheme. Note that, because the Lax scheme is very diffusive and uses a mesh that is staggered in time, a two-level scheme using such a mesh is often associated with a highly diffusive scheme. The a-ε scheme, which also uses a mesh staggered in time, demonstrates that such a scheme could be free from numerical dissipation.

In Section 4, the a-ε scheme was extended to become an Euler solver. This solver has the unusual property that numerical dissipation at any mesh point (j, n) can be controlled by a set of local parameters $(\hat{\varepsilon}_m)_j^n$, $m = 1, 2, 3$. As in the a-ε scheme, stability of the Euler solver is limited by the CFL condition and the requirement that, for all (j, n), $0 \leq (\hat{\varepsilon}_m)_j^n \leq 1$, $m = 1, 2, 3$. Note that an Euler solver using a mesh staggered in time is usually highly diffusive for a small CFL number. It was shown in Section 7 that the current solver is an exception. It can generate highly accurate shock tube solutions with the CFL number ranging from 0.88 to 0.022.

In Section 5, the a-μ scheme was extended to become a Navier–Stokes solver. Stability of this *explicit* solver is also limited only by the CFL condition. Despite the fact that it does not use (i) *any* techniques related to the high-resolution upwind methods, and (ii) *any* ad hoc parameter, it was shown in Section 7 that the current solver is capable of generating highly accurate shock tube solutions. Particularly, shock discontinuites can be resolved within one mesh interval.

A summary of the key results of the present work has been given. Behind these results is a continuous effort to maintain the simplicity, generality, and accuracy of the present method. This effort is summarized in the following remarks:

(a) *Simplicity*. The current numerical framework rests upon only two basic building blocks, i.e., the space-time conservation and solution elements. It uses only local discrete variables. Also, the set of discrete variables in any one of the numerical equations to be solved is associated with a single SE or a few immediately neighboring SEs. Thus, local flexibility is preserved and one needs only to deal with a very sparse matrix. Moreover, flux evaluation at an interface separating two CEs requires no interpolation or extrapolation. Nor does it require the use of an ad hoc flux model. Finally, partly because no characteristics-based techniques are used, a numerical scheme can be constructed by using only the simplest approximation techniques.

(b) *Generality*. A guiding principle in the design of the present method is to limit the use of special assumptions or techniques that would restrict its use in more general situations. Thus we do not use characteristics-based techniques, and we try to avoid using ad hoc techniques.

(c) *Accuracy*. Because (i) a physical solution of the conservation laws may involve shocks or high-gradient regions, and (ii) an accurate numerical simulation of such a solution is difficult to obtain without enforcing flux conservation, the present method requires that *a numerical solution satisfies* (i) *the differential form of the conservation laws uniformly within an SE, and* (ii) *the integral form over any space-time region that is the union of any combination of CEs*. In addition, accuracy of the present method is aided by treating both $(u_m)_j^n$ and $(u_{mx})_j^n$ as independent variables, instead of expressing $(u_{mx})_j^n$ as a finite-difference approximation involving $(u_m)_j^n$'s of neighboring mesh points. The latter approach may result in poor accuracy in a high-gradient region. Also, accuracy is enhanced by the fact that the flux at an interface separating two CEs is evaluated without interpolation or extrapolation. Moreover, because flux conservation is fundamentally a property in space-time, the current unified treatment of space and time may also contribute to a more accurate simulation of the conservation laws.

As a result of its simplicity and generality, the current framework is also very flexible in its ability to generate discretized equations such that number of equations can match number of unknowns. In [5], this flexibility is demonstrated in a discussion

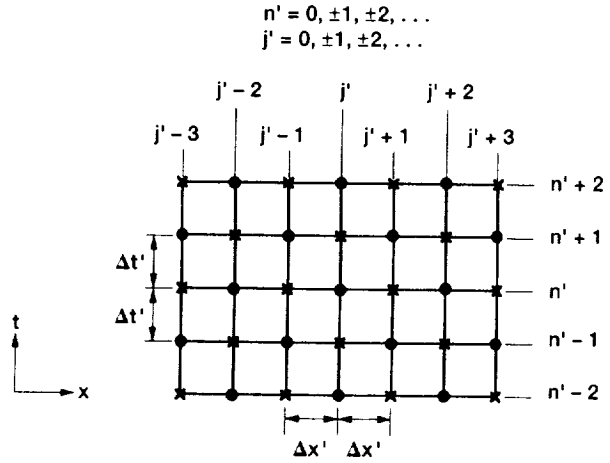

FIG. 18. A regular space-time mesh.

