Global spectral analysis: Review of numerical methods

Pierre Sagaut, V.K. Suman, P. Sundaram, M.K. Rajpoot, Y.G. Bhumkar, Soumyo Sengupta, A. Sengupta, T.K. Sengupta

Review

Global spectral analysis: Review of numerical methods

Pierre Sagaut [a,1,*], V.K. Suman [b,c,2], P. Sundaram [c,2], M.K. Rajpoot [e,3], Y.G. Bhumkar [f,3], Soumyo Sengupta [g,4], A. Sengupta [d,5], T.K. Sengupta [d,6]

[a] Aix Marseille Univ, CNRS, Centrale Marseille, M2P2 UMR 7340, Marseille Cedex 13, France
[b] Computational & Theoretical Fluid Dynamics, CSIR-NAL, Bangalore, India
[c] High Performance Computing Laboratory, Department of Aerospace Engineering, Indian Institute of Technology Kanpur, Kanpur 208016, India
[d] Department of Mechanical Engineering, IIT (ISM) Dhanbad, Dhanbad 826 004, Jharkhand, India
[e] Mathematics and Computing Laboratory, Department of Mathematical Sciences, Rajiv Gandhi Institute of Petroleum Technology, Jais, Amethi 229304, UP, India
[f] Scientific Computing Laboratory, School of Mechanical Sciences, IIT Bhubaneswar, Bhubaneswar 752050, Odisha, India
[g] CERFACS, 42 Avenue G. Coriolis, 31057, Toulouse Cedex 1, France

ARTICLE INFO

Keywords:
Computational fluid dynamics
Global spectral analysis
High accuracy method
Dispersion relation preserving scheme
Error dynamics
q-waves
Focusing
Pseudo-spectral method
High performance computing
Convection diffusion reaction analysis

ABSTRACT

The design and analysis of numerical methods are usually guided by the following: (a) von Neumann analysis using Fourier series expansion of unknowns, (b) the modified differential equation approach, and (c) a more generalized approach that analyzes numerical methods globally, using Fourier–Laplace transform to treat the total or disturbance quantities in terms of waves. This is termed as the global spectral analysis (GSA). GSA can easily handle non-periodic problems, by invoking wave properties of the field through the correct numerical dispersion relation, which is central to the design and analysis. This has transcended dimensionality of the problem, while incorporating various physical processes e.g. by studying convection, diffusion and reaction as the prototypical elements involved in defining the physics of the problem. Although this is used for fluid dynamical problems, it can also explain many multi-physics and multi-scale problems. This review describes this powerful tool of scientific computing, with new results originating from GSA: (i) providing a common framework to analyze both hyperbolic and dispersive wave problems; (ii) analyze numerical methods by comparing physical and numerical dispersion relation, which leads to the new class of dispersion relation preserving (DRP) schemes; (iii) developing error dynamics as a distinct tool, identifying sources of numerical errors involving both the truncation and round-off error. Such studies of error dynamics provide the epistemic tool of analysis rather than an aleatoric tool, which depends on uncertainty quantification for high performance computing (HPC). One of the central themes of GSA covers the recent advances in understanding numerical phenomenon like focusing, which defied analysis so far. An application of GSA shown here for the objective evaluation of the so-called DNS by pseudo-spectral method for spatial discretization along with time integration by two-stage Runge–Kutta method is performed. GSA clearly shows that this should not qualify as DNS for multiple reasons. A new design of HPC methods for peta- and exa-flop computing tools necessary for parallel computing by compact schemes are also described.

1. Introduction

Scientific computing has evolved significantly over a short period of time due to the early analysis and design of numerical methods by mathematicians. Such early attempts have been noted [1] as *basic advances in numerical technique made in previous centuries encouraged Richardson [2] to seek a solution of system of equations for weather forecasting using a desk calculator*. However, the results were not successful

* Corresponding author.
E-mail addresses: pierre.sagaut@univ-amu.fr (P. Sagaut), vksuman@iitk.ac.in (V.K. Suman), prasanna@iitk.ac.in (P. Sundaram), mrajpoot@rgipt.ac.in (M.K. Rajpoot), bhumkar@iitbbs.ac.in (Y.G. Bhumkar), soumyoss08@gmail.com (S. Sengupta), aditi@iitism.ac.in (A. Sengupta), tksengupta@iitism.ac.in (T.K. Sengupta).
[1] Professor.
[2] PhD student.
[3] Associate Professor.
[4] PostDoc candidate.
[5] Assistant Professor.
[6] Visiting Professor.

due to lack of proper analysis which showed the method [3] to be unconditionally unstable for the numerical mode invoked by the time discretization of the heat equation, as explained in [1,4].

von Neumann is credited with developing the Fourier analysis method [4–6] which has been used ever since. However, the use of Fourier analysis has few restrictions. First, the approach only works for spatially periodic problems, with constant coefficients appearing in the differential equation. Also, implicit in such an analysis is the fact that each Fourier mode acts independently, with no provisions for any interactions among all the Fourier modes. This prompted Zingg to note [7,8] that Fourier analysis is easy to apply, yet it is difficult to interpret, due to the fact that non-periodic problems are excluded (and hence no effects of boundary conditions are included).

There is also another method of analysis practiced which converts the discrete equation back to its equivalent differential equation, shown in [9]. In [10,11], the authors claim to show the similarity of these two approaches, and furthermore noted that the modified equation approach has been in practice since 1950s. Linear problems with variable coefficients have been studied in [12] and nonlinear problems are shown in [13]. However, Li and Yang [11] claim that the modified equation approach is *very heuristic, unfortunately just valid for solutions in smooth regions or at low frequency modes [9]: Therefore the connection with the von Neumann analysis is only restricted there.*

Some researchers have raised concerns about the utility of modified equation approach, as Li and Yang [11] have quoted from [14] that there is lack of theoretical foundation in this approach and the results obtained are viewed with apprehension. It is to be noted however that modified equation approach has been in use for the design of numerical methods for parabolic equations to increase accuracy [15] and stability [16]. It is known that in the quest of stability [16], one faces the problem of consistency [1,17]. Additionally for hyperbolic equations, Lax and Wendroff [18,19] have used modified equation approach for enhanced accuracy and stability. Having noted the importance of analyzing difference equations by its equivalent differential form, one also notes a vast source of literature with sufficient mathematical rigor from the Russian school, as reported in the monograph of Shokin [20] and Yanenko et al. [21], and references contained therein. In the modified equation approach, if one converts the difference equation into its equivalent differential equation form by retaining space and time derivatives, it is called the Γ- form analysis. If the discrete equation is reverted back in the differential representation, with all the truncation terms converted only in terms of the spatial derivatives, the resultant approach is known to be in the Π-form. The variants of these Γ- and Π-forms of the modified equation approach along with Fourier–Laplace representation of the unknowns are used in the development of the global spectral analysis (GSA) of numerical methods.

The goal of the present article is to summarize both the key elements (see Sections 2, 3 and 5) and the recent developments and results in the field of GSA of advanced numerical methods for the linear 1D convection equation (CE) (see Sections 4 and 8), the Navier–Stokes equation (NSE) (see Sections 7 and 8) and the rotating shallow water equations (Section 9). The present survey emphasizes several very important issues, such as the derivation of the true numerical phase speed and group speed of the numerical solution, along with the related stability analysis. It is of major importance here to point out that GSA accounts for all effects, i.e. space discretization, time integration and boundary conditions.

Another important point addressed in the present review deals with collective interactions between modes, in both linearized and nonlinear regime. In the linear case, these interactions can lead to instability because of focusing or side-band instabilities, that are not detected when carrying out the usual single-mode Fourier analysis. In practice, this is done by considering wave packets instead of monochromatic disturbances to perform the stability analysis. The analysis in terms of occurrence of spurious numerical caustics is discussed in Section 6.

The review of GSA is supplemented by an extension to the nonlinear case and the resonant mode analysis (Section 11). Also, the application of GSA is shown in the novel review of the Fourier spectral method for homogeneous isotropic turbulence; subdomain boundary closure for high performance computing using compact schemes and analysis of convection diffusion reaction equation.

2. Global spectral analysis (GSA)

2.1. A primer to GSA

Apart from strict boundary value problems, rest of scientific computing for solving partial differential equation can be viewed as space–time dependent problems, for which the unknowns can be written in their most general form as,

$$u(x,t) = \int\int U(\omega_0,k)e^{ik(x-ct)}d\omega_0 dk \tag{1}$$

where the wavenumber k and circular frequency ω_0 are related via the dispersion relation or the phase speed given as, $c = \omega_0/k$. The disturbance field is nothing but an ensemble of wave components with the parameters (k,ω_0) defining the continuum field. This representation applies equally to dynamical systems consisting of waves or eddies [1, 22]. The rudiments of GSA are demonstrated with the help of the space–time dependent one-dimensional (1D) CE (see the schematic view in Fig. 1), which is the simplest in appearance (admitting exact solution, in terms of initial condition), and yet provides one of the toughest tests for the accuracy of any numerical method [1,23]. This is given as,

$$\frac{\partial u}{\partial t} + c\frac{\partial u}{\partial x} = 0 \tag{2}$$

Substituting Eq. (1) in (2), one notices that the partial differential equation transforms into an algebraic relation,

$$\omega_0 = kc \tag{3}$$

This is the physical dispersion relation, and is the central element for the analysis and design of high accuracy DRP schemes. It is worth realizing that this relation can originate from governing differential equation (as is usually the case for hyperbolic partial differential equation), or from the boundary condition (which are noted to create dispersive waves). For example, for surface gravity waves [1, 24], the governing equation is a time-independent elliptic partial differential equation, while the dispersion relation originates from the time-dependent interface condition. Similar situations prevail in other interfacial instability problems, for which the equilibrium state can be time-independent, and the physical dispersion relation can indicate physical instability [25,26]. The physical implication of dispersion relation is simultaneous consideration of spatial and temporal scales. But in numerical computing of partial differential equations, this is often not satisfied rigorously, as can be seen in various methods known as the method of lines or fractional methods [27,28]. Such an approach can be even traced to Eq. (8) of [4], where the parabolic partial differential equation is converted into a boundary value problem, often credited with the beginning of the analysis of numerical methods.

It is worth noting that a wave equation can be derived from (2) by applying a time derivative operator, leading to

$$\frac{\partial^2 u}{\partial t^2} - c^2\frac{\partial^2 u}{\partial x^2} = -\left[\frac{\partial c}{\partial t} - c\frac{\partial c}{\partial x}\right]\frac{\partial u}{\partial x} \tag{4}$$

in the most general case in which a space–time-dependent advection speed $c(\mathbf{x},t)$ is considered. The right-hand side term originates in the variations of the advection speed. It can be interpreted as a source term responsible for some diffraction/refraction effects. In the case of a uniform steady c, an homogeneous wave equation is recovered. A space–time-dependent advection velocity can appear in many cases, e.g. when considering the advection of a passive scalar u by a velocity

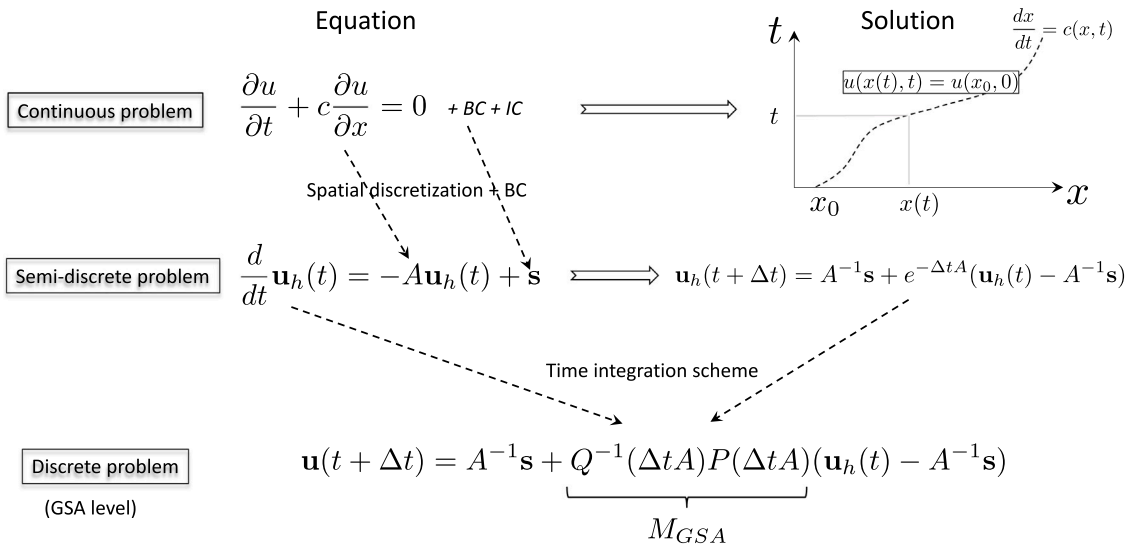

Fig. 1. Schematic view of GSA analysis level. The mass matrix is assumed to be the identity matrix for the sake of simplicity, i.e. $C = I_d$. The time integration scheme is described as a Padé approximant to the exact exponential solution, i.e. $e^{-M} \sim Q^{-1}(M)P(M)$, where Q and P are polynomials of the matrix M.

field c, or when investigating the linearized behavior of a small disturbance u about a base flow c as classically done in the hydrodynamic stability theory.

Despite the approximate attempt by Li and Yang [11] in relating Fourier analysis of von Neumann [4] and modified equation approach [9], more mathematically rigorous analysis can be performed by GSA, which has been introduced in [29–31]. This begins with the representation of numerical discretization in spectral space. To facilitate the description of the developed GSA, one represents the unknown $u(x,t)$, in the hybrid spectral plane as,

$$u(x,t) = \int \hat{U}(k,t)e^{ikx}dk \tag{5}$$

Here $\hat{U}$ is the Fourier amplitude and k is treated as the independent variable for analysis. The exact spatial derivative can be obtained from Eq. (5) as,

$$\left.\frac{\partial u}{\partial x}\right|_{exact} = \int ik\hat{U}e^{ikx}dk$$

One can write an equivalent spatial derivative obtained numerically as,

$$\left.\frac{\partial u}{\partial x}\right|_{num} = \int ik_{eq}\hat{U}e^{ikx}dk$$

Thus, the Fourier–Laplace amplitude is multiplied by ik_{eq} (instead of ik used in Fourier spectral method), in order to obtain the spatial derivative, originally made popular by Vichnevetsky and Bowles [32] and later on has been used in [29,33–35] among many other references.

There are two aspects of introducing k_{eq}, with the first providing a yardstick for comparing different discretization methods used in computing. Before we show some typical cases from finite difference methods, it is to be noted that in Chapter 12 of [1] and in [36], the finite volume and finite element methods have been similarly compared with many such discretization schemes. Ideally the ratio k_{eq}/k, should be equal to one and is treated as the measure of resolution of discretization schemes plotted as function of nondimensional wave number kh, with h as the uniform grid spacing.

The second aspect of k_{eq} is its mathematical implication for numerical computation, in expressing the numerical dispersion relation. If one ignores all errors due to temporal discretization, then using the form of the spatial derivative in terms of k_{eq} in Eq. (2), one obtains the numerical dispersion relation as,

$$\omega_N = k_{eq}c \tag{6}$$

It appears intuitive that c is held constant in the formulation of Eq. (2), and thus the above representation could be viewed as a semi-discrete analysis [37–39] used for numerical stability analysis. One of the greatest drawbacks of this approach is its inability to incorporate information about temporal discretization. Thus, no quantitative analysis is possible by this semi-discrete approach using the numerical dispersion relation given by Eq. (6). For example, if one were to compute a numerical group velocity [1,40–42] as,

$$V_{gN} = \frac{d\omega_N}{dk} = c\frac{dk_{eq}}{dk} \tag{7}$$

This numerical dispersion relation is wrong, as the group velocity is independent of time discretization scheme for Eq. (2). Propagation of wave-packet has been studied [1] showing that the group velocity is a strong function of both space and time discretizations considered together. It is somewhat ironical that authors in [43] have proposed this wrong numerical dispersion relation in Eq. (6) to derive a DRP scheme by considering spatial discretization alone. The authors [43] actually took a four-time level method for their DRP scheme, without noting that such multi-time level methods invoke spurious numerical modes, a topic which will be highlighted here, as reported in [1,44,45].

In contrast to using Eq. (6), researchers [30,46] have shown that while solving Eq. (2) numerically, the phase speed does not retain the constant value which is hard-coded. This appears paradoxical, but it has been clearly explained by the authors in [30,31,44] that the choice of space–time discretization methods immediately fixes the phase shift after every time step. This along with the time step determines the numerical phase speed (c_N), which will not be equal to the prescribed phase speed, c. This simple and subtle cause for $c_N \neq c$, is one of the central results of GSA.

Thus, the correct numerical dispersion relation and group velocity that accounts for both spatial discretization and time integration are written as,

$$\omega_N = kc_N \quad \text{and} \quad V_{gN} = c_N + k\frac{dc_N}{dk} \tag{8}$$

One notes that the wavenumber k is truly the independent variable, which along with the spatial and temporal discretization schemes fix all the numerical dispersion parameters and wave properties. The fact that c changes to c_N applies equally to coefficients of many other transport and diffusion equations, as have been noted already in introduction.

2.2. Rationale for GSA

The justification to use GSA is explained here with an example. We would like to demonstrate the utility of Eq. (8) for the 1D CE, with leap-frog and CD_2 schemes employed for time and space discretizations, respectively, on a uniformly spaced grid of spacing h. Using these discretizations in Eq. (2), one obtains the difference equation for a discrete node at (x_j, t^n) as,

$$\frac{u_j^{n+1} - u_j^{n-1}}{2\Delta t} + c\left(\frac{u_{j+1}^n - u_{j-1}^n}{2h}\right) = 0 \tag{9}$$

where Δt is the uniform time-step used for integrating Eq. (2).

We note that the numerical phase shift derived using Eq. (5) will be different from the expression given by the Π-form analysis [20]. This is due to the fact that the numerical dispersion relations used by these two approaches are different, as demonstrated.

In the discrete relation obtained in Eq. (9), the unknown is represented by the Fourier–Laplace transform as $u(x_j, t^n) = \iint \hat{U}(k, \omega_0)\, e^{(kx_j - \omega_0 t^n)} dk\ d\omega_0$. Substituting this into Eq. (2), we obtained the dispersion relation given in Eq. (3). The implication of existence of such a relationship is that the wavenumber and circular frequency both cannot be independent of each other. While this may appear as obvious, numerical analysis using Π-form overlooks this for discrete computations.

In Π-form analysis, the numerical phase speed is computed from the dispersion relation $\omega_\Pi = k_{eq} c$ as $c_{N_\Pi} = \frac{k_{eq}}{k} c$ where $\omega_\Pi = k c_{N_\Pi}$. For the chosen numerical method $k_{eq} = \frac{\sin(kh)}{h}$. Thus, the numerical phase speed from Π-form for the leap-frog and CD_2 scheme is given by

$$\frac{c_{N_\Pi}}{c} = \frac{\sin(kh)}{kh} \tag{10}$$

The same Π-form analysis has been performed in [32] and the above expression has been derived (cf. Eq. (2.13)). There are a few distinctive features of this result presented above, which requires highlighting. First, we note that the time-integration is performed by a three-time level method and hence one would expect two distinct numerical phase speeds, which is not given here. Secondly, and most importantly, the assumption that the numerical phase speed is dependent on time discretization is not used in this form of analysis. A consistent approach is used, as demonstrated next by GSA analysis which is based on Γ-form approach.

The importance of GSA comes into picture with its emphasis in representing the dependent variable in Eq. (5) using the wavenumber which is ascribed the role of the sole independent variable. Then the circular frequency and phase speed are calculated based on not only the governing differential equation, but also its discretization is considered in writing the numerical dispersion relation by Eq. (8). Based on different difference equations, one first obtains the numerical amplification following the hybrid representation of Eq. (5). Such numerical amplification factors would enforce phase shifts per time step, based on the discretization schemes, providing the glimpse of numerical phase speed. This is explained with the help of the leap-frog and CD_2 discretization schemes.

For the GSA analysis using Γ-form, the unknown variable u is represented in the hybrid spectral plane as already given by Eq. (5). Representing the initial condition for the governing equation in (2) as,

$$u(x_j, t = 0) = u_j^0 = \int U_0(k)\, e^{ikx_j} dk \tag{11}$$

where the subscript and superscript denotes the spatial and temporal indices respectively. The solution at any later time $t = n\Delta t$ is written using the definition of amplification factor as

$$u_j^n = \int U_0(k)\, [|G_j|]^n\, e^{i(kx_j - n\phi_j)} dk \tag{12}$$

where $G_j = \left(\frac{\hat{U}(k, t^n + \Delta t)}{\hat{U}(k, t^n)}\right)$ is the amplification factor and is, in general, a complex quantity i.e. $G_j = G_{rj} + iG_{ij}$. $|G_j|$ is the modulus as given by $|G_j| = (G_{rj}^2 + G_{ij}^2)^{1/2}$. The phase is calculated as $\tan\phi_j = -G_{ij}/G_{rj}$.

