18th AIAA Computational Fluid Dynamics Conference, 27 June 2007, Miami, FL AIAA 2007-4321

THE $a(3)$ SCHEME—A FOURTH-ORDER NEUTRALLY STABLE CESE SOLVER

Sin-Chung Chang
NASA Glenn Research Center, Cleveland, OH 44135
E-mail: sin-chung.chang@nasa.gov

The CESE development is driven by a belief that a solver should (i) enforce conservation laws in both space and time, and (ii) be built from a non-dissipative (i.e., neutrally stable) core scheme so that the numerical dissipation can be controlled effectively. To provide a solid foundation for a systematic CESE development of high order schemes, in this paper we describe a new 4th-order neutrally stable CESE solver of the advection equation $\partial u/\partial t + a\partial u/\partial x = 0$. The space-time stencil of this two-level explicit scheme is formed by one point at the upper time level and three points at the lower time level. Because it is associated with three independent mesh variables u_j^n, $(u_x)_j^n$, and $(u_{xx})_j^n$ (the numerical analogues of u, $\partial u/\partial x$, and $\partial^2 u/\partial x^2$, respectively) and four equations per mesh point, the new scheme is referred to as the $a(3)$ scheme. As in the case of other similar CESE neutrally stable solvers, the $a(3)$ scheme enforces conservation laws in space-time locally and globally, and it has the basic, forward marching, and backward marching forms. These forms are equivalent and satisfy a space-time inversion (STI) invariant property which is shared by the advection equation. Based on the concept of STI invariance, a set of algebraic relations is developed and used to prove that the $a(3)$ scheme must be neutrally stable when it is stable. Moreover it is proved rigorously that all three amplification factors of the $a(3)$ scheme are of unit magnitude for all phase angles if $|\nu| \leq 1/2$ ($\nu = a\triangle t/\triangle x$). This theoretical result is consistent with the numerical stability condition $|\nu| < 1/2$. Through numerical experiments, it is established that the $a(3)$ scheme generally is (i) 4th-order accurate for the mesh variables u_j^n and $(u_x)_j^n$; and 2nd-order accurate for $(u_{xx})_j^n$. However, in some exceptional cases, the scheme can achieve perfect accuracy aside from round-off errors.

1. Introduction

The space-time conservation element and solution element (CESE) method is a high-resolution and genuinely multidimensional method for solving conservation laws [1–73]. Its nontraditional features include: (i) a unified treatment of space and time; (ii) the introduction of conservation elements (CEs) and solution elements (SEs) as the vehicles for enforcing space-time flux conservation; (iii) a novel time marching strategy that has a space-time staggered stencil at its core and, as such, fluxes at an interface can be evaluated without using any interpolation or extrapolation procedure (which, in turn, leads to the method's ability to capture shocks without using Riemann solvers); (iv) the requirement that each scheme be built from a non-dissipative core scheme and, as a result, the numerical dissipation can be controlled effectively; and (v) the fact that mesh values of the physical dependent variables and their spatial derivatives are considered as independent marching variables to be solve for simultaneously. Note that CEs are non-overlapping space-time subdomains introduced such that (i) the computational domain can be filled by these subdomains; and (ii) flux conservation can be enforced over each of them and also over the union of any combination of them. On the other hand, SEs are space-time subdomains introduced such that (i) the boundary of each CE can be divided into several component parts with each of them belonging to a unique SE; and (ii) within a SE, any physical flux vector is approximated using simple smooth functions. In general, a CE does not coincide with a SE.

Without using flux-splitting or other special techniques, since its inception in 1991 [1] the unstructured-mesh compatible CESE method has been used to obtain numerous accurate 1D, 2D and 3D steady and unsteady flow solutions with Mach numbers ranging from 0.0028 to 10 [51]. The physical phenomena

modeled include traveling and interacting shocks, acoustic waves, vortex shedding, viscous flows, detonation waves, cavitation, flows in fluid film bearings, heat conduction with melting and/or freezing, electrodynamics, MHD vortex, hydraulic jump, crystal growth, and chromatographic problems [3–73]. In particular, the rather unique capability of the CESE method to resolve both strong shocks and small disturbances (e.g., acoustic waves) simultaneously [13,15,16] makes it an effective tool for attacking computational aeroacoustics (CAA) problems. Note that the fact that second-order CESE schemes can solve CAA problems accurately is an exception to the commonly-held belief that a second-order scheme is not adequate for solving CAA problems. Also note that, while numerical dissipation is needed for shock capturing, it may also result in annihilation of small disturbances. Thus a solver that can handle both strong shocks and small disturbances simultaneously must be able to overcome this difficulty.

In spite of its nontraditional features and potent capabilities, the core ideas of the CESE method are simple. In fact, all of its key features are the inescapable results of an honest pursuit driven by these simple ideas. The first and foremost is the belief that the method must be solid in physics. As such, in the CESE development, conservation laws are enforced locally and globally in their natural space-time unity forms for 1D, 2D and 3D cases. Moreover, because *direct* physical interaction generally occurs only among the immediate neighbors, use of the simplest stencil also becomes a CESE requirement. Obviously, this requirement is also very helpful in simplifying boundary-condition implementation.

The second idea emerges from the realization that stability and accuracy are two competing issues in time-accurate computations, i.e., too much numerical dissipation would degrade accuracy while too little of it will cause instability. In other words, to meet both accuracy and stability requirements, computation must be performed away from the edge ("cliff") of instability but not too far from it. This represents a real dilemma in numerical method development. As an example, schemes with high-order accuracy generally has high accuracy and low numerical dissipation. However, it is susceptible to instability. In fact, in dealing with complicated real-world problems, stability of these schemes often is difficult to maintain without resorting to ad hoc treatments. To confront this issue head-on, in CESE development, it is required that a solver be built from a non-dissipative (i.e., neutrally stable) core scheme. By definition, computations involving a neutrally stable scheme are performed right on the edge of instability and therefore the numerical results generated are non-dissipative. As such numerical dissipation can be controlled effectively if the deviation of a solver from its non-dissipative core scheme can be adjusted using some built-in parameters. Note that the above idea also plays an essential role in the recent successful development of a family of Courant number insensitive schemes [59,61,64,65,67].

Other CESE ideas are: (i) the flux at an interface be evaluated in a simple and consistent manner; (ii) genuinely multidimensional schemes be built as simple, consistent and straightforward extensions of 1D schemes; (iii) triangular and tetrahedral meshes be used in 2D and 3D cases, respectively, so that the method is compatible to the simplest unstructured meshes and thus can be used to solve problems with complex geometries; and (iv) logical structures and approximation techniques used be as simple as possible, and special techniques that has only limited applicability and may cause undesirable side effects be avoided. Fortunately for the CESE development, as it turns out, the realization of the above lesser ideas (i)–(iv) follows effortlessly from that of the first two core ideas.

The first model equation considered in the CESE development is the simple convection equation

$$\frac{\partial u}{\partial t} + a\frac{\partial u}{\partial x} = 0 \tag{1.1}$$

where the advection speed $a \neq 0$ is a constant. Let $x_1 = x$, and $x_2 = t$ be considered as the coordinates of a two-dimensional Euclidean space E_2. Then, because Eq. (1.1) can be expressed as $\nabla \cdot \vec{h} = 0$ with $\vec{h} \stackrel{\text{def}}{=} (au, u)$, Gauss' divergence theorem in the space-time E_2 implies that Eq. (1.1) is the differential form of the integral conservation law

$$\oint_{S(V)} \vec{h} \cdot d\vec{s} = 0 \tag{1.2}$$

As depicted in Fig. 1, here (i) $S(V)$ is the boundary of an arbitrary *space-time* region V in E_2, and (ii) $d\vec{s} = d\sigma\, \vec{n}$ with $d\sigma$ and $\vec{n}$, respectively, being the area and the unit outward normal vector of a surface element

on $S(V)$. Note that: (i) because $\vec{h} \cdot d\vec{s}$ is the *space-time* flux of $\vec{h}$ leaving the region V through the surface element $d\vec{s}$, Eq. (1.2) simply states that the total *space-time* flux of $\vec{h}$ leaving V through $S(V)$ vanishes; (ii) in E_2, $d\sigma$ is the length of a line segment on the simple closed curve $S(V)$; and (iii) all mathematical operations can be carried out as though E_2 were an ordinary two-dimensional Euclidean space.

It is well known that a solution to Eq. (1.1) represents *non-dissipative* data propagation along its characteristic lines defined by $dx/dt = a$. Moreover, Eq. (1.1) is invariant under space-time inversion (STI), i.e., it transforms back to itself if x and t are replaced by $-x$ and $-t$, respectively. (In physics, STI invariance generally is referred to as PT invariance where P denotes a mirror-image or spatial-reflection operation while T denotes a time-reversal operation). Thus a solution to Eq. (1.1) possesses the following properties: (i) it is completely determined by the data specified at an initial time level; (ii) its value at a space-time point has a finite domain of dependence (a point) at the initial time level; and (iii) the space-time inversion image of a solution to Eq. (1.1) is also a solution and vice versa. As such, in the initial CESE development, the focus is on the construction of an ideal core solver of Eq. (1.1) that enforces the conservation law Eq. (1.2) and also possesses all other properties of Eq. (1.1), i.e., it is a two-level, explicit, non-dissipative, and STI invariant solver. An in-depth account of this development and the resulting "a" scheme is given in [71]. As it turns out, the 2nd-order accurate a scheme (i) has a space-time stencil formed by one mesh point at the upper time level and two mesh points at the lower time level; and (ii) it is neutrally stable if $\nu^2 < 1$ where $\nu = a\Delta t/\Delta x$. Also, at each space-time mesh point (j, n), the a scheme is associated with two independent mesh variables u_j^n and $(u_x)_j^n$ (the numerical analogues of u and $\partial u/\partial x$, respectively) and two equations.

Until recently, with one exception (a three-level and 3rd-order accurate scheme reported on p. 80 of [1]), all CESE solvers of Eq. (1.1) are two-level and 2nd-order accurate extensions of the a scheme. To initiate a systematic CESE development of high-order schemes, in this paper we describe a new 4th-order accurate, conservation-law enforcing, and neutrally stable CESE solver of Eq. (1.1). As will be shown, the space-time stencil of this two-level explicit scheme is formed by one point at the upper time level and three points at the lower time level. Because it is associated with three independent mesh variables u_j^n, $(u_x)_j^n$ and $(u_{xx})_j^n$ (the numerical analogues of u, $\partial u/\partial x$, and $\partial^2 u/\partial x^2$, respectively) and three equations at each mesh point, hereafter the new scheme is referred to as the $a(3)$ scheme.

2. The $a(3)$ scheme

To proceed, consider the set Ω of space-time mesh points (j, n) (marked by dots and crosses in Fig. 2(a)) where

$$\Omega \stackrel{\text{def}}{=} \{(j, n) | j, n = 0, \pm 1, \pm 2, \pm 3, \ldots\} \tag{2.1}$$

We have

$$\Omega = \Omega_1 \cup \Omega_2 \tag{2.2}$$

where Ω_1 and Ω_2 are two disjoint sets defined by

$$\Omega_1 \stackrel{\text{def}}{=} \{(j, n) | j, n = 0, \pm 1, \pm 2, \pm 3, \ldots, \text{ and } (j + n) \text{ is an odd integer}\} \tag{2.3}$$

$$\Omega_2 \stackrel{\text{def}}{=} \{(j, n) | j, n = 0, \pm 1, \pm 2, \pm 3, \ldots, \text{ and } (j + n) \text{ is an even integer}\} \tag{2.4}$$

In Fig. 2(a), the mesh points $\in \Omega_1$ are marked by dots while those $\in \Omega_2$ are marked by crosses. Hereafter Ω_2 is referred to as the complement set of Ω_1 and vice versa. Obviously each of Ω_1 and Ω_2 represents a set of space-time staggered mesh points.

Each $(j, n) \in \Omega$ is associated with (i) a solution element (SE), denoted by $\text{SE}(j, n)$ (see Fig. 2(b) where $(j, n) \in \Omega_1$ is assumed), and (ii) two conservation elements (CEs), denoted by $\text{CE}_-(j, n)$ and $\text{CE}_+(j, n)$ (see Figs. 2(c) and 2(d) where $(j, n) \in \Omega_1$ is assumed), respectively. Each SE is the *interior* of a *space-time* region that includes a horizontal line segment, a vertical line segment, and their immediate neighborhood. On the other hand, each CE is a rectangular space-time region. Hereafter, (i) SEs or CEs associated with mesh points $\in \Omega_1$ ($\in \Omega_2$) may be referred to simply as SEs or CEs associated with Ω_1 (Ω_2).

As a preliminary for the following development, note that (see Figs. 2(a)–(d)):

(a) Two CEs which are associated with two mesh points, one $\in \Omega_1$ while another $\in \Omega_2$ may occupy the same space-time region. As an example, (i) $\mathrm{CE}_-(j,n)$ and $\mathrm{CE}_+(j-1,n)$ occupy the same space-time region; and (ii) $(j,n) \in \Omega_1 \Leftrightarrow (j-1,n) \in \Omega_2$. Hereafter the symbol"$\Leftrightarrow$" is used as a shorthand for the statement "if and only if".

(b) A pair of diagonally opposite vertices of a CE both belong to the same set Ω_1 or Ω_2 while another pair both belong to the complement set. As an example, points A and C belong to Ω_1 while points B and D belong to Ω_2.

(c) The CEs associated with each of Ω_1 and Ω_2 by themselves are nonoverlapping and can fill the space-time E_2.

(d) Among the line segments forming the boundary of the same space-time region occupied by both $\mathrm{CE}_-(j,n)$ and $\mathrm{CE}_+(j-1,n)$, (i) $\overline{AB}$ and $\overline{AD} \subset \mathrm{SE}(j,n)$; (ii) $\overline{CB}$ and $\overline{CD} \subset \mathrm{SE}(j-1,n-1)$; (iii) $\overline{BA}$ and $\overline{BC} \subset \mathrm{SE}(j-1,n)$; and (iv) $\overline{DA}$ and $\overline{DC} \subset \mathrm{SE}(j,n-1)$. Because $\overline{AB}$ and $\overline{BA}$ represent the same line segment, one can see that any line segment on this boundary is a subset of two SEs with one of them being associated with Ω_1 and another associated with Ω_2. *Hereafter, this ambiguity is removed by the following SE designation rule: any line segment designated as a boundary of a CE associated with Ω_1 (Ω_2) is designated as a subset of a SE associated with Ω_1 (Ω_2).* As an example, if $\overline{AB}$, $\overline{AD}$, $\overline{CB}$, and $\overline{CD}$ are designated as boundaries of $\mathrm{CE}_-(j,n)$, then because points A and C belong to Ω_1, the above rule implies that: (i) both $\overline{AB}$ and $\overline{AD}$ are designated as subsets of $\mathrm{SE}(j,n)$; and (ii) both $\overline{CB}$ and $\overline{CD}$ are designated as subsets of $\mathrm{SE}(j-1,n-1)$. On the other hand, if $\overline{BA}$, $\overline{BC}$, $\overline{DA}$, and $\overline{DC}$ are designated as boundaries of $\mathrm{CE}_+(j-1,n)$, then: (i) both $\overline{BA}$ and $\overline{BC}$ are designated as subsets of $\mathrm{SE}(j-1,n)$; and (ii) both $\overline{DA}$ and $\overline{DC}$ are designated as subsets of $\mathrm{SE}(j,n-1)$.

Let $(x,t) \in \mathrm{SE}(j,n)$. Then Eqs. (1.1) and (1.2) will be simulated numerically assuming that $u(x,t)$ and $\vec{h}(x,t)$, respectively, are approximated by

$$u^*(x,t\,;j,n) \stackrel{\text{def}}{=} u_j^n + (u_x)_j^n (x-x_j) + (u_t)_j^n (t-t^n) + \frac{1}{2}(u_{xx})_j^n (x-x_j)^2 + (u_{xt})_j^n (x-x_j)(t-t^n) + \frac{1}{2}(u_{tt})_j^n (t-t^n)^2 \tag{2.5}$$

and

$$\vec{h}^*(x,t\,;j,n) \stackrel{\text{def}}{=} \big(a u^*(x,t\,;j,n),\, u^*(x,t\,;j,n)\big) \tag{2.6}$$

Note that: (i) u_j^n, $(u_x)_j^n$, $(u_t)_j^n$, $(u_{xx})_j^n$, $(u_{xt})_j^n$, and $(u_{tt})_j^n$ are constants in $\mathrm{SE}(j,n)$, and the numerical analogues of the values of u, $\partial u/\partial x$, $\partial u/\partial t$, $\partial^2 u/\partial x^2$, $\partial^2 u/\partial x \partial t$, and $\partial^2 u/\partial t^2$ at the mesh point (j,n), respectively; (ii) (x_j, t^n) are the coordinates of the mesh point (j,n) where $x_j = j\Delta x$ and $t^n = n\Delta t$; (iii) $u^*(x,t\,;j,n)$ represents a 2nd-order Taylor's approximation of u; and (iv) Eq. (2.6) is the numerical analogy of the definition $\vec{h} = (au, u)$.

For any $(j,n) \in \Omega$, let $u = u^*(x,t\,;j,n)$ satisfy Eq. (1.1) for all $(x,t) \in \mathrm{SE}(j,n)$. Then one has

$$(u_t)_j^n = -a(u_x)_j^n, \quad (u_{xt})_j^n = -a(u_{xx})_j^n, \quad \text{and} \quad (u_{tt})_j^n = a^2(u_{xx})_j^n, \qquad (j,n) \in \Omega \tag{2.7}$$

Substituting Eq. (2.7) into Eq. (2.5), one has

$$u^*(x,t\,;j,n) = u_j^n + (u_x)_j^n \left[(x-x_j) - a\,(t-t^n)\right] + \frac{1}{2}(u_{xx})_j^n \left[(x-x_j) - a\,(t-t^n)\right]^2, \quad (j,n) \in \Omega \tag{2.8}$$

i.e., u_j^n, $(u_x)_j^n$, and $(u_{xx})_j^n$ are the only independent mesh variables associated with (j,n).

With the above preliminaries, next we derive the flux conservation relations that underline the $a(3)$ scheme.

2.1. Flux conservation relations

Let the flux of $\vec{h}^*$ conserve over all CEs, i.e.,

$$\oint_{S(CE_-(j,n))} \vec{h}^* \cdot d\vec{s} = 0, \qquad (j,n) \in \Omega \tag{2.9}$$

and

$$\oint_{S(CE_+(j,n))} \vec{h}^* \cdot d\vec{s} = 0, \qquad (j,n) \in \Omega \tag{2.10}$$

Because (i) with respect to $\mathrm{CE}_-(j,n)$, the outward unit normal vectors $\vec{n}$ at $\overline{AB}$, $\overline{AD}$, $\overline{CD}$, and $\overline{CB}$ are $(0,1)$, $(1,0)$, $(0,-1)$, and $(-1,0)$, respectively; and (ii) with respect to $\mathrm{CE}_+(j,n)$, the vectors $\vec{n}$ at $\overline{AF}$, $\overline{AD}$, $\overline{ED}$, and $\overline{EF}$ are $(0,1)$, $(-1,0)$, $(0,-1)$, and $(1,0)$, respectively, by using (i) the definitions given following Eq. (1.2), (ii) the above SE designation rule, and (iii) Eqs. (2.6) and (2.8), it can be shown that Eqs. (2.9) and (2.10) are equivalent to

$$(1+\nu)\left[u-(1-\nu)u_{\bar{x}}+\frac{2(1-\nu+\nu^2)}{3}u_{\bar{x}\bar{x}}\right]_j^n=(1+\nu)\left[u+(1-\nu)u_{\bar{x}}+\frac{2(1-\nu+\nu^2)}{3}u_{\bar{x}\bar{x}}\right]_{j-1}^{n-1}, \quad (j,n)\in\Omega \tag{2.11}$$

and

$$(1-\nu)\left[u+(1+\nu)u_{\bar{x}}+\frac{2(1+\nu+\nu^2)}{3}u_{\bar{x}\bar{x}}\right]_j^n=(1-\nu)\left[u-(1+\nu)u_{\bar{x}}+\frac{2(1+\nu+\nu^2)}{3}u_{\bar{x}\bar{x}}\right]_{j+1}^{n-1}, \quad (j,n)\in\Omega \tag{2.12}$$

respectively. Here: (i) $\nu \stackrel{\text{def}}{=} a\Delta t/\Delta x$ is the Courant number; (ii)

$$(u_{\bar{x}})_j^n \stackrel{\text{def}}{=} \frac{\Delta x}{2}(u_x)_j^n \quad \text{and} \quad (u_{\bar{x}\bar{x}})_j^n \stackrel{\text{def}}{=} \frac{(\Delta x)^2}{4}(u_{xx})_j^n \tag{2.13}$$

and (iii) to simplify notation, in the above and hereafter we adopt a convention that can be explained using an expression on the left side of Eq. (2.11) as an example, i.e.,

$$\left[u+(1+\nu)u_{\bar{x}}+\frac{2(1+\nu+\nu^2)}{3}u_{\bar{x}\bar{x}}\right]_j^n=u_j^n+(1+\nu)(u_{\bar{x}})_j^n+\frac{2(1+\nu+\nu^2)}{3}(u_{\bar{x}\bar{x}})_j^n$$

At this juncture, note that:

(a) Because

$$\frac{\partial u}{\partial \bar{x}}=\frac{\Delta x}{2}\frac{\partial u}{\partial x} \quad \text{and} \quad \frac{\partial^2 u}{\partial \bar{x}^2}=\frac{(\Delta x)^2}{4}\frac{\partial^2 u}{\partial x^2} \qquad \text{if} \qquad \bar{x}\stackrel{\text{def}}{=}\frac{x}{\Delta x/2}$$

the normalized parameters $(u_{\bar{x}})_j^n$ and $(u_{\bar{x}\bar{x}})_j^n$, respectively, can be interpreted as the numerical analogues of the values at (j,n) of the first and second derivatives of u with respect to the normalized coordinate $\bar{x}$.