on how the current framework can be used to discretize a 2D steady incompressible Navier–Stokes problem. In the same discussion, the important issue of boundary-condition implementation is also addressed.

Finally, the present paper is concluded with remarks on several extensions of the current basic solvers:

(a) In [6], the Euler solver discussed in Sections 4 and 7 was extended and applied to more complex flow problems involving shock tubes of finite or infinite length. The numerical results obtained clearly demonstrate the ability of the extended solver to resolve discontinuities accurately even in the presence of wave interactions and reflections.

(b) Several solvers developed in the present paper have been extended to solve two-dimensional time-marching problems [7, 8]. The construction of these extensions are simplified greatly by the use of a nontraditional space-time mesh. Its use results in the simplest stencil possible, i.e., a tetrahedron in a 3D space-time with a vertex at the upper time level and the other three at the lower time level. Other discussions of these 2D schemes were given in Section 1.

(c) Extensions to solve 2D steady, incompressible Navier–Stokes equations were discussed near the end of Section 2.

Note. To obtain the NASA Technical Memorandums referred to in the present paper, please contact the author.

APPENDIX A: AN ALTERNATIVE STABILITY ANALYSIS FOR THE LAX AND LEAPFROG/DUFORT–FRANKEL SCHEMES

With the use of the regular mesh depicted in Fig. 18, the Lax scheme for solving Eq. (2.22) can be expressed as

$$\frac{u_j^{n'+1} - (u_{j'+1}^{n'} + u_{j'-1}^{n'})/2}{\Delta t'} + a\frac{u_{j'+1}^{n'} - u_{j'-1}^{n'}}{2\Delta x'} = 0, \qquad \text{(A.1)}$$

where $j', n' = 0, \pm 1, \pm 2, \ldots$. The system of equations represented by Eq. (A.1) can be divided into two sets completely independent from each other. The first set involves only the variables associated with those mesh points marked by dots in Fig. 18, and the second set, by crosses. Thus, the solution to Eq. (A.1) contains two decoupled solutions. Traditionally the von Neumann stability analysis for the Lax scheme is performed without taking into account this decoupling nature. Consider a solution to Eq. (A.1) in which $u_{j'}^{n'} = 1$ for all mesh points (j', n') that are marked by dots, and $u_{j'}^{n'} = -1$ for all other (j', n'). In reality, this solution represents the union of two completely decoupled *constant* solutions. However, at any time level, the combined solution is represented by a Fourier component of the shortest wavelength $(=2\Delta x')$ in the traditional analysis. Therefore, two decoupled *constant* solutions may be wrongly perceived as a *rapidly-varying* solution. For the above reason, we shall consider each decoupled solution separately in the following von Neumann stability analysis.

Let $n = n'/2$, $j = j'/2$, $\Delta x = 2\Delta x'$, and $\Delta t = 2\Delta t'$. Then the mesh depicted in Fig. 18 is identical to that depicted in Fig. 2(a) except that those mesh points marked by crosses in Fig. 18 have no counterparts in Fig. 2(a). As a result, the decoupling nature of Eq. (A.1) will be removed if the Lax scheme is expressed using the staggered mesh depicted in Fig. 2(a), i.e., for all $(j, n) \in \Omega$,

$$\frac{u_j^n - (u_{j+1/2}^{n-1/2} + u_{j-1/2}^{n-1/2})/2}{\Delta t/2} + a\frac{u_{j+1/2}^{n-1/2} - u_{j-1/2}^{n-1/2}}{\Delta x} = 0. \qquad \text{(A.2)}$$