From the relation for ϕ, the numerical phase speed (c_{N_Γ}) is obtained by noting that ϕ_j is the phase shift per time step and is given below.

$$c_{N_\Gamma} = \frac{\phi_j}{k\Delta t} \tag{13}$$

The physical phase speed is c for all wavenumbers, but c_{N_Γ} is noted to depend on k. Thus, the numerical solution is dispersive, in contrast to the non-dispersive nature of 1D CE. Thus, both the Π- and Γ-form analyses show the numerical phase speed to depend upon the length scale k, with the important difference that the latter uses the actual temporal discretization used for computing. In contrast, the Π-form analysis, like the semi-discrete analysis ignores the information originating from the temporal discretization, making such results of very limited value, despite its very wide-spread use among practitioners.

Representing the variable u in the hybrid spectral plane given by Eq. (5) and substituting this into Eq. (9), we obtain the relation

$$\hat{U}_j^{n+1} - \hat{U}_j^{n-1} + \frac{c\Delta t}{h}\left(e^{ikh} - e^{-ikh}\right)\hat{U}_j^n = 0 \tag{14}$$

where the variables with hat superscript denote the spectral amplitudes. Noting the definition of numerical amplification factor G_j, defined as $G_j = \left(\frac{\hat{U}_j^{n+1}}{\hat{U}_j^n}\right) = \left(\frac{\hat{U}_j^n}{\hat{U}_j^{n-1}}\right)$, a quadratic relation for G_j is obtained and the two roots or amplification factors are determined as

$$G_{j\,1,2} = -iN_c \sin(kh) \pm \sqrt{1 - {N_c}^2 \sin^2(kh)} \tag{15}$$

where $N_c = \frac{c\Delta t}{h}$ denotes the CFL number. From the above relations, the numerical phase speeds c_{N_Γ} can be calculated from Eq. (13) as

$$\frac{c_{N_{\Gamma_{1,2}}}}{c} = \frac{\phi_j}{(kh)N_c} = \left(\frac{1}{(kh)N_c}\right)\tan^{-1}\left(\frac{N_c \sin(kh)}{\pm\sqrt{1 - {N_c}^2 \sin^2(kh)}}\right) \tag{16}$$

Vichnevetsky and Bowles [32] also employed the Γ-form analysis in determining the numerical phase speed. If the unknown is represented by the Fourier–Laplace transform whose form is as given earlier and substituting this in Eq. (9) one obtains the spectral plane representation for leap-frog and CD_2 schemes as,

$$e^{-i\omega_0\Delta t} - e^{i\omega_0\Delta t} + N_c\left(e^{ikh} - e^{-ikh}\right) = 0$$

which upon simplification yields

$$\sin(\omega_0\Delta t) = N_c \sin(kh)$$

This provides the frequency as

$$\omega_0 = \frac{1}{\Delta t}\sin^{-1}(N_c \sin(kh)) \tag{17}$$

Having obtained the expression for the circular frequency in terms of the wavenumber, numerical phase speed is computed by dividing the circular frequency by the wavenumber as, $c_{NVB} = \frac{\omega_0}{k}$, which upon simplification yields

$$\frac{c_{NVB}}{c} = \frac{1}{(kh)N_c}\sin^{-1}(N_c \sin(kh)) \tag{18}$$

We note that Vichnevetsky and Bowles obtain amplification factors given by Eq. (4.8c) in [32] and show the above expression for numerical phase speed in Table 4.3 in [32]. Although, at a first glance the numerical phase speeds computed from the Γ-form GSA analysis (Eq. (16)) and the expression given by Vichnevetsky and Bowles [32] (Eq. (18)) appear different, one notices on closer scrutiny that these two expressions are equivalent, following the trigonometric identity: $\sin^{-1}(x) = \tan^{-1}\left(\frac{x}{\sqrt{1-x^2}}\right)$.

Thus, one notes four phase speeds have been described in the above analysis: With c- as the physical phase speed; c_{N_Π} as the numerical

phase speed obtained from the Π-form analysis; $c_{N_{\Gamma_{1,2}}}$ are the numerical phase speeds from the GSA analysis based on Γ-form and $c_{N_{VB}}$ as the phase speed derived by Vichnevetsky and Bowles [32], also based on Γ-form analysis. The numerical phase speed(s) obtained from the Γ-form analysis is the correct approach providing the quantities consistent with the definition of phase speed, whereas c_{N_Π} and the same expression obtained for semi-discrete analysis is incorrect due to the use of wrong numerical dispersion relation. Finally, the expression for numerical phase speed provided by the analysis due to Vichnevetsky and Bowles [32] gives only one numerical phase speed ($c_{N_{VB}}$) which is identical to $c_{N_{\Gamma_1}}$. The discerning readers would note that the expression of $c_{N_{VB}}$ also contains the expression for the spurious numerical mode, provided one identifies the correct range of the argument for the expression of $c_{N_{VB}}$. However, the authors in [32] failed to identify and discuss the importance of the spurious mode in practical computations. One also notes that Haltiner and Williams [46] obtained the same expressions for $c_{N_{\Gamma_{1,2}}}$, but their implications in long range weather forecasting calculation was not seized upon. Although the Γ-form analysis has been introduced by Shokin [20], only a discussion on the analysis was made without a derivation of the numerical phase speeds. The foundation of GSA was founded and the implications noted in a series of articles by the authors in [1,29–31,47–49].

2.3. Elements of matrix analysis

Being a global method that accounts for spatial discretization, time integration and boundary conditions, GSA can be recast as a matrix analysis method. Starting from the linear parabolic equation

$$\frac{\partial u}{\partial t} + c\frac{\partial u}{\partial x} = \nu\frac{\partial^2 u}{\partial x^2}, \quad \nu \geq 0 \tag{19}$$

with the hyperbolic linear problem (2) recovered for a null diffusion, i.e. $\nu = 0$. The first step of matrix analysis consists of deriving the time-continuous semi-discrete problem by taking into account of only the spatial discretization that leads to:

$$C\frac{d\mathbf{u}_h(t)}{dt} = -A\mathbf{u}_h(t) + \mathbf{s}(t), \quad t > 0, \quad \mathbf{u}_h(0) = \mathbf{u}_0 \tag{20}$$

where matrices C and A are related to the spatial numerical scheme and the vector $\mathbf{s}(t)$ accounts for Dirichlet-like boundary conditions (Neumann and periodic conditions are inserted into A). C is the mass matrix, which simplifies as the identity matrix in finite differences methods, and is lumped into a diagonal matrix in many Finite Element Methods. Both C and A are assumed to be time-independent hereafter for the sake of simplicity. Assuming that the numerical method is such that the mass matrix C is invertible, one obtains the following ordinary matrix equation:

$$\frac{d\mathbf{u}_h(t)}{dt} = -C^{-1}A\mathbf{u}_h(t) + C^{-1}\mathbf{s}(t), \quad t > 0 \tag{21}$$

whose solution is

$$\mathbf{u}_h(t) = e^{-tC^{-1}A}\mathbf{u}_0 + e^{-tC^{-1}A}\int_0^t e^{-t'C^{-1}A}C^{-1}\mathbf{s}(t')dt', \quad t \geq 0 \tag{22}$$

For the case of time-independent $\mathbf{s}$, this solution simplifies as

$$\mathbf{u}_h(t) = A^{-1}\mathbf{s} + e^{-tC^{-1}A}(\mathbf{u}_0 - A^{-1}\mathbf{s}), \quad t \geq 0 \tag{23}$$

This solution can be interpreted as the sum of a steady-state solution plus a transient term, and thus can be rewritten as,

$$\mathbf{u}_h(t+dt) = A^{-1}\mathbf{s} + e^{-dtC^{-1}A}(\mathbf{u}_h(t) - A^{-1}\mathbf{s}) \tag{24}$$

The fully discrete problem is now obtained considering the time-integration method, which amounts to finding an approximation $M_{GSA}(t)$ of the exponential matrix $\exp(-tC^{-1}A)$ in Eq. (23) or $M_{GSA}(dt)$ for $\exp(-dtC^{-1}A)$ in Eq. (24). This can be done in several ways, among which using Taylor-series expansion, Padé approximants or even Chebyshev rational approximations of the matrix exponential function (e.g. see [50]). The case of the Padé approximant is of particular interest, since several popular time integration methods can be recast as particular cases of this approach. Writing the Padé approximant as

$$e^z = \frac{P_n(z)}{Q_m(z)} \tag{25}$$

where $P_n(z)$ and $Q_m(z)$ are nth order and mth order polynomial in z, respectively, the first values and related time-integration methods are given in Table 1.

As a consequence, a time-marching numerical method can be in the following compact form:

$$\mathbf{u}^{n+1} = M_{GSA}\mathbf{u}^n + \tilde{\mathbf{s}}, \quad \tilde{\mathbf{s}} = (I_d - M_{GSA})A^{-1}\mathbf{s} \tag{26}$$

where the vector $\tilde{\mathbf{s}}$ is related to boundary conditions and u_i^n denotes the computed value of u at node i at the nth time step. The dt dependency in M_{GSA} has been omitted for the sake of simplicity. A commonly used stability criterion is

$$\rho(M_{GSA}) \leq 1, \tag{27}$$

where $\rho(M_{GSA})$ denotes the spectral radius of M_{GSA}, i.e. the maximum eigenvalue modulus of M_{GSA}. This condition ensures that $\|\mathbf{u}^n\|$ remains bounded over arbitrary time, and then the error will also be bounded, provided $\tilde{\mathbf{s}}$ remains bounded. However, in [31], it has been shown that phase and dispersion errors do not allow $\tilde{\mathbf{s}}$ to remain bounded. In the special case of CE, $\tilde{\mathbf{s}} = 0$, and then the more restrictive condition $\rho(M_{GSA}) < 1$ guaranties that the numerical solution vanishes over very long integration time. If the physical solution also vanishes (which is true for the diffusive parabolic problem (19) with periodic boundary conditions), the difference between the physical and numerical solutions also vanishes over long times, but does not preclude transient error growth over finite time [51–55].

We can now think about the hyperbolic problem adopted to explain GSA. During the course of our discussion, it appears that there is some conflict in using GSA for the hyperbolic problem solved numerically for the following reasons.

(a) A truly hyperbolic problem can be analyzed as a Cauchy problem in an unbounded domain, as advocated in [32]. However, to compute it we need a finite domain and therefore, boundary conditions. One of the strong points of GSA is that one can analyze non-periodic problems incorporating boundary closure, as in the case of parabolic problem [4,48].

(b) There are hyperbolic problems with boundary conditions, for which one uses a method of characteristics, with characteristics providing necessary boundary conditions. This class of problems can be easily analyzed by GSA.

(c) For periodic hyperbolic problems, use of GSA is straightforward.

Thus, one must consider the problems of type (a) above for the use of GSA for Eq. (2). There is an indirect way to solve this by considering 1D CE in a finite domain and analyze the methods by GSA. We consider the solution of the following,

(i) The right running wave problem of Eq. (2) by considering propagation of a wave-packet, given as the initial condition. One places the packet and the domain in a way that the initial condition does not reach out to the numerical boundaries. Thus on the left or inflow boundary, the boundary condition can be the trivial solution (u = 0). Thus the inflow boundary will never be the source of any error, as has been analyzed in [31].

(ii) Let us say that we have N points in the domain with equi-spaced points. Then one can obtain the spatial derivatives at$t = 0$ at all the nodes, the usual way applying any explicit or implicit method [29].

(iii) Next, one solves Eq. (2) for the next time step at $t = \Delta t$ for all the points. One accepts the solution up to $j = (N-1)$. At $j = N$, one uses a backward first order formula using the solution at $j = (N-1)$ at the current time step and time advance the solution at $j = N$ by using $N_c = 1$ (the reason for which will be apparent, as explained in [1]). This will provide a solution at $j = N$ which is

Table 1
Padé approximants to e^z. From [50].

$P_n(z)/Q_m(z)$	m = 0	1	2	3	4
n = 0	$\frac{1}{1}$	$\frac{1}{1-z}$ (Backward Euler)	$\frac{2}{2-2z+z^2}$	$\frac{6}{6-6z+3z^2-z^3}$	$\frac{24}{24-24z+12z^2-4z^3+z^4}$
1	$\frac{1+z}{1}$ (Forward Euler)	$\frac{2+z}{2-z}$ (Crank-Nicholson)	$\frac{6+2z}{6-4z+z^2}$	$\frac{24+6z}{24-18z+6z^2-z^3}$	$\frac{120+24z}{120-96z+36z^2-8z^3+z^4}$
2	$\frac{2+2z+z^2}{2}$ (RK 2)	$\frac{6+4z+z^2}{6-2z}$	$\frac{12+6z+z^2}{12-6z+z^2}$	$\frac{60+24z+3z^2}{60-36z+9z^2-z^3}$	$\frac{360+120z+12z^2}{360-240z+72z^2-12z^3+z^4}$
3	$\frac{6+6z+3z^2+z^3}{2}$ (RK 3)	$\frac{24+18z+16z^2+z^3}{24-6z}$	$\frac{60+36z+9z^2+z^3}{60-24z+3z^2}$	$\frac{120+60z+12z^2+z^3}{120-60z+12z^2-z^3}$	$\frac{840+360z+60z^2+4z^3}{840-480z+120z^2-16z^3+z^4}$
4	$\frac{24+24z+12z^2+4z^3+z^4}{24}$	$\frac{120+96z+36z^2+8z^3+z^4}{120-24z}$	$\frac{360+240z+72z^2+12z^3+z^4}{360-120z+12z^2}$	$\frac{840+480z+120z^2+16z^3+z^4}{840-360z+60z^2-4z^3}$	$\frac{1680+840z+180z^2+20z^3+z^4}{1680-840z+180z^2-20z^3+z^4}$

'exact' for the time step corresponding to $N_c = 1$. Note that the time step for interior points will be dictated by GSA, ensuring $|G| = 1$, and $c_N/c = 1$ and $V_{gN}/c = 1$ [1,31]. This is the way, one also uses the Sommerfeld boundary condition, while solving the NSE that requires a convection speed at the boundary. In this test case, the convection speed is determined by the CFL condition.

This is the way, one can time march the solution ensuring the property of the exact, non-dissipative, non-dispersive nature of the 1D CE everywhere.

3. Global resolution in GSA: Diffusion and anti-diffusion

In the previous section, it is noted that GSA views the numerical dispersion relation (Eq. (8)) differently from the intuitive assumption of numerical dispersion given by Eq. (6). These appear as difference in interpretation of numerical discretization, by considering constant phase speed as input to the problem. Whereas, for a dispersive case, the phase speed is different from the constant value, and is dictated by the spatial and temporal discretization applied globally in the computational domain. Thus, it is important to understand more carefully the nature of spatial and temporal discretizations used.

Considering for the moment the solution of Eq. (2) for a structured grid with uniformly distributed points, one treats the unknowns as a vector, $\{u\}$, and the corresponding first spatial derivative as the vector, $\{u'\}$, with the prime indicating the spatial derivative. Without going into specific nature of the scheme chosen to obtain the spatial derivative, one can represent the generic spatial discretization scheme as,

$$[A]\{u'\} = \frac{1}{h}[B]\{u\} \tag{28}$$

where in usual discretization schemes, $[A]$ and $[B]$ matrices are band-limited with constant elements. While this is for the general implicit discretization scheme, one can also alternately express an equivalent explicit discretization scheme using $[C] = [A]^{-1}[B]$ as,

$$\{u'\} = \frac{1}{h}[C]\{u\} \tag{29}$$

It is to be noted that despite $[A]$ and $[B]$ matrices being very band-limited (tri-diagonal or penta-diagonal non-zero entries), the $[C]$ matrix may have many non-zero entries and is not necessarily band-limited. Thus, the derivative at the jth-node can, in general, depend upon the values of the function at all N-nodes, and can be alternately written as,

$$u'_j = \frac{1}{h}\sum_{l=1}^{N} C_{jl}u_l \tag{30}$$

If one project all the functions on the right hand side to the jth node in spectral form as, $u_l = \int U(k)e^{ik(x_l-x_j)}e^{ikx_j}dk$, then one can rewrite Eq. (30) as,

$$u'_j = \frac{1}{h}\int \sum_{l=1}^{N} C_{jl}U(k)e^{ik(x_l-x_j)}e^{ikx_j}dk \tag{31}$$

This immediately enables one to note,

$$ik_{eq}|_{x=x_j} = \frac{1}{h}\sum_{l=1}^{N} C_{jl}e^{ik(x_l-x_j)} \tag{32}$$

There are few noteworthy features of Eq. (32), which highlight the subsequent discussion on GSA. Adoption of matrix notation in Eqs. (28) and (29), enables one to incorporate the boundary condition treatment in the constituent matrices, which will provide global information of k_{eq} at all the nodes where solution is sought. Full-domain analysis of numerical methods is the central strength of GSA. Secondly, one notes that for central difference schemes (both explicit and implicit), viewed as a full-domain problem with non-periodic boundary conditions, the $[A]$, $[B]$ and $[C]$ matrices are non-Hermitian. For this reason alone, the GKS-stability analysis [56] is inapplicable, as propounded by Gustaffson, Kreiss and Sundström. This problem has also been highlighted in [29]. Thirdly, one also notices from Eq. (32) that k_{eq} is in general complex, i.e. $k_{eq} = k_{real} + ik_{imag}$. Then the real part of k_{eq} contributes to the first spatial derivative, so that one can rewrite Eq. (2) as,

$$\frac{\partial u}{\partial t} + \int ik_{real}c\hat{U}e^{ikx}dk = \int ik_{imag}c\hat{U}e^{ikx}dk \tag{33}$$

Compare this with the convection–diffusion equation (CDE) given by,

$$\frac{\partial u}{\partial t} + c\frac{\partial u}{\partial x} = \alpha\frac{\partial^2 u}{\partial x^2} \tag{34}$$

where α is the coefficient of diffusion and for many physical phenomena is characterized as a positive quantity. For $\alpha \geq 0$, the right-hand side of Eq. (34) will have a negative sign, when represented in the spectral plane, as $\frac{\partial^2 u}{\partial x^2} = -\int k^2\hat{U}(k)e^{ikx}dk$, for this reason diffusion connotes often as dissipation. In contrast, negative value of α is referred to as anti-diffusion [1,29,34,57,58] and its unintended presence in numerical computation leads to catastrophic breakdown of the solution process, as the action of this term is to pump in energy to the system. Thus, if the imaginary part of k_{eq} turns out to be positive, it leads to anti-diffusion and the numerical solution will break down.

3.1. High accuracy schemes

High accuracy schemes are those which have extensive resolution across the wavenumbers. This is represented by k_{eq}/k having the ideal value of unity over as large a range of kh as possible in the resolved Nyquist limit of π [34]. Quite often in the literature, instead of high accuracy schemes, one comes across high order schemes to denote higher resolution. However, it has been explained in [1,29] that one can derive optimal compact schemes with lower formal order, yet one will have extremely high resolution, as has been derived in [59] for solving Eq. (2) with periodic boundary condition.

To avoid numerical diffusion and anti-diffusion introduced in discretizing first derivative, one prefers using central explicit and compact schemes [1,29]. In [60,61], the first and second derivatives at the sub-domain boundary have been calculated using eighth order central difference (CD8) scheme with stencils given by,

$$\begin{aligned} u'_i &= \frac{4}{5h}(u_{i+1}-u_{i-1}) - \frac{1}{5h}(u_{i+2}-u_{i-2}) + \frac{4}{105h}(u_{i+3}-u_{i-3}) \\ &\quad - \frac{1}{280h}(u_{i+4}-u_{i-4}) \\ u''_i &= \frac{8}{5h^2}(u_{i+1}+u_{i-1}) - \frac{1}{5h^2}(u_{i+2}+u_{i-2}) + \frac{8}{315h^2}(u_{i+3}+u_{i-3}) \\ &\quad - \frac{1}{560h^2}(u_{i+4}+u_{i-4}) - \end{aligned} \tag{35}$$

$$\frac{205}{72h^2}u_i \tag{36}$$

One also comes across a class of combined compact difference (CCD) schemes [62–65], which discretize first (indicated with a prime) and second derivatives (indicated with double prime), simultaneously. The interior stencils of this scheme for a periodic problem are given by [63],

$$\frac{7}{16h}(u'_{i+1}+u'_{i-1})+\frac{u'_i}{h}-\frac{1}{16}(u''_{i+1}-u''_{i-1})=\frac{15}{16h^2}(u_{i+1}-u_{i-1}) \tag{37}$$

$$\frac{9}{8h}(u'_{i+1}-u'_{i-1})-\frac{1}{8}(u''_{i+1}+u''_{i-1})+u''_i=\frac{3}{h^2}(u_{i+1}-2u_i+u_{i-1}) \tag{38}$$

To solve non-periodic problems, the following explicit boundary closure schemes have been proposed in [64,65] for the nodes at $j=2$ and $(N-1)$

$$u'_2=\frac{1}{h}\left[\left(\frac{2\beta_2}{3}-\frac{1}{3}\right)u_1-\left(\frac{8\beta_2}{3}+\frac{1}{2}\right)u_2+(4\beta_2+1)u_3-\left(\frac{8\beta_2}{3}+\frac{1}{6}\right)u_4+\frac{2\beta_2}{3}u_5\right] \tag{39}$$

$$u'_{N-1}=-\frac{1}{h}\left[\left(\frac{2\beta_{N-1}}{3}-\frac{1}{3}\right)u_N-\left(\frac{8\beta_{N-1}}{3}+\frac{1}{2}\right)u_{N-1}+(4\beta_{N-1}+1)u_{N-2}-\left(\frac{8\beta_{N-1}}{3}+\frac{1}{6}\right)u_{N-3}+\frac{2\beta_{N-1}}{3}u_{N-4}\right] \tag{40}$$

$$u''_2=(u_1-2u_2+u_3)/h^2 \tag{41}$$

$$u''_{N-1}=(u_N-2u_{N-1}+u_{N-2})/h^2 \tag{42}$$

with $\beta_2=-0.025$ and $\beta_{N-1}=0.09$, as given in [29,64,65].