(b) By definition, points B and D depicted in Fig. 2(c) do not belong to either $\mathrm{SE}(j,n)$ or $\mathrm{SE}(j-1,n-1)$. This fact, however, does not pose a problem for flux evaluation over $S(CE_-(j,n))$ because the values of $\vec{h}^*$ at isolated points do not contribute to the flux of $\vec{h}^*$ over a finite line segment. Similarly, the fact that points D and F depicted in Fig. 2(d) do not belong to $\mathrm{SE}(j,n)$ and $\mathrm{SE}(j+1,n-1)$ does not pose a problem for flux evaluation over $S(CE_+(j,n))$.

(c) According to the SE designation rule, each line segment such as $\overline{AB}$ depicted in Fig. 2(c) can be assigned with two different fluxes of $\vec{h}^*$, one is associated with Ω_1 (hereafter referred to as the Ω_1-flux) and another associated with Ω_2 (hereafter referred to as the Ω_2-flux). As such, among those local conservation relations Eqs. (2.9) and (2.10), those associated with $(j,n)\in\Omega_1$ are completely decoupled from those associated with $(j,n)\in\Omega_2$. Because Eqs. (2.9) and (2.10) are equivalent to Eqs. (2.11) and (2.12), respectively, it follows that each of the two systems of equations defined by Eqs. (2.11) and (2.12) is formed by two decoupled subsystems, one is associated with Ω_1 while another associated with Ω_2.

(d) Moreover, because (i) the vector $\vec{h}^*$ at any interface separating two neighboring CEs associated with the same set Ω_1 (Ω_2) is evaluated using the information from the same SE, and (ii) the unit outward normal vector on the surface element pointing outward from one of these two neighboring CEs is exactly the

negative of that pointing outward from another CE, one concludes that the flux leaving one of these CEs through the interface is the negative of that leaving another CE through the same interface. Due to this interface flux cancelation and the fact that the CEs associated with each of Ω_1 and Ω_2 by themselves are nonoverlapping and can fill the space-time E_2, the local conservation relations Eqs. (2.9) and (2.10) associated with $(j,n) \in \Omega_1$ ($(j,n) \in \Omega_2$) lead to a global conservation relation, i.e., *the total Ω_1- (Ω_2-) flux of $\vec{h}^*$ leaving the boundary of any space-time region that is the union of any combination of CEs associated with the same set Ω_1 (Ω_2) vanishes.*

Let $1-\nu^2 \neq 0$, i.e. $1+\nu \neq 0$ and $1-\nu \neq 0$. Then Eqs. (2.11) and (2.12) reduce to

$$\left[u-(1-\nu)u_{\bar{x}}+\frac{2(1-\nu+\nu^2)}{3}u_{\bar{x}\bar{x}}\right]_j^n = \left[u+(1-\nu)u_{\bar{x}}+\frac{2(1-\nu+\nu^2)}{3}u_{\bar{x}\bar{x}}\right]_{j-1}^{n-1}, \quad (j,n)\in\Omega \tag{2.14}$$

and

$$\left[u+(1+\nu)u_{\bar{x}}+\frac{2(1+\nu+\nu^2)}{3}u_{\bar{x}\bar{x}}\right]_j^n = \left[u-(1+\nu)u_{\bar{x}}+\frac{2(1+\nu+\nu^2)}{3}u_{\bar{x}\bar{x}}\right]_{j+1}^{n-1}, \quad (j,n)\in\Omega \tag{2.15}$$

respectively. Obviously, each of the two systems of equations defined by Eqs. (2.14) and (2.15) is also formed by two decoupled subsystems. Moreover, each component equation in Eq. (2.14) represents a stronger condition than the corresponding equation in Eq. (2.11) in the sense that the former implies the latter for any given ν while the latter implies the former only if an extra condition (i.e., $\nu \neq -1$ for this case) is imposed. Similarly, each component equation in Eq. (2.15) represents a stronger condition than the corresponding equation in Eq. (2.12). These stronger conditions will be used in the construction of the $a(3)$ scheme.

As a preliminary to a later development, next we will take a side tour and introduce the concept of invariance under space-time inversion.

2.2. Invariance under space-time inversion

Let $u = u(x,t)$ be a solution to Eq. (1.1) in the domain $-\infty < x,t < +\infty$, i.e.,

$$\frac{\partial u(x,t)}{\partial t} + a\frac{\partial u(x,t)}{\partial x} \equiv 0, \qquad -\infty < x,t < +\infty \tag{2.16}$$

Let

$$x' \stackrel{\text{def}}{=} -x \quad \text{and} \quad t' \stackrel{\text{def}}{=} -t \tag{2.17}$$

and

$$\hat{u}(x,t) \stackrel{\text{def}}{=} u(-x,-t) \tag{2.18}$$

Then (i) Eq. (2.16) $\Leftrightarrow$

$$\frac{\partial u(x',t')}{\partial t'} + a\frac{\partial u(x',t')}{\partial x'} \equiv 0, \qquad -\infty < x',t' < +\infty \tag{2.19}$$

and (ii)

$$\frac{\partial}{\partial t'} = -\frac{\partial}{\partial t} \quad \text{and} \quad \frac{\partial}{\partial x'} = -\frac{\partial}{\partial x} \tag{2.20}$$

Thus Eq. (2.16) $\Leftrightarrow$

$$\frac{\partial \hat{u}(x,t)}{\partial t} + a\frac{\partial \hat{u}(x,t)}{\partial x} \equiv 0, \qquad -\infty < x,t < +\infty \tag{2.21}$$

In other words, if $u = u(x,t)$ is a solution to Eq. (1.1), so must be $u = \hat{u}(x,t)$ and vice versa. Because the one-to-one mapping

$$(x,t) \leftrightarrow (-x,-t), \qquad -\infty < x,t < +\infty \tag{2.22}$$

represents a space-time inversion (STI) operation, hereafter (i) a pair of functions such as u and $\hat{u}$ will be referred to as the STI images of each other; and (ii) a partial differential equation (PDE) such as Eq. (1.1) is said to be STI invariant if the STI image of a solution is also a solution and vice versa.

Next let

$$u^{(k,\ell)}(x,t) \stackrel{\text{def}}{=} \frac{\partial^{k+\ell} u(x,t)}{\partial x^k \partial t^\ell} \quad \text{and} \quad \hat{u}^{(k,\ell)}(x,t) \stackrel{\text{def}}{=} \frac{\partial^{k+\ell} \hat{u}(x,t)}{\partial x^k \partial t^\ell}, \qquad -\infty < x,t < +\infty;\ k,\ell = 0,1,2,\ldots \tag{2.23}$$

Then, with the aid of the chain rule, Eqs. (2.17), (2.18), and (2.23) imply that

$$\begin{aligned} \hat{u}^{(k,\ell)}(x,t) &= \frac{\partial^{k+\ell} u(-x,-t)}{\partial x^k \partial t^\ell} = (-1)^{k+\ell} \frac{\partial^{k+\ell} u(x',t')}{\partial x'^k \partial t'^\ell} \qquad -\infty < x,t < +\infty;\ k,\ell = 0,1,2,\ldots \\ &= (-1)^{k+\ell} u^{(k,\ell)}(x',t') = (-1)^{k+\ell} u^{(k,\ell)}(-x,-t) \end{aligned} \tag{2.24}$$

i.e.,

$$\hat{u}^{(k,\ell)}(x,t) = \begin{cases} u^{(k,\ell)}(-x,-t) & \text{if } (k+\ell) \text{ is even} \\ -u^{(k,\ell)}(-x,-t) & \text{if } (k+\ell) \text{ is odd} \end{cases} \tag{2.25}$$

According to Eq. (2.23), $u^{(0,0)} = u$ and $\hat{u}^{(0,0)} = \hat{u}$. Thus Eq. (2.18) is a special case of Eq. (2.24) with $k = \ell = 0$.

In the following, the concept of STI invariance will be introduced for the $a(3)$ scheme. As a preliminary, note that: (i)

$$(j,n) \leftrightarrow (-j,-n) \tag{2.26}$$

is the numerical analogue of the STI mapping Eq. (2.22); and (ii) u_j^n, $(u_x)_j^n$, $(u_t)_j^n$, $(u_{xx})_j^n$, $(u_{xt})_j^n$, and $(u_{tt})_j^n$ are the numerical analogues of the values of u, $\partial u/\partial x$, $\partial u/\partial t$, $\partial^2 u/\partial x^2$, $\partial^2 u/\partial x \partial t$, and $\partial^2 u/\partial t^2$, at the mesh point (j,n), respectively. Thus, motivated by Eq. (2.25), the one-to-one mapping

$$\begin{aligned} &u_j^n \leftrightarrow u_{-j}^{-n}; \quad (u_x)_j^n \leftrightarrow -(u_x)_{-j}^{-n}; \quad (u_t)_j^n \leftrightarrow -(u_t)_{-j}^{-n} \\ &(u_{xx})_j^n \leftrightarrow (u_{xx})_{-j}^{-n}; \quad (u_{xt})_j^n \leftrightarrow (u_{xt})_{-j}^{-n}; \quad (u_{tt})_j^n \leftrightarrow (u_{tt})_{-j}^{-n} \end{aligned} \qquad (j,n) \in \Omega \tag{2.27}$$

is taken as the numerical analogue of the one-to-one mapping

$$u^{(k,\ell)}(x,t) \leftrightarrow \hat{u}^{(k,\ell)}(x,t), \qquad -\infty < x,t < +\infty;\ k,\ell = 0,1,2,3 \tag{2.28}$$

For the independent mesh variables, by using Eq. (2.13), Eq. (2.27) reduces to

$$\begin{pmatrix} u_j^n \\ (u_{\bar{x}})_j^n \\ (u_{\bar{x}\bar{x}})_j^n \end{pmatrix} \leftrightarrow \begin{pmatrix} u_{-j}^{-n} \\ -(u_{\bar{x}})_{-j}^{-n} \\ (u_{\bar{x}\bar{x}})_{-j}^{-n} \end{pmatrix}, \qquad (j,n) \in \Omega \tag{2.29}$$

Eq. (2.29) can be expressed as

$$\vec{q}(j,n) \leftrightarrow U\vec{q}(-j,-n), \qquad (j,n) \in \Omega \tag{2.30}$$

where

$$\vec{q}(j,n) \stackrel{\text{def}}{=} \begin{pmatrix} u_j^n \\ (u_{\bar{x}})_j^n \\ (u_{\bar{x}\bar{x}})_j^n \end{pmatrix}, \qquad (j,n) \in \Omega \tag{2.31}$$

and

$$U \stackrel{\text{def}}{=} \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{2.32}$$

The matrix U is unitary. In fact it is a real matrix with

$$U = U^{-1} \tag{2.33}$$

Hereafter (i) M^{-1} denotes the inverse of any nonsingular square matrix M; (ii) for each (j,n), $U\vec{q}(-j,-n)$ is referred to as the STI image of $\vec{q}(j,n)$; and (iii) the set formed by $U\vec{q}(-j,-n)$, $(j,n) \in \Omega$ is also referred to as the image of the set formed by $\vec{q}(j,n)$, $(j,n) \in \Omega$. According to Eq. (2.33), $\vec{q}(j,n) = UU\vec{q}(-(-j),-(-n))$. Thus $\vec{q}(j,n)$ is the STI image of $U\vec{q}(-j,-n)$ as an individual (j,n) or as the set defined over Ω. In the following, we will show that by itself each of the four subsystems of equations associated with Eqs. (2.14) and (2.15) is STI invariant, i.e., *the subsystem maps onto an equivalent subsystem under the mapping Eq. (2.29)*.

As an example, consider the subsystem of equations formed by the component equations associated with Ω_1 in Eq. (2.14). Let it be denoted as Eq. (2.14a). Under the mapping Eq. (2.29), Eq. (2.14a) maps onto

$$\left[u + (1-\nu)u_{\bar{x}} + \frac{2(1-\nu+\nu^2)}{3}u_{\bar{x}\bar{x}}\right]_{-j}^{-n} = \left[u - (1-\nu)u_{\bar{x}} + \frac{2(1-\nu+\nu^2)}{3}u_{\bar{x}\bar{x}}\right]_{-(j-1)}^{-(n-1)}, \quad (j,n) \in \Omega_1 \tag{2.34}$$

At this juncture, note that, in addition to changing the sign of each $u_{\bar{x}}$, mapping Eq. (2.29) requires that the upper and lower indices j, n, $j-1$, and $n-1$ in Eq. (2.14a) be replaced by their negatives, respectively. *This is different from simply replacing the symbols j and n everywhere with $-j$ and $-n$, respectively*. Moreover, to simplify argument, hereafter system B is referred to as the STI image of system A if A maps onto B under the mapping Eq. (2.29), e.g., the subsystem Eq. (2.34) is the STI image of Eq. (2.14a). Let

$$j^* \stackrel{\text{def}}{=} 1-j \quad \text{and} \quad n^* \stackrel{\text{def}}{=} 1-n, \qquad (j,n) \in \Omega_1 \tag{2.35}$$

Then, by using the fact that $(j^*+n^*)+(j+n) \equiv 2$ and therefore $(j^*,n^*) \in \Omega_1 \Leftrightarrow (j,n) \in \Omega_1$, Eq. (2.34) can be cast into the form

$$\left[u - (1-\nu)u_{\bar{x}} + \frac{2(1-\nu+\nu^2)}{3}u_{\bar{x}\bar{x}}\right]_{j^*}^{n^*} = \left[u + (1-\nu)u_{\bar{x}} + \frac{2(1-\nu+\nu^2)}{3}u_{\bar{x}\bar{x}}\right]_{j^*-1}^{n^*-1}, \quad (j^*,n^*) \in \Omega_1 \tag{2.36}$$

By comparing Eqs. (2.14a) and (2.36), one can see that the subsystem Eq. (2.14a) is identical to its STI image Eq. (2.34) (which is identical to Eq. (2.36)). Thus, under the mapping Eq. (2.29), Eq. (2.14a) maps onto itself, i.e., the subsystem Eq. (2.14a) is STI invariant. QED.

The STI invariance of another three subsystems associated with Eqs. (2.14) and (2.15) can be established in a similar manner. As such the system formed by all component equations in each of Eqs. (2.14) and (2.15) is STI invariant.

The three mesh variables at any $(j,n) \in \Omega$ are linked to those at $(j-1,n-1)$ and $(j+1,n-1)$ by two component equations in Eqs. (2.14) and (2.15), respectively. In order that the three mesh variables at (j,n) can be determined in terms of those mesh variables at the $(n-1)$th time level, in the next subsection we introduce an extra STI invariant condition that links the mesh variables at (j,n) with those at the mesh point $(j,n-1)$.

2.3. A family of STI invariant solvers

Consider the following system of equations:

$$[u + \alpha u_{\bar{x}} + \beta u_{\bar{x}\bar{x}}]_j^n = [u - \alpha u_{\bar{x}} + \beta u_{\bar{x}\bar{x}}]_j^{n-1}, \qquad (j,n) \in \Omega \tag{2.37}$$

where α and β are parameters independent of (j,n). By definition, $(j,n) \in \Omega_1$ (Ω_2) $\Leftrightarrow (j,n-1) \in \Omega_2$ (Ω_1). Thus, unlike Eqs. (2.14) and (2.15), *the mesh variables associated with Ω_1 are linked to those associated with Ω_2 through Eq. (2.37)*. However, as will be shown, like a subsystem associated with Eq. (2.14) or Eq. (2.15), the system of equations Eq. (2.37) is STI invariant for any pair of α and β.

The STI image of the system Eq. (2.37) is

$$[u - \alpha u_{\bar{x}} + \beta u_{\bar{x}\bar{x}}]_{-j}^{-n} = [u + \alpha u_{\bar{x}} + \beta u_{\bar{x}\bar{x}}]_{-j}^{-(n-1)}, \qquad (j,n) \in \Omega \tag{2.38}$$

Let

$$j' \stackrel{\text{def}}{=} -j \quad \text{and} \quad n' \stackrel{\text{def}}{=} 1 - n, \qquad (j,n) \in \Omega \tag{2.39}$$

Then because $(j', n') \in \Omega \Leftrightarrow (j,n) \in \Omega$, Eq. (2.38) can be cast into the form

$$[u + \alpha u_{\bar{x}} + \beta u_{\bar{x}\bar{x}}]_{j'}^{n'} = [u - \alpha u_{\bar{x}} + \beta u_{\bar{x}\bar{x}}]_{j'}^{n'-1}, \qquad (j', n') \in \Omega \tag{2.40}$$

By comparing Eqs. (2.37) and (2.40), one can see that the system Eq. (2.37) is identical to its STI image Eq. (2.38) (which is identical to Eq. (2.40)). Thus, under the mapping Eq. (2.29), Eq. (2.37) maps onto itself, i.e., the system Eq. (2.37) is STI invariant. QED.

Because each of Eqs. (2.14) and (2.15) is STI invariant. one can see that, for any pair of α and β, the system formed by Eqs. (2.14), (2.15), and (2.37) is STI invariant.

Next, the three mesh variables at any $(j,n) \in \Omega$ will be solved in terms of those at $(j-1, n-1)$, $(j, n-1)$ and $(j+1, n-1)$ using Eqs. (2.14), (2.15), and (2.37). Let

$$\Delta \stackrel{\text{def}}{=} \frac{4}{3}(1 + \alpha\nu) - 2\beta \tag{2.41}$$

and assume $\Delta \neq 0$. Then it can be shown that Eqs. (2.14), (2.15), and (2.37) $\Leftrightarrow$

$$\begin{aligned} u_j^n = {} & \frac{4}{3\Delta}\big[u - \alpha u_{\bar{x}} + \beta u_{\bar{x}\bar{x}}\big]_j^{n-1} \\ & + \frac{1}{\Delta}\Big[\Big(\frac{2\alpha\nu}{3} - \beta\Big)(1-\nu) - \frac{2\alpha}{3}\Big]\Big[u - (1+\nu)u_{\bar{x}} + \frac{2(1+\nu+\nu^2)}{3}u_{\bar{x}\bar{x}}\Big]_{j+1}^{n-1} \\ & + \frac{1}{\Delta}\Big[\Big(\frac{2\alpha\nu}{3} - \beta\Big)(1+\nu) + \frac{2\alpha}{3}\Big]\Big[u + (1-\nu)u_{\bar{x}} + \frac{2(1-\nu+\nu^2)}{3}u_{\bar{x}\bar{x}}\Big]_{j-1}^{n-1} \end{aligned} \qquad (j,n) \in \Omega \tag{2.42}$$

$$\begin{aligned} (u_{\bar{x}})_j^n = {} & \frac{4\nu}{3\Delta}\big[u - \alpha u_{\bar{x}} + \beta u_{\bar{x}\bar{x}}\big]_j^{n-1} \\ & + \frac{1}{\Delta}\Big[\frac{2(1-\nu+\nu^2)}{3} - \beta\Big]\Big[u - (1+\nu)u_{\bar{x}} + \frac{2(1+\nu+\nu^2)}{3}u_{\bar{x}\bar{x}}\Big]_{j+1}^{n-1} \\ & - \frac{1}{\Delta}\Big[\frac{2(1+\nu+\nu^2)}{3} - \beta\Big]\Big[u + (1-\nu)u_{\bar{x}} + \frac{2(1-\nu+\nu^2)}{3}u_{\bar{x}\bar{x}}\Big]_{j-1}^{n-1} \end{aligned} \qquad (j,n) \in \Omega \tag{2.43}$$

and

$$\begin{aligned} (u_{\bar{x}\bar{x}})_j^n = {} & -\frac{2}{\Delta}\big[u - \alpha u_{\bar{x}} + \beta u_{\bar{x}\bar{x}}\big]_j^{n-1} \\ & + \frac{1-\nu+\alpha}{\Delta}\Big[u - (1+\nu)u_{\bar{x}} + \frac{2(1+\nu+\nu^2)}{3}u_{\bar{x}\bar{x}}\Big]_{j+1}^{n-1} \\ & + \frac{1+\nu-\alpha}{\Delta}\Big[u + (1-\nu)u_{\bar{x}} + \frac{2(1-\nu+\nu^2)}{3}u_{\bar{x}\bar{x}}\Big]_{j-1}^{n-1} \end{aligned} \qquad (j,n) \in \Omega \tag{2.44}$$

For any pair of α and β with $\Delta \neq 0$, Eqs. (2.42)–(2.44) represent a solver for Eq. (1.1). In the next subsection, we pick out the pair of α and β with which the solver will have the smallest truncation error (i.e., the highest order of truncation error) for Eq. (2.42).