With the aid of Eq. (2.13), Eq. (A.2) can be simplified as

$$u_j^n = \tfrac{1}{2}[(1+\nu)u_{j-1/2}^{n-1/2} + (1-\nu)u_{j+1/2}^{n-1/2}]. \qquad \text{(A.3)}$$

By applying Eq. (A.3) successively, one has

$$u_j^{n+1} = \tfrac{1}{4}[(1+\nu)^2 u_{j-1}^{n-1} + 2(1-\nu^2)u_j^n + (1-\nu)^2 u_{j+1}^{n-1}]. \qquad \text{(A.4)}$$

In contrast to Eq. (2.19), Eq. (A.4) implies that u_j^{n+1} *does not* approach u_j^n as $\Delta t \to 0$. Moreover, by substituting

$$u_j^n = [G(\nu, \theta)]^n e^{ij\theta} \quad (i \overset{\text{def}}{=} \sqrt{-1},\ -\pi < \theta \le \pi) \qquad \text{(A.5)}$$

into Eq. (A.4), one concludes that the amplification factor of the Lax scheme is given by

$$G(\nu, \theta) = |\cos(\theta/2) - i\nu\sin(\theta/2)|^2. \qquad \text{(A.6)}$$

A comparison among Eqs. (3.12), (3.15), and (A.6) reveals that $G_{\pm}^{(2)} = G^{(2)} = G(\nu, \theta)$ when $\varepsilon = 1$.

Because u_j^{n+1} does not approach u_j^n as $\Delta t \to 0$. It follows from Eq. (A.5) that $G(\nu, \theta)$ cannot approach 1 as $\nu \to 0$. As

a matter of fact, $G(\nu, \theta) \to \cos^2(\theta/2)$ as $\nu \to 0$. In turn, this implies that the Lax scheme is highly diffusive when $|\nu|$ is small.

With the use of the regular mesh depicted in Fig. 18, the Leapfrog/DuFort–Frankel scheme for solving Eq. (2.1) can be expressed as

$$\frac{u_{j'}^{n'+1} - u_{j'}^{n'-1}}{2\Delta t'} + a\frac{u_{j'+1}^{n'} - u_{j'-1}^{n'}}{2\Delta x'} - \mu\frac{u_{j'-1}^{n'} + u_{j'-1}^{n'} - u_{j'}^{n'+1} - u_{j'}^{n'-1}}{(\Delta x')^2} = 0. \tag{A.7}$$

where $j', n' = 0, \pm 1, \pm 2, \ldots$. Even though Eq. (A.7) is a three-level scheme while Eq. (A.1) is a two-level scheme, they have the same decoupling nature. The decoupling of Eq. (A.7) can be removed if the scheme is expressed with respect to the staggered mesh depicted in Fig. 2(a), i.e., for all $(j, n) \in \Omega$,

$$\frac{u_j^n - u_j^{n-1}}{\Delta t} + a\frac{u_{j+1/2}^{n-1/2} - u_{j-1/2}^{n-1/2}}{\Delta x} - \mu\frac{u_{j+1/2}^{n-1/2} + u_{j-1/2}^{n-1/2} - u_j^n - u_j^{n-1}}{(\Delta x/2)^2} = 0. \tag{A.8}$$

With the aid of Eq. (2.13), Eq. (A.8) can be simplified as

$$(1 + \xi)u_j^n = (1 - \xi)u_j^{n-1} + (\nu + \xi)u_{j-1/2}^{n-1/2} - (\nu - \xi)u_{j+1/2}^{n-1/2}; \tag{A.9}$$

Eq. (A.9) can also be expressed in a two-level form, i.e.,

$$\mathbf{u}(j, n) = L_-\mathbf{u}(j - \tfrac{1}{2}, n - \tfrac{1}{2}) + L_+\mathbf{u}(j + \tfrac{1}{2}, n - \tfrac{1}{2}). \tag{A.10}$$