The steps of GSA are provided for a new combined compact difference noted in shorthand as NCCD-scheme for solving non-periodic problems in [1,64,65]. For the sake of analysis, one writes the NCCD-scheme as

$$\frac{1}{h}[A_1]\{u'\}+[B_1]\{u''\}=\frac{1}{h^2}[R_1]\{u\}$$

$$\frac{1}{h}[A_2]\{u'\}+[B_2]\{u''\}=\frac{1}{h^2}[R_2]\{u\}$$

These can be alternatively written in an equivalent explicit form for the first and second derivatives as,

$$\{u'\}=\frac{1}{h}\,[C]\{u\}$$

$$\{u''\}=\frac{1}{h^2}\,[C_2]\{u\}$$

$$\text{where}\quad [C]=([A_1]-[B_1][B_2]^{-1}[A_2])^{-1}\,([R_1]-[B_1][B_2]^{-1}[R_2])\,h \tag{43}$$

$$\text{and}\quad [C_2]=([B_2]-[A_2][A_1]^{-1}[B_1])^{-1}\,([R_2]-[A_2][A_1]^{-1}[R_1])\,h^2 \tag{44}$$

The sixth order compact scheme in [33] is shown next, for an internal node given by,

$$\alpha_6u'_{i-1}+u'_i+\alpha_6u'_{i+1}=\frac{a_6}{2h}(u_{i+1}-u_{i-1})+\frac{b_6}{4h}(u_{i+2}-u_{i-2}) \tag{45}$$

which must satisfy, $1+2\alpha_6=a_6+b_6$ for consistency of the scheme, and it has been deduced that for sixth order accuracy these coefficients become, $\alpha_6=1/3$, $a_6=14/9$ and $b_6=1/9$. This will be referred to as Lele6 scheme for the purpose of identification. This scheme has been used in [66] to solve non-periodic problems with the following boundary closure scheme,

$$j=1:\quad 2u'_1+4u'_2=\frac{-5u_1+4u_2+u_3}{h} \tag{46}$$

$$j=2:\quad u'_1+4u'_2+u'_3=\frac{3}{h}(u_3-u_1) \tag{47}$$

The stencils for $j=N-1$ and N, are similar to Eqs. (46) and (47). The near-boundary point stencil at $j=2$ is fourth order accurate. We will refer to this as Adams' scheme, consisting of Eqs. (45) to (47).

In comparison to higher order, high accuracy compact schemes, Haras and Ta'asan [59] initiated an optimal search for extremely high accuracy central schemes in spectral plane for 1D CE with periodic boundary condition. This is equivalent to minimizing integrated error for first spatial derivative over the resolved length scales up to the Nyquist limit for a basic representation as given in Eq. (45), while treating the consistency condition as the constraint.

This lead from [59] has been further extended for the same problem, but with non-periodic boundary conditions to derive the OUCS3 scheme in [29]. For the interior points the stencil is written as,

$$r_{-1}u'_{i-1}+u'_i+r_{+1}u'_{i+1}=\frac{1}{h}\left(s_{-2}u_{i-2}+s_{-1}u_{i-1}+s_0u_i+s_{+1}u_{i+1}+s_{+2}u_{i+2}\right) \tag{48}$$

The boundary closure schemes are those already given in the following for the boundary closure at $j=1$ and 2,

$$u'_1=\frac{1}{2h}[-3u_1+4u_2-u_3] \tag{49}$$

$$u'_2=\frac{1}{h}\left[\left(\tfrac{2\beta_2}{3}-\tfrac{1}{3}\right)u_1-\left(\tfrac{8\beta_2}{3}+\tfrac{1}{2}\right)u_2\right]+\frac{1}{h}\left[(4\beta_2+1)\,u_3-\left(\tfrac{8\beta_2}{3}+\tfrac{1}{6}\right)u_4+\tfrac{2\beta_2}{3}u_5\right] \tag{50}$$

Similar boundary closure can be obtained for $j=N$ and $(N-1)$ from Eqs. (49) and (50), respectively, with right hand side terms' sign made opposite and introduce β_{N-1}, instead of β_2. The optimal values of these two parameters have been reported in [1,29] with $\beta_2=-0.025$ and $\beta_{N-1}=0.09$. The upwinding of the scheme is introduced via a fourth diffusion term with a coefficient, η_3. The resultant scheme, with optimized parameters from [59] are given by, $r_{\pm1}=D_H\pm\frac{\eta_3}{60}$; $s_{\pm2}=\pm\frac{F_H}{4}+\frac{\eta_3}{300}$; $s_{\pm1}=\pm\frac{E_H}{2}+\frac{\eta_3}{30}$ and $s_0=-\frac{11\eta_4}{150}$, with $D_H=0.3793894912$; $E_H=1.57557379$; $F_H=0.183205192$. This is formally only a second order scheme, but it will be demonstrated that the OUCS3 scheme is superior over many other explicit and implicit schemes.

Another fifth order upwind scheme has been proposed in [39], with stencils given as,

$$j=1:\quad 6u'_1+18u'_2=\frac{-17u_1+9u_2+9u_3-u_4}{h} \tag{51}$$

$$j=2:\quad u'_1+4u'_2+u'_3=\frac{3}{h}(u_3-u_1) \tag{52}$$

$$3\le j\le N-2:\quad b_{j-1}\,u'_{j-1}+b_j\,u'_j+b_{j+1}\,u'_{j+1}=\frac{1}{h}\sum_{l=-2}^{2}a_{j+l}\,u_{j+l} \tag{53}$$

where $a_{j\pm2}=\pm\frac{5}{3}+\frac{5\eta_z}{6}$; $a_{j\pm1}=\pm\frac{140}{3}+\frac{20\eta_z}{3}$; $a_j=-15\eta_z$, $b_{j\pm1}=20\pm\eta_z$ and $b_j=60$.

The boundary closure schemes for $j=N-1$ and N can be analogously written using Eqs. (51) and (52), respectively. The quantity η_z is the upwind coefficient that is explicitly added as the diffusion term $\frac{\eta_z}{6!}h^5\,\frac{\partial^6u}{\partial x^6}$. It is noted that if one fixes $\eta_z=0$, then one recovers the Lele6 or Adams' scheme. This class of discrete schemes will be referred to as the Zhong's scheme.

Zhong [39] tested the design of the fifth order upwind compact scheme by investigating numerical instability of the scheme for Eq. (2) by semi-discrete matrix analysis for $\eta_z=-2,-1$ and 0. It is pertinent to note that the author performed semi-discrete analysis, due to the confusion between correct numerical dispersion relation in Eq. (8), and the wrong approach based on treating the phase speed as constant resulting in wrong dispersion relation given in Eq. (6).

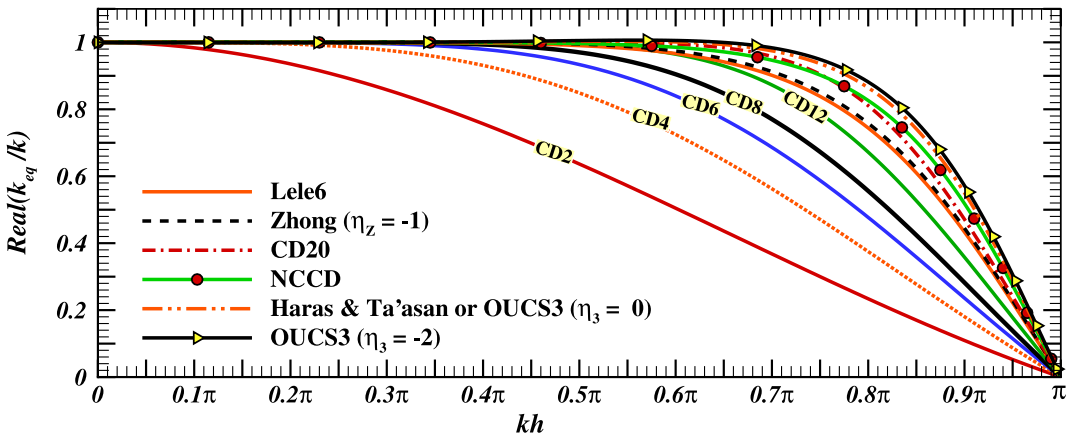

Fig. 2. Spectral resolution of different central explicit and implicit spatial discretization schemes for first derivative by plotting the real part of k_{eq}/k as function of kh. The schemes are calibrated for interior nodes, unaffected by boundary closure.

3.2. Resolution of explicit and implicit spatial discretization

In the discussion so far, we have introduced a few typical high accuracy schemes. The resolution of these schemes are investigated by looking at $\frac{k_{eq}}{k}$, as can be obtained from Eq. (32). In Fig. 2, the resolution of the compact schemes (Lele6 or Adams, Zhong, OUCS3 or Haras–Ta'asan and NCCD) are compared with various central difference explicit schemes. In this figure only the central node resolution is shown for the sake of clarity. It is noted that the OUCS3 ($\eta_3 = -2$), Haras–Ta'asan scheme for non-periodic problem (with $\eta_3 = 0$) and NCCD scheme have resolution which is better than twentieth order explicit central difference scheme. This once again justifies the superior performance of compact schemes, as compared to very high order explicit scheme. It is also noted that the order of schemes is not very relevant, as one notes from this figure that the formally second order OUCS3 scheme displays higher resolution (k_{eq}/k attains the ideal unity value over a large range of kh) as compared to the sixth order Lele6 or Adams' scheme.

The spatial discretization of first derivative invokes numerical diffusion and/ or anti-diffusion for non-central scheme. This can be noted from the imaginary part of $\frac{k_{eq}}{k}$, as obtained from Eq. (32). One notes that the GSA obtained for the full domain, will make the $[C]$ matrix non-symmetric due to one-sided stencils used for the boundary and near-boundary points. This is clearly evident in Fig. 3, where representative points are displayed as part of full-domain analysis [29]. The resolution (on the left frames) and diffusion/anti-diffusion of the spatial discretization (on the right frames) are shown for the compact schemes (Lele6 or Adams, Zhong, OUCS3 or Haras–Ta'asan and NCCD). In the top frame, Adams' scheme results are shown with the resolution shown to be variable from $j = 1$ to points in the interior. Although the first node ($j = 1$) shows large phase distortion, this is of hardly any concern, as the governing equation is never going to be discretized at the boundary node. However, the major concern is about the presence of anti-diffusion noted in the imaginary part of k_{eq}/k, as shown on top right frame for Adams' scheme. This is evident for the nodes which are near the inflow of the domain, with maximum effects near the Nyquist limit ($kh = \pi$). It is to be realized that the problem of anti-diffusion arises due to boundary closure schemes. As the global schemes are implicit, any problem created at one node percolates over the full domain, with maximum effects noted for near-boundary nodes.

One notes from the second row from the top in Fig. 3(b), the resolution and diffusion properties of spatial discretization for first derivative for the Haras and Ta'asan scheme [59], which uses the interior stencil and boundary closure schemes given in Eqs. (46) to (48), with $\eta_3 = 0$. Once again, one notices small variations for resolution from one node to other (except that is for $j = 1$), as noted from the real part of k_{eq}/k for this scheme. This scheme uses the same boundary closure as in Adams' schemes (Eqs. (46) and (47)), and yet the effects of anti-diffusion is severer for the Haras–Ta'asan scheme. This is despite the fact that the interior stencil for Haras–Ta'asan scheme has displayed much superior resolution in Fig. 2, as compared to Adams' or Lele6 scheme.

In the third row from the top in Fig. 3(c), the resolution and diffusion properties are shown for the Zhong's scheme [39], where one notes stronger effects near the boundary for resolution shown by the real part of k_{eq}/k for almost the complete domain. It is equally noted in the imaginary part of k_{eq}/k that the anti-diffusion is significantly higher for the Zhong's scheme, affecting more number of points. These two frames are drawn for $\eta_Z = -1$, and if one wants to control the anti-diffusion more, then one needs to take larger value of η_Z in magnitude.

This leads to the conclusion that the problems of compact high accuracy schemes originate in the use of implicit schemes for boundary closure. This observation was used in [1,29] by replacing implicit boundary closure schemes by explicit boundary closure schemes. Noting that explicit schemes have only local effects, one can control anti-diffusion by such replacements. The results are evident in the bottom two rows of Fig. 3 showing two such carefully designed schemes, OUCS3 and NCCD schemes, showing their properties in discretizing first derivative. Two such boundary closure schemes are given in Eqs. (49) and (50). In Fig. 3(d), the real and imaginary parts of k_{eq}/k have some specific features depending upon the interior and boundary closure stencils. Since the stencils for $j = 1$ and N are symmetric, one notes the real part to be identical, while the imaginary part show numerical diffusion and anti-diffusion in identical magnitude. However, the closure given by Eq. (50) is dispersive due to effects of β_2, and also β_{N-1} are of different sign and magnitudes, which results in different values for the real and imaginary parts of k_{eq}/k for this pair of points. As these boundary closures are essentially lower order schemes, the resolution degradation is noted for near boundary points. However, the main benefit is noted in reducing the intensity and extent of anti-diffusion for OUCS3 scheme. It is to be noted that the authors in [29], additionally suggested using explicit fourth order diffusion terms with specified coefficient values, so that one has a practical compact scheme. In [1], it is also noted that in actual applications, one would discard the derivative obtained by compact scheme at $j = 2$, by a simple CD2 scheme, so that there is no anti-diffusion for any nodes.

The above concept of using explicit boundary closure schemes have been also used for the proposed NCCD scheme [1,64,65], whose results are shown in Fig. 3(e) as the bottom frames. The resolution for the near-boundary points are similar for NCCD scheme with those for the OUCS3 scheme. However, the anti-diffusion property of NCCD scheme for near-boundary points are even better than that is noted for the OUCS3 scheme. This aspect has been highlighted in computing the lid-driven

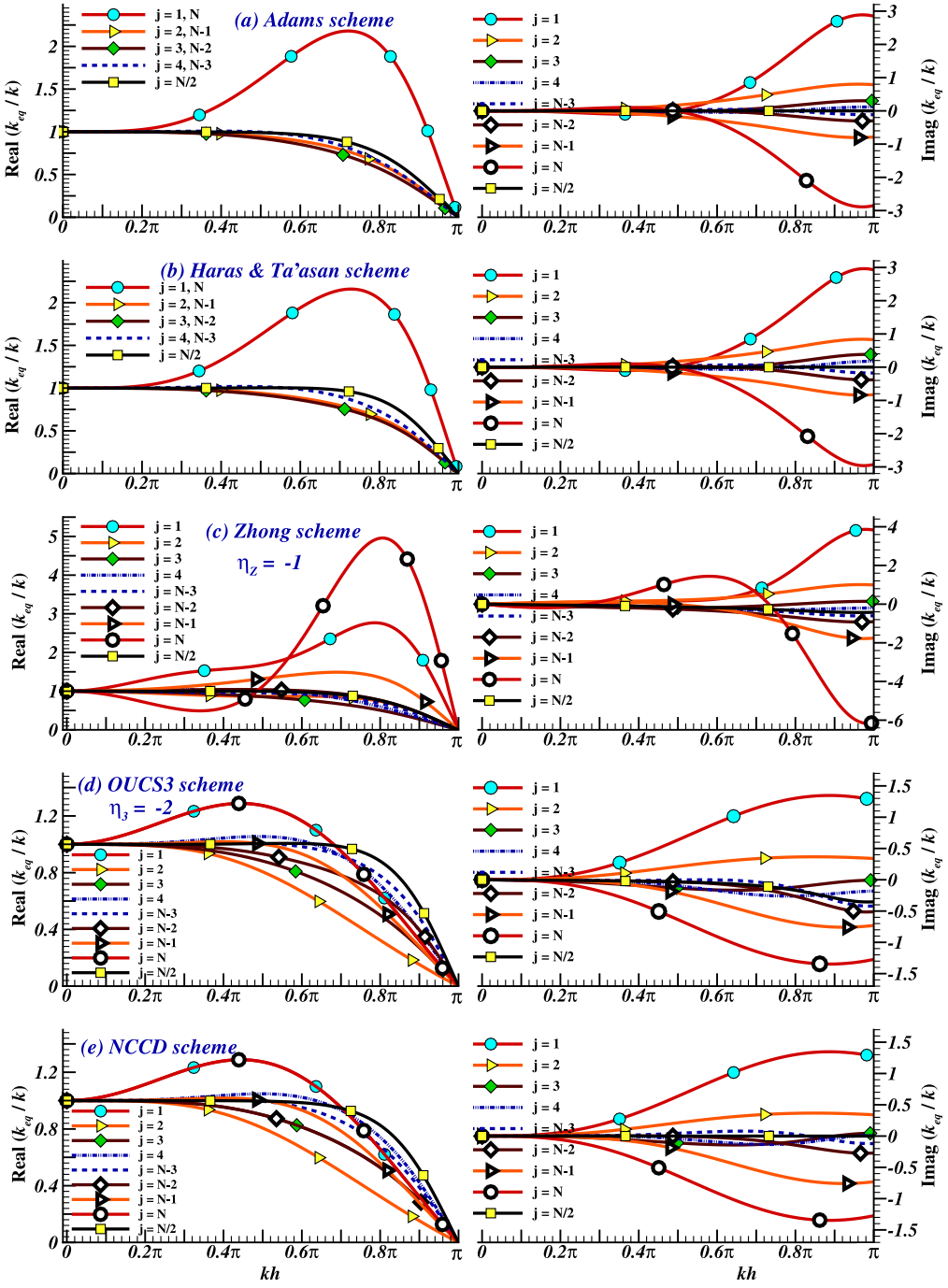

Fig. 3. Numerical resolution and diffusion/anti-diffusion introduced for different central and upwind compact schemes used for spatial discretization of first derivative are shown by plotting the real (left) and imaginary parts of k_{eq}/k as function of kh for the indicated nodes.

cavity problem by solving the NSE in [64,65] by NCCD scheme in capturing gyrating polygonal core vortex at the center of the cavity. It has been noted in [1] that Lele6 scheme with different closure schemes failed due to aliasing error.

3.3. Resolution of schemes for second derivative

One can compare the efficacy of different schemes in representing second derivatives by again displaying the quantities in spectral plane. Using the representation for the unknown given in Eq. (5) (omitting the time as the other independent variable, without any scope for confusion), the second spatial derivative can be written using Eq. (5) as, $u''_j|_{exact} = -\int k^2 U(k) e^{ikx_j}\, dk$. As it has been done for first derivative, one can also notationally represent numerical second derivative by,

$$u''_j|_{num} = -\int k^{(2)}_{eq} U(k) e^{ikx_j}\, dk$$

Thus, the resolution of second derivative can be represented in the spectral plane by plotting $-k^{(2)}_{eq}/k^2$ as a function of kh.

Some of the methods of discretization for the second derivative is described next, with the CD8 scheme already given in Eq. (36). For the NCCD scheme which provides both the first and second derivatives, the interior stencils are given as in Eqs. (37) and (38), and the boundary closure schemes given in Eqs. (39) to (42). Additionally, the compact scheme given in [33] for second derivative are given for the boundary

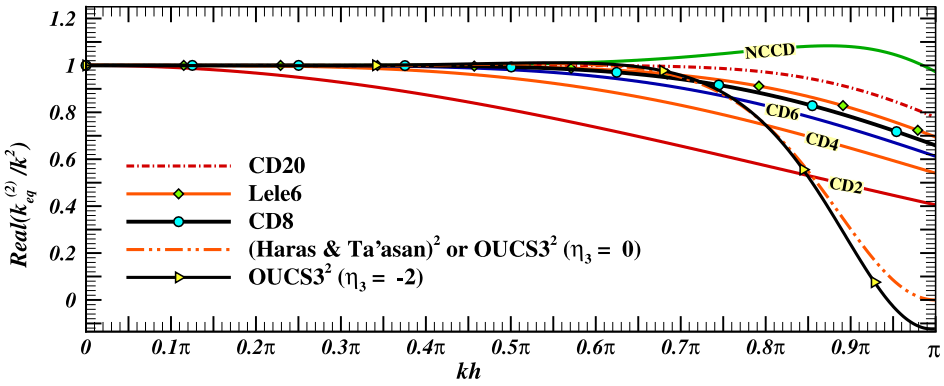

Fig. 4. Spectral resolution of different central explicit and implicit spatial discretization schemes for second derivative shown by plotting the real part of $k_{eq}^{(2)}/k^2$ as function of kh. The CD-schemes are explicitly central differenced, and here shown up to twentieth order accurate. The Haras–Ta'asan and OUCS3 schemes for first derivative have been applied twice to obtain second derivative and marked as $(.)^2$ in the legend.

and interior nodes as [1],

$$j = 1: \quad u_1'' + 11u_2'' = (13u_1 - 27u_2 + 15u_3 - u_4)/h^2 \tag{54}$$

$$j = 2: \quad u_1'' + 10u_2'' + u_3'' = 12\,(u_3 - 2u_2 + u_1)/h^2 \tag{55}$$

$$3 \le j \le N-2: \quad \alpha\, u_{j-1}'' + u_j'' + \alpha\, u_{j+1}'' = \frac{b}{4h^2}\,(u_{j-2} - 2u_j + u_{j+2}) + \frac{a}{h^2}\,(u_{j-1} - 2u_j + u_{j+1}) \tag{56}$$

with $\alpha = 2/11$, $a = 12/11$ and $b = 3/11$ required for formal sixth order accuracy based on truncation error. This will also be identified as Lele6 implicit closure scheme. To show the importance of using explicit boundary closure scheme, we also consider another variant of this scheme, for which one will use Eq. (56) for the interior nodes and use Eqs. (42) for $j = N$ and CD2-scheme stencil for $j = N-1$. This will be referred to as Lele6 explicit closure scheme.