2.4. A study of truncation error

Because, at each (j,n), Eqs. (2.42)–(2.44) represent a system of three equations for three *independent* mesh variables, Eqs. (2.42)–(2.44) represent a numerical analogue of a system of three coupled partial differential equations (PDEs) with three dependent variables. (Eq. (1.1) is one of these PDEs). As such, in the

following study, three different symbols $\tilde{u}$, $\tilde{v}$, and $\tilde{w}$ will be used to denote the analytical versions of u_j^n, and the *non-normalized* variables $(u_x)_j^n$ and $(u_{xx})_j^n$, respectively. Specifically, let $\tilde{u}(x,t)$, $\tilde{v}(x,t)$, and $\tilde{w}(x,t)$ be functions having all the derivatives needed. Thus one can define

$$\hat{v}(x,t) \stackrel{\text{def}}{=} \tilde{v}(x,t) - \frac{\partial \tilde{u}(x,t)}{\partial x} \quad \text{and} \quad \hat{w}(x,t) \stackrel{\text{def}}{=} \tilde{w}(x,t) - \frac{\partial^2 \tilde{u}(x,t)}{\partial x^2} \tag{2.45}$$

Also, as an example, one can define

$$\left(\frac{\partial^{\ell+m}\tilde{u}}{\partial x^\ell \partial t^m}\right)_j^n \stackrel{\text{def}}{=} \frac{\partial^{\ell+m}\tilde{u}}{\partial x^\ell \partial t^m}(j\Delta x, n\Delta t) \qquad \ell, m = 0, 1, 2, \ldots \tag{2.46}$$

Next we will consider the "analytical" version of Eq. (2.42) which results from replacing (i) u_j^n, $(u_x)_j^n$, and $(u_{xx})_j^n$, respectively, with $\tilde{u}_j^n$, $\tilde{v}_j^n$, and $\tilde{w}_j^n$, for each (j,n); and (ii) the index n with $n+1$ everywhere. By using Eq. (2.13) and the fact that $(j,n+1) \in \Omega \Leftrightarrow (j,n) \in \Omega$, the analytical form can be expressed as

$$\begin{aligned}
(e_1)_j^n \stackrel{\text{def}}{=} \frac{1}{\Delta t}\Bigg\{ & \tilde{u}_j^{n+1} - \frac{4}{3\Delta}\Big[\tilde{u} - \frac{\alpha\Delta x}{2}\tilde{v} + \frac{\beta(\Delta x)^2}{4}\tilde{w}\Big]_j^n \\
& - \frac{1}{\Delta}\Big[\Big(\frac{2\alpha\nu}{3} - \beta\Big)(1-\nu) - \frac{2\alpha}{3}\Big]\Big[\tilde{u} - \frac{(1+\nu)\Delta x}{2}\tilde{v} + \frac{(1+\nu+\nu^2)\Delta x^2}{6}\tilde{w}\Big]_{j+1}^n \\
& - \frac{1}{\Delta}\Big[\Big(\frac{2\alpha\nu}{3} - \beta\Big)(1+\nu) + \frac{2\alpha}{3}\Big]\Big[\tilde{u} + \frac{(1-\nu)\Delta x}{2}\tilde{v} + \frac{(1-\nu+\nu^2)\Delta x^2}{6}\tilde{w}\Big]_{j-1}^n \Bigg\} = 0 \\
& (j,n) \in \Omega;\ \Delta \neq 0
\end{aligned} \tag{2.47}$$

By applying Taylor's formula, it can be shown that

$$\begin{aligned}
(e_1)_j^n \stackrel{\text{def}}{=} \Bigg\{ & \Big(\frac{\partial\tilde{u}}{\partial t} + a\frac{\partial\tilde{u}}{\partial x}\Big) + \frac{4(\alpha-\nu)}{3\Delta}\frac{\partial\tilde{u}}{\partial x}\frac{\Delta x}{\Delta t} + \Big(\frac{\partial^2\tilde{u}}{\partial t^2} - a^2\frac{\partial^2\tilde{u}}{\partial x^2}\Big)\frac{\Delta t}{2} - \frac{1}{\Delta}\Big[\frac{2\alpha\nu^3}{3} + \beta(1-\nu^2)\Big]\frac{\partial\hat{v}}{\partial x}\frac{(\Delta x)^2}{\Delta t} \\
& + \frac{2\nu(\nu-\alpha)}{3\Delta}\frac{\partial^2\tilde{u}}{\partial x^2}\frac{(\Delta x)^2}{\Delta t} + \Big(\frac{\partial^3\tilde{u}}{\partial t^3} + a^3\frac{\partial^3\tilde{u}}{\partial x^3}\Big)\frac{(\Delta t)^2}{6} - \frac{\alpha}{3\Delta}\frac{\partial^2\hat{v}}{\partial x^2}\frac{(\Delta x)^3}{\Delta t} \\
& + \frac{1}{3\Delta}\Big[\frac{2\alpha}{3}(1+\nu^2+\nu^4) - \beta\nu^3\Big]\frac{\partial\hat{w}}{\partial x}\frac{(\Delta x)^3}{\Delta t} + \frac{\alpha(1+4\nu^2) - 3\beta\nu - 2\nu^3}{9\Delta}\frac{\partial^3\tilde{u}}{\partial x^3}\frac{(\Delta x)^3}{\Delta t} \\
& + \Big(\frac{\partial^4\tilde{u}}{\partial t^4} - a^4\frac{\partial^4\tilde{u}}{\partial x^4}\Big)\frac{(\Delta t)^3}{24} - \frac{1}{6\Delta}\Big[\frac{2\alpha\nu^3}{3} + \beta(1-\nu^2)\Big]\frac{\partial^3\hat{v}}{\partial x^3}\frac{(\Delta x)^4}{\Delta t} + \frac{\beta}{6\Delta}\frac{\partial^2\hat{w}}{\partial x^2}\frac{(\Delta x)^4}{\Delta t} \\
& + \frac{1}{12\Delta}\Big[\frac{2\nu^4}{3} + (1+2\nu^2-\nu^4)\Big(\beta - \frac{2\alpha\nu}{3}\Big)\Big]\frac{\partial^4\tilde{u}}{\partial x^4}\frac{(\Delta x)^4}{\Delta t}\Bigg\}_j^n + O\big[(\Delta t)^4\big] \\
& + \frac{1}{\Delta}\Big[\frac{2\alpha}{3}(1-\nu+\nu^2) + \beta(1-\nu)\Big]\Big[O\big[(\Delta x)^5\big]/\Delta t + O\big[(\Delta x)^4\big] + O\big[\Delta t(\Delta x)^3\big]\Big] \\
& + \frac{1}{\Delta}\Big[\frac{2\alpha}{3}(1+\nu+\nu^2) - \beta(1+\nu)\Big]\Big[O\big[(\Delta x)^5\big]/\Delta t + O\big[(\Delta x)^4\big] + O\big[\Delta t(\Delta x)^3\big]\Big] \\
& (j,n) \in \Omega\ ; \Delta \neq 0
\end{aligned} \tag{2.48}$$

Note that $(e_1)_j^n$ defined in Eq. (2.47) is normalized by the factor $(1/\Delta t)$ so that the lowest-order terms in the above Taylor's expansion contain the leading term $(\partial\tilde{u}/\partial t + a\partial\tilde{u}/\partial x)$ which is independent of Δt and Δx. Also, in Eq. (2.48) a term is denoted by $O[(\Delta t)^{\ell_1}(\Delta x)^{\ell_2}]$ if there exists a constant $C > 0$ and two fixed integers $\ell_1 \geq 0$ and $\ell_2 \geq 0$ such that the absolute value of this term $< C(\Delta t)^{\ell_1}(\Delta x)^{\ell_2}$ for all sufficiently small

Δt and Δx. Note that, *in determining the order of magnitude of a term such as* $O\left[(\Delta x)^5\right]$ *in Eq. (2.48), the parameters* α *and* β *are not assumed to be constants independent of* Δt *and* Δx. In fact, to reduce the truncation error of the $a(3)$ scheme, they will be chosen to be functions of ν (see Eqs. (2.58)) and thus vary with the ratio $\Delta t/\Delta x$.

In the following, let $u = \tilde{u}(x,t)$, $v = \tilde{v}(x,t)$, and $w = \tilde{w}(x,t)$ be a solution to the system of PDEs formed by Eq. (1.1) and

$$v - \frac{\partial u}{\partial x} = 0 \quad \text{and} \quad w - \frac{\partial^2 u}{\partial x^2} = 0 \tag{2.49}$$

i.e.,

$$\frac{\partial \tilde{u}}{\partial t} + a\frac{\partial \tilde{u}}{\partial x} \equiv 0, \quad \tilde{v} - \frac{\partial \tilde{u}}{\partial x} \equiv 0, \quad \text{and} \quad \tilde{w} - \frac{\partial^2 \tilde{u}}{\partial x^2} \equiv 0 \tag{2.50}$$

In other words, here the scheme Eqs. (2.42)–(2.44) is considered as a solver of the system of PDEs Eqs. (1.1) and (2.49). Eqs. (2.45) and (2.50) imply that

$$\frac{\partial^{\ell+m}\hat{v}}{\partial x^\ell \partial t^m} \equiv 0 \quad \text{and} \quad \frac{\partial^{\ell+m}\hat{w}}{\partial x^\ell \partial t^m} \equiv 0 \qquad \ell, m = 0, 1, 2, \ldots \tag{2.51}$$

$$\frac{\partial^2 \tilde{u}}{\partial t^2} - a^2\frac{\partial^2 \tilde{u}}{\partial x^2} \equiv \left(\frac{\partial}{\partial t} - a\frac{\partial}{\partial x}\right)\left(\frac{\partial \tilde{u}}{\partial t} + a\frac{\partial \tilde{u}}{\partial x}\right) \equiv 0 \tag{2.52}$$

$$\frac{\partial^3 \tilde{u}}{\partial t^3} + a^3\frac{\partial^3 \tilde{u}}{\partial x^3} \equiv \left(\frac{\partial^2}{\partial t^2} - a\frac{\partial^2}{\partial t \partial x} + a^2\frac{\partial^2}{\partial x^2}\right)\left(\frac{\partial \tilde{u}}{\partial t} + a\frac{\partial \tilde{u}}{\partial x}\right) \equiv 0 \tag{2.53}$$

$$\frac{\partial^4 \tilde{u}}{\partial t^4} - a^4\frac{\partial^4 \tilde{u}}{\partial x^4} \equiv \left(\frac{\partial^2}{\partial t^2} + a^2\frac{\partial^2}{\partial x^2}\right)\left(\frac{\partial}{\partial t} - a\frac{\partial}{\partial x}\right)\left(\frac{\partial \tilde{u}}{\partial t} + a\frac{\partial \tilde{u}}{\partial x}\right) \equiv 0 \tag{2.54}$$

Note that the first equation in Eq. (2.50), and Eqs. (2.52)–(2.54) are all special cases of

$$\frac{\partial^{\ell+m}}{\partial x^\ell \partial t^m}\left[\frac{\partial^k \tilde{u}}{\partial t^k} + (-1)^{k-1}a^k\frac{\partial^k \tilde{u}}{\partial x^k}\right] \equiv 0, \qquad \ell, m = 0, 1, 2, \ldots;\ k = 1, 2, 3 \ldots \tag{2.55}$$

With the hint provided by Eqs. (2.52)–(2.54), Eqs. (2.55) can be proved using the first equation in Eq. (2.50) and elementary algebra.

By using Eqs. (2.46) and (2.51)–(2.54), one can see that $(e_1)_j^n$ reduces to

$$\begin{aligned}
(e_1)_j^n = &\left\{ \frac{4(\alpha-\nu)}{3\Delta}\frac{\partial \tilde{u}}{\partial x}\frac{\Delta x}{\Delta t} + \frac{2\nu(\nu-\alpha)}{3\Delta}\frac{\partial^2 \tilde{u}}{\partial x^2}\frac{(\Delta x)^2}{\Delta t} + \frac{\alpha(1+4\nu^2) - 3\beta\nu - 2\nu^3}{9\Delta}\frac{\partial^3 \tilde{u}}{\partial x^3}\frac{(\Delta x)^3}{\Delta t} \right. \\
&\left. + \frac{1}{12\Delta}\Big[\frac{2\nu^4}{3} + (1 + 2\nu^2 - \nu^4)\Big(\beta - \frac{2\alpha\nu}{3}\Big)\Big]\frac{\partial^4 \tilde{u}}{\partial x^4}\frac{(\Delta x)^4}{\Delta t} \right\}_j^n + O\big[(\Delta t)^4\big] \\
&+ \frac{1}{\Delta}\Big[\frac{2\alpha}{3}(1 - \nu + \nu^2) + \beta(1-\nu)\Big]\Big[O\big[(\Delta x)^5\big]/\Delta t + O\big[(\Delta x)^4\big] + O\big[\Delta t(\Delta x)^3\big]\Big] \\
&+ \frac{1}{\Delta}\Big[\frac{2\alpha}{3}(1 + \nu + \nu^2) - \beta(1+\nu)\Big]\Big[O\big[(\Delta x)^5\big]/\Delta t + O\big[(\Delta x)^4\big] + O\big[\Delta t(\Delta x)^3\big]\Big] \\
&(j,n) \in \Omega\ ;\Delta \neq 0
\end{aligned} \tag{2.56}$$

By definition, the expression on the right side of Eq. (2.56) represents the truncation error of Eqs. (2.42) if the scheme Eqs. (2.42)–(2.44) are considered as a solver of the system of PDEs Eqs. (1.1) and (2.49).

Here the values of α and β will be chosen so that the truncation error will reach the highest order. From Eq. (2.56), one can see that the coefficients of the three lowest-order terms in the truncation error vanish if

$$\alpha - \nu = 0 \quad \text{and} \quad \alpha(1+4\nu^2) - 3\beta\nu - 2\nu^3 = 0 \tag{2.57}$$

For the case $\nu \neq 0$, Eq. (2.57) $\Leftrightarrow$

$$\alpha = \nu \quad \text{and} \quad \beta = \frac{1+2\nu^2}{3} \tag{2.58}$$

Next the $a(3)$ scheme will be defined as the special solver with α and β being chosen according to Eq. (2.58).

2.5. The basic and forward marching forms of the $a(3)$ scheme

Assuming Eq. (2.58), Eqs. (2.37), (2.41)–(2.44) and (2.56) reduce to

$$\left[u + \nu u_{\bar{x}} + \frac{1+2\nu^2}{3} u_{\bar{x}\bar{x}}\right]_j^n = \left[u - \nu u_{\bar{x}} + \frac{1+2\nu^2}{3} u_{\bar{x}\bar{x}}\right]_j^{n-1} \tag{2.59}$$

$$\Delta = 2/3 \tag{2.60}$$

$$\begin{aligned} u_j^n = {} & 2\Big[u - \nu u_{\bar{x}} + \frac{1+2\nu^2}{3} u_{\bar{x}\bar{x}}\Big]_j^{n-1} - \frac{1+\nu}{2}\Big[u - (1+\nu)u_{\bar{x}} + \frac{2(1+\nu+\nu^2)}{3} u_{\bar{x}\bar{x}}\Big]_{j+1}^{n-1} \\ & - \frac{1-\nu}{2}\Big[u + (1-\nu)u_{\bar{x}} + \frac{2(1-\nu+\nu^2)}{3} u_{\bar{x}\bar{x}}\Big]_{j-1}^{n-1} \end{aligned} \qquad (j,n) \in \Omega \tag{2.61}$$

$$\begin{aligned} (u_{\bar{x}})_j^n = {} & 2\nu\Big[u - \nu u_{\bar{x}} + \frac{1+2\nu^2}{3} u_{\bar{x}\bar{x}}\Big]_j^{n-1} + \frac{1-2\nu}{2}\Big[u - (1+\nu)u_{\bar{x}} + \frac{2(1+\nu+\nu^2)}{3} u_{\bar{x}\bar{x}}\Big]_{j+1}^{n-1} \\ & - \frac{1+2\nu}{2}\Big[u + (1-\nu)u_{\bar{x}} + \frac{2(1-\nu+\nu^2)}{3} u_{\bar{x}\bar{x}}\Big]_{j-1}^{n-1} \end{aligned} \qquad (j,n) \in \Omega \tag{2.62}$$

$$\begin{aligned} (u_{\bar{x}\bar{x}})_j^n = {} & -3\Big[u - \nu u_{\bar{x}} + \frac{1+2\nu^2}{3} u_{\bar{x}\bar{x}}\Big]_j^{n-1} + \frac{3}{2}\Big[u - (1+\nu)u_{\bar{x}} + \frac{2(1+\nu+\nu^2)}{3} u_{\bar{x}\bar{x}}\Big]_{j+1}^{n-1} \\ & + \frac{3}{2}\Big[u + (1-\nu)u_{\bar{x}} + \frac{2(1-\nu+\nu^2)}{3} u_{\bar{x}\bar{x}}\Big]_{j-1}^{n-1} \end{aligned} \qquad (j,n) \in \Omega \tag{2.63}$$

and

$$\begin{aligned} (e_1)_j^n = {} & \frac{1}{24}\left(\frac{\partial^4 \tilde{u}}{\partial x^4}\right)_j^n \Big[\frac{(\Delta x)^4}{\Delta t} + 2a^2 \Delta t (\Delta x)^2 + a^4 (\Delta t)^3\Big] + O\big[(\Delta t)^4\big] + O\big[(\Delta x)^4\big] \\ & + O\big[\Delta t (\Delta x)^3\big] + O\big[(\Delta t)^2 (\Delta x)^2\big] + \frac{O\big[(\Delta x)^5\big]}{\Delta t} \end{aligned} \qquad (j,n) \in \Omega \tag{2.64}$$

Note that: (i) the forms of the last four terms in Eq. (2.64) have been simplified using the definition $\nu = a\Delta t/\Delta x$; and (ii) the expression on the right side of Eq. (2.64) represents the truncation error of Eq. (2.61) if the scheme formed by Eqs. (2.61)–(2.63) is considered as a solver of the system of PDEs Eqs. (1.1) and (2.49).

Next we convert Eq. (2.62) into its analytical form by replacing (i) u_j^n, $(u_x)_j^n$, and $(u_{xx})_j^n$, respectively, with $\tilde{u}_j^n$, $\tilde{v}_j^n$, and $\tilde{w}_j^n$, for each (j,n); and (ii) the index n with $n+1$ everywhere. By using (i) Eq. (2.13), (ii) $\nu = a\Delta t/\Delta x$, and (iii) the fact that $(j,n+1) \in \Omega \Leftrightarrow (j,n) \in \Omega$, then after a normalization by the factor $1/2$, the analytical form can be expressed as

$$\begin{aligned} (e_2)_j^n \overset{\text{def}}{=} {} & \frac{1}{2}\tilde{v}_j^{n+1} - \frac{2a\Delta t}{(\Delta x)^2}\Big[\tilde{u} - \frac{a\Delta t}{2}\tilde{v} + \frac{(\Delta x)^2 + 2a^2(\Delta t)^2}{12}\tilde{w}\Big]_j^n \\ & - \frac{1}{2\Delta x}\Big(1 - \frac{2a\Delta t}{\Delta x}\Big)\Big[\tilde{u} - \frac{\Delta x + a\Delta t}{2}\tilde{v} + \frac{(\Delta x)^2 + a\Delta t\Delta x + a^2(\Delta t)^2}{6}\tilde{w}\Big]_{j+1}^n \\ & + \frac{1}{2\Delta x}\Big(1 + \frac{2a\Delta t}{\Delta x}\Big)\Big[\tilde{u} + \frac{\Delta x - a\Delta t}{2}\tilde{v} + \frac{(\Delta x)^2 - a\Delta t\Delta x + a^2(\Delta t)^2}{6}\tilde{w}\Big]_{j-1}^n = 0 \end{aligned} \qquad (j,n) \in \Omega \tag{2.65}$$

Similarly, the analytical form of Eq. (2.63) can be expressed as

$$
\begin{aligned}
(e_3)_j^n \overset{\text{def}}{=} & \frac{1}{6}\tilde{w}_j^{n+1} + \frac{2}{(\Delta x)^2}\Big[\tilde{u} - \frac{a\Delta t}{2}\tilde{v} + \frac{(\Delta x)^2 + 2a^2(\Delta t)^2}{12}\tilde{w}\Big]_j^n \\
& - \frac{1}{(\Delta x)^2}\Big[\tilde{u} - \frac{\Delta x + a\Delta t}{2}\tilde{v} + \frac{(\Delta x)^2 + a\Delta t\Delta x + a^2(\Delta t)^2}{6}\tilde{w}\Big]_{j+1}^n \qquad (j,n) \in \Omega \qquad (2.66) \\
& - \frac{1}{(\Delta x)^2}\Big[\tilde{u} + \frac{\Delta x - a\Delta t}{2}\tilde{v} + \frac{(\Delta x)^2 - a\Delta t\Delta x + a^2(\Delta t)^2}{6}\tilde{w}\Big]_{j-1}^n = 0
\end{aligned}
$$