Here

$$\mathbf{u}(j, n) \stackrel{\text{def}}{=} \begin{pmatrix} u_j^n \\ u_{j+1/2}^{n-1/2} \end{pmatrix} \tag{A.11}$$

for all $(j, n) \in \Omega$ with $n > 0$, and

$$L_+ \stackrel{\text{def}}{=} \begin{pmatrix} \dfrac{\nu + \xi}{1 + \xi} & \dfrac{1 - \xi}{1 + \xi} \\ 0 & 0 \end{pmatrix}, \quad L_- \stackrel{\text{def}}{=} \begin{pmatrix} \dfrac{-(\nu - \xi)}{1 + \xi} & 0 \\ 1 & 0 \end{pmatrix}. \tag{A.12}$$

By applying Eq. (A.10) successively, one has

$$\mathbf{u}(j, n + 1) = (L_-)^2\mathbf{u}(j - 1, n) + (L_-L_+ + L_+L_-)\mathbf{u}(j, n) + (L_+)^2\mathbf{u}(j + 1, n). \tag{A.13}$$

To perform the von Neumann stability analysis for Eq. (A.13), let

$$\mathbf{u}(j, n) = \mathbf{u}^\star(n, \theta)e^{ij\theta} \quad (i \stackrel{\text{def}}{=} \sqrt{-1}, -\pi < \theta \le \pi), \tag{A.14}$$

where $\mathbf{u}^\star(n, \theta)$ is a 2×1 column matrix. Substituting Eq. (A.14) into Eq. (A.13), one obtains

$$\mathbf{u}^\star(n + 1, \theta) = [L(\nu, \xi, \theta)]^2\mathbf{u}^\star(n, \theta), \tag{A.15}$$

where

$$L(\nu, \xi, \theta) \stackrel{\text{def}}{=} e^{-i\theta/2}L_- + e^{i\theta/2}L_+. \tag{A.16}$$

According to Eq. (A.15), $[L(\nu, \xi, \theta)]^2$ is the amplification matrix. Substituting Eq. (A.12) into Eq. (A.16), one has

$$L(\nu, \xi, \theta) = \begin{pmatrix} \dfrac{2[\xi\cos(\theta/2) - i\nu\sin(\theta/2)]}{1 + \xi} & \dfrac{(1 - \xi)e^{-i\theta/2}}{1 + \xi} \\ e^{i\theta/2} & 0 \end{pmatrix}. \tag{A.17}$$

The amplification factors $A_\pm$ given in Eq. (2.21) are the eigenvalues of the amplification matrix $[L(\nu, \xi, \theta)]^2$.