In Fig. 4, the resolution of the interior points are compared among various explicit and implicit schemes, with appropriate boundary closure scheme by plotting $-k_{eq}^{(2)}/k^2$ as a function of kh. The central explicit schemes have been shown for representative formal order of accuracy up to twentieth order. The CD8 central scheme has been used for high performance computing in [61]. One also notes another possibility of computing second derivative by using compact scheme for the first derivative twice for the Haras–Ta'asan and OUCS3 schemes. However, this brings down the resolution to zero value, at the Nyquist limit, which will appear inferior, as compared to even CD2 scheme above $kh = 0.85\pi$. The Lele6 scheme provides good resolution up to $kh = \pi$ and provide identical resolution for both implicit and implicit closure for the interior nodes. This scheme has marginal superior resolution as compared to CD8 scheme. In contrast, CD20 has overall better resolution than the rest of the schemes shown, except for the NCCD scheme. NCCD scheme is unique in that for the complete range of kh, nowhere the quotient $-k_{eq}^{(2)}/k^2$ is less than one. It actually shows that the numerical second derivative is slightly more than the physical value and its importance has been highlighted in [64,65] from the physical and numerical reason. As the scales near the Nyquist limit are most important from the point of view of resolving the enstrophy and dissipation, its proper resolution cannot be over-estimated. Also, at and near the maximum resolved scales, aliasing error deposits spurious values there, and the displayed feature of NCCD scheme helps alleviate this major source of error. This has been shown and explained in [1,65].

In Fig. 5, results of GSA for the representative nodes are shown, by plotting the real and imaginary parts of $k_{eq}^{(2)}/k^2$ as function of kh. In Fig. 5(a) the real and imaginary parts of $k_{eq}^{(2)}/k^2$ are shown as function of kh for representative nodes. We note the attribute of GSA by which the effectiveness of $k_{eq}^{(2)}/k^2$ does not vanish due to full domain analysis, which is noted for the deep interior points. Results are shown for the upwind constant value of $\eta_3 = -2$, and induced dispersion caused by the imaginary part of $k_{eq}^{(2)}/k^2$ are mostly noted for near-boundary points, while such dispersion is absent for interior nodes for a significant range of kh. Of specific interest is the comparison between implicit and explicit boundary closures used with Lele's scheme for second derivative. In the implicit closure, the use of one sided scheme at $j = 1$ and N creates an unphysical bias of the computed second derivative. This creates a strong deviation of $k_{eq}^{(2)}/k^2$ from the ideal unity value, shown in Fig. 5(b) and (c), as function of kh. Also, the closure being implicit in frame (b), one notices strong distortion from ideal effectiveness percolating to the next inner points. However, from $j = 3$ onwards one notices very desirable effectiveness. Similar non=ideal behavior is noted mostly at $j = 1$ and N. In contrast to the implicit boundary closure case, in frame (c) one notices significantly reduced distortion for real and imaginary parts of $k_{eq}^{(2)}/k^2$ as function of kh, which is reduced by an order of magnitude. This clearly shows that even though the compact schemes are implicit, one would violate the physical nature of information propagation, if one chooses implicit boundary closure. In Fig. 5(d), the resolution and dispersion properties are shown for NCCD scheme. The resolution of NCCD scheme for the second derivative is noted to be even better than that provided by the Lele explicit closure scheme. The dispersion error of NCCD scheme is slightly inferior, as compared to the explicit closure scheme, and this once again should convince discerning readers that high accuracy compact schemes with good interior stencils (as shown in Fig. 4) would perform better with explicit boundary closure schemes. Any loss of accuracy of explicit closure confines itself locally.

4. Properties of space-time discretization: Canonical 1D CE

So far the spatial discretization methods have been described with typical measure for their effectiveness. While one can analyze time discretization schemes, as if ordinary differential equation is being solved, we are interested in solving space–time dependent equations, for which the concept of dispersion relation has been introduced in Section 2, identifying correct way of interpreting dispersion relation for the 1D CE. In the quest for tracking numerical error, it is essential that various sources of it should be understood. Starting with early work in [4], one of the main preoccupations of analysts has been to somehow prevent numerical instability. The authors in [4] clearly have articulated this with their famous von Neumann stability analysis that *our concern here is with stability rather than with accuracy*. For parabolic partial differential equations, such as that for heat conduction, one is often interested in the time-asymptotic solution and for which overtly stable method may *converge* faster to steady state. However, if one were to be also interested in the transient state, then correct physical stability is to be sought, and numerical stability should mimic this, as shown

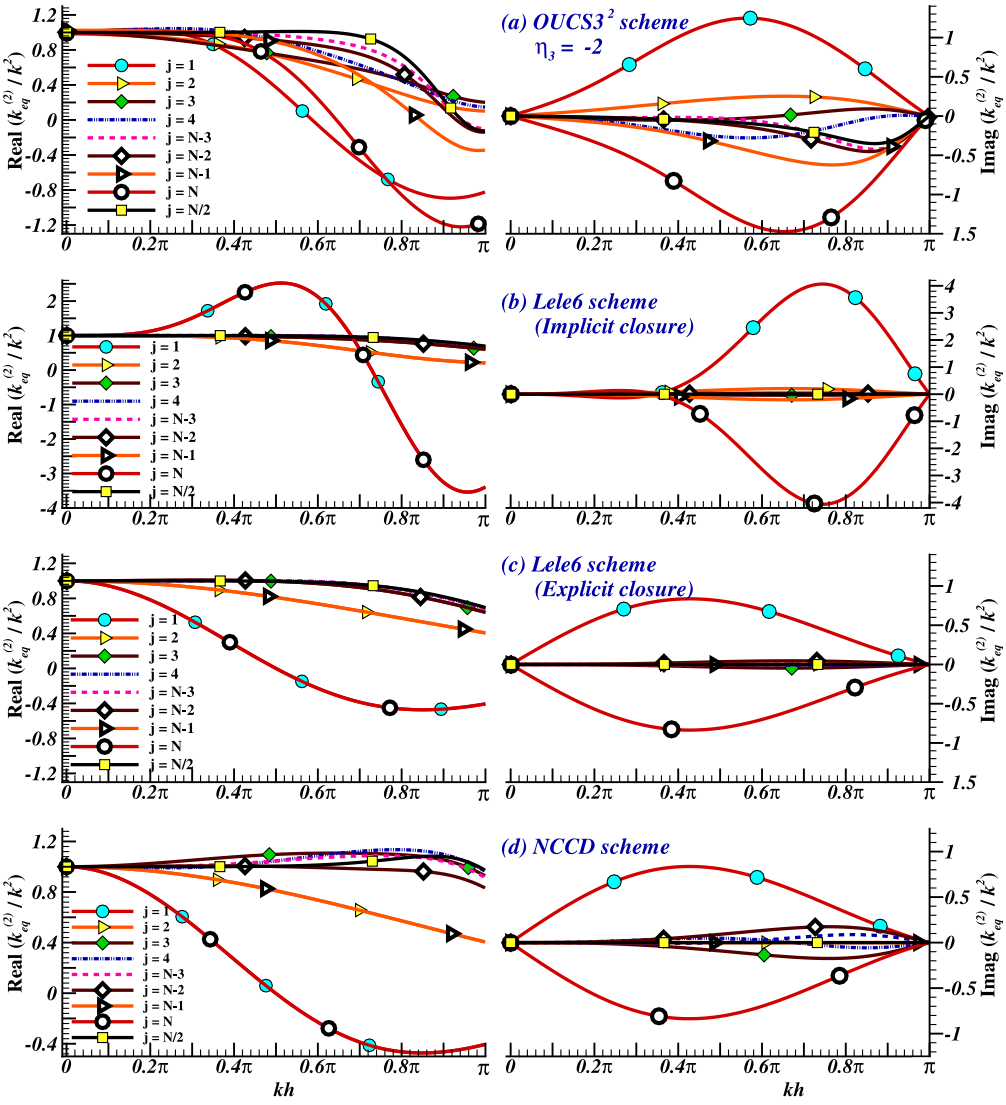

Fig. 5. Spectral resolution and dispersion of non-periodic, explicit-central and implicit spatial discretization schemes for second derivative shown by plotting the real and imaginary parts of $k_{eq}^{(2)}/k^2$ as function of kh. The OUCS3 scheme for first derivative have been applied twice and marked as OUCS3^2 to obtain second derivative and all other schemes are specifically designed to discretize second derivative.

in [48] for heat equation. This makes a search for good model space–time dependent equation very necessary. A good case in point is the consideration of 1D CE [1,23,31,34] or the 1D convection diffusion equation [49], which admits an exact solution. For example, the 1D CE shows the solution to be neutrally stable. To test a method for neutral stability a precise analysis of the method is required as opposed to a qualitative answer such as the method is stable or unstable. An unstable method will readily display its pathology. It is the so-called stable method, which adds to the confusion. Thus, using 1D CE as the canonical problem is a very useful approach and is followed next.

In writing this in Eq. (2), one notices the presence of a first derivative with respect to time. That in turn needs the time integration to be a two-step method. In this context, it has been shown that four stage, Runge–Kutta (RK4) method to be very efficient [31] given for a semi-discrete equation in the form as,

$$\frac{\partial u}{\partial t} = L(u),$$

with L representing the operator after performing all spatial discretizations. The four steps of RK4 method are given in terms of the time step Δt for the time integration by,

First Stage: $u^{(1)} = u^{(n)} + \frac{\Delta t}{2} L[u^{(n)}],$

Second Stage: $u^{(2)} = u^{(n)} + \frac{\Delta t}{2} L[u^{(1)}],$

Third Stage: $u^{(3)} = u^{(n)} + \Delta t L[u^{(2)}],$

Fourth Stage: $u^{(n+1)} = u^{(n)} + \frac{\Delta t}{6}\left[L[u^{(n)}] + 2L[u^{(1)}] + 2L[u^{(2)}] + L[u^{(3)}]\right]$

For the space–time advancement of the unknown given in Eq. (5) from t to $t + \Delta t$ is notationally represented by the numerical amplification factor given as, $G(kh, N_c) = U(kh, t + \Delta t)/U(kh, t)$, with N_c as the CFL number equal to $c\Delta t/h$. For the RK4-time integration method, this is given for the jth node by [1,32],

$$G_j = 1 - A_j + \frac{A_j^2}{2} - \frac{A_j^3}{6} + \frac{A_j^4}{24} \tag{57}$$

where $A_j = N_c \sum_{l=1}^{N} C_{jl}\, e^{ik(x_l - x_j)}$. This equation for RK4 scheme is for any spatial discretization of non-periodic problems obtained by GSA for the full-domain analysis, as given for some explicit and implicit spatial schemes in [1,67,68]. While $|G_j|$ as the nodal amplification factor is a source of error, additional error can arise due to dispersion, which can be severe as compared to error caused by stable algorithm.

The dispersion error is obtained using GSA, that identifies its primary source as due to the constant prescribed phase speed becoming wavenumber dependent. If the initial condition is represented for Eq. (2) to be given by Eq. (11), then the solution at any time, $t = n\Delta t$, is written using the definition of amplification factor G_j given by Eq. (12). We also recollect that $|G_j| = (G_{rj}^2 + G_{ij}^2)^{1/2}$ and $\tan\phi_j = -G_{ij}/G_{rj}$, with G_{rj} and G_{ij} as the real and imaginary parts of G_j, respectively. One calculates ϕ_j appropriately, by considering signs of G_{rj} and G_{ij}.

The numerical phase speed (c_N) is obtained from ϕ_j as the phase shift per time step so that $n\phi_j = kc_N n\Delta t$ as given by Eq. (13). The physical phase speed is c for all wavenumber, but c_N is noted to depend on k. Thus, the numerical solution is dispersive, in contrast to the non-dispersive nature of 1D CE, with the computed solution is given by,

$$\bar{u}_N = \int U_0(k)\,[|G|]^{t/\Delta t}\, e^{ik(x-c_N t)}dk \tag{58}$$

The numerical dispersion relation is given in Eq. (8), instead of the wrong dispersion relation given in Eq. (6). Having obtained the correct dispersion relation and the non-dimensional numerical phase speed, numerical group velocity at the jth-node can be expressed as

$$\left[\frac{V_{gN}}{c}\right]_j = \frac{1}{hN_c}\frac{d\phi_j}{dk} \tag{59}$$

It is important to understand the roles played by these numerical parameters in ensuring the accuracy of scientific computing. It has been very effectively achieved in [31], where the concept of error propagation was introduced in the correct perspective. This subject of error dynamics has been subsequently shown for the diffusion equation [48], CDE [49], the KdV equation [69] and the NSE [70]. Following demonstration is following the presentation in [1] for 1D CE.

Defining the numerical error by, $e(x,t) = u(x,t) - \bar{u}_N(x,t)$, one can derive the governing equation for $e(x,t)$ in the manner shown next. Using Eq. (58) one obtains the expressions for $\frac{\partial \bar{u}_N}{\partial x}$ and $\frac{\partial \bar{u}_N}{\partial t}$, with the help of which one writes the error dynamics equation as,

$$\begin{aligned}\frac{\partial e}{\partial t} + c\frac{\partial e}{\partial x} &= -[1 - \frac{c_N}{c}]c\frac{\partial \bar{u}_N}{\partial x} \\ &\quad - \int \frac{V_{gN} - c_N}{k}\left[\int ik' U_0\,[|G|]^n\, e^{ik'(x-c_N t)}dk'\right]dk \\ &\quad - \int \frac{\mathrm{Ln}\,|G|}{\Delta t} U_0\,[|G|]^n\, e^{ik(x-c_N t)}\,dk\end{aligned} \tag{60}$$

This error dynamics equation is different in form and concept that can be deduced following the assumption of von Neumann analysis [4]. In the latter, the right-hand side of Eq. (60) is equated to zero, on the premise that the basic governing 1D CE is linear, so signal and error follow the identical governing equation. Quite revealingly the von Neumann analysis does not even quantify the error due to stability/instability as given by the last term on the right-hand side of Eq. (60). Of course, another major advantage of adopted GSA lies in its ability to quantify the dispersion and phase error, as given by the right hand side terms, which are absent from any discussion on von Neumann analysis. From Eq. (60), one readily notes that $|G|$, c_N/c and V_{gN}/c are the main metrics which contribute to error for 1D CE.

To understand the utility of Eq. (60), one must inspect the numerical properties of a specific combination of spatial and temporal discretization methods. As noted already, for 1D CE for the first derivative with respect to time, RK4 time integration is adopted, as it provides high accuracy, without invoking spurious numerical modes. The obtained properties are shown next in Figs. 6 to 8, where the correct interpretation of numerical dispersion relation is explained. The error metrics $|G|$, c_N/c and V_{gN}/c are used for the comparison, as obtained by GSA and those obtained following the wrong dispersion relation given by Eq. (6). In the wrong approach, the numerical phase speed is taken as that is given by the physical phase speed. Hence, for this approach, c_N/c is identically equal to one and is therefore not shown. One must also appreciate the fact that the numerical dispersion relation is one of interpretation, after the space and time discretization have been fixed. Thus, in both these view points, the discrete equation remains the same. As a consequence, the numerical amplification rate shown are one and the same, when plotted in the (N_c, kh)-plane.

In Fig. 6, the spatial discretization considered is the CD8 scheme, as used in [61]. For all the plotted quantities in various frames, the maximum and minimum values are printed to understand the dynamic range of the quantities plotted. As mentioned above, the $|G|$ contour plots obtained by following the GSA and incorrect dispersion relation approach shows the same portrait in both the frames. As for the CE, ideally the amplification factor should be neutrally stable, this figure shows that one must choose a very small value of N_c close to the origin, so that $|G| \approx 1$. One notes from Eq. (60), that the quotient c_N/c not being equal to one contributes to forcing of the error, when $\frac{\partial \bar{u}_N}{\partial x}$ takes large non-negligible values. Such conditions prevail when there is a sharp front in low speed flow, or when there is a sharp discontinuity, as in a shock for transonic/supersonic flows. Large phase error is noted for high wave numbers at low values of N_c, or for moderate wavenumbers at high values of N_c. It is the group velocity, for which one notices major differences between the correct and incorrect numerical dispersion relations. While the range of maximum and minimum values are distinctly different, there is the value of kh for which both these show zero group velocity. Above this range the group velocity becomes negative, and these are called as the q-waves, following the nomenclature in [32,36,71]. For central schemes, this is related to $\frac{dk_{eq}}{dk} = 0$, which is the case for both of these interpretations that occurs for kh slightly greater than 2. However, for all other combinations of kh and N_c values the numerical group velocity has to be obtained by GSA accurately/correctly.

In Fig. 7, the spatial discretization is chosen as the sixth order central compact scheme, used along with RK4 time integration scheme for the solution of 1D CE. Once again, the $|G|$ contours are same from both these perspectives. As the Lele's scheme [33] is more accurate as compared to CD8 scheme, this will improve the resolution and provide better neutral stability of the schemes shown in Fig. 7. At the same time, the maximum value of $|G|$ will be higher for Lele-RK4 scheme, as compared to CD8-RK4 scheme. The numerical phase speed obtained by GSA is also noted to be better for the Lele-RK4 scheme, as compared to the CD8-RK4 scheme. One also notices that q-waves appear for higher value of kh for the scheme in Fig. 7, showing better numerical properties for Lele-RK4 scheme.

In Fig. 8, the same three error metrics ($|G|$, c_N/c, V_{gN}/c) are compared when the spatial discretization is replaced by OUCS3 scheme. In comparing the spatial discretizing of first derivatives, it has been noted that the OUCS3 scheme provides extremely high accuracy. This aspect is noted in the property charts shown in Fig. 8. Because of the upwind nature of this scheme, the numerical amplification factor shows attenuation as compared to the previous two methods. However, the dispersion and phase error properties of OUCS3-RK4 scheme is superior, as compared to CD8-RK4 and Lele-RK4 schemes. The onset kh value for q-waves is also the highest for this scheme. It is noted that dispersion errors are the major source of problems for high accuracy computing, and thus OUCS3-RK4 scheme is found to be preferable as compared to other high order schemes.