By using Taylor's formula and Eq. (2.45), Eqs. (2.65) and (2.66) imply that

$$
\begin{aligned}
(e_2)_j^n = & \Big\{\hat{v} + \Big[\frac{\partial \hat{v}}{\partial t} - a\frac{\partial \hat{v}}{\partial x} + \frac{\partial}{\partial x}\Big(\frac{\partial \tilde{u}}{\partial t} + a\frac{\partial \tilde{u}}{\partial x}\Big)\Big]\frac{\Delta t}{2} + \frac{\partial^2 \hat{v}}{\partial t^2}\frac{(\Delta t)^2}{4} + \frac{\partial^2 \hat{v}}{\partial x^2}\frac{\big[(\Delta x)^2 - 2a^2(\Delta t)^2\big]}{4} \\
& + \frac{\partial \hat{w}}{\partial x}\frac{\big[a^2(\Delta t)^2 - (\Delta x)^2\big]}{6} + \frac{\partial}{\partial x}\Big(\frac{\partial^2 \tilde{u}}{\partial t^2} - a^2\frac{\partial^2 \tilde{u}}{\partial x^2}\Big)\frac{(\Delta t)^2}{4} - \frac{(1+\nu^2)}{12}\frac{\partial^3 \tilde{u}}{\partial x^3}(\Delta x)^2\Big\}_j^n \qquad (2.67) \\
& + O\big[(\Delta t)^3\big] + O\big[(\Delta t)^2\Delta x\big] + O\big[\Delta t(\Delta x)^2\big] + O\big[(\Delta x)^3\big] \qquad (j,n) \in \Omega
\end{aligned}
$$

and

$$
\begin{aligned}
(e_3)_j^n = & \Big\{\frac{\partial \hat{v}}{\partial x} + \Big[\frac{\partial^2}{\partial x^2}\Big(\frac{\partial \tilde{u}}{\partial t} + a\frac{\partial \tilde{u}}{\partial x}\Big) + \frac{\partial \hat{w}}{\partial t} - 2a\frac{\partial \hat{w}}{\partial x} + 3a\frac{\partial^2 \hat{v}}{\partial x^2}\Big]\frac{\Delta t}{6} \\
& + \Big[\frac{\partial^2}{\partial x^2}\Big(\frac{\partial^2 \tilde{u}}{\partial t^2} - a^2\frac{\partial^2 \tilde{u}}{\partial x^2}\Big) + \frac{\partial^2 \hat{w}}{\partial t^2} - 2a^2\frac{\partial^2 \hat{w}}{\partial x^2}\Big]\frac{(\Delta t)^2}{12} + \Big(\frac{\partial^3 \hat{v}}{\partial x^3} - \frac{\partial^2 \hat{w}}{\partial x^2}\Big)\frac{(\Delta x)^2}{6} \qquad (j,n) \in \Omega \qquad (2.68) \\
& - \frac{(1+\nu^2)}{12}\frac{\partial^4 \tilde{u}}{\partial x^4}(\Delta x)^2\Big\}_j^n + O\big[(\Delta t)^3\big] + O\big[(\Delta t)^2\Delta x\big] + O\big[\Delta t(\Delta x)^2\big] + O\big[(\Delta x)^3\big]
\end{aligned}
$$

Assuming Eqs. (2.51) and (2.55), $(e_2)_j^n$ and $(e_3)_j^n$ reduce to

$$
(e_2)_j^n = -\left(\frac{\partial^3 \tilde{u}}{\partial x^3}\right)_j^n \frac{\big[(\Delta x)^2 + a^2(\Delta t)^2\big]}{12} + O\big[(\Delta t)^3\big] + O\big[(\Delta t)^2\Delta x\big] + O\big[\Delta t(\Delta x)^2\big] + O\big[(\Delta x)^3\big] \quad (j,n) \in \Omega \quad (2.69)
$$

and

$$
(e_3)_j^n = -\left(\frac{\partial^4 \tilde{u}}{\partial x^4}\right)_j^n \frac{\big[(\Delta x)^2 + a^2(\Delta t)^2\big]}{12} + O\big[(\Delta t)^3\big] + O\big[(\Delta t)^2\Delta x\big] + O\big[\Delta t(\Delta x)^2\big] + O\big[(\Delta x)^3\big] \quad (j,n) \in \Omega \quad (2.70)
$$

respectively.

Hereafter, for any ν, let *the system of equations* defined by Eqs. (2.14), (2.15), and (2.59) be referred to as the basic form of the $a(3)$ scheme while that defined by Eqs. (2.61)–(2.63) be referred to as the forward marching form of the $a(3)$ scheme. Because (i) Eqs. (2.14), (2.15), and (2.37) $\Leftrightarrow$ Eqs. (2.42)–(2.44) if $\Delta \neq 0$, and (ii) Eqs. (2.59)–(2.63) are special cases of Eqs. (2.37), and (2.41)–(2.44), respectively, the basic form of the $a(3)$ scheme $\Leftrightarrow$ its forward marching form. Thus the essential conditions represented by these or other equivalent forms may be referred to simply as the $a(3)$ scheme.

With the above definitions, the expressions on the right sides of Eqs. (2.64), (2.69) and (2.70) represent the truncation errors of Eqs. (2.61)–(2.63), respectively, if the forward marching form of the $a(3)$ scheme is considered as a solver of the system of PDEs Eqs. (1.1) and (2.49). According to Eqs. (2.69) and (2.70), $(e_2)_j^n \to 0$ and $(e_3)_j^n \to 0$ as $\Delta t, \Delta x \to 0$, regardless how Δt and Δx are related when $\Delta t, \Delta x \to 0$. On

the other hand, Eq. (2.64) implies that $(e_1)_j^n \to 0$ as $\Delta t, \Delta x \to 0$ only if the mesh refinement procedure is subjected to the condition

$$\frac{(\Delta x)^4}{\Delta t} \to 0 \quad \text{as} \quad \Delta t, \Delta x \to 0 \tag{2.71}$$

Thus the $a(3)$ scheme is consistent with the system of PDEs Eqs. (1.1) and (2.49) if and only if Eq. (2.71) is satisfied.

At this juncture, we offer the following remarks:

(a) Let $\Delta t/\Delta x$ be held as constant as $\Delta t, \Delta x \to 0$. Then for this mesh refinement procedure, Eqs. (2.64), (2.69), and (2.70) imply that the truncation errors for Eqs. (2.61)–(2.63), respectively, are third order, second order, and second order in Δt and Δx.

(b) Because (i) each of the two decoupled subsystems in each of Eqs. (2.14) and (2.15) is STI invariant by itself, and (ii) the system Eq. (2.37) is also STI invariant if α and β are parameters independent of (j, n), by the definition of STI invariance one can easily see that the basic form of the $a(3)$ scheme is STI invariant.

(c) Let $\vec{q}(j, n) = \vec{q}_o(j, n)$, $(j, n) \in \Omega$, be a solution to the basic form. Then, by substituting $\vec{q}(j, n) = \vec{q}_o(j, n)$ into the basic form, one obtains a system of identities involving $\vec{q}_o(j, n)$, $(j, n) \in \Omega$. Due to the STI invariance of the basic form, the above system of identities is equivalent to that obtained by substituting $\vec{q}(j, n) = U\vec{q}_o(-j, -n)$ into the basic form. As such $\vec{q}(j, n) = \vec{q}_o(j, n)$, $(j, n) \in \Omega$, represent a solution to the basic form $\Leftrightarrow$ $\vec{q}(j, n) = U\vec{q}_o(-j, -n)$, $(j, n) \in \Omega$, represent another solution to the basic form. In other words, *the STI image of a solution to the basic form is also a solution and vice versa*. Obviously this conclusion is valid for other STI invariant forms of the $a(3)$ scheme.

Next, the forward marching form Eqs. (2.61)–(2.62) will be cast into a matrix form. Let

$$\vec{c}_0(\nu) \stackrel{\text{def}}{=} \begin{pmatrix} 1 \\ -\nu \\ (1+2\nu^2)/3 \end{pmatrix}, \quad \vec{c}_+(\nu) \stackrel{\text{def}}{=} \begin{pmatrix} 1 \\ -(1+\nu) \\ (2/3)(1+\nu+\nu^2) \end{pmatrix}, \quad \vec{c}_-(\nu) \stackrel{\text{def}}{=} \begin{pmatrix} 1 \\ 1-\nu \\ (2/3)(1-\nu+\nu^2) \end{pmatrix} \tag{2.72}$$

$$\vec{d}_0(\nu) \stackrel{\text{def}}{=} \begin{pmatrix} 2 \\ 2\nu \\ -3 \end{pmatrix}, \quad \vec{d}_+(\nu) \stackrel{\text{def}}{=} \begin{pmatrix} -(1+\nu)/2 \\ (1-2\nu)/2 \\ (3/2) \end{pmatrix}, \quad \vec{d}_-(\nu) \stackrel{\text{def}}{=} \begin{pmatrix} -(1-\nu)/2 \\ -(1+2\nu)/2 \\ (3/2) \end{pmatrix} \tag{2.73}$$

$$Q_0(\nu) \stackrel{\text{def}}{=} \vec{d}_0(\nu)\,[\vec{c}_0(\nu)]^t = \begin{pmatrix} 2 & -2\nu & (2/3)(1+2\nu^2) \\ 2\nu & -2\nu^2 & (2/3)\nu(1+2\nu^2) \\ -3 & 3\nu & -(1+2\nu^2) \end{pmatrix} \tag{2.74}$$

$$Q_+(\nu) \stackrel{\text{def}}{=} \vec{d}_+(\nu)\,[\vec{c}_+(\nu)]^t = \begin{pmatrix} -(1+\nu)/2 & (1+\nu)^2/2 & -(1+\nu)(1+\nu+\nu^2)/3 \\ (1-2\nu)/2 & -(1-2\nu)(1+\nu)/2 & (1-2\nu)(1+\nu+\nu^2)/3 \\ 3/2 & -(3/2)(1+\nu) & 1+\nu+\nu^2 \end{pmatrix} \tag{2.75}$$

and

$$Q_-(\nu) \stackrel{\text{def}}{=} \vec{d}_-(\nu)\,[\vec{c}_-(\nu)]^t = \begin{pmatrix} -(1-\nu)/2 & -(1-\nu)^2/2 & -(1-\nu)(1-\nu+\nu^2)/3 \\ -(1+2\nu)/2 & -(1+2\nu)(1-\nu)/2 & -(1+2\nu)(1-\nu+\nu^2)/3 \\ 3/2 & (3/2)(1-\nu) & 1-\nu+\nu^2 \end{pmatrix} \tag{2.76}$$

Hereafter $\vec{c}^{\,t}$ denote the transpose of any column or row matrix $\vec{c}$. By using Eqs. (2.31) and (2.74)–(2.76), the forward marching form can be cast into the matrix form:

$$\vec{q}(j, n) = Q_0(\nu)\vec{q}(j, n-1) + Q_+(\nu)\vec{q}(j+1, n-1) + Q_-(\nu)\vec{q}(j-1, n-1), \qquad (j, n) \in \Omega \tag{2.77}$$

Here the reader is warned that the notations $Q_+(\nu)$ and $Q_-(\nu)$ used in earlier CESE papers are now replaced by $Q_-(\nu)$ and $Q_+(\nu)$, respectively. As such, *the terms $Q_-(\nu)\vec{q}(j+1, n-1)$ and $Q_+(\nu)(j-1, n-1)$ in Eq. (3.48)*

of [71] appear here as $Q_+(\nu)\vec{q}(j+1,n-1)$ *and* $Q_-(\nu)(j-1,n-1)$*, respectively*. Also note that each of $Q_0(\nu)$, $Q_+(\nu)$, and $Q_-(\nu)$ is in the form of $\vec{d}\,\vec{c}^{\,t}$ where $\vec{c}$ and $\vec{d}$ are 3×1 column matrix. Thus each is a matrix of rank one (see pp. 80-82 in [74]). Rank-one matrices are singular and have many interesting properties. As an example, the eigenvalues of $Q_0(\nu)$ are 0, 0, and $[\vec{c}_0(\nu)]^t\,\vec{d}_0(\nu)$ with $\vec{d}_0(\nu)$ being the eigenvector of the last eigenvalue.

To facilitate the proof of the STI invariance of the forward marching form, first we will introduce some basic concept. Note that, *for any set of variables* x_ℓ, y_ℓ, $\ell = 1, 2$, the conditions

$$x_1 + y_1 = x_2 - y_2 \quad \text{and} \quad x_1 - y_1 = x_2 + y_2 \tag{2.78}$$

$\Leftrightarrow$

$$x_1 = x_2 \quad \text{and} \quad y_1 = -y_2 \tag{2.79}$$

Thus, the image of Eq. (2.78) under *any* one-to-one mapping

$$(x_\ell, y_\ell) \leftrightarrow (x'_\ell, y'_\ell), \qquad \ell = 1, 2 \tag{2.80}$$

i.e.,

$$x'_1 + y'_1 = x'_2 - y'_2 \quad \text{and} \quad x'_1 - y'_1 = x'_2 + y'_2 \tag{2.81}$$

$\Leftrightarrow$ the image of Eq. (2.79) under the same mapping, i.e.,

$$x'_1 = x'_2 \quad \text{and} \quad y'_1 = -y'_2 \tag{2.82}$$

where the variables x'_ℓ and y'_ℓ, $\ell = 1, 2$, may or may not be related to x_ℓ, y_ℓ, $\ell = 1, 2$. Moreover, in case that these two sets of variables are related, the condition Eq. (2.78) (or its equivalent Eq. (2.79)) may or may not be equivalent to the condition Eq. (2.81) (or its equivalent Eq. (2.82)). If the mapping Eq. (2.80) is such that Eq. (2.78) $\Leftrightarrow$ the image under this mapping (i.e., Eq. (2.81)), then Eq. (2.79) (the equivalent of Eq. (2.78)) $\Leftrightarrow$ Eq. (2.82) (the equivalent of Eq. (2.81)). Eq. (2.80) with $x'_\ell = x_\ell$ and $y'_\ell = y_\ell$, $\ell = 1, 2$, is an example of such mapping while Eq. (2.80) with $x'_\ell = y_\ell$ and $y'_\ell = x_\ell$, $\ell = 1, 2$, is not.

To prove the STI invariance of the forward marching form, Note that: (i) the basic form of the $a(3)$ scheme $\Leftrightarrow$ its forward marching form for *any choice* of $\vec{q}(j,n)$, $(j,n) \in \Omega$; and (ii) the STI images of the basic and forward marching forms, respectively, are obtained from the basic and forward marching forms through the mapping Eq. (2.30), i.e., through replacing $\vec{q}(j,n)$ in the basic form and the forward marching form with $U\vec{q}(-j,-n)$, $(j,n) \in \Omega$. From the above observations and the illustration given in the last paragraph, one concludes that the STI image of the basic form $\Leftrightarrow$ that of the forward marching form. Because the basic form is STI invariant, i.e., the STI image of the basic form $\Leftrightarrow$ the basic form itself, Now we arrive at the conclusion that the forward marching form $\Leftrightarrow$ the basic form $\Leftrightarrow$ the STI image of the basic form $\Leftrightarrow$ the STI image of the forward marching form. Thus the forward marching form $\Leftrightarrow$ its STI image, i.e., the forward marching form is STI invariant. QED.

With the above preliminaries, the backward marching form of the $a(3)$ scheme will be developed in Sec. 2.6.

2.6. The backward marching forms of the $a(3)$ scheme

The STI invariance of the forward marching form of the $a(3)$ scheme implies that Eq. (2.77) $\Leftrightarrow$ its STI image, i.e.,

$$U\vec{q}(-j,-n) = Q_0(\nu)U\vec{q}(-j,-n+1) + Q_+(\nu)U\vec{q}(-j-1,-n+1) + Q_-(\nu)U\vec{q}(-j+1,-n+1), \quad (j,n) \in \Omega \tag{2.83}$$

Moreover, by multiplying Eq. (2.83) from left using the matrix U and using Eq. (2.33), one concludes that Eq. (2.83) $\Leftrightarrow$

$$\vec{q}(-j,-n) = \hat{Q}_0(\nu)\vec{q}(-j,-n+1) + \hat{Q}_-(\nu)\vec{q}(-j-1,-n+1) + \hat{Q}_+(\nu)\vec{q}(-j+1,-n+1), \qquad (j,n) \in \Omega \tag{2.84}$$

where

$$\hat{Q}_0(\nu) \stackrel{\text{def}}{=} UQ_0(\nu)U = \begin{pmatrix} 2 & 2\nu & (2/3)(1+2\nu^2) \\ -2\nu & -2\nu^2 & -(2/3)\nu(1+2\nu^2) \\ -3 & -3\nu & -(1+2\nu^2) \end{pmatrix} \tag{2.85}$$

$$\hat{Q}_-(\nu) \stackrel{\text{def}}{=} UQ_+(\nu)U = \begin{pmatrix} -(1+\nu)/2 & -(1+\nu)^2/2 & -(1+\nu)(1+\nu+\nu^2)/3 \\ -(1-2\nu)/2 & -(1-2\nu)(1+\nu)/2 & -(1-2\nu)(1+\nu+\nu^2)/3 \\ 3/2 & (3/2)(1+\nu) & 1+\nu+\nu^2 \end{pmatrix} \tag{2.86}$$

and

$$\hat{Q}_+(\nu) \stackrel{\text{def}}{=} UQ_-(\nu)U = \begin{pmatrix} -(1-\nu)/2 & (1-\nu)^2/2 & -(1-\nu)(1-\nu+\nu^2)/3 \\ (1+2\nu)/2 & -(1+2\nu)(1-\nu)/2 & (1+2\nu)(1-\nu+\nu^2)/3 \\ 3/2 & -(3/2)(1-\nu) & 1-\nu+\nu^2 \end{pmatrix} \tag{2.87}$$

By replacing the "dummy" indices $-j$ and $-n$ everywhere in Eq. (2.84) with j and n, respectively, one can see that the system Eq. (2.84) is identical to the system

$$\vec{q}(j,n) = \hat{Q}_0(\nu)\vec{q}(j,n+1) + \hat{Q}_+(\nu)\vec{q}(j+1,n+1) + \hat{Q}_-(\nu)\vec{q}(j-1,n+1), \quad (j,n) \in \Omega \tag{2.88}$$

Because the mesh variables at (j,n) can be determined in terms of those at $(j-1,n+1)$, $(j,n+1)$, and $(j+1,n+1)$ using Eq. (2.88), hereafter Eq. (2.88) (which is equivalent to other forms of the $a(3)$ scheme) will be referred to as the backward marching form of the $a(3)$ scheme.

According to Eqs. (2.74) and (2.85), $\hat{Q}_0(\nu) = U\vec{d}_0(\nu)\,[\vec{c}_0(\nu)]^t\,U$. Because $U\vec{d}_0(\nu)$ and $[\vec{c}_0(\nu)]^t\,U$ are 3×1 column matrix and 1×3 row matrix, respectively, $\hat{Q}_0(\nu)$ is a rank-one matrix. Similarly, $\hat{Q}_-(\nu)$ and $\hat{Q}_+(\nu)$ are also rank-one matrices.

Eq. (2.88) was derived using the STI invariance of the forward marching form of the $a(3)$ scheme. Alternatively, it can also be derived from the basic form. To proceed, note that: (i) by replacing the indices j and n everywhere in Eq. (2.14) with $j+1$ and $n+1$ and using the fact that $(j,n) \in \Omega \Leftrightarrow (j-1,n-1) \in \Omega \Leftrightarrow (j+1,n+1) \in \Omega$, one can see that the system Eq. (2.14) is identical to the system

$$\left[u + (1-\nu)u_{\bar{x}} + \frac{2(1-\nu+\nu^2)}{3}u_{\bar{x}\bar{x}}\right]_j^n = \left[u - (1-\nu)u_{\bar{x}} + \frac{2(1-\nu+\nu^2)}{3}u_{\bar{x}\bar{x}}\right]_{j+1}^{n+1}, \quad (j,n) \in \Omega \tag{2.89}$$

(ii) by replacing the indices j and n everywhere in Eq. (2.15) with $j-1$ and $n+1$ and using the fact that $(j,n) \in \Omega \Leftrightarrow (j+1,n-1) \in \Omega \Leftrightarrow (j-1,n+1) \in \Omega$, one can see that the system Eq. (2.15) is identical to the system

$$\left[u - (1+\nu)u_{\bar{x}} + \frac{2(1+\nu+\nu^2)}{3}u_{\bar{x}\bar{x}}\right]_j^n = \left[u + (1+\nu)u_{\bar{x}} + \frac{2(1+\nu+\nu^2)}{3}u_{\bar{x}\bar{x}}\right]_{j-1}^{n+1}, \quad (j,n) \in \Omega \tag{2.90}$$

and (iii) by replacing the index n everywhere in Eq. (2.59) with $n+1$ and using the fact that $(j,n) \in \Omega \Leftrightarrow (j,n-1) \in \Omega \Leftrightarrow (j,n+1) \in \Omega$, one can see that the system Eq. (2.59) is identical to the system

$$\left[u - \nu u_{\bar{x}} + \frac{1+2\nu^2}{3}u_{\bar{x}\bar{x}}\right]_j^n = \left[u + \nu u_{\bar{x}} + \frac{1+2\nu^2}{3}u_{\bar{x}\bar{x}}\right]_j^{n+1} \qquad (j,n) \in \Omega \tag{2.91}$$

As such the system Eqs. (2.89)–(2.91) are identical to Eqs. (2.14), (2.15), (2.59), respectively.