APPENDIX B: A SAMPLE PROGRAM FOR SOLVING Sod's SHOCK TUBE PROBLEM

```
      implicit real*8(a-h,o-z)
      dimension  q(3,1000), qn(3,1000), qx(3,1000), qt(3,1000),
     *           s(3,1000), vxl(3), vxr(3), xx(1000)
c
      it = 100
      dt = 0.4d-2
      dx = 0.1d-1
      ga = 1.4d0
      rhol = 1.d0
      ul = 0.d0
      pl = 1.d0
      rhor = 0.125d0
      ur = 0.d0
      pr = 0.1d0
      ia = 1
c
      hdt = dt/2.d0
      tt = hdt*dfloat(it)
      qdt = dt/4.d0
      hdx = dx/2.d0
      qdx = dx/4.d0
      dtx = dt/dx
      a1 = ga - 1.d0
      a2 = 3.d0 - ga
      a3 = a2/2.d0
      a4 = 1.5d0*a1
      q(1,1) = rhol
      q(2,1) = rhol*ul
      q(3,1) = pl/a1 + 0.5d0*rhol*ul**2
      itp = it + 1
      do 5 j = 1,itp
      q(1,j+1) = rhor
      q(2,j+1) = rhor*ur
      q(3,j+1) = pr/a1 + 0.5d0*rhor*ur**2
      do 5 i = 1,3
      qx(i,j) = 0.d0
5     continue
c
      open (unit=8,file='for008')
      write (8,10) tt,it,ia
      write (8,20) dt,dx,ga
      write (8,30) rhol,ul,pl
      write (8,40) rhor,ur,pr
c
      m = 2
      do 400 i = 1,it
      do 100 j = 1,m
      w2 = q(2,j)/q(1,j)
      w3 = q(3,j)/q(1,j)
      f21 = -a3*w2**2
      f22 = a2*w2
      f31 = a1*w2**3 - ga*w2*w3
      f32 = ga*w3 - a4*w2**2
      f33 = ga*w2
      qt(1,j) = -qx(2,j)
      qt(2,j) = -(f21*qx(1,j) + f22*qx(2,j) + a1*qx(3,j))
      qt(3,j) = -(f31*qx(1,j) + f32*qx(2,j) + f33*qx(3,j))
      s(1,j) = qdx*qx(1,j) + dtx*(q(2,j) + qdt*qt(2,j))
```

```
      s(2,j) = qdx*qx(2,j) + dtx*(f21*(q(1,j) + qdt*qt(1,j)) +
     *  f22*(q(2,j) + qdt*qt(2,j)) + a1*(q(3,j) + qdt*qt(3,j)))
      s(3,j) = qdx*qx(3,j) + dtx*(f31*(q(1,j) + qdt*qt(1,j)) +
     *  f32*(q(2,j) + qdt*qt(2,j)) + f33*(q(3,j) + qdt*qt(3,j)))
100   continue
      mm = m - 1
      do 200 j = 1,mm
      do 200 k =1,3
      qn(k,j+1) = 0.5d0*(q(k,j) + q(k,j+1) + s(k,j) - s(k,j+1))
      vxl(k) = (qn(k,j+1) - q(k,j) - hdt*qt(k,j))/hdx
      vxr(k) = (q(k,j+1) + hdt*qt(k,j+1) - qn(k,j+1))/hdx
      qx(k,j+1) = (vxl(k)*(dabs(vxr(k)))**ia + vxr(k)*(dabs(vxl(k)))
     *      **ia)/((dabs(vxl(k)))**ia + (dabs(vxr(k)))**ia + 1.d-60)
200   continue
      do 300 j = 2,m
      do 300 k = 1,3
      q(k,j) = qn(k,j)
300   continue
      m = m + 1
400   continue
c
      t2 = dx*dfloat(itp)
      xx(1) = -0.5d0*t2
      do 500 j = 1,itp
      xx(j+1) = xx(j) + dx
500   continue
      do 600 j = 1,m
      x = q(2,j)/q(1,j)
      z = a1*(q(3,j) - 0.5d0*x**2*q(1,j))
      write (8,50) xx(j),q(1,j),x,z
600   continue
c
      close (unit=8)
10    format(' t = ',g14.7,' it = ',i4,' ia = ',i4)
20    format(' dt = ',g14.7,' dx = ',g14.7,' gamma = ',g14.7)
30    format(' rhol = ',g14.7,' ul = ',g14.7,' pl = ',g14.7)
40    format(' rhor = ',g14.7,' ur = ',g14.7,' pr = ',g14.7)
50    format(' x =',f8.4,' rho =',g14.7,' u =',g14.7,' p =',g14.7)
      stop
      end
400   continue
c
      t2 = dx*dfloat(itp)
      xx(1) = -0.5d0*t2
      do 500 j = 1,itp
      xx(j+1) = xx(j) + dx
500   continue
      do 600 j = 1,m
      x = q(2,j)/q(1,j)
      z = a1*(q(3,j) - 0.5d0*x**2*q(1,j))
      write (8,50) xx(j),q(1,j),x,z
600   continue
c
      close (unit=8)
10    format(' t = ',g14.7,' it = ',i4,' ic = ',i4)
20    format(' dt = ',g14.7,' dx = ',g14.7,' gamma = ',g14.7)
30    format(' rhol = ',g14.7,' ul = ',g14.7,' pl = ',g14.7)
40    format(' rhor = ',g14.7,' ur = ',g14.7,' pr = ',g14.7)
50    format(' x =',f8.4,' rho =',g14.7,' u =',g14.7,' p =',g14.7)
      stop
      end
```