4.1. Proof of GSA dispersion relation

We have noted above that the existence and interpretation of dispersion is one of concept, as both the GSA-based and the traditional methods use the same discrete equation that is actually solved numerically. As a consequence, the numerical amplification factor is one and the same for both the methods. It is the supposition that the numerical and physical phase speed for Eq. (2) are same, is at the root of the problem. Because application of GSA in Section 2 shows that the numerical phase speed cannot be assumed to be a constant with respect to wavenumber, while the discrete equation enforces a phase shift via

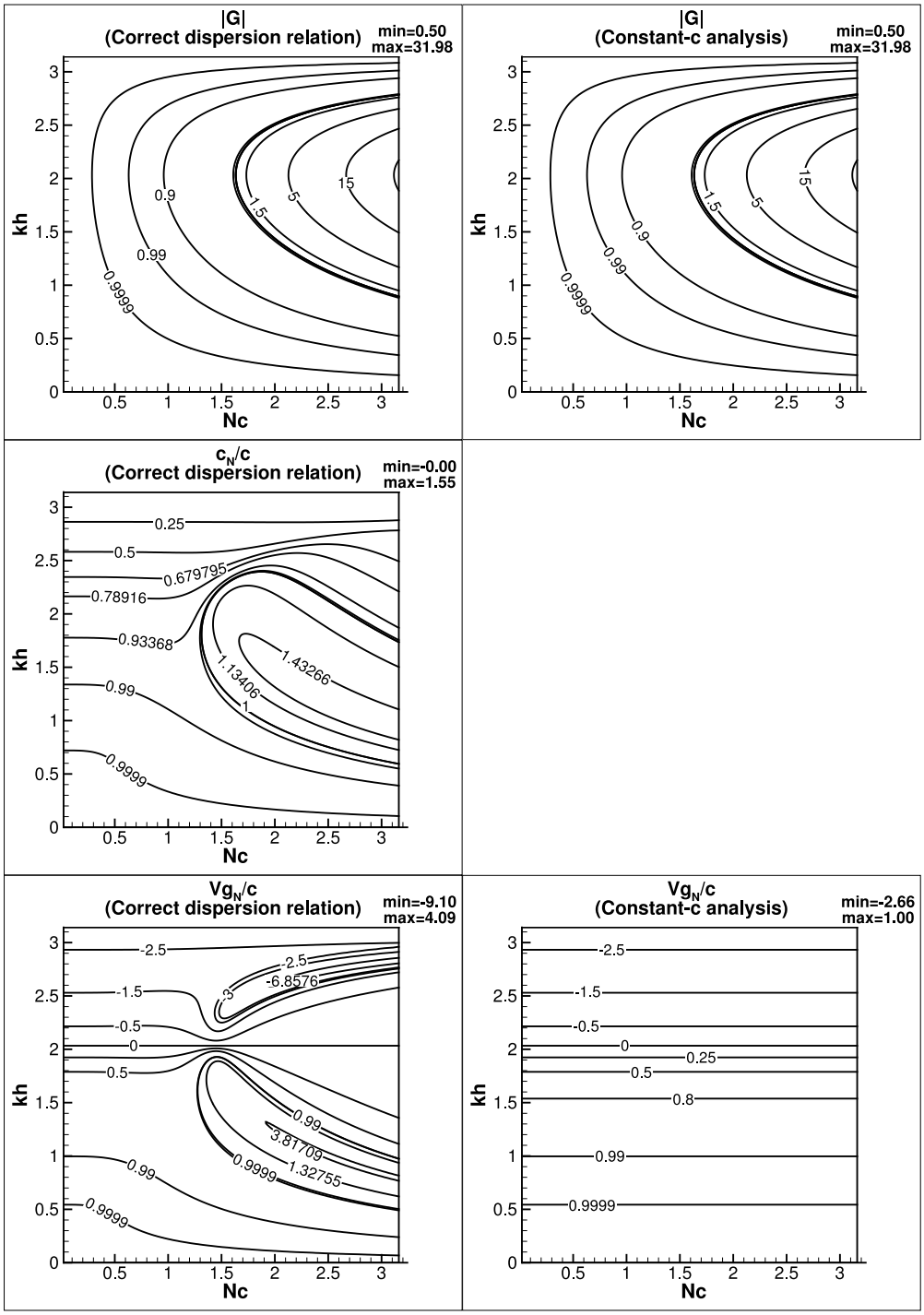

Fig. 6. Numerical properties of CD8 spatial discretization and RK4 time integration schemes in solving Eq. (2). The properties obtained using correct dispersion relation, Eq. (8) are shown in the left frames. Use of the incorrect dispersion relation (i.e. using $c_N \cong c$) in Eq. (6) leads to the properties shown on the right frames. Shown are $|G|$, c_N/c and V_{g_N}/c in various frames and compared for a typical interior point.

the discrete equation. This enabled us to show in Figs. 6 to 8 that numerical phase speed to be wavenumber dependent, and consequently the group velocity provided by Eqs. (6) and (8) are qualitatively and quantitatively different.

Thus, it is possible to design a test in which an initial wave-packet will be allowed to propagate following Eq. (2). The wave-packet is considered such that it can be traced distinctly with a small range of kh, such that the property of the wave-packet can be identified by a mean location without very large spread about this mean. Also, to compare the correct dispersion relation based on GSA, with the incorrect dispersion relation used by many authors by treating the numerical phase speed as constant [43], we need this wave-packet to have properties which are easily visualized. Propagation speed of the wave-packet given by the group velocity computed from Eqs. (6) and (8) is the appropriate quantity to check, as the numerical amplification factors are same, while the numerical phase speed is the subject of

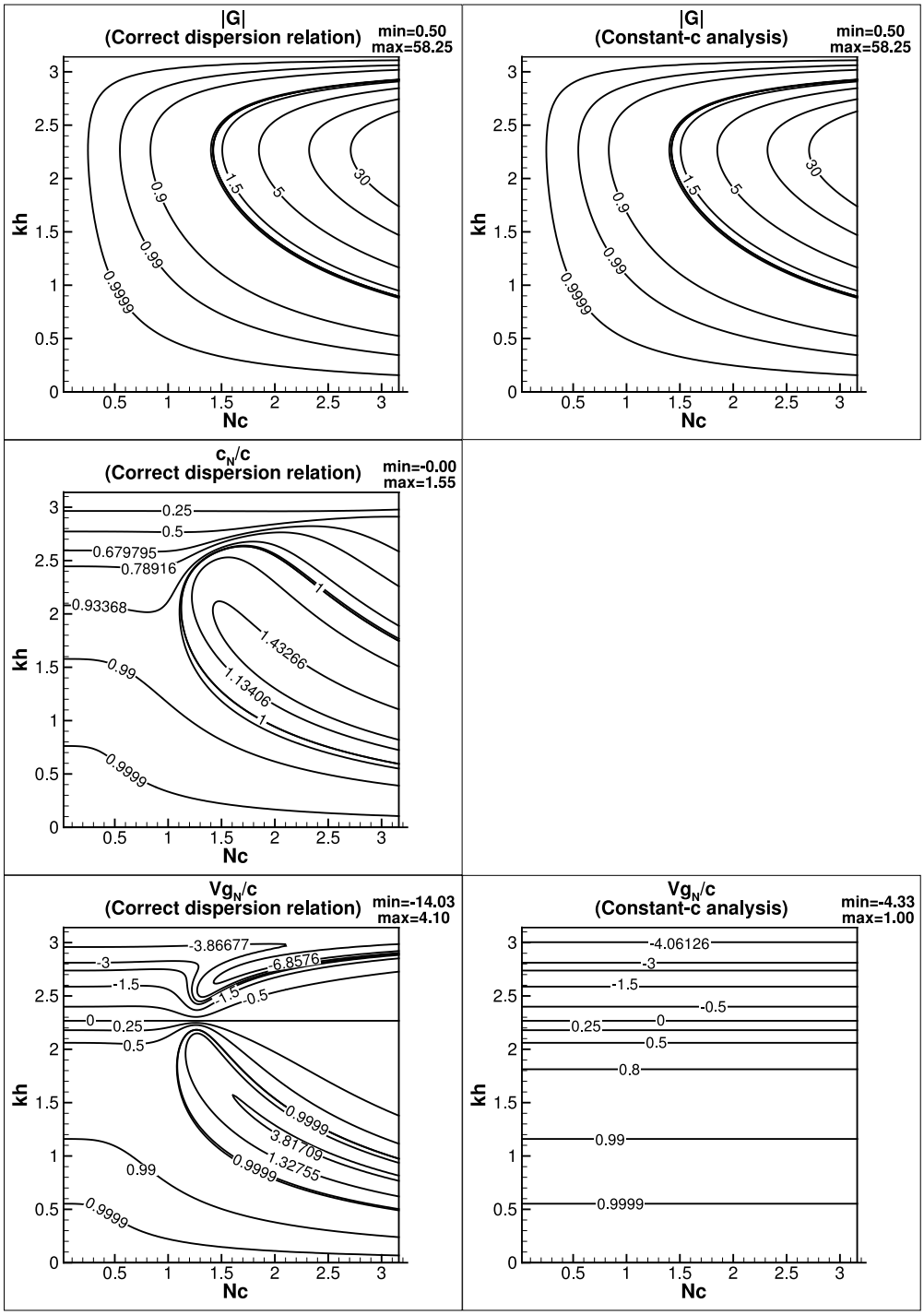

Fig. 7. Numerical properties of Lele's compact discretization and RK4 time integration schemes used for solving Eq. (2). The properties obtained using correct dispersion relation, Eq. (8) are shown in the left frames. Use of the incorrect dispersion relation (i.e. using $c_N \cong c$) in Eq. (6) leads to the properties shown on the right frames. Shown are $|G|$, c_N/c and V_{g_N}/c in various frames and are compared for a typical interior point.

contention. To perform the test, OUCS3-RK4 method is chosen with $\eta_3 = 0$.

For the ease of such a comparison and to demonstrate existence of numerical q-wave, we carefully choose the wave packet as,

$$u = e^{-x^2} \sin(k_{in} x) \tag{61}$$

where the input wavenumber is fixed from $k_{in} = kh_{in}/h$, with kh_{in} considered as 0.7981π. The periodic domain considered is given by, $-2.5\pi < x \leq 2.5\pi$ with equi-distant 2505 points, so that the grid spacing is given by, $h = 0.002\pi$. In Eq. (2), the physical phase speed is taken as $c = \pi/2$ and the time step is taken as $\Delta t = 0.00511$, so that the CFL number is $N_c = 1.2775$. The reason for the choice of these numerical parameters are explained with the help of the property charts shown in Fig. 9, with the top two frames showing superposition of numerical amplification factors and numerical group velocity obtained by GSA and by considering numerical phase as identical to physical phase speed. The combinations of kh_{in} and N_c is marked by a circle in frames (a) and (b). Also note the vertical dash–dot–dot line plotted tangential to $|G| = 1$ line, to the right of which, some length scales are found to be unstable. Thus, this dash–dot–dot line provides the critical value

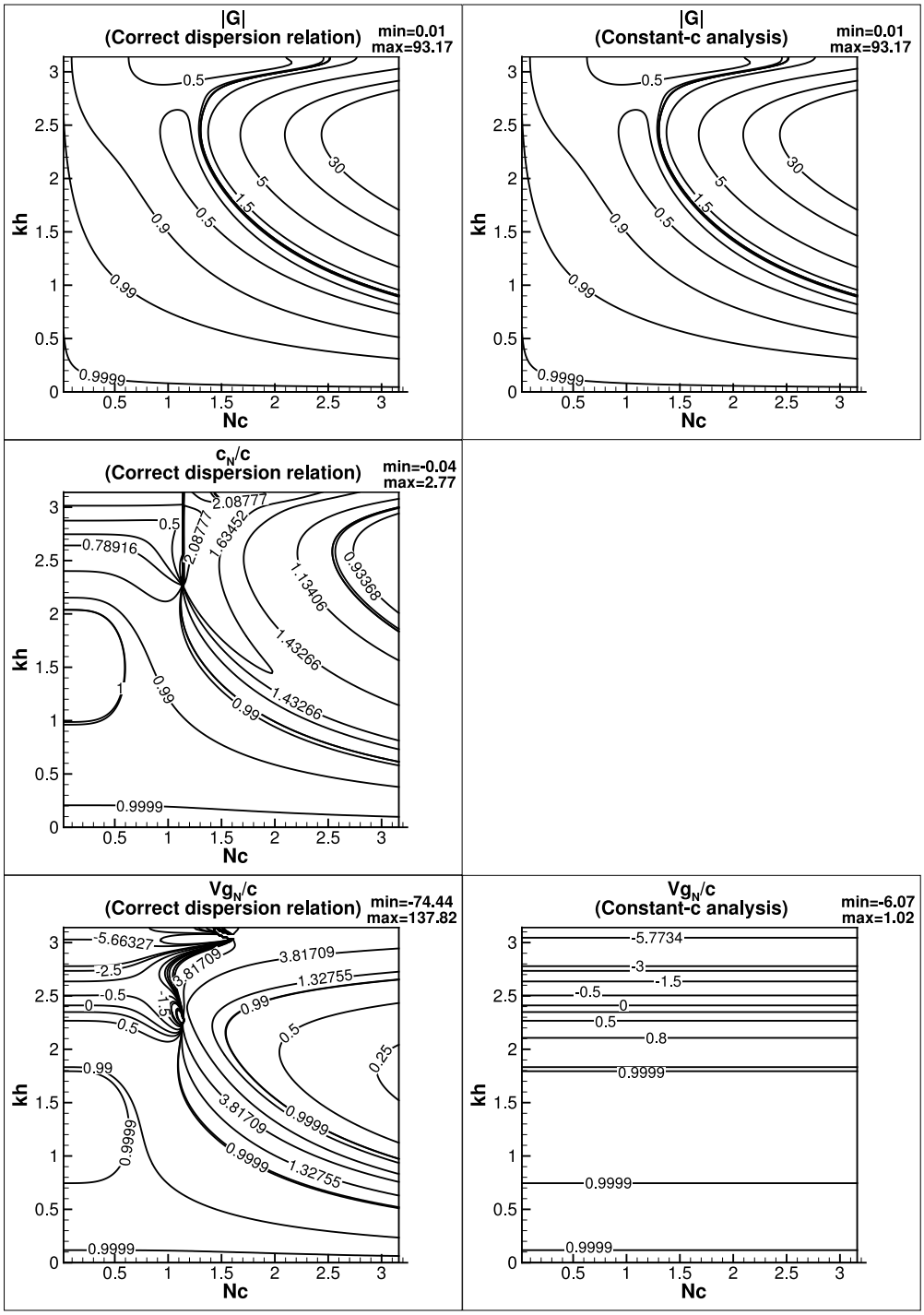

Fig. 8. Numerical properties of OUCS3 spatial discretization and RK4 time integration schemes used in solving Eq. (2). The properties obtained using correct dispersion relation, Eq. (8) are shown in the left frames. Use of the incorrect dispersion relation (i.e. using $c_N \cong c$) in Eq. (6) leads to the properties shown on the right frames. Shown are $|G|$, c_N/c and V_{g_N}/c in various frames and are compared for a typical interior point.

of N_c above which some length scales are inherently unstable, and omnipresent background numerical disturbances will magnify without being explicitly excited in the problem, an issue known as focusing in the literature [45,57,72–75].

In Fig. 9(a), $|G_N|$-contours are shown by thick lines and flood, while the thin contour lines depict V_{g_N}/c, as obtained from GSA using the correct dispersion relation given by Eq. (8). One also notices that only a small range of N_c is chosen for investigation, while the full range for wavenumber is chosen ($0 \leq kh \leq \pi$), with properties obtained using a uniformly distributed (1000×1000) values of kh and N_c. The intention for this choice of numerical parameter values is to consider the propagation of the wave-packet which does not attenuate in very few time steps. For both the dispersion relation cases the $|G_N|$-contours are the same. It is the numerical group velocity value that will be different, due to the conceptual error in writing Eq. (6). In Fig. 9(b), the numerical group velocity obtained from the wrong dispersion relation shows it to be independent of N_c, i.e. time integration method. However, for very low values of N_c close to zero, both the GSA and incorrect dispersion relation provide V_{g_N}/c-values, which are visually indistinguishable. Hence, in the range of Fig. 9(a), a region of interest

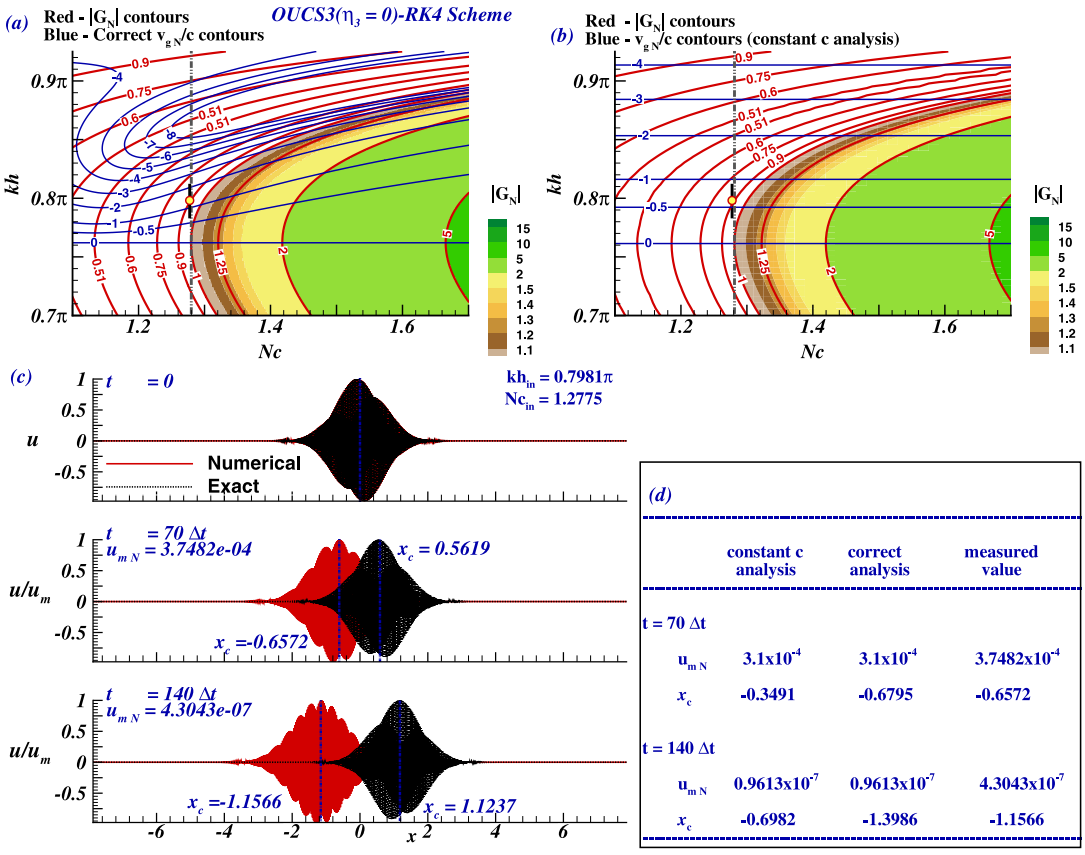

Fig. 9. The iso-contours of numerical amplification factor and group velocity is plotted in the (kh, N_c)-plane obtained by (a) present GSA and (b) treating phase speed as constant. The thick vertical strip shows the spectral extent of the chosen wave-packet. (c) Numerical solutions of Eq. (2) obtained by OUCS3-RK4 (with $\eta_3 = 0$) scheme are compared with exact solution. The numerical results are normalized with its time-dependent maximum value indicated as $u_{m\ N}$. The vertical lines show the approximate center of wave-packet (x_c). The $u_{m\ N}$ and x_c obtained by GSA and constant phase speed analysis is compared with numerical results.

is where $|G_N|$ is very close to one for the chosen wave-packet, while V_{gN}/c values given by correct and incorrect dispersion relations are distinctly different for Eq. (2). For the chosen numerical parameter of $kh = 0.7981\pi$ and $N_c = 1.2775$, numerical properties in solving Eq. (2) are given by, $|G| = 0.891$, the numerical group velocity calculated by correct dispersion relation is given by, $V_{gNc} = -1.2093$, while numerical group velocity obtained by constant phase speed assumption is given by, $V_{gNw} = -0.6213$.

The spectral extent of the wave packet is marked by a black vertical strip in Fig. 9(a) and (b), with the filled circle marking the center of the packet. The numerical simulation of Eq. (2) for the initial condition given by Eq. (61) using the above grid parameters and time steps are shown in Fig. 9(c). It is to be noted that for chosen parameters, at every time step, the magnitude of the numerical solution reduces 10.9% of the current value. Thus, the plotted numerical results in Fig. 9(c) are normalized with the time-dependent maximum value (denoted as u_{mN}). The group velocity of the numerical result can be calculated by measuring the speed of propagation of the wave-packet. The variable x_c shows the approximate center of the wave-packet, which is used to approximately calculate the numerical group velocity and compared with V_{gNc} and V_{gNw}. The table in frame (d) compares the numerically measured $u_{m\ N}$ and x_c with the values obtained by present GSA and constant phase speed analysis. It is to be noted that the measured value of u_{mN} from the simulation is higher than the estimated value from analyses. While u_{mN} from the analyses is obtained for a wave with $kh_{in} = 0.7981\pi$, the numerical simulation is performed for a wave-packet centered at kh_{in}. The spectral extent of the chosen wave-packet marked in the top two frames of Fig. 9 shows that different wavenumbers of the wave-packet decays with different $|G|$. Due to this effect, the estimated u_{mN} from analyses slightly differ from the numerical simulation. Further, in the chosen wave-packet, the wavenumbers less that kh_{in} have $|G| > 0.891$ and are dominant over time. The corresponding group velocity of these wavenumbers are less in magnitude. Thus it is expected that the analyses will slightly overestimate the magnitude of the x_c. The table in Fig. 9(d) conforms that present analysis give the appropriate x_c. The constant phase speed analysis highly underestimates the magnitude of the x_c which indicates that the numerical wave-packet follows the properties of the present GSA.