For *each* $(j,n) \in \Omega$, Eqs. (2.14), (2.15), and (2.59) form a linear system of three equations for the three mesh variables u_j^n, $(u_{\bar{x}})_j^n$, and $(u_{\bar{x}\bar{x}})_j^n$. Eqs. (2.89)–(2.91) form another system. Moreover, one can see that, under the mesh variable mapping

$$\begin{aligned} &\vec{q}(j,n) \leftrightarrow U\vec{q}(j,n), \quad \vec{q}(j,n-1) \leftrightarrow \vec{q}(j,n+1), \\ &\vec{q}(j+1,n-1) \leftrightarrow U\vec{q}(j-1,n+1), \quad \text{and} \quad \vec{q}(j-1,n-1) \leftrightarrow \vec{q}(j+1,n+1) \end{aligned} \tag{2.92}$$

Eqs. (2.89)–(2.91), respectively, are the images of Eqs. (2.14), (2.15), and (2.59) and vice versa. By using the concept introduced earlier in a discussion involving Eqs. (2.78)–(2.82), one concludes that the solution to Eqs. (2.89)–(2.91) must be the image of Eq. (2.77) (i.e., the solution to Eqs. (2.14), (2.15) and (2.59)) under the same mapping. In other words, the solution to Eqs. (2.89)–(2.91) is

$$U\vec{q}(j,n) = Q_0(\nu)U\vec{q}(j,n+1) + Q_+(\nu)U\vec{q}(j-1,n+1) + Q_-(\nu)U\vec{q}(j+1,n+1), \qquad (j,n) \in \Omega \tag{2.93}$$

By multiplying Eq. (2.93) from left using the matrix U and using Eqs. (2.33) and (2.85)–(2.87), one has Eq. (2.88). QED.

As a preliminary for the developments in Sec. 3, in the following, important algebraic relations involving $Q_0(\nu)$, $Q_+(\nu)$, $Q_-(\nu)$, $\hat{Q}_0(\nu)$, $\hat{Q}_+(\nu)$, and $\hat{Q}_-(\nu)$ will be extracted from the STI invariance of the $a(3)$ scheme.

2.7. Algebraic relations associated with STI invariance

Let $(j_o, n_o) \in \Omega$ be any given fixed mesh point. Let $\vec{q}(j_o, n_o)$, $\vec{q}(j_o \pm 1, n_o)$, and $\vec{q}(j_o \pm 2, n_o)$, respectively, be the arbitrary initial data specified at (j_o, n_o), $(j_o \pm 1, n_o)$, and $(j_o \pm 2, n_o)$, respectively. Let $\vec{q}(j_o, n_o + 1)$, and $\vec{q}(j_o \pm 1, n_o + 1)$ be specified in terms of the mesh variables at the n_oth time level using the forward marching form Eq. (2.77), i.e.,

$$\vec{q}(j_o, n_o + 1) = Q_0(\nu)\,\vec{q}(j_o, n_o) + Q_+(\nu)\,\vec{q}(j_o + 1, n_o) + Q_-(\nu)\,\vec{q}(j_o - 1, n_o) \tag{2.94}$$

$$\vec{q}(j_o + 1, n_o + 1) = Q_0(\nu)\,\vec{q}(j_o + 1, n_o) + Q_+(\nu)\,\vec{q}(j_o + 2, n_o) + Q_-(\nu)\,\vec{q}(j_o, n_o) \tag{2.95}$$

and

$$\vec{q}(j_o - 1, n_o + 1) = Q_0(\nu)\,\vec{q}(j_o - 1, n_o) + Q_+(\nu)\,\vec{q}(j_o, n_o) + Q_-(\nu)\,\vec{q}(j_o - 2, n_o) \tag{2.96}$$

On the other hand, because Eq. (2.77) $\Leftrightarrow$ Eq. (2.88), $\vec{q}(j_o, n_o + 1)$, $\vec{q}(j_o \pm 1, n_o + 1)$, and $\vec{q}(j_o, n_o)$ must also be linked by Eq. (2.88), i.e.,

$$\vec{q}(j_o, n_o) = \hat{Q}_0(\nu)\ \vec{q}(j_o, n_o + 1) + \hat{Q}_+(\nu)\,\vec{q}(j_o + 1, n_o + 1) + \hat{Q}_-(\nu)\,\vec{q}(j_o - 1, n_o + 1) \tag{2.97}$$

Substituting Eqs. (2.94)–(2.96) into (2.97), one has

$$\begin{aligned} &\big[\hat{Q}_0(\nu)Q_0(\nu) + \hat{Q}_+(\nu)Q_-(\nu) + \hat{Q}_-(\nu)Q_+(\nu) - I\big]\vec{q}(j_o, n_o) \\ &+ \big[\hat{Q}_0(\nu)Q_+(\nu) + \hat{Q}_+(\nu)Q_0(\nu)\big]\vec{q}(j_o + 1, n_o) + \big[\hat{Q}_0(\nu)Q_-(\nu) + \hat{Q}_-(\nu)Q_0(\nu)\big]\vec{q}(j_o - 1, n_o) \\ &+ \hat{Q}_+(\nu)Q_+(\nu)\,\vec{q}(j_o + 2, n_o) + \hat{Q}_-(\nu)Q_-(\nu)\,\vec{q}(j_o - 2, n_o) = \vec{0} \end{aligned} \tag{2.98}$$

where I is the 3×3 identity matrix and $\vec{0}$ is the 3×1 null column matrix.

Because Eq. (2.98) must be valid for any choice of $\vec{q}(j_o, n_o)$, $\vec{q}(j_o \pm 1, n_o)$, and $\vec{q}(j_o \pm 2, n_o)$, the coefficients matrices in front of these column matrices must be null identically, i.e.,

$$\hat{Q}_0(\nu)Q_0(\nu) + \hat{Q}_+(\nu)Q_-(\nu) + \hat{Q}_-(\nu)Q_+(\nu) = I \tag{2.99}$$

$$\hat{Q}_0(\nu)Q_+(\nu) + \hat{Q}_+(\nu)Q_0(\nu) = \mathbf{0} \tag{2.100}$$

$$\hat{Q}_0(\nu)Q_-(\nu) + \hat{Q}_-(\nu)Q_0(\nu) = \mathbf{0} \tag{2.101}$$

$$\hat{Q}_+(\nu)Q_+(\nu) = \mathbf{0} \tag{2.102}$$

and

$$\hat{Q}_-(\nu)Q_-(\nu) = \mathbf{0} \tag{2.103}$$

where $\mathbf{0}$ is the 3×3 null matrix. As an example, one can prove Eq. (2.99) by substituting into Eq. (2.98) each of the following sets of the initial data: (i) $\vec{q}(j_o \pm 1, n_o) = \vec{q}(j_o \pm 2, n_o) = \vec{0}$ and $\vec{q}(j_o, n_o) = (1, 0, 0)^t$,

(ii) $\vec{q}(j_o \pm 1, n_o) = \vec{q}(j_o \pm 2, n_o) = \vec{0}$ and $\vec{q}(j_o, n_o) = (0,1,0)^t$, and (iii) $\vec{q}(j_o \pm 1, n_o) = \vec{q}(j_o \pm 2, n_o) = \vec{0}$ and $\vec{q}(j_o, n_o) = (0,0,1)^t$.

Similarly, by substituting the backward marching relations

$$\vec{q}(j_o, n_o - 1) = \hat{Q}_0(\nu)\,\vec{q}(j_o, n_o) + \hat{Q}_+(\nu)\,\vec{q}(j_o + 1, n_o) + \hat{Q}_-(\nu)\,\vec{q}(j_o - 1, n_o) \tag{2.104}$$

$$\vec{q}(j_o + 1, n_o - 1) = \hat{Q}_0(\nu)\,\vec{q}(j_o + 1, n_o) + \hat{Q}_+(\nu)\,\vec{q}(j_o + 2, n_o) + \hat{Q}_-(\nu)\,\vec{q}(j_o, n_o) \tag{2.105}$$

and

$$\vec{q}(j_o - 1, n_o - 1) = \hat{Q}_0(\nu)\,\vec{q}(j_o - 1, n_o) + \hat{Q}_+(\nu)\,\vec{q}(j_o, n_o) + \hat{Q}_-(\nu)\,\vec{q}(j_o - 2, n_o) \tag{2.106}$$

into the forward marching relation

$$\vec{q}(j_o, n_o) = Q_0(\nu)\;\vec{q}(j_o, n_o - 1) + Q_+(\nu)\,\vec{q}(j_o + 1, n_o - 1) + Q_-(\nu)\,\vec{q}(j_o - 1, n_o - 1) \tag{2.107}$$

one has

$$\begin{aligned}
&\big[Q_0(\nu)\hat{Q}_0(\nu) + Q_+(\nu)\hat{Q}_-(\nu) + Q_-(\nu)\hat{Q}_+(\nu) - I\big]\vec{q}(j_o, n_o) \\
&+ \big[Q_0(\nu)\hat{Q}_+(\nu) + Q_+(\nu)\hat{Q}_0(\nu)\big]\vec{q}(j_o + 1, n_o) + \big[Q_0(\nu)\hat{Q}_-(\nu) + Q_-(\nu)\hat{Q}_0(\nu)\big]\vec{q}(j_o - 1, n_o) \\
&+ Q_+(\nu)\hat{Q}_+(\nu)\,\vec{q}(j_o + 2, n_o) + Q_-(\nu)\hat{Q}_-(\nu)\,\vec{q}(j_o - 2, n_o) = \vec{0}
\end{aligned} \tag{2.108}$$

Because Eq. (2.107) must be valid for any choice of $\vec{q}(j_o, n_o)$, $\vec{q}(j_o \pm 1, n_o)$, and $\vec{q}(j_o \pm 2, n_o)$, one concludes that

$$Q_0(\nu)\hat{Q}_0(\nu) + Q_+(\nu)\hat{Q}_-(\nu) + Q_-(\nu)\hat{Q}_+(\nu) = I \tag{2.109}$$

$$Q_0(\nu)\hat{Q}_+(\nu) + Q_+(\nu)\hat{Q}_0(\nu) = \mathbf{0} \tag{2.110}$$

$$Q_0(\nu)\hat{Q}_-(\nu) + Q_-(\nu)\hat{Q}_0(\nu) = \mathbf{0} \tag{2.111}$$

$$Q_+(\nu)\hat{Q}_+(\nu) = \mathbf{0} \tag{2.112}$$

and

$$Q_-(\nu)\hat{Q}_-(\nu) = \mathbf{0} \tag{2.113}$$

By using Eqs. (2.32) and (2.85)–(2.87), it can be shown that: (i) Eq. (2.99) $\Leftrightarrow$ Eq. (2.109) $\Leftrightarrow$

$$Q_0(\nu)UQ_0(\nu) + Q_-(\nu)UQ_-(\nu) + Q_+(\nu)UQ_+(\nu) = U \tag{2.114}$$

(ii) Eq. (2.100) $\Leftrightarrow$ Eq. (2.111) $\Leftrightarrow$

$$Q_0(\nu)UQ_+(\nu) + Q_-(\nu)UQ_0(\nu) = \mathbf{0} \tag{2.115}$$

(iii) Eq. (2.101) $\Leftrightarrow$ Eq. (2.110) $\Leftrightarrow$

$$Q_0(\nu)UQ_-(\nu) + Q_+(\nu)UQ_0(\nu) = \mathbf{0} \tag{2.116}$$

(iv) Eq. (2.102) $\Leftrightarrow$ Eq. (2.113) $\Leftrightarrow$

$$Q_-(\nu)UQ_+(\nu) = \mathbf{0} \tag{2.117}$$

and (v) Eq. (2.103) $\Leftrightarrow$ Eq. (2.112) $\Leftrightarrow$

$$Q_+(\nu)UQ_-(\nu) = \mathbf{0} \tag{2.118}$$

2.8. Other invariant properties and related algebraic relations

By using Eqs. (2.32) and (2.74)–(2.76), one can show that

$$Q_0(-\nu) = UQ_0(\nu)U, \quad Q_-(-\nu) = UQ_+(\nu)U, \quad \text{and} \quad Q_+(-\nu) = UQ_-(\nu)U \tag{2.119}$$

By using Eqs. (2.85)–(2.87), one can also show that Eq. (2.119) $\Leftrightarrow$

$$\hat{Q}_0(-\nu) = U\hat{Q}_0(\nu)U, \quad \hat{Q}_-(-\nu) = U\hat{Q}_+(\nu)U, \quad \text{and} \quad \hat{Q}_+(-\nu) = U\hat{Q}_-(\nu)U \tag{2.120}$$

As will be shown, the above relations are linked with other invariant properties of the $a(3)$ scheme.

Let the advection speed a in Eq. (1.1) be considered as a variable parameter. Let $u = u(x,t;a)$ be a solution to Eq. (1.1), in the domain $-\infty < x,t,a < +\infty$, i.e.,

$$\frac{\partial u(x,t;a)}{\partial t} + a\frac{\partial u(x,t;a)}{\partial x} \equiv 0, \qquad -\infty < x,t,a < +\infty \tag{2.121}$$

Let

$$x' \stackrel{\text{def}}{=} -x, \quad t' \stackrel{\text{def}}{=} t, \quad \text{and} \quad a' = -a, \qquad -\infty < x,t,a < +\infty \tag{2.122}$$

and

$$\hat{u}(x,t;a) \stackrel{\text{def}}{=} u(-x,t;-a) \tag{2.123}$$

Then (i) Eq. (2.121) $\Leftrightarrow$

$$\frac{\partial u(x',t';a')}{\partial t'} + a'\frac{\partial u(x',t';a')}{\partial x'} \equiv 0, \qquad -\infty < x',t',a' < +\infty \tag{2.124}$$

and (ii)

$$\frac{\partial}{\partial t'} = \frac{\partial}{\partial t} \quad \text{and} \quad \frac{\partial}{\partial x'} = -\frac{\partial}{\partial x} \tag{2.125}$$

Thus one concludes that Eq. (2.121) $\Leftrightarrow$

$$\frac{\partial \hat{u}(x,t;a)}{\partial t} + a\frac{\partial \hat{u}(x,t;a)}{\partial x} \equiv 0, \qquad -\infty < x,t < +\infty \tag{2.126}$$

In other words, if $u = u(x,t;a)$ is a solution to Eq. (1.1), so must be $u = \hat{u}(x,t;a)$ and vice versa. Because the one-to-one mapping

$$(x,t,a) \leftrightarrow (-x,t,-a), \qquad -\infty < x,t,a < +\infty \tag{2.127}$$

represents a combined spatial-reflection (parity) and advection direction reversal (ADR) operation, hereafter (i) a pair of functions such as u and $\hat{u}$ will be referred to as the PADR images of each other; and (ii) a PDE such as Eq. (1.1) is said to be PADR invariant if the PADR image of a solution is also a solution and vice versa.

Because $\nu = a\Delta t/\Delta x$, the numerical analogue of Eq. (2.127) is

$$(j,n) \leftrightarrow (-j,n) \quad \text{and} \quad \nu \leftrightarrow -\nu \tag{2.128}$$

Motivated by an argument similar to that leads to Eq. (2.30) for STI mapping, the PADR mapping for the $a(3)$ scheme is defined by

$$\vec{q}(j,n) \leftrightarrow U\vec{q}(-j,n) \quad \text{and} \quad \nu \leftrightarrow -\nu, \qquad (j,n) \in \Omega \tag{2.129}$$

Thus the PADR image Eq. (2.77) is

$$U\vec{q}(-j,n) = Q_0(-\nu)U\vec{q}(-j,n-1) + Q_+(-\nu)U\vec{q}(-j-1,n-1) + Q_-(-\nu)U\vec{q}(-j+1,n-1), \qquad (j,n) \in \Omega \tag{2.130}$$

By using Eqs. (2.32) and (2.119), it can be shown that Eq. (2.130) $\Leftrightarrow$

$$\vec{q}(-j,n) = Q_0(\nu)\vec{q}(-j,n-1) + Q_-(\nu)\vec{q}(-j-1,n-1) + Q_+(\nu)\vec{q}(-j+1,n-1), \qquad (j,n) \in \Omega \quad (2.131)$$

By replacing the dummy index $-j$ with j everywhere in Eq. (2.131) and using the fact that $(-j,n) \in \Omega \Leftrightarrow (j,n) \in \Omega$, one concludes that Eq. (2.131) $\Leftrightarrow$ Eq. (2.77). Thus Eq. (2.77) is PADR invariant, i.e., it is equivalent to its PADR image.

By exchanging the roles of x and t, one can define invariance under a combined time reversal and advection direction reversal operation. Because (i) this operation is equivalent to a STI operation followed by a PADR operation or vice versa, and (ii) Eq. (1.1) and the $a(3)$ scheme are invariant under both STI and PADR operations, one concludes that Eq. (1.1) and the $a(3)$ scheme are also invariant under the new operation. In fact, invariance of the $a(3)$ scheme under this new operation can be proved using Eq. (2.120) (which is equivalent to Eq. (2.119)).

3. von Neumann analysis

Let $G(\nu,\theta)$ be a 3×3 nonsingular complex matrix function of ν and the phase angle θ such that

$$\vec{q}(j,n) = e^{ij\theta}\,[G(\nu,\theta)]^n\,\vec{b}, \qquad (j,n) \in \Omega;\ -\infty < \nu,\theta < +\infty \qquad (i \equiv \sqrt{-1}) \quad (3.1)$$

is a solution to Eq. (2.77) for all possible complex constant 3×1 column matrices $\vec{b}$. (Note: because $[G(\nu,\theta)]^n \stackrel{\text{def}}{=} \left\{[G(\nu,\theta)]^{-1}\right\}^{|n|}$ for an integer $n < 0$, $[G(\nu,\theta)]^n$ is not defined if $n < 0$ unless $[G(\nu,\theta)]^{-1}$ exists, i.e., $G(\nu,\theta)$ is nonsingular.) By substituting Eq. (3.1) into Eq. (2.77), one has

$$\left[G(\nu,\theta) - Q_0(\nu) - e^{i\theta}Q_+(\nu) - e^{-i\theta}Q_-(\nu)\right][G(\nu,\theta)]^n\,\vec{b} = 0, \qquad n = 0, \pm1, \pm2, \ldots \quad (3.2)$$

Because (i) $[G(\nu,\theta)]^0 = I$, and (ii) $\vec{b}$ can be any complex constant 3×1 column matrix, Eq. (3.2) $\Leftrightarrow$

$$G(\nu,\theta) = Q_0(\nu) + e^{i\theta}Q_+(\nu) + e^{-i\theta}Q_-(\nu) \quad (3.3)$$

By definition, $G(\nu,\theta)$ is the amplification matrix of the forward marching form of the $a(3)$ scheme. Because $Q_0(\nu)$, $Q_+(\nu)$, and $Q_-(\nu)$ are real matrices, Eq. (3.3) implies that

$$G(\nu,-\theta) = \overline{G(\nu,\theta)} \quad (3.4)$$

Hereafter $\overline{M}$ denotes the complex conjugate of any matrix M. Also, with the aid of Eq. (2.119) and the relation $U = U^{-1}$, one has

$$G(-\nu,\theta) = UG(\nu,-\theta)U = UG(\nu,-\theta)U^{-1} \quad (3.5)$$

In the following, we will show that the $a(3)$ scheme must be neutrally stable when it is stable.

3.1. Neutral stability of the $a(3)$ scheme

By using Eqs. (2.114)–(2.118), one can show easily that

$$\begin{aligned} &U\left[Q_0(\nu) + e^{i\theta}Q_-(\nu) + e^{-i\theta}Q_+(\nu)\right]U\left[Q_0(\nu) + e^{i\theta}Q_+(\nu) + e^{-i\theta}Q_-(\nu)\right] \\ &= \left[Q_0(\nu) + e^{i\theta}Q_+(\nu) + e^{-i\theta}Q_-(\nu)\right]U\left[Q_0(\nu) + e^{i\theta}Q_-(\nu) + e^{-i\theta}Q_+(\nu)\right]U = I \end{aligned} \quad (3.6)$$

Thus $G(\nu,\theta)$ defined in Eq. (3.3) is nonsingular and its inverse is

$$[G(\nu,\theta)]^{-1} = U\left[Q_0(\nu) + e^{i\theta}Q_-(\nu) + e^{-i\theta}Q_+(\nu)\right]U \quad (3.7)$$

Indeed, with the aid of Eq. (2.85)–(2.87), Eq. (3.7) is what one obtains after substituting Eq. (3.1) into the backward marching form Eq. (2.88). Moreover, by using Eqs. (2.32), (3.3), (3.4), and (3.7), one has

$$[G(\nu,\theta)]^{-1} = U\overline{G(\nu,\theta)}U^{-1} \tag{3.8}$$

For each (ν,θ), the three eigenvalues $G(\nu,\theta)$ will be denoted as $\sigma_\ell(\nu,\theta)$, $\ell = 1,2,3$, and referred to as the amplification factors of the $a(3)$ scheme. Because $G(\nu,\theta)$ is nonsingular,

$$\sigma_\ell(\nu,\theta) \neq 0, \qquad \ell = 1,2,3;\ -\infty < \nu,\theta < +\infty \tag{3.9}$$

(see part (i) of Theorem 1 given below). Also, as will be shown, $\sigma_\ell(\nu,\theta)$, $\ell = 1,2,3$, satisfy the following set condition:

$$\left\{\frac{1}{\sigma_1(\nu,\theta)},\ \frac{1}{\sigma_2(\nu,\theta)},\ \frac{1}{\sigma_3(\nu,\theta)}\right\} = \left\{\overline{\sigma_1(\nu,\theta)},\ \overline{\sigma_2(\nu,\theta)},\ \overline{\sigma_3(\nu,\theta)}\right\}, \qquad -\infty < \nu,\theta < +\infty \tag{3.10}$$

Hereafter $\overline{z}$ denotes the complex conjugate of any complex number z.