REFERENCES

1. S. C. Chang, and W. M. To, NASA TM 104495, August 1991 (unpublished).
2. S. C. Chang, "On An Origin of Numerical Diffusion: Violation of Invariance under Space-Time Inversion," in *Proceedings, 23rd Conference on Modeling and Simulation, April 30–May 1, 1992, Pittsburgh, PA*, edited by W. G. Vogt and M. H. Mickle Part 5, p. 2727, NASA TM 105776.
3. S. C. Chang and W. M. To, "A Brief Description of a New Numerical Framework for Solving Conservation Laws—The Method of Space-Time Conservation Element and Solution Element," in *Proceedings of the Thirteenth International Conference on Numerical Methods in Fluid Dynamics, Rome, Italy, 1992*, edited by M. Napolitano and F. Sabetta, Lecture Notes in Physics, Vol. 414, (Springer-Verlag, New York/Berlin, 1992), p. 396; NASA TM 105757.
4. J. R. Scott and S. C. Chang, *Int. J. Comput. Fluid Dynamics*, to appear.
5. S. C. Chang, NASA TM 106226, August 1993 (unpublished).
6. X. Y. Wang, C. Y. Chow, and S. C. Chang, NASA TM 106806, December 1994, *J. Comput. Phys.*, submitted.
7. S. C. Chang, C. Y. Wang, and C. Y. Chow, NASA TM 106758, December 1994 (unpublished).
8. X. Y. Wang, C. Y. Chow, and S. C. Chang, in preparation
9. J. R. Scott, *Int. J. Comput. Fluid Dynamics*, submitted.
10. L. H. Dill, A. Himansu, and J. R. Scott, in preparation.
11. B. D., Greenspan and J. R. Scott, in preparation.
12. D. A. Anderson, J. C. Tannehill, and R. H. Pletcher, *Computational Fluid Mechanics and Heat Transfer* (Hemisphere, Washington, DC/New York, 1984).
13. A. J. Baker, *Finite Element Computational Fluid Mechanics* (Hemisphere, Washington, DC/New York, 1983).
14. C. Canuto, M. Y. Hussaini, A. Quarteroni, and T. A. Zang, *Spectral Methods in Fluid Dynamics* (Springer-Verlag, New York, 1988).
15. M. Vinokur, *J. Comput. Phys.* **81,** 1 (1989).
16. R. J. LeVeque, *Numerical Methods for Conservation Laws* (Birkhäuser, Basel, 1990).
17. P. L. Roe, *J. Comput. Phys.* **43,** 357 (1981).
18. B. van Leer, *Lecture Notes in Physics*, Vol. 170, (Springer-Verlag, New York/Berlin, 1982), p. 501.
19. S. Osher and S. Chakravarthy, *J. Comput. Phys.* **50,** 447 (1983).
20. B. van Leer, *J. Comput. Phys.* **23,** 276 (1977).
21. M. J. Smith and R. W. Stoker, AIAA Paper 93-0150, Reno, Nevada, January 1993 (unpublished).
22. P. L. Roe, "A Survey of Upwind Differencing Techniques," in *Proceedings, Eleventh International Conference on Numerical Methods in Fluid Dynamics, 1988*, Lecture Notes in Physics, Vol. 323 (Springer-Verlag, New York/Berlin, 1989), p. 69.
23. R. F. Warming, R. M. Beam, and B. J. Hyett, *Math. Comput.* **29,** 1037 (1975).
24. V. V. Rusanov, *Zh. Vychisl. Mat. i Math. Fiz.* **3,** 508 (1963).
25. B. P., Leonard, NASA TM 100916, September 1988 (unpublished).
26. H. C. Yee, R. F. Warming, and A. Harten, AIAA Paper 83-1902 (unpublished).
27. H. Nessyahu and E. Tadmor, *J. Comput. Phys.* **87,** 408 (1990).
28. B. van Leer, *J. Compu. Phys.* **23,** 276 (1977).
29. G. D. van Albada, B. van Leer, and W. W. Roberts, *Astronom. Astrophys.* **108,** 76 (1982).
30. B. Noble and J. W. Daniel, *Linear Algebra and Its Applications*, 2nd ed. (Prentice–Hall, Englewood Cliffs, NJ, 1977).
31. G. A. Sod, *J. Comput. Phys.* **27,** 1 (1978).