4.2. Proof of q-waves in computing: Identification and demonstration for 1D CE using GSA

In performing numerical computations some researchers have reported an interesting phenomena where spurious numerical waves are created in addition to physical waves [32,71,76,77]. The spurious numerical waves are termed as q-waves and they are called so as they propagate in the opposite direction of the physical waves. Trefethen [40] conjectured that these q-waves are related to the group velocity property based on earlier work by Vichnevetsky and Pfeiffer [75] but provided no quantitative measure for their occurrence except for suggesting that they occur close to the Nyquist limit. It is also noted that Vichnevetsky and co-authors [32,75] qualitatively explained the spurious waves for CD_2 and Galerkin FEM schemes for the 1D convection equation using semi-discrete analysis. Sengupta et al. [36] finally explained the existence of q-waves for finite difference, finite volume and finite element methods using numerical properties derived for the model convection equation using GSA for the first time. Furthermore, the authors conclusively demonstrated an excellent correspondence with GSA using numerical simulations of 1D, 2D convection equation and 2D linearized rotating shallow water equations. Following

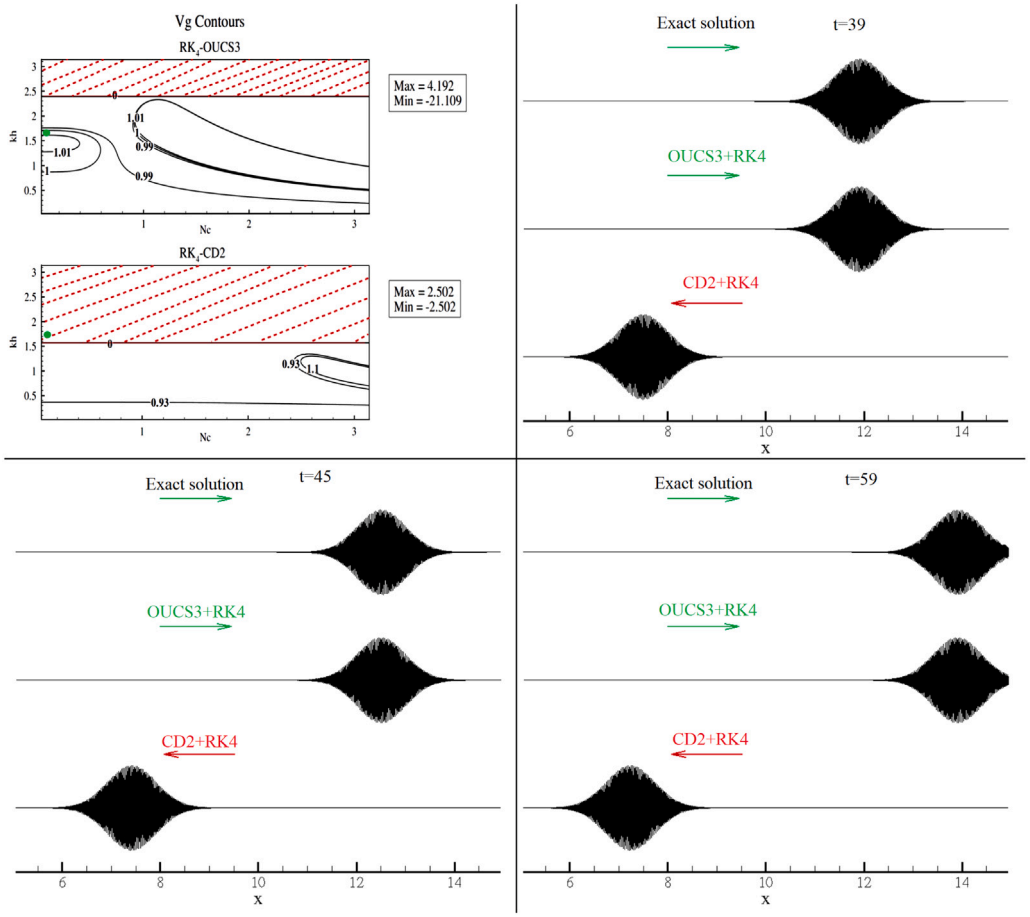

Fig. 10. Analytical proof and demonstration of q-waves using GSA for 1D linear convection equation. Top left panel shows the ratio of numerical group velocity to physical group velocity for the indicated numerical schemes. Top right and bottom panels show the numerical solution of 1D convection equation at the given time instants for the same schemes demonstrating the presence of q-waves.

the success of GSA, q-waves have also been demonstrated using the same analysis for convection–diffusion [49] and convection–diffusion–reaction [47] systems thus establishing their omnipresent nature in numerical computing of wave phenomena.

In this subsection, a proof of q-waves in numerical computing is presented using GSA and their existence is demonstrated using solution of linear 1D convection equation. For the purpose of demonstration, two different spatial explicit second order central difference scheme (CD_2) and an optimized upwind compact scheme (OUCS3) are used for spatial derivatives while the classical fourth order Runge–Kutta RK_4 method is employed for time integration. For the two schemes the ratio of numerical group velocity to its physical counterpart (Vg) is determined using GSA as given by Eq. (59) and the contours are plotted in the top left panel of Fig. 10. From the plot one immediately notes the existence of upstream propagating waves with $Vg < 0$. For the 1D convection equation, the information should travel downstream when $c > 0$ in Eq. (2). Thus, these upstream propagating waves are spurious in nature and are therefore the q-waves.

Having established the existence of q-waves using GSA, a numerical solution of the 1D convection equation is performed in order to demonstrate their presence in computing. An initial solution in the form of a wave packet is considered with its central wavenumber chosen as indicated by a green dot in the Vg contours in the top left panel of Fig. 10. According to GSA, the chosen initial solution should lead to the appearance of q-waves for the RK_4-CD_2 scheme whereas they are noted to be absent for the high accuracy compact scheme case i.e. RK_4-OUCS3 scheme. The top right and bottom panels of Fig. 10 show the evolution of numerical solution plotted at different times for the two numerical schemes. In Fig. 11, a multimedia link showing the animation of numerical solutions of the 1D convection equation using the two schemes is presented. It is evident from the results that the solution from the RK_4-OUCS3 scheme propagates in the physical direction whereas for the RK_4-CD_2 scheme upstream propagation is noted which is in accordance with GSA thereby demonstrating the presence of q-waves in numerical computing.

The present demonstration unequivocally establishes the presence of spurious, upstream propagating q-waves and highlights the power of GSA in identifying and accurately quantifying these waves. Although the demonstration is shown for the 1D linear convection equation, it should be noted that these q-waves are ubiquitous in numerical computation of wave phenomena as shown in [36,47,49].

5. Multi time-level methods

The four stage Runge–Kutta method, introduced earlier, belongs to the class of higher order time integration schemes known as single-step multistage methods. Another class of methods also exist in the literature for higher order integration, called the multi time-level schemes. As the name indicates, these schemes involve at least three levels for time integration. Popular methods such as Adams–Bashforth [78–80], Leap-frog [5,80], EXT2 [81,82] and Gear schemes belong to this class.

Here, only three-time level methods are considered, particularly second order Adams–Bashforth method (AB_2), to illustrate the performance of multi time-level methods. Any generic three-time level method for solving the equation $\frac{\partial u}{\partial t} = L(u)$ can be written as

$$u_j^{n+1} = \kappa_1 u_j^n + \kappa_2 u_j^{n-1} + \gamma_1 \Delta t L(u_j^n) + \gamma_2 \Delta t L(u_j^{n-1}) \tag{62}$$

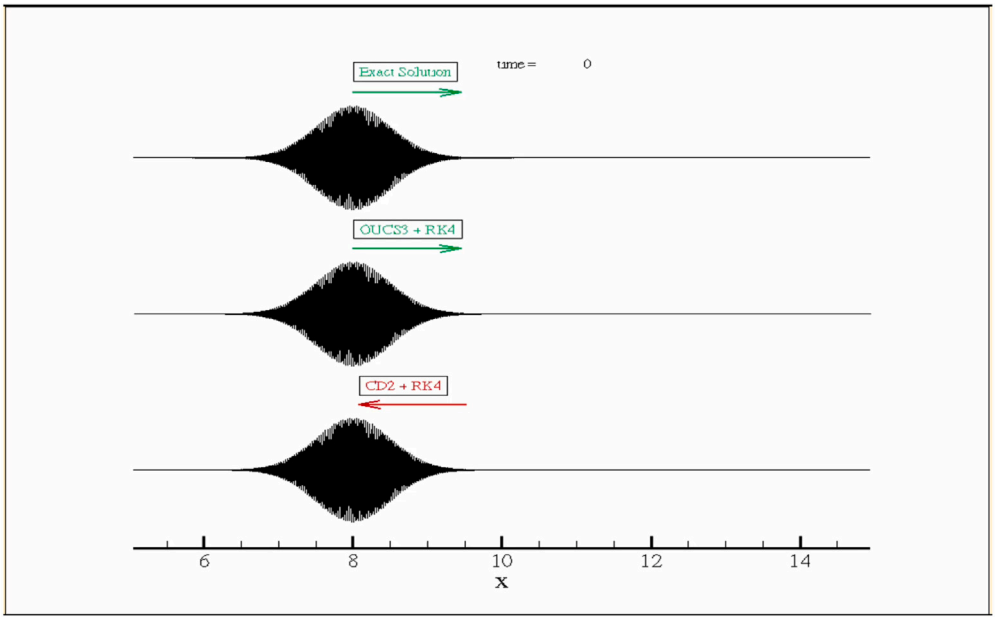

Fig. 11. Demonstration of q-waves for the numerical solution of 1D convection equation. q-waves are displayed by the numerical method employing RK_4-CD_2 where as the solution using optimized compact scheme RK_4-OUCS3 shows physical waves. (Multimedia View).

where, $n+1$, n and, $n-1$ are the three time levels. The parameters κ_1, κ_2, γ_1 and, γ_2 give rise to a family of three-time level schemes. One such scheme, the second order AB_2 scheme, is obtained by setting $\kappa_1 = 1$, $\kappa_2 = 0$, $\gamma_1 = \frac{3}{2}$ and, $\gamma_2 = -\frac{1}{2}$. The AB_2 scheme is given below as

$$u_j^{n+1} = u_j^n + \frac{3}{2}\Delta t L(u_j^n) - \frac{1}{2}\Delta t L(u_j^{n-1}) \tag{63}$$

For the model CE given in Eq. (2), one obtains the discrete equation by noting the generic representation of the first derivative scheme (Eq. (30)) as

$$u_j^{n+1} = u_j^n - \frac{3}{2}N_c\sum_{l=1}^{N} C_{jl}u_l^n + \frac{1}{2}N_c\sum_{l=1}^{N} C_{jl}u_l^{n-1} \tag{64}$$

where, superscript n indicates the time index, subscripts l and j indicate the index of the spatial nodes and N_c is the CFL number as defined earlier.

Employing the space–time representation in the hybrid spectral plane given by Eq. (5), and by noting the definition of numerical amplification factor as $G = \frac{\hat{U}(kh,t+\Delta t)}{\hat{U}(kh,t)}$, one obtains the numerical amplification factor for the AB_2 scheme as

$$G_j = 1 - \frac{3}{2}N_c\sum_{l=1}^{N} C_{jl}e^{ik(x_l-x_j)} + \frac{1}{2}N_c\left(\sum_{l=1}^{N} C_{jl}e^{ik(x_l-x_j)}\right)\frac{1}{G_j} \tag{65}$$

We note that the above equation is a quadratic equation and it has two roots, G_{j1} and G_{j2}, which are given by

$$G_{j1,2} = \frac{1}{2}\left[1 - \frac{3}{2}N_c\sum_{l=1}^{N} C_{jl}P_{lj} \pm \sqrt{\left(1 - \frac{3}{2}N_c\sum_{l=1}^{N} C_{jl}P_{lj}\right)^2 + 4\left(\frac{1}{2}N_c\sum_{l=1}^{N} C_{jl}P_{lj}\right)}\right] \tag{66}$$

where, $P_{lj} = e^{ik(x_l-x_j)}$ is the projection matrix. Of the two roots, one notes that G_{j1} corresponds to the physical mode whereas G_{j2} is the numerical mode. The basis of this classification is due to the requirement of the physical mode to have numerical amplification factor $|G_j|$ equal to one in the continuum limit (as $kh \to 0$ then $|G_j| \to 1$). The presence of two modes is a characteristic feature of all three-time level methods. One will note from the ensuing discussion the spurious nature of the numerical mode which is an additional source of contribution to the numerical error of the scheme.

Due to the formulation of the multi time-level methods, these cannot be employed for the initial advancement of numerical solution. Rather, one uses single-step methods for the first time step. This is called boot-strapping in literature. It should also be noted that the numerical solution is a superposition of the contributions due to the physical and numerical modes for the multi time-level methods. For the AB_2 method, this distribution of the numerical solution between the two modes can be evaluated as described next. It should be noted that the same procedure is applicable for any generic multi time-level methods. The contribution for the two modes is denoted by spectral weights M and N for the physical and computational modes, respectively. The weights are constrained by the relation $M + N = 1$. Noting that the initial stage employs a single-step method for solution evaluation, one represents the solution in the hybrid spectral plane as

$$u_j^{(1)} = \int \hat{U}(kh,\Delta t)e^{ikx_j}\,dk = \int \hat{U}(kh,0)G_{jE}(kh,N_c)e^{ikx_j}\,dk \tag{67}$$

where, G_{jE} is the numerical amplification factor of the corresponding single-step method. If RK_4 scheme is employed then this quantity is given by Eq. (57). From the above equation, it is noted that the amplitude $\hat{U}(kh,\Delta t) = \hat{U}(kh,0)G_{jE}$.

The solution from the next time step is obtained using the AB_2 time integration method. Noting that the solution at $t = 2\Delta t$ is split into a numerical and physical mode as given by $\hat{U}(kh,2\Delta t) = (MG_{1j} + NG_{2j})\hat{U}(kh,\Delta t)$, the solution is obtained as

$$u_j^{(2)} = \int (MG_{1j} + NG_{2j})\hat{U}(kh,\Delta t)e^{ikx_j}\,dk \tag{68}$$

From the discrete equation Eq. (63), one obtains at $t = 2\Delta t$ the following equation

$$\hat{U}(kh,2\Delta t) = \left(1 - \frac{3}{2}N_c\sum_{l=1}^{N} C_{jl}P_{lj}\right)\hat{U}(kh,\Delta t) + \left(\frac{1}{2}N_c\sum_{l=1}^{N} C_{jl}P_{lj}\right)\hat{U}(kh,0) \tag{69}$$

Equating Eqs. (68) and (69) by noting that $\hat{U}(kh,\Delta t) = \hat{U}(kh,0)G_{jE}$ and using the constraint $M + N = 1$, a system of equations is obtained for M and N. Solving these equations for M and N, one obtains,

$$M = \frac{(1 - \frac{3}{2}N_c\sum_{l=1}^{N} C_{jl}P_{lj}) + (\frac{1}{2G_{jE}}N_c\sum_{l=1}^{N} C_{jl}P_{lj}) - G_{2j}}{G_{1j} - G_{2j}} \tag{70}$$

$$N = \frac{(1 - \frac{3}{2} N_c \sum_{l=1}^{N} C_{jl} P_{lj}) + (\frac{1}{2G_{jE}} N_c \sum_{l=1}^{N} C_{jl} P_{lj}) - G_{1j}}{G_{2j} - G_{1j}} \tag{71}$$

Having obtained the spectral weights M and N, the numerical phase, phase speed and group velocity can be determined. Due to the presence of two modes, each mode has its corresponding numerical amplification factor $|G_{j1}M|, |G_{j2}N|$, phase (ϕ_{N1}, ϕ_{N2}), phase speed (c_{N1}, c_{N2}) and group velocity (V_{gN1}, V_{gN2}), respectively. The numerical phase, phase speed and group velocity are given by

$$\phi_{N1}\big|_j = -\tan^{-1}\left(\frac{(MG_{j1})_i}{(MG_{j1})_r}\right) \; ; \; \phi_{N2}\big|_j = -\tan^{-1}\left(\frac{(NG_{j2})_i}{(NG_{j2})_r}\right) \tag{72}$$

$$\left[\frac{c_{N1}}{c}\right]_j = \left.\frac{\phi_{N1}}{N_c kh}\right|_j \; ; \; \left[\frac{c_{N2}}{c}\right]_j = \left.\frac{\phi_{N2}}{N_c kh}\right|_j \tag{73}$$

$$\left[\frac{V_{gN1}}{c}\right]_j = \frac{1}{N_c}\left.\frac{d\phi_{N1}}{dkh}\right|_j \; ; \; \left[\frac{V_{gN2}}{c}\right]_j = \frac{1}{N_c}\left.\frac{d\phi_{N2}}{dkh}\right|_j \tag{74}$$

where the subscripts r and i in Eq. (72) denote the real and imaginary parts of the complex quantity.

The properties obtained from GSA of AB_2 time integration scheme for the model 1D CE are shown in Figs. 12–14 where CD_8, sixth order Lele's compact scheme and OUCS3 scheme are employed for spatial discretizations, respectively. The properties, viz. the numerical amplification factor, phase speed and group velocity, are presented for the physical and numerical modes for the first time application of the AB_2 scheme. For bootstrapping, RK4 method is employed. As in the previous section, these figures also show the comparison between the correct interpretation of the dispersion relation from GSA and the wrong approach. It should be noted that the wrong dispersion relation, by its very construction, cannot distinguish between the physical and numerical modes. Hence, the numerical phase speed and group velocity are identical for both modes.

In Fig. 12(a) and 12(b), properties are presented at $t = 2\Delta t$ for the physical and numerical modes respectively, with CD8 scheme as the spatial discretization method. For this discretization, the method is unstable as the physical mode has $|G_1| \geq 1$, even for very low values of N_c (not shown here). Hence, this method is unsuitable for the solution of the CE. From the numerical amplification contours, it is evident that the numerical mode $|G_2 N|$ is highly damped compared to the physical mode $|G_1 M|$, for the considered range of N_c values. One notes high damping of the spurious mode at low values of N_c and it decreases progressively with increasing N_c. The ratio of the numerical phase speed to exact phase speed $(\frac{c_N}{c})$ contours show overprediction of phase speed for low values of wavenumber kh for the physical mode and the error increases for higher wavenumbers. For the spurious mode, however, the error is large even for very small wavenumbers. The ratio of numerical group velocity to exact group velocity $\frac{V_{gN}}{c}$ shows the signal to propagate at a faster speed compared to the exact speed at low wavenumbers for the physical mode. The propagation speed decreases progressively for increasing wavenumbers and one notes the appearance of q-waves ($V_{gN} < 0$) beyond $kh = 2.03$. This is a typical feature of numerical schemes. Interestingly, the spurious mode shows the signal to propagate in the opposite direction at low wavenumbers and at higher wavenumbers it propagates in the correct direction albeit, at erroneous speeds.

Fig. 13(a) and 13(b), show the numerical properties for the physical and spurious modes for the sixth order Lele's scheme as the spatial discretization method. As noted in the case of CD8 scheme, this discretization is also unstable due to the amplification factor for the physical mode being greater than 1 (not shown here). The same observations for the CD8 case are noted here. The sixth order spatial discretization improves the numerical phase speed and group velocity properties of the physical mode. However, the improvement is inconsequential due to the unstable nature of the scheme.

Fig. 14(a) and 14(b), show the numerical properties for the physical and spurious modes for the optimized upwind compact scheme OUCS3. Unlike the instability noted for CD8 and Lele's schemes, a small region of stability is noted. The same observations for numerical phase speed and group velocity for the CD8 case are noted in this case too. However, an interesting observation can be made in this case where the numerical amplification factor for the physical and spurious modes changes discontinuously after a certain value of N_c. This is attributed to the change in sign of the imaginary part of the complex quantity under the square root term in Eq. (66) which leads to a discontinuity in the square root term. This is not noted for the central schemes — CD8 and Lele's scheme and hence, the discontinuity is absent.

The GSA of the AB_2 scheme demonstrates the typical nature of multi time-level integration schemes which involves spurious mode(s) in the computation of solutions. The spurious mode is omnipresent; it exists even at small values of N_c although its amplitude is small. Due to the distribution of the solution between physical and numerical modes, and noting the spurious mode to propagate in the direction opposite to the correct propagation direction, accurate propagation of energy is impossible. This demonstrates a major limitation of multi time-level integration methods and is the reason why such methods are undesirable for high accuracy computing.

In the next section, we show the existence of the spurious mode by a demonstration case using the AB_2 scheme for time integration. The numerical parameters are specifically chosen to highlight the numerical instability which arises when the spurious mode is the dominant mode for solution propagation.

5.1. Evidence for existence of numerical mode in Adams–Bashforth method

The existence of numerical modes for three-time level methods has been conclusively shown for the mid-point leapfrog method in [1], when Euler and RK_4 time integration methods are used to boot-strap the mid-point leapfrog scheme at the first time-step. In this method, the numerical amplification factor of the physical and spurious modes are perfectly neutral, with opposite group velocities. As a consequence, the given initial condition splits into two branches, as graphically demonstrated in [1].

For AB_2 time integration scheme, the strong asymmetry of the numerical amplification factors for the physical and numerical modes has been shown in [44], with the latter contributing very insignificantly. This prompted Lilly [79] to recommend the AB_2 scheme to be of practical use, provided the time of computing is kept to a minimum. From the property charts shown in [44], this allows the use of $AB_2 - CD_2$ in this context, as the numerical instability of the physical mode is marginal, allowing one to compute for some time. However, one should keep in mind that the numerical mode also exists however insignificant its contribution may be. For example, for the results shown in Figs. 13 and 14, for AB_2 method used with Lele's compact scheme and $OUCS3$, both the physical and numerical mode can be unstable. Here, we will establish the existence of the numerical mode unambiguously, by considering the third order upwind (UD_3) scheme for spatial discretization with AB_2. The numerical experiment is designed for a scenario in which the contribution from the spurious mode is dominant.