As a preliminary, first we introduce the following matrix theorems:

Theorem 1. Let A be a nonsingular $N \times N$ matrix with the eigenvalues λ_ℓ, $\ell = 1,2,\ldots,N$. Then (i) $\lambda_\ell \neq 0$, $\ell = 1,2,\ldots,N$; and (ii) the eigenvalues of A^{-1} are $1/\lambda_\ell$, $\ell = 1,2,\ldots,N$.

Theorem 2. Let A be a $N \times N$ matrix with the eigenvalues λ_ℓ, $\ell = 1,2,\ldots,N$. Then the eigenvalues of $\overline{A}$, the complex conjugate of A, are $\overline{\lambda_\ell}$, $\ell = 1,2,\ldots,N$.

Theorem 3. Let A and B be two similar $N \times N$ matrices, i.e., there exists a nonsingular $N \times N$ matrix S so that $B = S^{-1}AS$. Then A and B have the same eigenvalues, counting multiplicity.

The proof of Theorems 1 and 2 is given in Appendix A while that of Theorem 3 is given on p. 45 of [76].

To prove Eq. (3.10), note that part (ii) of Theorem 1 implies that, for any (ν,θ), the eigenvalues of $[G(\nu,\theta)]^{-1}$ are $1/\sigma_\ell(\nu,\theta)$, $\ell = 1,2,3$. Next, by using Theorems 2 and 3, and the fact that $(U^{-1})^{-1} = U$, one can see that the eigenvalues of the matrix on the right side of Eq. (3.8) are $\overline{\sigma_\ell(\nu,\theta)}$, $\ell = 1,2,3$. Thus Eq. (3.10) now is an immediate result of Eq. (3.8). QED.

An immediate result of Eq. (3.10) is

$$\frac{1}{\sigma_1(\nu,\theta)} \cdot \frac{1}{\sigma_2(\nu,\theta)} \cdot \frac{1}{\sigma_3(\nu,\theta)} = \overline{\sigma_1(\nu,\theta)} \cdot \overline{\sigma_2(\nu,\theta)} \cdot \overline{\sigma_3(\nu,\theta)}$$

i.e.,

$$|\sigma_1(\nu,\theta)| \cdot |\sigma_2(\nu,\theta)| \cdot |\sigma_3(\nu,\theta)| = 1 \tag{3.11}$$

For any given ν, stability of the $a(3)$ scheme requires that

$$|\sigma_\ell(\nu,\theta)| \leq 1, \qquad \ell = 1,2,3 \tag{3.12}$$

Thus Eq. (3.11) implies that, for any given ν, the $a(3)$ scheme must be neutrally stable, i.e.,

$$|\sigma_\ell(\nu,\theta)| = 1, \qquad \ell = 1,2,3 \tag{3.13}$$

if it is stable. As such, *Eq. (3.8) does not imply neutral stability of the* $a(3)$ *scheme*. However, it does imply that the scheme can only be neutrally stable (i.e., non-dissipative) if it is stable. Here we have reached this conclusion without using the explicit form of $\sigma_\ell(\nu,\theta)$, $\ell = 1,2,3$.

At this juncture, note that one can obtain

$$\sigma_\ell(-\nu,\theta) = \sigma_\ell(\nu,-\theta) = \overline{\sigma_\ell(\nu,\theta)}, \qquad \ell = 1,2,3 \tag{3.14}$$

by using Eqs. (3.4) and (3.5) along with Theorems 2 and 3.

Eq. (3.10) and (3.14) are the fundamental relations governing the eigenvalues of $G(\nu,\theta)$. In the following, we explore other properties of these eigenvalues.

3.2. Characteristic equation of $G(\nu,\theta)$

By using Eqs. (2.74)–(2.76) and (3.3), one has

$$G(\nu,\theta) = \begin{pmatrix} 2-\cos\theta-i\nu\sin\theta & 2\nu(\cos\theta-1)+i(1+\nu^2)\sin\theta & \frac{2(1+2\nu^2)}{3}(1-\cos\theta)-\frac{2i\nu(2+\nu^2)}{3}\sin\theta \\ 2\nu(1-\cos\theta)+i\sin\theta & (2\nu^2-1)\cos\theta-2\nu^2+i\nu\sin\theta & \frac{2\nu(1+2\nu^2)}{3}(1-\cos\theta)+\frac{2i(1-\nu^2)}{3}\sin\theta \\ 3(\cos\theta-1) & 3\nu(1-\cos\theta)-3\,i\sin\theta & 2(1+\nu^2)\cos\theta-1-2\nu^2+2\,i\nu\sin\theta \end{pmatrix}$$

$$-\infty<\nu,\theta<+\infty \tag{3.15}$$

It follows from Eq. (3.15) that (i)

$$\det[G(\nu,\theta)]=-1, \qquad -\infty<\nu,\theta<+\infty \tag{3.16}$$

and (ii) any eigenvalue σ of $G(\nu,\theta)$ must be a root of the characteristic equation:

$$\det[\sigma I-G(\nu,\theta)]\equiv\sigma^3+h(\nu,\theta)\sigma^2+\overline{h(\nu,\theta)}\sigma+1=0 \tag{3.17}$$

where

$$h(\nu,\theta)\stackrel{\text{def}}{=}-1+4\nu^2(1-\cos\theta)-2i\nu\sin\theta \tag{3.18}$$

The reader may be surprised by the simple result Eq. (3.16). However, by using Eq. (3.3) and the fact that each of $Q_0(\nu)$, $Q_+(\nu)$, and $Q_-(\nu)$ has the form $\vec{d}\,\vec{c}^{\,t}$ with $\vec{c}$ and $\vec{d}$ being 3×1 column vectors, an application of the fundamental definition of determinant (in which the Levi-Civita antisymmetric symbol is used) leads to the conclusion that $\det[G(\nu,\theta)]$ must be independent of θ, i.e., $\det[G(\nu,\theta)]=\det[G(\nu,0)]$. As such, Eq. (3.16) now follows from the fact that $G(\nu,0)=U$ (see Eqs. (3.15) and (2.32)) and $\det(U)=-1$. Hereafter, for simplicity, the arguments ν and θ may be omitted if no confusion would arise.

Because σ_1, σ_2, and σ_3 are the eigenvalues of G, Eq. (3.17) implies that

$$\sigma^3+h\sigma^2+\overline{h}\sigma+1\equiv(\sigma-\sigma_1)(\sigma-\sigma_2)(\sigma-\sigma_3) \tag{3.19}$$

for any complex variable σ. On the other hand, because Eq. (3.10) $\Leftrightarrow$

$$\{\sigma_1(\nu,\theta),\,\sigma_2(\nu,\theta),\,\sigma_3(\nu,\theta)\}=\left\{\frac{1}{\overline{\sigma_1(\nu,\theta)}},\,\frac{1}{\overline{\sigma_2(\nu,\theta)}},\,\frac{1}{\overline{\sigma_3(\nu,\theta)}}\right\}, \qquad -\infty<\nu,\theta<+\infty \tag{3.20}$$

$1/\overline{\sigma_1}$, $1/\overline{\sigma_2}$, and $1/\overline{\sigma_3}$ must also be the eigenvalues. Thus

$$\sigma^3+h\sigma^2+\overline{h}\sigma+1\equiv\left(\sigma-\frac{1}{\overline{\sigma_1}}\right)\left(\sigma-\frac{1}{\overline{\sigma_2}}\right)\left(\sigma-\frac{1}{\overline{\sigma_3}}\right) \tag{3.21}$$

for any complex variable σ. In the following, independently, it will be shown that indeed Eq. (3.19) implies Eq. (3.21).

Proof. Let $\sigma = 0$. Then Eq. (3.19) implies that

$$\sigma_1\,\sigma_2\,\sigma_3 = -1 \tag{3.22}$$

Eq. (3.9) is an immediate result of Eq. (3.22). Also, one can see that Eq. (3.21) $\Leftrightarrow$ Eq. (3.22) if $\sigma = 0$.

Let $\sigma \neq 0$. Then, by replacing σ with $1/\overline{\sigma}$ in Eq. (3.19), one has

$$\frac{1}{\overline{\sigma}^3} + \frac{h}{\overline{\sigma}^2} + \frac{\overline{h}}{\overline{\sigma}} + 1 = \left(\frac{1}{\overline{\sigma}} - \sigma_1\right)\left(\frac{1}{\overline{\sigma}} - \sigma_2\right)\left(\frac{1}{\overline{\sigma}} - \sigma_3\right) \tag{3.23}$$

Also, by using Eq. (3.22), one has

$$\overline{\sigma}^3 = (-\overline{\sigma}/\sigma_1)(-\overline{\sigma}/\sigma_2)(-\overline{\sigma}/\sigma_3) \tag{3.24}$$

Because the product of the expressions on the left sides of Eq. (3.23) and (3.24) equals to that on the right sides, we have

$$\overline{\sigma}^3 + \overline{h}\,\overline{\sigma}^2 + h\overline{\sigma} + 1 = \left(\overline{\sigma} - \frac{1}{\sigma_1}\right)\left(\overline{\sigma} - \frac{1}{\sigma_2}\right)\left(\overline{\sigma} - \frac{1}{\sigma_3}\right) \tag{3.25}$$

Eq. (3.21) is the complex conjugate form of Eq. (3.25). QED.

Moreover, according to Eq. (3.18),

$$h(-\nu,\theta) = h(\nu,-\theta) = \overline{h(\nu,\theta)} \tag{3.26}$$

Thus Eq. (3.14) can also be derived directly from Eq. (3.17).

In Sec. 3.3, we will prove the following proposition:

$$|\sigma_\ell(\nu,\theta)| = 1, \qquad \ell = 1,2,3; \quad -\infty < \theta < +\infty \quad \text{if} \quad |\nu| \leq 1/2 \tag{3.27}$$

3.3. A proof of proposition Eq. (3.27)

First we introduce the following well-established algebraic theorem:

Theorem 4. Let $\sigma_1, \sigma_2, \ldots, \sigma_{N'}$ be the distinct roots of an algebraic equation of Nth order, i.e.,

$$\sigma^N + a_1\sigma^{N-1} + a_2\sigma^{N-2} + \ldots + a_{N-1}\sigma + a_N = 0 \tag{3.28}$$

where $a_1, a_2, \ldots, a_N$ are complex constant coefficients and σ is a complex variable. For each $\ell = 1, 2, \ldots, N'$, let $m_\ell \geq 1$ denote the multiplicity of the root σ_ℓ. Then

$$\sum_{\ell=1}^{N'} m_\ell = N \tag{3.29}$$

According to the above theorem, for any given (ν,θ), the roots of the cubic equation Eq. (3.17) must fall into one of the following three mutually exclusive cases: (a) there is one triple root (multiplicity = 3); (b) there are one double root (multiplicity = 2) and one simple root (multiplicity = 1) and (c) there are three simple roots.

Consider case (a). Then $\sigma_1 = \sigma_2 = \sigma_3$. Let σ_o denote the common value of σ_1, σ_2, and σ_3. Then Eqs. (3.9) and (3.20) imply that (i) $\sigma_o \neq 0$ and (ii) $1/\overline{\sigma_o}$ must also be a triple root of Eq. (3.17). Thus the only choice that will not contradict Theorem 4 is that $\sigma_o = 1/\overline{\sigma_o}$, i.e., $|\sigma_o| = |\sigma_1| = |\sigma_2| = |\sigma_3| = 1$.

Consider case (b). Without any loss of generality, one can assume $\sigma_1 = \sigma_2 \neq \sigma_3$. Again let σ_o denote the common value of σ_1 and σ_2. Then Eqs. (3.9) and (3.20) imply that (i) $\sigma_o \neq 0$; (ii) $\sigma_3 \neq 0$; and (iii) $1/\overline{\sigma_o}$

and $1/\overline{\sigma_3}$ must also be a double root and a simple root of Eq. (3.17), respectively. Thus the only choice that will not contradict Theorem 4 is that $\sigma_o = 1/\overline{\sigma_o}$ and $\sigma_3 = 1/\overline{\sigma_3}$, i.e., $|\sigma_o| = |\sigma_1| = |\sigma_2| = |\sigma_3| = 1$.

The conclusions reached above imply the following lemma:

Lemma 1. For any given (ν, θ), the roots of Eq. (3.17) must all be of unit magnitude if any one of them is a multiple root.

Thus, to prove Eq. (3.27), we need only to consider case (c), i.e., the case with

$$\sigma_1 \neq \sigma_2, \quad \sigma_1 \neq \sigma_3, \quad \text{and} \quad \sigma_2 \neq \sigma_3 \tag{3.30}$$

To proceed, each σ_ℓ is expressed in its polar form, i.e.,

$$\sigma_\ell = r_\ell e^{i\phi_\ell}, \qquad \ell = 1, 2, 3 \tag{3.31}$$

where, because of Eq. (3.9)

$$r_\ell = |\sigma_\ell| > 0, \qquad \ell = 1, 2, 3 \tag{3.32}$$

Moreover, we assume that

$$-\pi < \phi_\ell \leq \pi, \qquad \ell = 1, 2, 3 \tag{3.33}$$

Thus, for each σ_ℓ, the value of $e^{i\phi_\ell}$ is unique. It follows from Eqs. (3.31) and (3.32) that

$$1/\overline{\sigma_\ell} = (1/r_\ell)e^{i\phi_\ell}, \qquad \ell = 1, 2, 3 \tag{3.34}$$

Also, by using Eqs. (3.31)–(3.33), Eqs. (3.30) can be expressed as the following ordered pair inequalities:

$$(r_1, \phi_1) \neq (r_2, \phi_2), \quad (r_1, \phi_1) \neq (r_3, \phi_3), \quad \text{and} \quad (r_2, \phi_2) \neq (r_3, \phi_3) \tag{3.35}$$

The distribution of ϕ_1, ϕ_2, and ϕ_3 must fall into one of the following mutually exclusive cases: (c1) all have distinct values; (c2) two of them have the same value while the third assumes a different value; and (c3) all have the same values. In the following, these sub-cases will be discussed separately.

Consider case (c1) where

$$\phi_1 \neq \phi_2, \quad \phi_1 \neq \phi_3, \quad \text{and} \quad \phi_2 \neq \phi_3 \tag{3.36}$$

Because Eqs. (3.20) and (3.34) imply that $(1/r_\ell)e^{i\phi_\ell}$, $\ell = 1, 2, 3$, must also be roots of Eq. (3.17), Eq. (3.36) implies that the only choice that will not contradict Theorem 4 is that $r_\ell = 1/r_\ell$, i.e., $r_\ell = 1$, $\ell = 1, 2, 3$. Thus, *for case (c1), again we have* $|\sigma_1| = |\sigma_2| = |\sigma_3| = 1$.

Consider case (c2) where, without any loss of generality, one can assume that

$$\phi_1 = \phi_2 \neq \phi_3 \tag{3.37}$$

Because of (3.35), Eq. (3.37) implies that

$$r_1 \neq r_2 \tag{3.38}$$

By using Eqs. (3.37) and (3.38) along with the fact that $(1/r_\ell)e^{i\phi_\ell}$, $\ell = 1, 2, 3$, must also be roots of Eq. (3.17), one concludes that the only choice that will not contradict Theorem 4 is that $r_1 r_2 = 1$, $r_1 \neq 1$, $r_2 \neq 1$, and $r_3 = 1$. Thus, *for case (c2), (i) one of the roots is of unit magnitude while the other two are not; and (ii) the product of the magnitudes of the two roots which are not of unit magnitude is one*.

Consider case (c3) where

$$\phi_1 = \phi_2 = \phi_3 \tag{3.39}$$

Because of Eq. (3.35), Eq. (3.39) implies that

$$r_1 \neq r_2, \quad r_1 \neq r_3, \quad \text{and} \quad r_2 \neq r_3 \tag{3.40}$$

By using an argument similar to that invoked in the discussion of case (c2), one concludes that, *for case (c3), again (i) one of the roots is of unit magnitude while the other two are not; and (ii) the product of the magnitudes of the two roots which are not of unit magnitude is one.*

As a result of the above discussions, we have the following lemma:

Lemma 2. For any given (ν, θ), the case with at least one of the roots of Eq. (3.17) not being of unit magnitude may occur only if it meets the following conditions: (i) one and only one of r_1, r_2, and r_3 is of unit magnitude; and (ii) the two roots that are not of unit magnitude share the same phase angle and the product of their magnitudes is one.

Consider any case that meets the conditions referred to in Lemma 2. Then, without any loss of generality, one may assume that

$$r_1 r_2 = 1, \quad r_1 \neq 1, \quad r_3 = 1, \quad \text{and} \quad \phi_1 = \phi_2 = \phi \tag{3.41}$$

where ϕ denotes the common value of ϕ_1 and ϕ_2. Moreover, by using Eqs. (3.31) and (3.34), Eqs. (3.18), (3.19), and (3.21) imply that

$$r_1 r_2 r_3 e^{i(\phi_1+\phi_2+\phi_3)} = \frac{1}{r_1 r_2 r_3} e^{i(\phi_1+\phi_2+\phi_3)} = -1 \tag{3.42}$$

$$\begin{aligned} r_1 r_2 e^{i(\phi_1+\phi_2)} + r_1 r_3 e^{i(\phi_1+\phi_3)} + r_2 r_3 e^{i(\phi_2+\phi_3)} &= \frac{1}{r_1 r_2} e^{i(\phi_1+\phi_2)} + \frac{1}{r_1 r_3} e^{i(\phi_1+\phi_3)} + \frac{1}{r_2 r_3} e^{i(\phi_2+\phi_3)} \\ &= -1 + 4\nu^2(1-\cos\theta) + 2i\nu\sin\theta \end{aligned} \tag{3.43}$$

and

$$r_1 e^{i\phi_1} + r_2 e^{i\phi_2} + r_3 e^{i\phi_3} = \frac{1}{r_1} e^{i\phi_1} + \frac{1}{r_2} e^{i\phi_2} + \frac{1}{r_3} e^{i\phi_3} = 1 - 4\nu^2(1-\cos\theta) + 2i\nu\sin\theta \tag{3.44}$$

Because of Eq. (3.32), Eq. (3.42) $\Leftrightarrow$

$$r_1 r_2 r_3 = 1 \tag{3.45}$$

and

$$e^{i(\phi_1+\phi_2+\phi_3)} = -1 \tag{3.46}$$

By using Eqs. (3.45) and (3.46), Eq.(3.43) $\Leftrightarrow$

$$r_1 e^{-i\phi_1} + r_2 e^{-i\phi_2} + r_3 e^{-i\phi_3} = \frac{1}{r_1} e^{-i\phi_1} + \frac{1}{r_2} e^{-i\phi_2} + \frac{1}{r_3} e^{-i\phi_3} = 1 - 4\nu^2(1-\cos\theta) - 2i\nu\sin\theta \tag{3.47}$$

Because Eq. (3.47) is the complex conjugate of Eq. (3.44), we conclude that Eqs. (3.44)–(3.46) $\Leftrightarrow$ Eqs. (3.42)–(3.44), i.e., Eqs. (3.44)–(3.46) represent all the independent constraints imposed on r_ℓ and ϕ_ℓ, $\ell = 1, 2, 3$.

Note that Eq. (3.45) is satisfied by Eq. (3.41) automatically. However, Eqs. (3.41) and (3.46) imply that

$$e^{i\phi_3} = -e^{-i(\phi_1+\phi_2)} = -e^{-2i\phi} \tag{3.48}$$

Let

$$\rho \stackrel{\text{def}}{=} r_1 \tag{3.49}$$

Then, with the aid of Eqs. (3.41) and (3.48), Eq. (3.44) can be expressed as

$$f(\rho) e^{i\phi} - e^{-2i\phi} = 1 - 4\nu^2(1-\cos\theta) + 2i\nu\sin\theta, \qquad -\pi < \phi \leq \pi; \ \rho > 0 \text{ and } \rho \neq 1 \tag{3.50}$$

where

$$f(\rho) \stackrel{\text{def}}{=} \rho + \frac{1}{\rho}, \qquad \rho > 0 \text{ and } \rho \neq 1 \tag{3.51}$$

Eq. (3.50) $\Leftrightarrow$

$$f(\rho)\cos\phi - \cos(2\phi) = 1 - 4\nu^2(1-\cos\theta), \qquad -\pi < \phi \leq \pi; \ \rho > 0 \text{ and } \rho \neq 1 \tag{3.52}$$

and

$$f(\rho)\sin\phi + \sin(2\phi) = 2\nu\sin\theta, \qquad -\pi < \phi \leq \pi;\ \rho > 0 \text{ and } \rho \neq 1 \tag{3.53}$$

Thus, given any (ν, θ), Eqs. (3.52) and (3.53) must admit a solution for ρ and ϕ in the specified domain if the case Eq. (3.41) indeed exists.

To explore Eqs. (3.52) and (3.53), note that (i)

$$[f(\rho)\cos\phi - \cos(2\phi)]^2 + [f(\rho)\sin\phi + \sin(2\phi)]^2 = \left[1 - 4\nu^2(1-\cos\theta)\right]^2 + [2\nu\sin\theta]^2 \tag{3.54}$$

is a direct result of Eqs. (3.52) and (3.53); (ii)

$$[f(\rho)\cos\phi - \cos(2\phi)]^2 + [f(\rho)\sin\phi + \sin(2\phi)]^2 \equiv [f(\rho)-1]^2 + 2f(\rho)\,[1-\cos(3\phi)] \tag{3.55}$$

and (iii)

$$\left[1 - 4\nu^2(1-\cos\theta)\right]^2 + [2\nu\sin\theta]^2 \equiv 1 - 4\nu^2(1-4\nu^2)(1-\cos\theta)^2 \tag{3.56}$$

Next, because (i) the minimum of $f(\rho)$ in the domain $\rho > 0$ occurs at $\rho = 1$ and (ii) $f(1) = 2$, we have

$$f(\rho) > 2 \quad \text{if} \quad \rho > 0 \text{ and } \rho \neq 1 \tag{3.57}$$

Combining Eqs. (3.55) and (3.57), and using the fact that $1 - \cos(3\phi) \geq 0$ for all ϕ, one has

$$[f(\rho)\cos\phi - \cos(2\phi)]^2 + [f(\rho)\sin\phi + \sin(2\phi)]^2 > 1 \quad \text{if} \quad \rho > 0 \text{ and } \rho \neq 1 \tag{3.58}$$

On the other hand, because $1 - 4\nu^2 \geq 0$ if $|\nu| \leq 1/2$, Eq. (3.56) implies that

$$\left[1 - 4\nu^2(1-\cos\theta)\right]^2 + [2\nu\sin\theta]^2 \leq 1 \quad \text{if} \quad |\nu| \leq 1/2 \tag{3.59}$$

Combining Eqs. (3.58) and (3.59), one arrives at the conclusion that, for all θ,

$$[f(\rho)\cos\phi - \cos(2\phi)]^2 + [f(\rho)\sin\phi + \sin(2\phi)]^2 > \left[1 - 4\nu^2(1-\cos\theta)\right]^2 + [2\nu\sin\theta]^2 \tag{3.60}$$

i.e., Eq. (3.54) cannot be satisfied, if (i) $|\nu| \leq 1/2$; and (ii) $\rho > 0$ and $\rho \neq 1$. Because Eq. (3.54) is a direct result of Eqs. (3.52) and (3.53), this implies that, for any θ, Eqs. (3.52) and (3.53) admit no solution for ρ and ϕ in the specified domain, i.e., the case Eq. (3.41) does not exist. In turn, this implies that, for all θ, the roots of Eq. (3.17) are all of unit magnitude if $|\nu| \leq 1/2$, i.e., Eq. (3.27) has been proved. QED.

Note that, by itself, Eq. (3.27) does not imply that the $a(3)$ scheme is stable when $|\nu| \leq 1/2$. According to the discussion given in Sec. 4 of [1] and Secs. 3 and 4 of [64], the fact that all its amplification factors are of unit magnitude for all phase angles at a value of ν insures that the $a(3)$ scheme is stable at this particular value of ν only if the amplification matrix $G(\nu, \theta)$ is nondefective for all θ. By definition [76], for any given (ν, θ), the 3×3 matrix $G(\nu, \theta)$ is nondefective if the dimension of its eigenspace (i.e., the vector space spanned by its eigenvectors) is three. On the other hand, $G(\nu, \theta)$ is defective if the dimension of its eigenspace is less than three. It can be shown that $G(\nu, \theta)$ is defective if and only if its Jordan canonical form [76] contains a Jordan block with its dimension > 1. As such $G(\nu, \theta)$ is defective if and only if (i) it has an eigenvalue with multiplicity > 1 and (ii) the Jordan block associated with the eigenvalue has a dimension > 1. For the defective case, stability requires a stronger condition, i.e., the magnitude of the eigenvalue associated with the Jordan block with its dimension > 1 must be less than 1.

Eq. (3.27) and its proof strongly suggest that the stability boundary of the $a(3)$ scheme is defined by $|\nu| = 1/2$. In fact this conjecture has been verified numerically. We conclude this section by studying this special case.

3.4. The $|\nu| = 1/2$ case

Let $\nu = 1/2$. Then Eq. (3.17) reduces to

$$\sigma^3 - e^{i\theta}\sigma^2 - e^{-i\theta}\sigma + 1 \equiv \left(\sigma - e^{i\theta}\right)\left(\sigma - e^{-i\theta/2}\right)\left(\sigma + e^{-i\theta/2}\right) = 0 \qquad (\nu = 1/2) \tag{3.61}$$

Thus the eigenvalues σ are

$$\sigma = \sigma_{\pm}(\theta) \stackrel{\text{def}}{=} \pm e^{-i\theta/2} \quad \text{and} \quad \sigma = \sigma_0(\theta) \stackrel{\text{def}}{=} e^{i\theta} \qquad (\nu = 1/2) \tag{3.62}$$

On the other hand, by using Eqs. (3.14) and (3.62), one concludes that the eigenvalues σ for the case $\nu = -1/2$ are

$$\sigma = \overline{\sigma_{\pm}(\theta)} = \pm e^{i\theta/2} \quad \text{and} \quad \sigma = \overline{\sigma_0(\theta)} = e^{-i\theta} \qquad (\nu = -1/2) \tag{3.63}$$

For each of the above two cases, Eqs. (3.62) and (3.63) imply that the three eigenvalues are distinct if

$$\theta \neq \pm\frac{2n\pi}{3}, \qquad n = 0, \pm 1, \pm 2, \ldots \tag{3.64}$$

Also, because the analytical amplification factor is $e^{-i\nu\theta}$ for any (ν, θ) (see p.4 of [61]), for the case $\nu = 1/2$ ($\nu = -1/2$), one of the roots of Eq. (3.17), i.e., $\sigma_+(\theta)$ ($\overline{\sigma_+(\theta)}$), is identical to the analytical amplification factor.

Consider the plane wave solution

$$u(x,t) = e^{ik(x-at)} \tag{3.65}$$

The period associated with this solution is

$$T = \frac{2\pi}{|ka|} \tag{3.66}$$

Let n, the number of total marching steps, and Δt be chosen such that

$$n\Delta t = NT, \qquad N = 1, 2, 3, \ldots \tag{3.67}$$

Then one has

$$n = \frac{2\pi N}{|\theta||\nu|} \tag{3.68}$$

where

$$\theta = k\Delta x \tag{3.69}$$

is the variation of phase angle over the interval Δx. For the case $|\nu| = 1/2$, Eq. (3.68) reduces to

$$n = \frac{4\pi N}{|\theta|} \tag{3.70}$$

Eqs. (3.62), (3.63), and (3.70) imply that

$$[\sigma_{\pm}(\theta)]^n = \left[\overline{\sigma_{\pm}(\theta)}\right]^n = (\pm 1)^n \tag{3.71}$$

and

$$[\sigma_0(\theta)]^n = \left[\overline{\sigma_0(\theta)}\right]^n = 1 \tag{3.72}$$

Thus

$$[\sigma_{\pm}(\theta)]^n = \left[\overline{\sigma_{\pm}(\theta)}\right]^n = 1, \qquad \text{if } n \text{ is even} \tag{3.73}$$

On the other hand, Eq. (3.68) implies that

$$\left(e^{-i\nu\theta}\right)^n = 1 \tag{3.74}$$

Eqs. (3.72)–(3.74) and the fact that $e^{-i\nu\theta}$ is the analytical amplification factor imply that the numerical solution generated by the $a(3)$ scheme in a simulation involving a periodic boundary condition, aside from round-off errors, should be identical to the exact solution if n and Δt are chosen according to Eq. (3.67), and n is even, and (ii) the phase angles of the Fourier components involved in the simulation observe the condition Eq. (3.64) (see Sec. 5 in [1]). This prediction has been verified numerically (see Sec. 4).

4. Numerical results

To assess the accuracy of the $a(3)$ scheme, consider the model problem with the PDE

$$\frac{\partial u}{\partial t} + \frac{\partial u}{\partial x} = 0 \tag{4.1}$$

and the exact solution

$$u = u_e(x,t) \stackrel{\text{def}}{=} \sin\left[2\pi(x-t)\right] \equiv \frac{1}{2i}\left[e^{i2\pi(x-t)} - e^{-i2\pi(x-at)}\right] \tag{4.2}$$

We have

$$a = L = T = 1 \tag{4.3}$$

where L = wavelength and T = period. Let (i)

$$u_{xe}(x,t) \stackrel{\text{def}}{=} \frac{\partial u_e(x,t)}{\partial x} \quad \text{and} \quad u_{xxe}(x,t) \stackrel{\text{def}}{=} \frac{\partial^2 u_e(x,t)}{\partial x^2} \tag{4.4}$$

and (ii) the spatial domain of unit length be divided into K uniform intervals. Thus

$$\Delta x = 1/K, \quad \Delta t = \nu \Delta x \quad \text{and} \quad t = n\Delta t \tag{4.5}$$

where n = number of time steps, and t = total marching time.

It has been shown numerically that the $a(3)$ scheme is stable if

$$|\nu| < 1/2 \tag{4.6}$$

On the other hand, (i) the a scheme is stable if

$$|\nu| < 1 \tag{4.7}$$

and (ii) the $a(4)$ scheme [72], another high order solver of Eq. (1.1), is stable if

$$|\nu| < 1/3 \tag{4.8}$$

In Tables 1–4, the numerical errors of several computations using the $a(3)$ and a schemes are presented in terms of the following error norms for the *non-normalized* mesh variables:

$$E(K,n,\nu) \stackrel{\text{def}}{=} \sqrt{\frac{1}{K}\sum_{j=0}^{K-1}[u_j^n - u_e(x_j,t^n)]^2} \tag{4.9}$$

$$E_x(K,n,\nu) \stackrel{\text{def}}{=} \sqrt{\frac{1}{K}\sum_{j=0}^{K-1}[(u_x)_j^n - u_{xe}(x_j,t^n)]^2} \tag{4.10}$$

and

$$E_{xx}(K,n,\nu) \stackrel{\text{def}}{=} \sqrt{\frac{1}{K}\sum_{j=0}^{K-1}[(u_{xx})_j^n - u_{xxe}(x_j,t^n)]^2} \tag{4.11}$$

The numerical errors of several simulations with $\nu = 0.1$ and $t = 9.876$ are given in Table 1. For the a scheme, as the values of K and n become larger, the values of E and E_x are both reduced by a factor of about 4 as both K and n double their values, i.e., the scheme is 2nd order in accuracy for both u_j^n and $(u_x)_j^n$. On the other hand, for the $a(3)$ scheme, the values of E, E_x, and E_{xx} are reduced by the factors 16, 16, and 4, respectively. Thus the $a(3)$ scheme is 4th order in accuracy for both u_j^n and $(u_x)_j^n$ while only 2nd order in accuracy for $(u_{xx})_j^n$. From the results shown, one can see that the $a(3)$ scheme is much more accurate than the a scheme. As an example, for the case with $K = 25$ and $n = 2469$, the value of E for the a scheme is larger than that for the $a(3)$ scheme by a factor of 3450! Because the a scheme is only 2nd order in accuracy for u_j^n, it is estimated that the accuracy of u_j^n achieved by the $a(3)$ scheme with $K = 25$ and $n = 2469$ is identical to that achieved by the a scheme with $K = 25 \times \sqrt{3450} \approx 1468$ and $n = 2469 \times \sqrt{3450} \approx 145029$.

In Table 2, the cases considered have $\nu = 0.1$ and $t = 10.00 = 10T$. For these cases where t is an integer multiple of the period T, it is seen that the $a(3)$ scheme is 4th order in accuracy for u_j^n, $(u_x)_j^n$, *and* $(u_{xx})_j^n$.

In Table 3, the cases considered have $\nu = 0.5$ and $t = 49.38$, For these cases where the value of ν is right at the stability boundary of the $a(3)$ scheme, aside from round-off errors, the numerical values of $(u_x)_j^n$ generated using the $a(3)$ scheme are all identical to their exact solution values, respectively, if n are *even* integers.

In Table 4, the cases considered have $\nu = 0.5$ and $t = 50.00 = 50T$, For these cases where (i) the value of ν is right at the stability boundary of the $a(3)$ scheme and (ii) t is an integer multiple of T, aside from round-off errors, the numerical values of u_j^n, $(u_x)_j^n$, and $(u_{xx})_j^n$ generated using the $a(3)$ scheme are all identical to their exact solution values, respectively. Note that: (i) n and Δt are chosen according to Eq. (3.67) and n is even for each of these cases, and (ii) the exact solution are a linear combination of two plane wave solutions with $|\theta| = 2\pi\Delta x \leq 2\pi/25$, i.e., θ observes the condition Eq. (3.64). Thus the results shown in Table 4 confirm the prediction made at the end of Sec. 3.

5. Conclusions and discussions

A thorough and rigorous discussion of a new high order neutrally stable CESE solver of Eq. (1.1) has been presented. Because this two-level explicit scheme is associated with three independent mesh variables and three equations per mesh point, it is referred to as the $a(3)$ scheme. As in the case of other similar CESE neutrally stable solvers [1,5,11,72], the $a(3)$ scheme enforces conservation laws locally and globally, and it has the basic, forward marching, and backward marching forms. These forms are equivalent and satisfy the STI invariant property defined in Sec. 2.

Based on the concept of STI invariance, a set of algebraic relations (Eqs. (2.114)–(2.118)) involving the coefficient matrices $Q_0(\nu)$, $Q_+(\nu)$ and $Q_-(\nu)$ is developed in Sec. 2. As it turns out, these relations can be used to construct a simple proof for the fact that the $a(3)$ scheme is neutrally stable (i.e., non-dissipative) when it is stable. Another set of algebraic relations (Eq. (2.119)) involving the same matrices and its relation to other invariance property are also discussed in Sec. 2.

In addition to establishing the neutral stability of $a(3)$ scheme, in Sec. 3, it is also proved analytically that all three amplification factors of the $a(3)$ scheme are of unit magnitude for all phase angles if $|\nu| \leq 1/2$. This amazing theoretical result is completely consistent with the numerical stability condition $|\nu| < 1/2$.

It is shown in Sec. 4 that the $a(3)$ scheme generally is (i) 4th-order accurate for the mesh variables u_j^n and $(u_x)_j^n$; and 2nd-order accurate for $(u_{xx})_j^n$. However, in some exceptional cases, the scheme can achieve perfect accuracy aside from round-off errors. The stability bound $|\nu| < 1/2$ of the $a(3)$ scheme is lower than the stability bound $|\nu| < 1$ of the a scheme. but high than the stability bound $|\nu| < 1/3$ of the $a(4)$ scheme.

The CESE development has been driven by a basic idea that each practical scheme be built from a non-dissipative core scheme so that the numerical dissipation can be controlled effectively. As such, development of the $a(3)$ and $a(4)$ schemes provides a solid foundation for the development of other more practical high order CESE schemes.

Appendix A. Proof for Theorems 1 and 2 in Sec. 3

First we prove Theorem 1. According to Eq. (8.20) on p.265 of [75], the fact that λ_ℓ, $\ell = 1, 2, \ldots, N$ are the eigenvalues of the $N \times N$ matrix $A \Leftrightarrow$

$$\det(A - \lambda I) = (\lambda_1 - \lambda)(\lambda_2 - \lambda)\cdots(\lambda_N - \lambda) \tag{A.1}$$

where λ is any complex variable and I is the $N \times N$ identity matrix. Let $\lambda = 0$. Eq. (A.1) implies that

$$\det(A) = \lambda_1\lambda_2\cdots\lambda_N \tag{A.2}$$

By definition, A is nonsingular $\Leftrightarrow \det(A) \neq 0$. Thus part (i) follows from Eq. (A.2).

According to Eq. (A.1), to prove part (ii) we need only to show that

$$\det\left(A^{-1} - \lambda I\right) = \left(\frac{1}{\lambda_1} - \lambda\right)\left(\frac{1}{\lambda_2} - \lambda\right)\cdots\left(\frac{1}{\lambda_N} - \lambda\right) \tag{A.3}$$

for any complex variable λ. Because $\det(BC) = \det(B)\det(C)$ for any two $N \times N$ matrices B and C, we have $\det(A)\det\left(A^{-1}\right) = \det\left(AA^{-1}\right) = \det(I) = 1$, i.e., $\det(A^{-1}) = 1/\det(A)$. Thus Eq. (A.2) implies that

$$\det\left(A^{-1}\right) = \frac{1}{\lambda_1\lambda_2\cdots\lambda_N} \tag{A.4}$$

By comparing Eqs. (A.3) and (A.4), one concludes that Eq. (A.3) is valid if $\lambda = 0$.

Let $\lambda \neq 0$. Then

$$A^{-1} - \lambda I = -\lambda I A^{-1}\left(A - \frac{1}{\lambda}I\right) \tag{A.5}$$

Thus

$$\det\left(A^{-1} - \lambda I\right) = \det(-\lambda I)\det\left(A^{-1}\right)\det\left(A - \frac{1}{\lambda}I\right) \tag{A.6}$$

With the aid of (i) Eqs. (A.1) and (A.4) and (ii) the fact that $\det(-\lambda I) = (-\lambda)^N$, Eq. (A.6) implies that

$$\begin{aligned}\det\left(A^{-1} - \lambda I\right) &= \frac{(-\lambda)^N}{\lambda_1\lambda_2\cdots\lambda_N}\left(\lambda_1 - \frac{1}{\lambda}\right)\left(\lambda_2 - \frac{1}{\lambda}\right)\cdots\left(\lambda_N - \frac{1}{\lambda}\right) = \\ &\left(\frac{-\lambda}{\lambda_1}\right)\left(\lambda_1 - \frac{1}{\lambda}\right)\left(\frac{-\lambda}{\lambda_2}\right)\left(\lambda_2 - \frac{1}{\lambda}\right)\cdots\left(\frac{-\lambda}{\lambda_N}\right)\left(\lambda_N - \frac{1}{\lambda}\right) = \left(\frac{1}{\lambda_1} - \lambda\right)\left(\frac{1}{\lambda_2} - \lambda\right)\cdots\left(\frac{1}{\lambda_N} - \lambda\right)\end{aligned} \tag{A.7}$$

i.e., Eq. (A.3) is also valid if $\lambda \neq 0$. Thus part (ii) of Theorem 1 has been proved. QED.

According to Eq. (A.1), to prove Theorem 2, we need only to show that

$$\det(\overline{A} - \lambda I) = (\overline{\lambda_1} - \lambda)(\overline{\lambda_2} - \lambda)\cdots(\overline{\lambda_N} - \lambda) \tag{A.8}$$

Because $\det(\overline{M}) = \overline{\det(M)}$, by using Eq. (A.1), we have

$$\begin{aligned}\det(\overline{A} - \lambda I) &= \det\left(\overline{A - \overline{\lambda} I}\right) = \overline{\det(A - \overline{\lambda} I)} \\ &= \overline{(\lambda_1 - \overline{\lambda})(\lambda_2 - \overline{\lambda})\cdots(\lambda_N - \overline{\lambda})} = (\overline{\lambda_1} - \lambda)(\overline{\lambda_2} - \lambda)\cdots(\overline{\lambda_N} - \lambda)\end{aligned} \tag{A.9}$$

i.e., Eq. (A.8) has been proved. QED.