The UD_3 method displays unusual numerical properties in Figs. 15 and 16. The numerical amplification rates, $|G_1 M|$ and $|G_2 N|$, are shown in Fig. 15. For the physical mode, one notes the method to be strictly unstable for a critical CFL of $(N_c)_{cr} = 0.66051$, above which the method displays violent instabilities. For the numerical mode, one notices another critical CFL of $(N_c)_{cr} = 0.20866$ above which the numerical mode displays strong numerical instabilities. Thus, this method is special as the contribution of the numerical mode is significant for a wide range of kh and N_c. For some such combinations, numerical mode may be more dominant than the physical mode. One such point is identified in Fig. 15, given by: $kh = 1.0$ and $N_c = 2.2$ for which $|G_1 M| = 0.020608$ (denoted by point P) and $|G_2 N| = 2.99656$ (denoted by point Q).

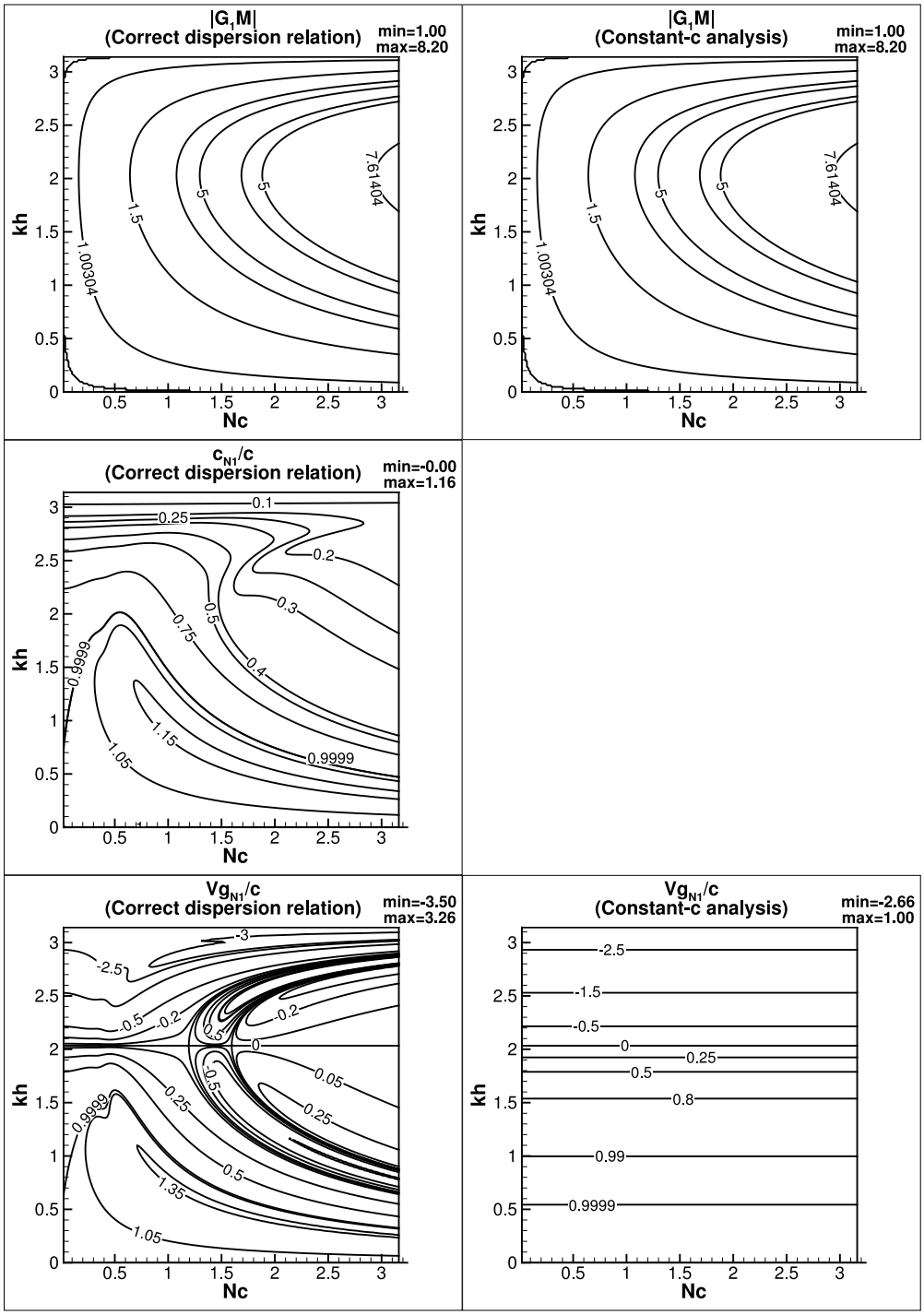

Fig. 12(a). Numerical properties of the physical mode obtained by CD8 spatial discretization and AB2 time integration schemes in solving Eq. (2). The properties obtained using correct dispersion relation, Eq. (8) are shown in the left frames. Use of the incorrect dispersion relation (i.e. using $c_N \cong c$) in Eq. (6) leads to the properties shown on the right frames. Shown are $|G_1 M|$, c_{N1}/c and V_{gN1}/c in various frames and compared for a typical interior point.

In Fig. 16, the numerical group velocity V_{gN}/c contours are shown for both physical and numerical modes. The point of interest (P) has $V_{gN1}/c = -0.10903$. This would have been of concern to us as the physical mode is propagating in the unphysical direction and upwind scheme does not accommodate this, however on observing the low value of $|G_1 M| = 0.020608$, one can be assured that the physical mode shall not exist for long in any simulation. For the spurious mode, the point of interest is marked as Q for which $V_{gN2}/c = 0.371861$.

The existence of the numerical mode has been established in Fig. 17 by solving the 1D CE using the combination of AB_2 method for time integration and UD_3 for spatial discretization. The numerical properties for the chosen values of kh and N_c are shown in Fig. 15, in terms of the product of the amplification factor and the fraction of the signal being propagated by the corresponding mode and in Fig. 16, the numerical group velocity of the respective modes are shown. The wave-packet propagation in Fig. 17 is described by a periodic signal convoluted by

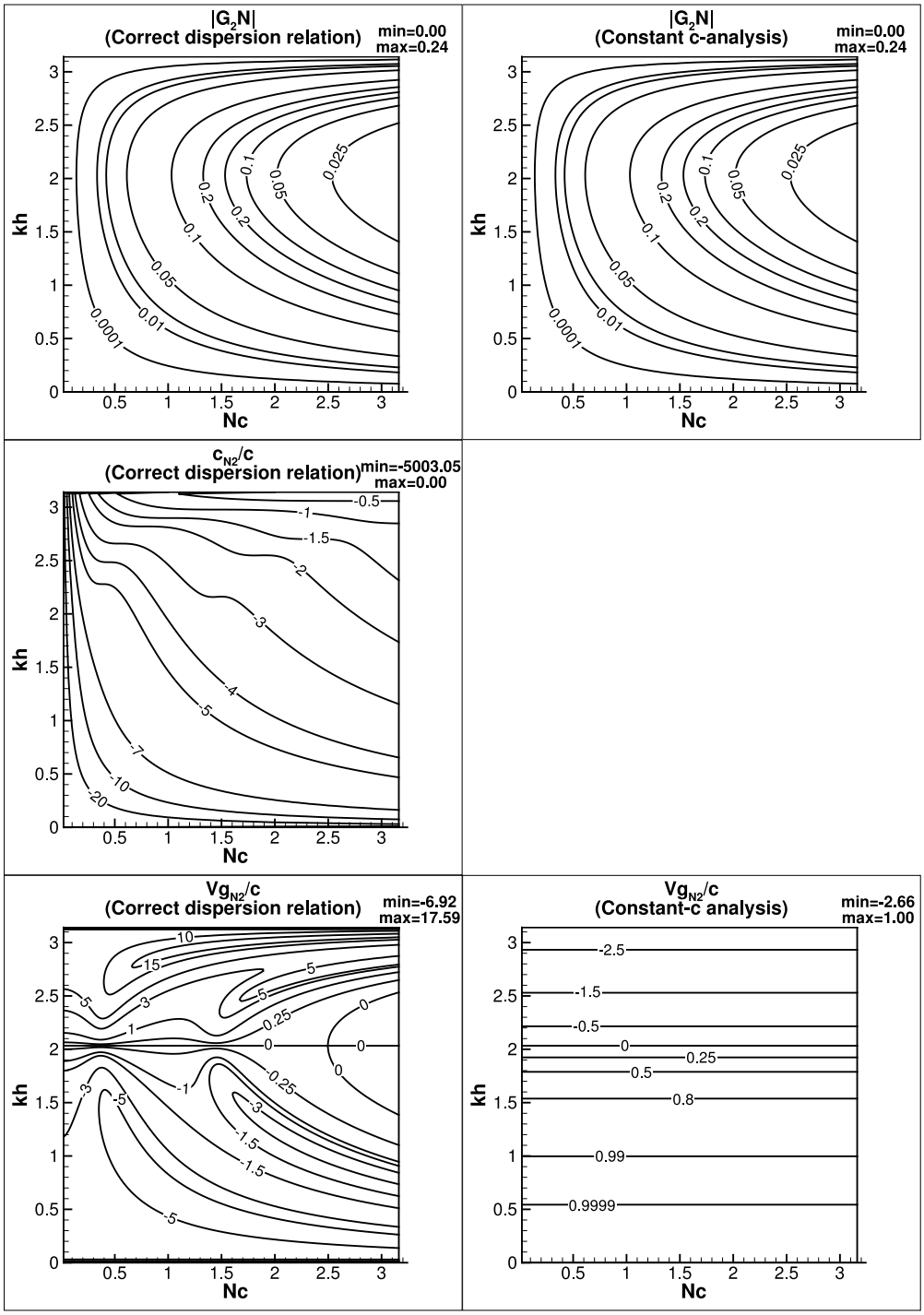

Fig. 12(b). Numerical properties of the spurious mode obtained by CD8 spatial discretization and AB2 time integration schemes in solving Eq. (2). The properties obtained using correct dispersion relation, Eq. (8) are shown in the left frames. Use of the incorrect dispersion relation (i.e. using $c_N \cong c$) in Eq. (6) leads to the properties shown on the right frames. Shown are $|G_2N|$, c_{N2}/c and V_{gN2}/c in various frames and compared for a typical interior point.

the Gaussian as,

$$u(x,0) = e^{-\alpha(x-x_0)^2} sin(k_0 x) \tag{75}$$

where $\alpha = 0.01$, $x_0 = 250$, $c = 300$ in a domain of $0 < x < 500$ with 144 000 points. The first time step is obtained by Euler time integration scheme. From the second time step onwards, AB_2 time integration is employed with periodic boundaries to avoid the problem of reflection from the domain boundaries. We will highlight aspects of numerical instability cropping up due to this issue in the next section. Here our primary and only intention is to show the existence of the numerical mode and reflections from the wall complicate interpretation of computational solution. Due to this focus on the primary goal, we will also display the solution only up to the first time step of AB_2 scheme.

A Fourier transform of the initial wave-packet is conducted to ensure localization of the wave-packet at the chosen value of $kh = 1$. This becomes particularly important for $AB_2 - UD_3$, as is evident from the peculiar numerical properties' rapid variations with kh in Figs. 15

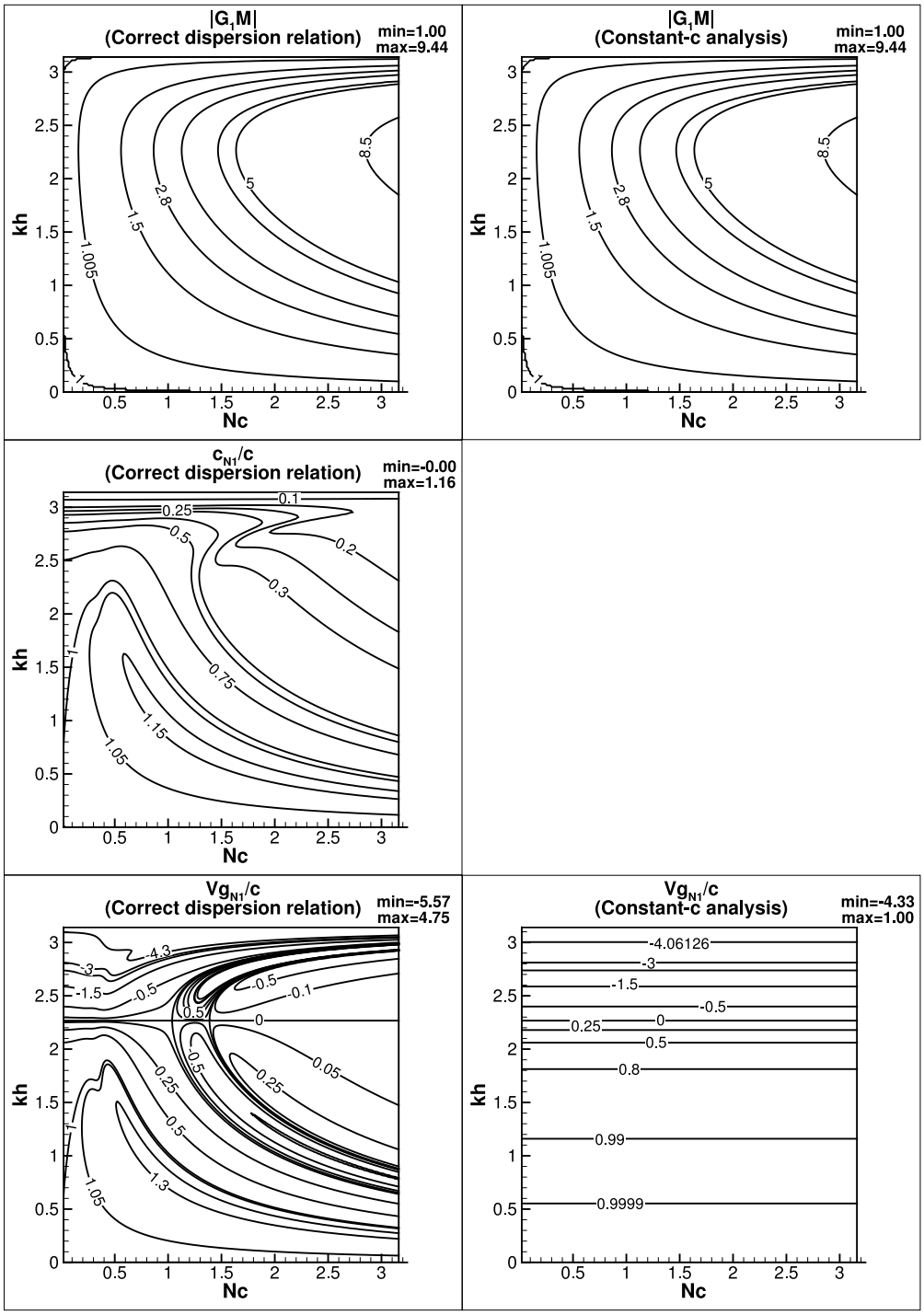

Fig. 13(a). Numerical properties of the physical mode obtained by sixth order Lele's compact scheme for spatial discretization and AB2 time integration schemes in solving Eq. (2). The properties obtained using correct dispersion relation, Eq. (8) are shown in the left frames. Use of the incorrect dispersion relation (i.e. using $c_N \cong c$) in Eq. (6) leads to the properties shown on the right frames. Shown are $|G_1 M|$, c_{N1}/c and V_{gN1}/c in various frames and compared for a typical interior point.

and 16. In Fig. 17, the initial solution is shown in the top frame, while the computed solution at $t = \Delta t$ and $2\Delta t$ are shown in the middle and bottom frames. For the computed solution shown in the middle frame, the maximum u value attained is 2.19338 which can be attributed to the unstable nature of the Euler time integration. In the subsequent frame at $t = 2\Delta t$, the maximum value of the wave-packet is noted as 6.56574 which is found to be 2.99343 times the value at $t = \Delta t$ and this is the value of $|G_2 N|$ recorded in Fig. 15 for the chosen kh and N_c combination. This fact is conclusive proof of the branching of the signal into physical and computational modes by the AB_2 method.

Also from Fig. 16, for the numerical mode, the value of the normalized group velocity is given by $V_{gN2}/c = 0.371861$. As noted already that following the growth of the initial solution by Euler time integration, the solution is split into physical and numerical modes. However, the quantum of signal going to the physical mode is about only 2% of the solution at $t = \Delta t$. Also, this small part of the signal

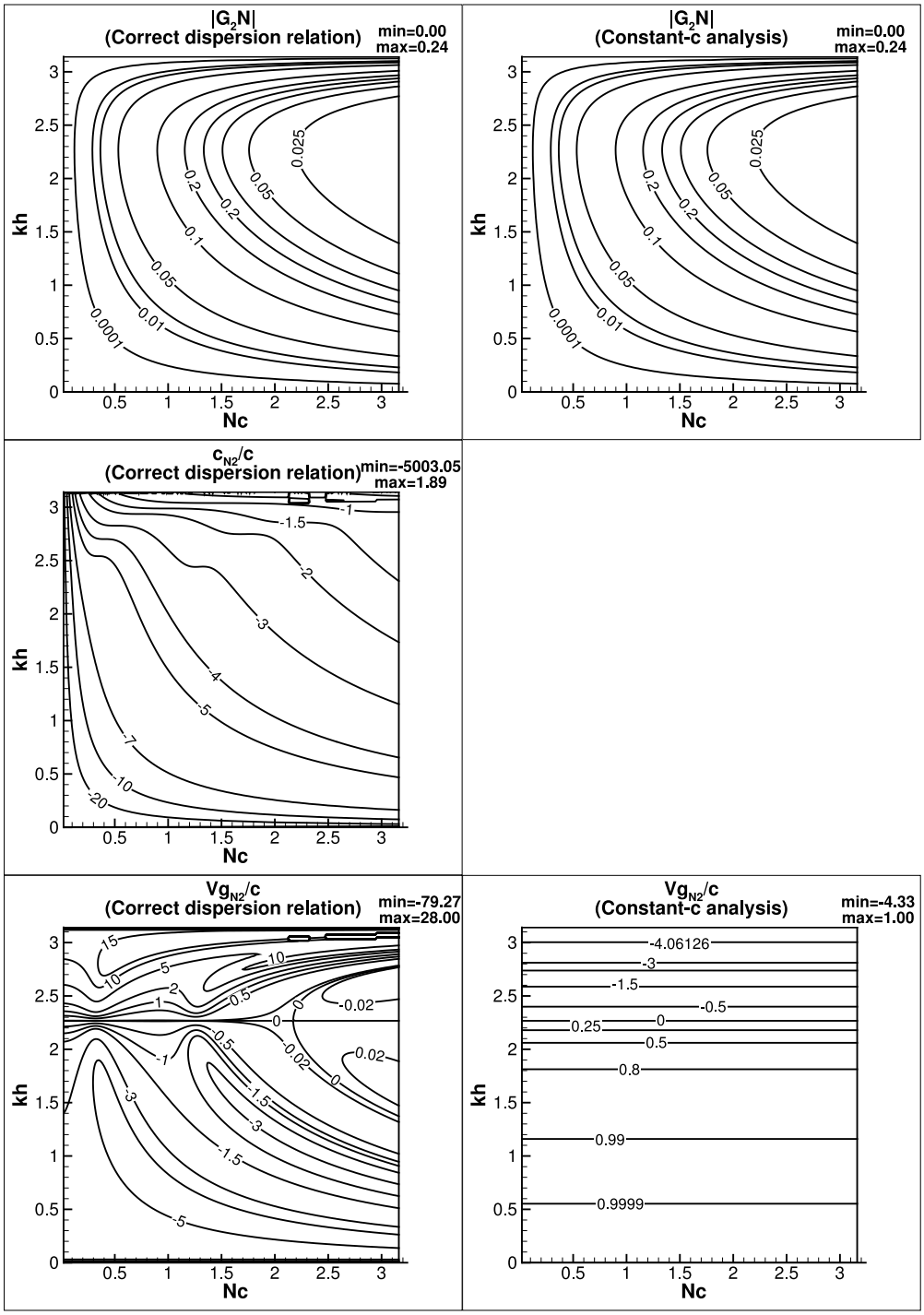

Fig. 13(b). Numerical properties of the spurious mode obtained by sixth order Lele's compact scheme for spatial discretization and AB2 time integration schemes in solving Eq. (2). The properties obtained using correct dispersion relation, Eq. (8) are shown in the left frames. Use of the incorrect dispersion relation (i.e. using $c_N \cong c$) in Eq. (6) leads to the properties shown on the right frames. Shown are $|G_2N|$, c_{N2}/c and V_{gN2}/c in various frames and compared for a typical interior point.

will travel in the wrong direction (as given by $V_{gN1}/c = -0.10903$). In contrast, the wave-packet will travel in the correct physical direction, carried by the numerical mode with V_{gN2}/c, in an unstable manner as $|G_2N| = 2.99656$. Thus, the signal would have traveled a distance of 2.840604×10^{-3}, if it was purely monochromatic. From Fig. 17, it is noted that the signal has traveled a distance of 2.82×10^{-3}. This also convinces one that the numerical mode is responsible for carrying a part of the signal in all applications of the AB_2 scheme. In the framework of the $AB_2 - UD_3$ method, it is established as the most dominant mode carrying the signal.