References

1. S.C. Chang and W.M. To, *A New Numerical Framework for Solving Conservation Laws–The Method of Space-Time Conservation Element and Solution Element*, NASA TM 104495, August 1991.
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, USA*, edited by W.G. Vogt and M.H. Mickle, Part 5, p. 2727. Also published as 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 414, (Springer-Verlag, New York/Berlin, 1992), p. 396.
4. S.C. Chang, X.Y. Wang, and C.Y. Chow, The method of Space-Time Conservation Element and Solution Element–Application to One-Dimensional and Two-Dimensional Time-Marching Flow Problems, AIAA Paper 95-1754-CP, appears in *A Collection of Technical Papers, Part 2, pp. 1258–1291, 12th AIAA CFD Conference, June 19-22, 1995, San Diego, California*, Also published as NASA TM 106915 (1995).
5. S.C. Chang, The method of space-time conservation element and solution Element—A new approach for solving the Navier-Stokes and Euler equations, *J. Comput. Phys.*, **119**, 295 (1995).
6. S.C. Chang, S.T. Yu, A. Himansu, X.Y. Wang, C.Y. Chow, and C.Y. Loh, The method of space-time conservation element and solution element—A new paradigm for numerical solution of conservation laws, in *Computational Fluid Dynamics Review 1998* edited by M.M. Hafez and K. Oshima (World Scientific, Singapore), Vol. 1, p. 206.
7. T. Molls and F. Molls, Space-Time Conservation Method Applied to Saint Venant Equations, *J. of Hydraulic Engr.*, **124(5)**, 501 (1998).
8. C. Zoppou and S. Roberts, Space-Time Conservation Method Applied to Saint Venant Equations: A Discussion, *J. of Hydraulic Engr.*, **125(8)**, 891 (1999).
9. S.C. Chang, X.Y. Wang, and C.Y. Chow, The space-time conservation element and solution element method: A new high-resolution and genuinely multidimensional paradigm for solving conservation laws, *J. Comput. Phys.*, **156**, 89 (1999).
10. X.Y. Wang, and S.C. Chang, A 2D non-splitting unstructured triangular mesh Euler solver based on the space-time conservation element and solution element method, *Computational Fluid Dynamics Journal*, **8(2)**, 309 (1999).
11. S.C. Chang, X.Y. Wang and W.M. To, Application of the space-time conservation element and solution element method to one-dimensional convection-diffusion problems, *J. Comput. Phys.*, **165**, 189 (2000).
12. J. Qin, S.T. Yu, Z.C. Zhang, and M.C. Lai, Direct Calculations of Cavitating Flows by the Space-Time CE/SE Method, *J. Fuels & Lubricants, SAE Transc.*, **108(4)**, 1720 (2000).
13. C.Y. Loh, L.S. Hultgren and S.C. Chang, Wave computation in compressible flow using the space-time conservation element and solution element method, *AIAA J.*, **39(5)**, 794 (2001).
14. Z.C. Zhang, S.T. Yu, and S.C. Chang, A Space-Time Conservation Element and Solution Element Method for Solving the Two- and Three-Dimensional Unsteady Euler Equations Using Quadrilateral and Hexahedral Meshes, *J. Comput. Phys.*, **175**, 168 (2002).
15. K.B.M.Q. Zaman, M.D. Dahl, T.J. Bencic, and C.Y. Loh, Investigation of A 'Transonic Resonance' with Convergent-Divergent Nozzles, *J. Fluid Mech.*, **463**, 313 (2002).
16. C.Y. Loh and K.B.M.Q. Zaman, Numerical Investigation of 'Transonic Resonance' with A Convergent-Divergent Nozzle, *AIAA J.*, **40(12)**, 2393 (2002).
17. S. Motz, A. Mitrovic, and E.-D. Gilles, Comparison of Numerical Methods for the Simulation of Dispersed Phase Systems, *Chemical Engineering Science*, **57**, 4329 (2002).
18. S. Cioc and T.G. Keith, Application of the CE/SE Method to One-Dimensional Flow in Fluid Film Bearings, *STLE Tribology Transactions*, **45**, 167 (2002).

19. S. Cioc and T.G. Keith, Application of the CE/SE Method to Two-Dimensional Flow in Fluid Film Bearings, *International J. of Numer. Methods for Heat & Fluid Flow*, **13(2)**, 216 (2003).
20. S. Cioc, F. Dimofte, T.G. Keith, and D.P. Fleming, Computation of Pressurized Gas Bearings Using the CE/SE Method, *STLE Tribology Transactions*, **46(1)**, 128 (2003).
21. S. Cioc, F. Dimofte, and T.G. Keith, Application of the CE/SE Method to Wave Journal Bearings, *STLE Tribology Transactions*, **46(2)**, 179 (2003).
22. A. Ayasoufi and T.G. Keith, Application of the Conservation Element and Solution Element Method in Numerical Modeling of Heat Conduction with Melting and/or Freezing, *International J. of Numer. Methods for Heat & Fluid Flow*, **13(4)**, 448 (2003).
23. A. Ayasoufi and T.G. Keith, Application of the Conservation Element and Solution Element Method in Numerical Modeling of Axisymmetric Heat Conduction with Melting and/or Freezing, *JSME International J. Series B*, **47(1)**, 115 (2004).
24. A. Ayasoufi and T.G. Keith, Application of the Conservation Element and Solution Element Method in Numerical Modeling of Three-dimensional Heat Conduction with Melting and/or Freezing, *Transactions of the ASME, J. of Heat Transfer*, **126(6)**, 937 (2004).
25. Y.I. Lim, S.C. Chang, and S.B. Jorgensen, A Novel Partial Differential Algebraic Equation (PDAE) Solver: Iterative Space-Time Conservation Element/Solution Element (CE/SE) Method, *Computers and Chemical Engineering*, **28**, 1309 (2004)
26. Y.I. Lim and S.B. Jorgensen, A Fast and Accurate Numerical Method for Solving Simulated Moving Bed (SMB) Chromatographic Separation Problems, *Chemical Engineering Science*, **59**, 1931 (2004).
27. Y.I. Lim, An Optimization Strategy for Nonlinear Simulated Moving Bed Chromatography: Multi-level Optimization Procedure (MLOP), *Korea J. Chem. Eng.*, 21(4), 836 (2004).
28. C.K. Kim, S.T. John Yu, and Z.C. Zhang, Cavity Flow in Scramjet Engine by the Space-Time Conservation Element and Solution Element Method, *AIAA J.*, **42(5)**, 912 (2004).
29. M. Zhang, S.T. John Yu, S.C. Lin, S.C. Chang, and I. Blankson, Solving Magnetohydrodynamic Equations Without Special Treatment for Divergence-Free Magnetic Field, *AIAA J.*, **42(12)**, 2605 (2004).
30. K.S. Im, M.C. Lai, S.T. John Yu, and Robert R. Matheson, Jr., Simulation of Spray Transfer Process in Electrostatic Rotary Bell Sprayer, *ASME J. of Fluid Engineering*, **126(3)**, 449 (2004).
31. S. Jerez, J.V. Romero, and M.D. Rosello, A Semi-Implicit Space-Time CE-SE Method to Improve Mass Conservation through Tapered Ducts in Internal Combustion Engines, *Math. and Computer Modeling*, **40**, 941 (2004).
32. S.C. Chang, Y. Wu, V. Yang, and X.Y. Wang, Local Time Stepping Procedures for the Space-Time Conservation Element and Solution Element Method, *International J. Comput. Fluid Dynamics*, **19(5)**, 359 (2005).
33. Y.I. Lim, S.B. Jorgensen, and I.H. Kim, Computer-Aided Model Analysis for Ionic Strength-Dependent Effective Charge of Protein in Ion-Exchange Chromatography, Bio-Chem. Eng. J., **25(2)**, 125 (2005).
34. T.I. Tseng and R.J. Yang, Simulation of the Mach Reflection in Supersonic Flows by the CE/SE Method, *Shock Waves*, **14(4)**, 307 (2005).
35. B. Wang, H. He, and S.-T.J. Yu, Direct Calculation of Wave Implosion for Detonation Initiation, *AIAA J*, **43(10)**, 2157 (2005).
36. T.I. Tseng and R.J. Yang, Numerical Simulation of Vorticity Production in Shock Diffraction, accepted for publication in *AIAA J*.
37. M. Zhang, S.-T. Yu, S.C.H. Lin, S.C. Chang, and I. Blankson, Solving the MHD Equations By the Space-Time Conservation Element and Solution Element Method, *J. Comput. Phys.*, **214**, 599 (2006).
38. S. Qamar and G. Warnecke, A Space-Time Conservation Method for Hyperbolic Systems with Stiff and Non Stiff Source Terms, *Commun. Comput. Phys.*, **1(3)**, 449 (2006).
39. X.Y. Wang, C.Y. Chow, and S.C.Chang, *Numerical Simulation of Flows Caused by Shock-Body Interaction*, AIAA Paper 96-2004 (1996).
40. C.Y. Loh, L.S. Hultgren and S.C. Chang, *Vortex Dynamics Simulation in Aeroacoustics by the Space-Time Conservation Element and Solution Element Method*, AIAA Paper 99-0359 (1999)

41. X.Y. Wang, S.C. Chang and P.C.E. Jorgenson, *Accuracy Study of the Space-Time CE/SE Method for Computational Aeroacoustics Problems Involving Shock Waves*, AIAA Paper 2000-0474 (2000).
42. C.Y. Loh, L.S. Hultgren, S.C. Chang and P.C.E. Jorgenson, *Noise Computation of a Supersonic Shock-Containing Axisymmetric Jet by the CE/SE Method*, AIAA Paper 2000-0475 (2000).
43. C.Y. Loh, X.Y. Wang, S.C. Chang, and P.C.E. Jorgenson, Computation of Feedback Aeroacoustic System by the CE/SE Method, in *Proceedings of the First International Conference on Computational Fluid Dynamics, Kyoto, Japan, 10-14 July, 2000*, edited by N. Satofuka, (Springer-Verlag Berlin Heidelberg 2001), p. 555.
44. C.Y. Loh, L.S. Hultgren and P.C.E. Jorgenson, *Near Field Screech Noise Computation for An Underexpanded Supersonic Jet by the CE/SE Method*, AIAA Paper 2001-2252 (2001).
45. X.Y. Wang, S.C. Chang, and P.C.E. Jorgenson, Numerical Simulation of Aeroacoustic Field in a 2D Cascade Involving a Downstream Moving Grid Using the Space-Time CE/SE method, in *Proceedings of the First International Conference on Computational Fluid Dynamics, Kyoto, Japan, 10-14 July, 2000*, edited by N. Satofuka, (Springer-Verlag Berlin Heidelberg 2001), p. 543.
46. X.Y. Wang, S.C. Chang, A. Himansu, and P.C.E. Jorgenson, *Gust Acoustic Response of A Single Airfoil Using the Space-Time CE/SE Method*, AIAA Paper 2002-0801 (2002).
47. S.T. Yu and S.C. Chang, *Treatments of Stiff Source Terms in Conservation Laws by the Method of Space-Time Conservation Element and Solution Element*, AIAA Paper 97-0435 (1997).
48. S.T. Yu and S.C. Chang, Applications of the Space-Time Conservation Element / Solution Element Method to Unsteady Chemically Reactive Flows," AIAA Paper 97-2099, in *A Collection of Technical Papers, 13th AIAA CFD Conference*, June 29-July 2, 1997, Snowmass, CO.
49. S.T. Yu, S.C. Chang, P.C.E. Jorgenson, S.J. Park and M.C. Lai, "Treating Stiff Source Terms in Conservation Laws by the Space-Time Conservation Element and Solution Element Method," in *Proceedings of the 16th International Conference on Numerical Method in Fluid Dynamics, Arcachon, France, 6-10 July, 1998*, edited by C.H. Bruneau, (Springer-Verlag Berlin Heidelberg 1998), p. 433.
50. X.Y. Wang and S.C. Chang, A 3D structured/unstructured Euler solver based on the space-time conservation element and solution element method, in *A Collection of Technical Papers, 14th AIAA CFD Conference, June 28–July 1, 1999, Norfolk, Virginia*, AIAA Paper 99-3278.
51. N.S. Liu and K.H. Chen, *Flux: An Alternative Flow Solver for the National Combustion Code*, AIAA Paper 99-1079.
52. G. Cook, *High Accuracy Capture of Curved Shock Front Using the Method of Conservation Element and Solution Element*, AIAA Paper 99-1008.
53. S.C. Chang, Y. Wu, X.Y. Wang, and V. Yang, Local Mesh Refinement in the Space-Time CE/SE Method, in *Proceedings of the First International Conference on Computational Fluid Dynamics, Kyoto, Japan, 10-14 July, 2000*, edited by N. Satofuka, (Springer-Verlag Berlin Heidelberg 2001), p. 61.
54. S.C. Chang, Z.C. Zhang, S.T. John Yu, and P.C.E. Jorgenson, A Unified Wall Boundary Treatment for Viscous and Inviscid Flows in the CE/SE Method, in *Proceedings of the First International Conference on Computational Fluid Dynamics, Kyoto, Japan, 10-14 July, 2000*, edited by N. Satofuka, (Springer-Verlag Berlin Heidelberg 2001), p. 671.
55. Z.C. Zhang, S.T. John Yu, S.C. Chang, and P.C.E. Jorgenson, Calculations of Low-Mach-Number Viscous Flows without Preconditioning by the Space-Time CE/SE method, in *Proceedings of the First International Conference on Computational Fluid Dynamics, Kyoto, Japan, 10-14 July, 2000*, edited by N. Satofuka, (Springer-Verlag Berlin Heidelberg 2001), p. 127.
56. A. Himansu, P.C.E. Jorgenson, X.Y. Wang, and S.C. Chang, Parallel CE/SE Computational via Domain Decomposition, in *Proceedings of the First International Conference on Computational Fluid Dynamics, Kyoto, Japan, 10-14 July, 2000*, edited by N. Satofuka, (Springer-Verlag Berlin Heidelberg 2001), p. 423.
57. Y. Wu, V. Yang, and S.C. Chang, Space-Time Method for Chemically Reacting Flows with Detailed Kinetics, in *Proceedings of the First International Conference on Computational Fluid Dynamics, Kyoto, Japan, 10-14 July, 2000*, edited by N. Satofuka, (Springer-Verlag Berlin Heidelberg 2001), p. 207.
58. I.S. Chang, *Unsteady Rocket Nozzle Flows*, AIAA Paper 2002-3884.
59. S.C. Chang, *Courant Number Insensitive CE/SE Schemes*, AIAA Paper 2002-3890 (2002).

60. I.S. Chang, *Unsteady Underexpanded Jet Flows*, AIAA Paper 2003-3885.
61. S.C. Chang and X.Y. Wang, *Multidimensional Courant Number Insensitive CE/SE Euler Solvers for Applications Involving Highly Nonuniform Meshes*, AIAA Paper 2003-5280.
62. B.S. Venkatachari, G.C. Cheng, and S.C. Chang, *Development of A Transient Viscous Flow Solver Based on Conservation Element-Solution Element Framework*, AIAA Paper 2004-3413.
63. B.S. Venkatachari, G.C. Cheng, and S.C. Chang, *Courant Number Insensitive Transient Viscous Flow Solver Based on CE/SE Framework*, AIAA Paper 2005-00931.
64. S.C. Chang, *Explicit von Neumann Stability Conditions for the c-τ Scheme—A Basic Scheme in the Development of the CE-SE Courant Number Insensitive Schemes*, NASA TM 2005-213627, April 2005.
65. J.C. Yen and D.A. Wagner, *Computational Aeroacoustics Using a Simplified Courant Number Insensitive CE/SE Method*, AIAA Paper 2005-2820.
66. I.S. Chang, C.L. Chang, and S.C. Chang, *Unsteady Navier-Stokes Rocket Nozzle Flows*, AIAA Paper 2005-4353.
67. S.C. Chang, *Courant Number and Mach Number Insensitive CE/SE Euler Solvers*, AIAA Paper 2005-4355
68. S.C. Chang, A. Himansu, C.Y. Loh, X.Y. Wang, and S.T. Yu, Robust and Simple Non-Reflecting Boundary Conditions for the Euler Equations–A New Approach Based on the Space-Time CE/SE Method, in *Proceedings, NSF-CBMS Regional Research Conference on Mathematical Methods in Nonlinear Wave Propagation, North Carolina A&T State University, Greensboro, North Carolina, May 15-19, 2005*, edited by D.P. Clemence and G. Tang, p. 155-190, Vol. 379 in *Contemporary Mathematics*, American Mathematical Society (2005).
69. I.S. Chang, C.L. Chang, and S.C. Chang, *3D Unsteady Navier-Stokes Rocket Nozzle Flows*, AIAA Paper 2006-4775.
70. C.L. Chang, *Time-accurate, Unstructured-Mesh Navier-Stokes Computations with the Space-time CESE Method*, AIAA Paper 2006-4780.
71. S.C. Chang, *On Space-Time Inversion Invariance and Its Relation to Non-Dissipativeness of a CESE Core Scheme*, AIAA Paper 2006-4779.
72. S.C.Chang, *The $a(4)$ Scheme—A High Order Neutrally Stable CESE Solver*, AIAA Paper 2007-5820.
73. Other CESE references are posted on: http://www.grc.nasa.gov/www/microbus.
74. G. Strang, *Introduction to Applied Mathematics*, Wellesley-Cambridge Press, 1986.
75. B. Noble and J.W. Daniel, *Applied linear Algebra*, Prentice-Hall Inc. (1977).
76. R.A. Horn and C.R. Johnson, *Matrix Analysis*, Cambridge University Press, 1985.

TABLE 1.—NUMERICAL RESULTS OF THE $a(3)$ AND a SCHEMES

		$v = 0.1$		$t = 9.876$	
		$K = 25$, $n = 2{,}469$	$K = 50$, $n = 4{,}938$	$K = 100$, $n = 9{,}876$	$K = 200$, $n = 19{,}752$
E	$a(3)$	0.131×10^{-3}	0.143×10^{-4}	0.883×10^{-6}	0.549×10^{-7}
	a	0.452	0.115	0.287×10^{-1}	0.716×10^{-2}
E_x	$a(3)$	0.445×10^{-1}	0.977×10^{-3}	0.611×10^{-4}	0.382×10^{-5}
	a	2.90	0.732	0.182	0.454×10^{-1}
E_{xx}	$a(3)$	0.225	0.169	0.406×10^{-1}	0.100×10^{-1}

TABLE 2.—NUMERICAL RESULTS OF THE $a(3)$ AND a SCHEMES

		$v = 0.1$		$t = 10.00$	
		$K = 25$, $n = 2{,}500$	$K = 50$, $n = 5{,}000$	$K = 100$, $n = 10{,}000$	$K = 200$, $n = 20{,}000$
E	$a(3)$	0.228×10^{-3}	0.110×10^{-4}	0.628×10^{-6}	0.384×10^{-7}
	a	0.469	0.118	0.292×10^{-1}	0.727×10^{-2}
E_x	$a(3)$	0.154×10^{-1}	0.992×10^{-3}	0.623×10^{-4}	0.390×10^{-5}
	a	2.89	0.728	0.182	0.455×10^{-1}
E_{xx}	$a(3)$	0.473	0.316×10^{-1}	0.199×10^{-2}	0.124×10^{-3}

TABLE 3.—NUMERICAL RESULTS OF THE $a(3)$ AND a SCHEMES

		$v = 0.5$		$t = 49.38$	
		$K = 25$, $n = 2{,}469$	$K = 50$, $n = 4{,}938$	$K = 100$, $n = 9{,}876$	$K = 200$, $n = 19{,}752$
E	$a(3)$	0.168×10^{-3}	0.471×10^{-5}	0.294×10^{-6}	0.183×10^{-7}
	a	1.34	0.429	0.109	0.271×10^{-1}
E_x	$a(3)$	0.583×10^{-1}	0.856×10^{-12}	0.261×10^{-11}	0.678×10^{-11}
	a	8.73	2.73	0.686	0.171
E_{xx}	$a(3)$	1.01	0.942×10^{-1}	0.235×10^{-1}	0.587×10^{-2}

TABLE 4.—NUMERICAL RESULTS OF THE $a(3)$ AND a SCHEMES

		$v = 0.5$		$t = 50.00$	
		$K = 25$, $n = 2{,}500$	$K = 50$, $n = 5{,}000$	$K = 100$, $n = 10{,}000$	$K = 200$, $n = 20{,}000$
E	$a(3)$	0.362×10^{-13}	0.140×10^{-12}	0.229×10^{-12}	0.262×10^{-12}
	a	1.35	0.440	0.111	0.275×10^{-1}
E_x	$a(3)$	0.162×10^{-12}	0.845×10^{-12}	0.261×10^{-11}	0.682×10^{-11}
	a	8.73	2.73	0.689	0.172
E_{xx}	$a(3)$	0.172×10^{-9}	0.282×10^{-8}	0.185×10^{-7}	0.840×10^{-7}

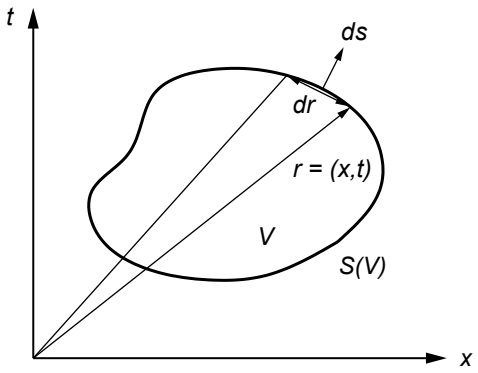
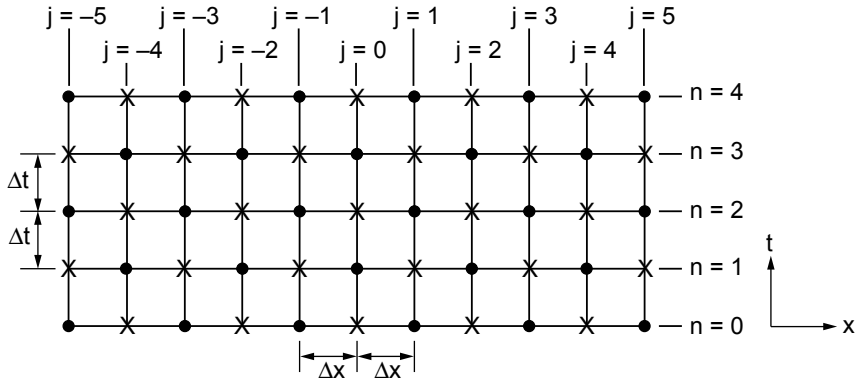
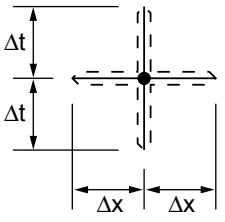


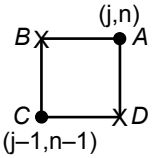
Figure 1.—A surface element on the boundary $S(V)$ of an arbitrary space-time volume V.



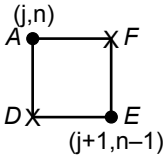
2(a).—The space-time mesh.



2(b).—SE(j,n).



2(c).—$CE_{-}(j,n)$.



2(d).—$CE_{+}(j,n)$.

Figure 2.—The SEs and CEs.