6. Beyond classical linearized stability analysis of nonlinear PDEs: introduction to local energy growth, side-band instabilities and focusing

The use of linear equations to study the stability and accuracy of discretized non-linear equations relies on the key assumptions that the

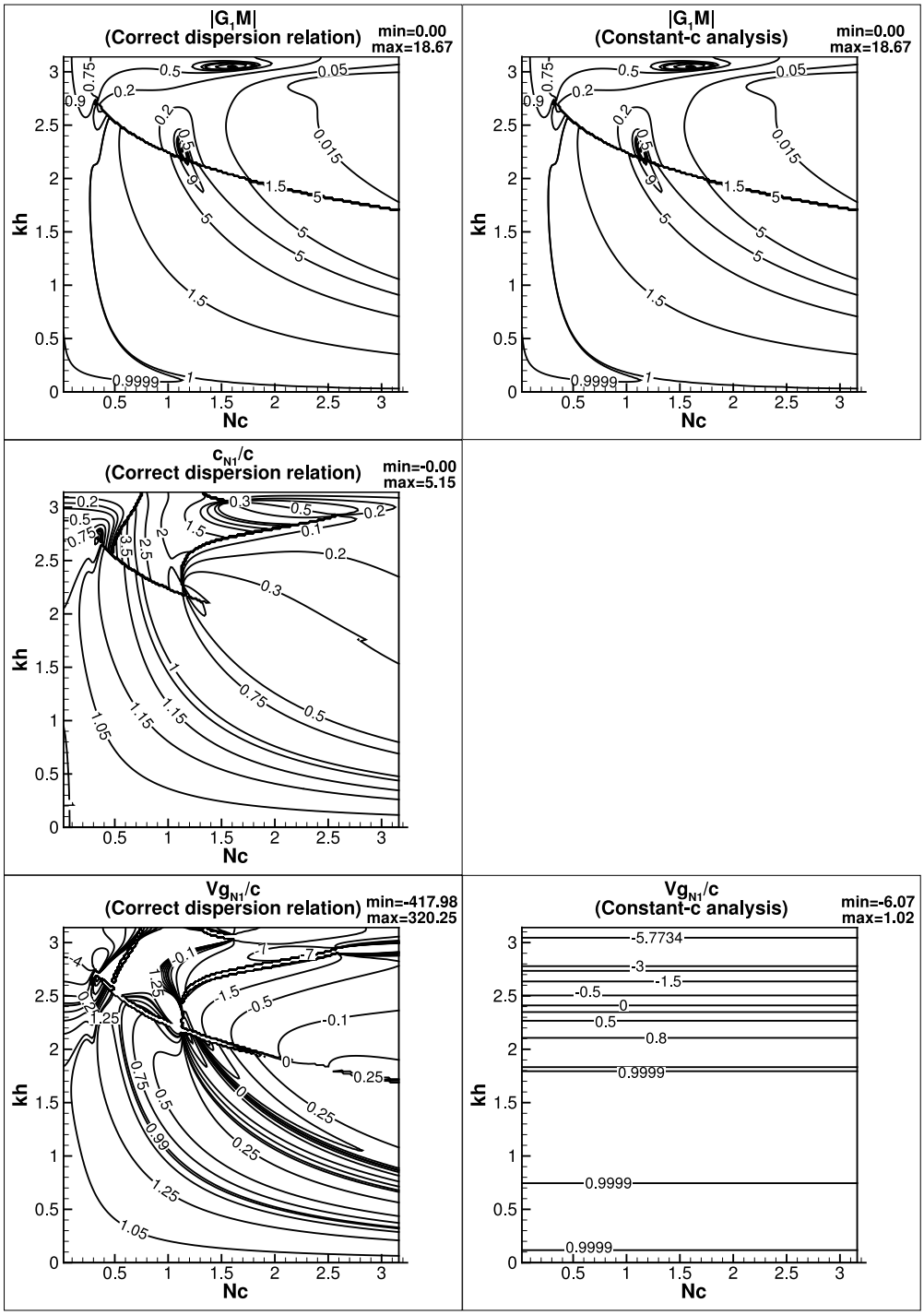

Fig. 14(a). Numerical properties of the physical mode obtained by optimized compact scheme OUCS3 for spatial discretization and AB2 time integration schemes in solving Eq. (2). The properties obtained using correct dispersion relation, Eq. (8) are shown in the left frames. Use of the incorrect dispersion relation (i.e. using $c_N \cong c$) in Eq. (6) leads to the properties shown on the right frames. Shown are $|G_1 M|$, c_{N1}/c and V_{gN1}/c in various frames and compared for a typical interior point.

errors, that are modeled as fluctuations around the exact solution, are small, allowing for the use of a linearized system. This is illustrated considering the nonlinear advection equation

$$\frac{\partial u}{\partial t} + (c+u)\frac{\partial u}{\partial x} = 0 \tag{76}$$

where c is related to a base flow. In common cases, u is interpreted as a fluctuation about the base flow solution. The associated linearized problem often used to perform numerical scheme analysis is equivalent to (2), i.e.

$$\frac{\partial u}{\partial t} + c\frac{\partial u}{\partial x} = 0 \tag{77}$$

under the key assumption that u is small in some sense, i.e. for a given norm:

$$\|u\| \ll c \tag{78}$$

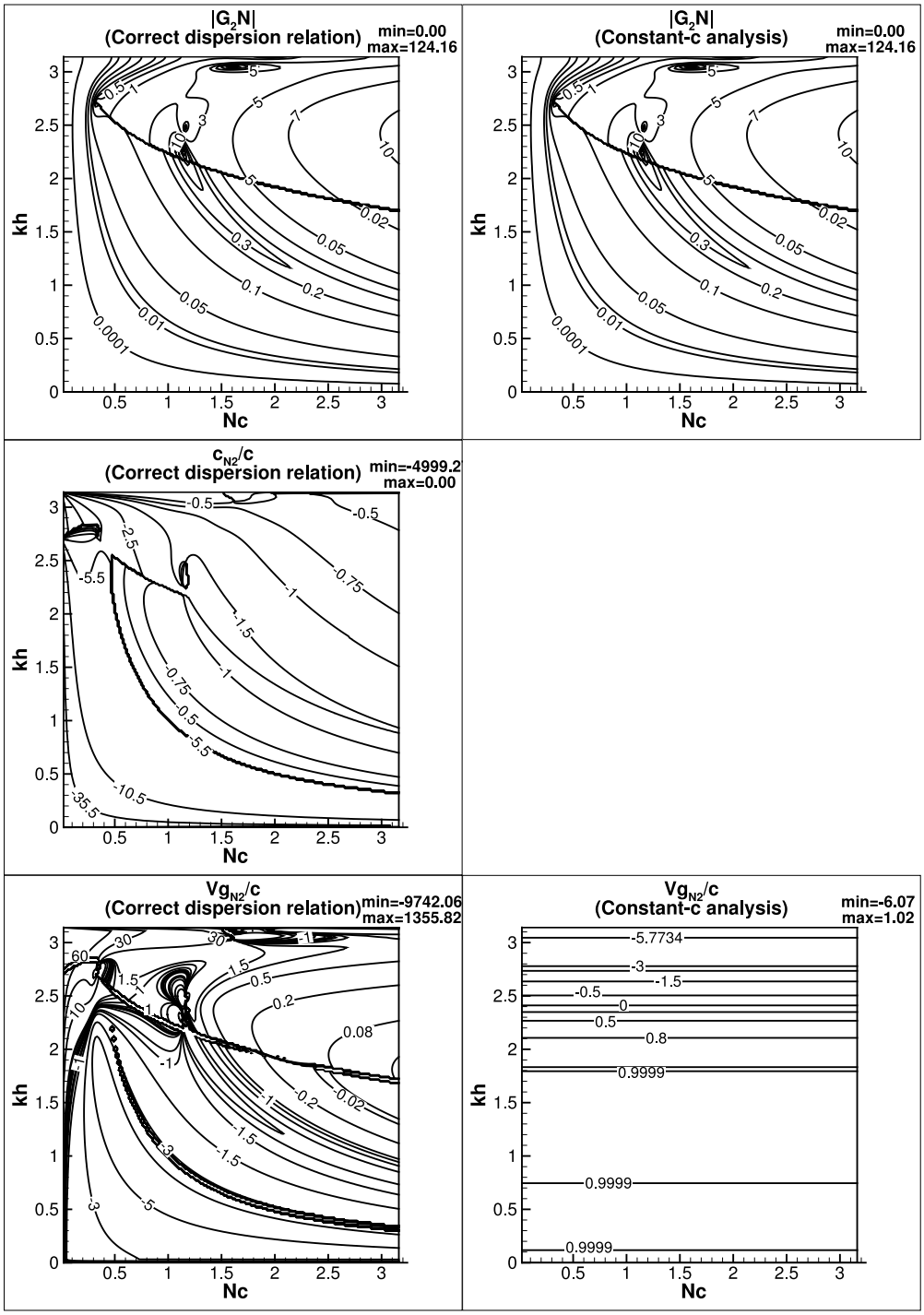

Fig. 14(b). Numerical properties of the spurious mode obtained by optimized compact scheme OUCS3 compact scheme for spatial discretization and AB2 time integration schemes in solving Eq. (2). The properties obtained using correct dispersion relation, Eq. (8) are shown in the left frames. Use of the incorrect dispersion relation (i.e. using $c_N \cong c$) in Eq. (6) leads to the properties shown on the right frames. Shown are $|G_2N|$, c_{N2}/c and V_{gN2}/c in various frames and compared for a typical interior point.

It is worth keeping in mind since we are dealing with finite discrete solutions when analyzing realistic numerical solutions, the choice of the norm is not a key problem, since all norms are equivalent in this finite discrete case.

It has been observed by many authors, e.g. [72,83,84], that some computed numerical solutions of Eq. (76) become unstable, while the stability analysis based on (77) predicts a stable solution, even considering very small initial noise, i.e. satisfying the key condition (78) at $t = 0$ and in the absence of spurious source terms which may play the role of error energy source.

While the nonlinear nature of the instability that is responsible for the exponential growth of the error is commonly admitted in such cases, the question arises of the existence of linear mechanisms that may lead to the triggering of nonlinear instability in linearly stable cases. The key point here is the breakdown of hypothesis (78), i.e. the occurrence of instantaneous local high values of u that are strong

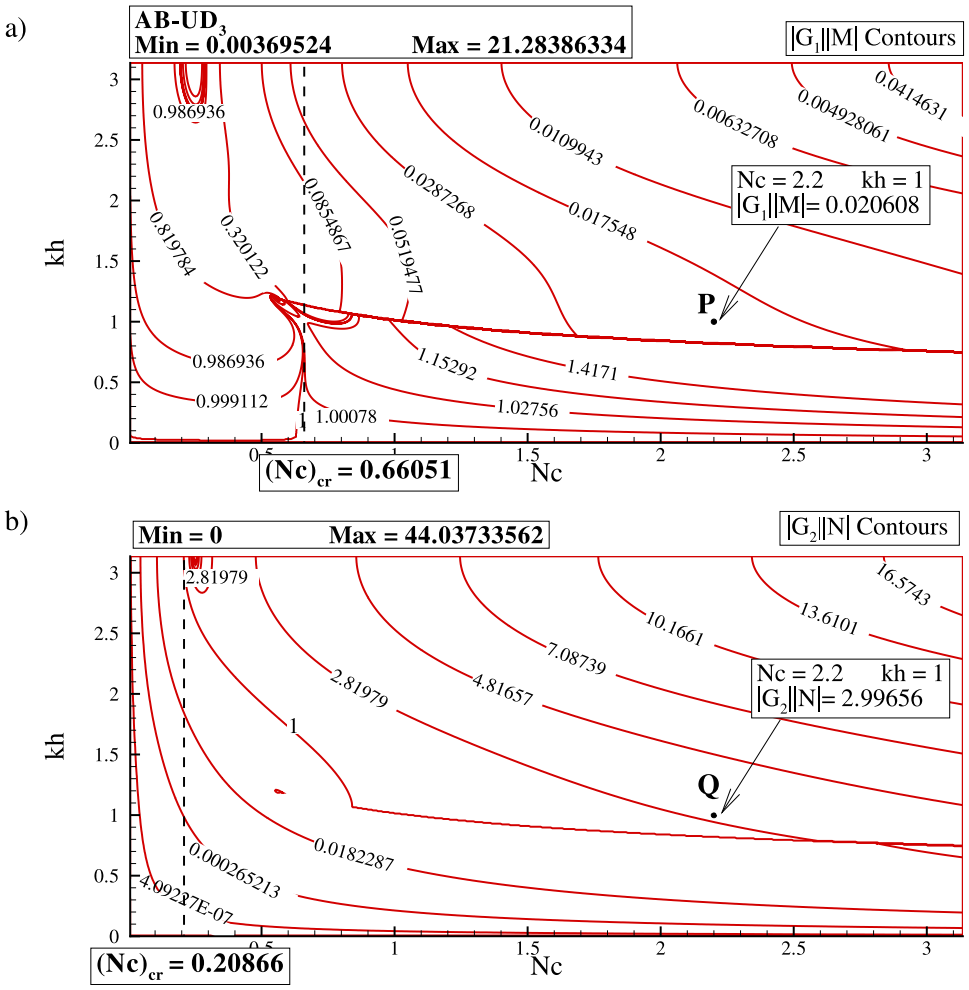

Fig. 15. The split numerical amplification factors for the (a) physical ($|G_1 M|$) and (b)numerical modes ($|G_2 N|$) of the AB_2 time integration method used with UD_3 for spatial discretization.

enough to locally trigger dominant nonlinear mechanisms, whose stability can be investigated thanks to dedicated methods (mostly based on the dynamical system theory).

In the absence of spurious error source terms, and for energy-preserving schemes for which $\|u\|_2 = const$, the occurrence of local high values of u, i.e. high values of $\|u\|_\infty$ is observed to be due to the local instantaneous concentration of energy of u, a phenomenon referred to as focusing [72]. While nonlinear focusing mechanisms are commonly reported in classical physics and numerical analysis, the possibility of linear transient error growth in linearly stable discrete solutions has been addressed by a few authors only, mostly using matrix analysis and extensions of the classical Fourier-based Von Neumann analysis.

6.1. Focusing of energy due to dispersive errors

It has been shown via GSA analysis that both the numerical phase speed c_N and the group velocity V_{gN} are scale- and frequency-dependent, even if the physical phase speed c is uniform and constant. Therefore, the intrinsic numerical errors transform the original continuous non-dispersive problem into a dispersive one. Starting from that observation, several researchers have proposed to apply the mathematical tools and the theoretical concepts developed to describe wave propagation through dispersive media to analyze the dispersive effects on the numerical solution. More precisely, these works aim at interpreting the local pile up of energy and error observed in numerical simulations by looking at the similarities with the physics of continuous dispersive waves.

The local focusing of wave energy is classically described using the caustics theory, which originates in the ray theory in geometrical optics and geometrical acoustics, see e.g. [85]. The ray theory is a Lagrangian approach to the wave propagation, that provides models for the propagation of energy and phase of the wave along characteristic lines. The key elements of ray and caustic theories are now reminded to illustrate the dispersive mechanisms that may lead to a local instantaneous very large growth of the energy of the solution.

The ray theory is based on the local approximation of waves whose amplitude and direction of propagation vary slowly over one wavelength as a plane wave. Rays are characteristic lines along which the energy is transported, precisely defined as the bicharacteristics of the Helmholtz equation.

Considering the propagation of a scalar monochromatic wave in a linear isotropic medium, the solution can be written as

$$u(\mathbf{x}, t) = U(\mathbf{x}) e^{-i(\omega t - k\psi(\mathbf{x}))} \tag{79}$$

where k and ω denote the wave number and the frequency, respectively. $\psi(\mathbf{x})$ is referred to as the eikonal function, and whose isosurfaces $\psi(\mathbf{x}) = const$ are the wavefronts. In the general case, there exist an infinite number of couple $(U(k), \psi(\mathbf{x}))$ at fixed (k, ω). In the present case, it is assumed that $u(\mathbf{x}, t)$ obeys the following scalar Helmholtz wave equation

$$\frac{1}{c(\mathbf{x})^2} \frac{\partial^2 u}{\partial t^2} + \nabla^2 u = 0 \tag{80}$$

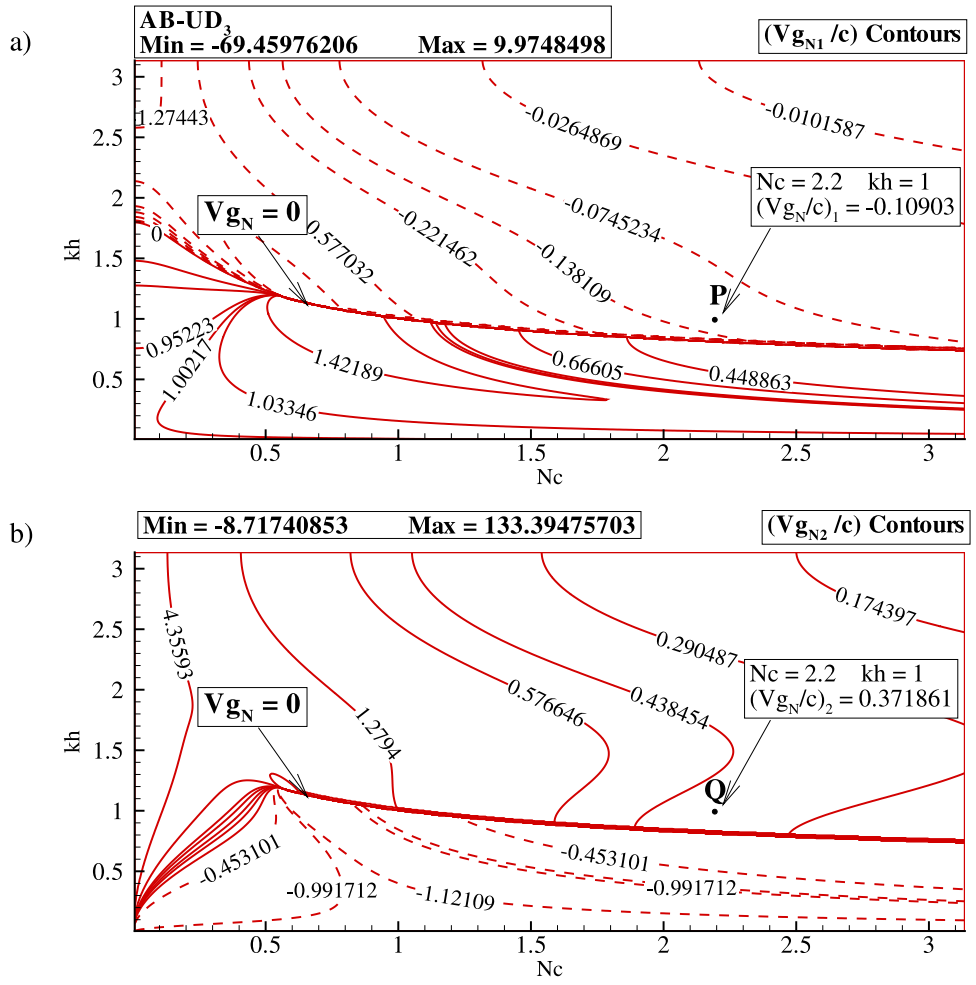

Fig. 16. The normalized numerical group velocities for the (a) physical (V_{gN1}/c) and (b)numerical modes (V_{gN2}/c) of the AB_2 time integration method used with UD_3 for spatial discretization.

where the propagation speed $c(\mathbf{x}) = c_0/n(\mathbf{x})$ may be space dependent in heterogeneous media. Here, c_0 and $n(\mathbf{x})$ are related to a reference propagation speed and local variations due to medium heterogeneities, respectively. Plane wave solutions are exactly preserved for uniform $c(\mathbf{x}) = c_0$. It is worth noting that (80) is an approximation of (4) in which the rhs term is neglected (this is reasonable for slowly and weakly varying c), but first dispersive effects are captured by accounting for the local value of $c(\mathbf{x})$ in the wave propagation operator. Inserting (79) into (80), one obtains:

$$\nabla^2 U + 2ik(\nabla\psi\cdot\nabla U) + iUk\nabla^2\psi - Uk^2(\nabla\psi)^2 + \frac{n(\mathbf{x})^2}{c_0^2}\omega^2 U = 0 \tag{81}$$

which can be rewritten by considering the real and imaginary parts separately and introducing the wavelength $\lambda = 2\pi/k$ as

$$\nabla^2 U - Uk^2\left[(\nabla\psi)^2 + n^2\right]U = 0 \tag{82}$$

and

$$k\left[2(\nabla\psi\cdot\nabla U) + U\nabla^2\psi\right] = 0 \tag{83}$$

Geometrical theories address small perturbations about plane wave solutions, due to small perturbations in the propagation speed. Therefore, $c(\mathbf{x})$ is assumed to exhibit small amplitude variations at the wavelength scale $\lambda = 2\pi/k$. Restricting the analysis this way, the solution is approximated as an asymptotic series in negative powers of the wave number $k = \omega/c$:

$$u(\mathbf{x}) = \left(U_0(\mathbf{x}) + \frac{U_1(\mathbf{x})}{ik} + \frac{U_2(\mathbf{x})}{(ik)^2} + \cdots\right)e^{ik\psi(\mathbf{x})} \tag{84}$$

Inserting this expansion in (82), one recovers at the leading order in k the Eikonal equation

$$(\nabla\psi)^2 = n^2 \tag{85}$$

along with the transfer equations in the amplitudes that originate in (83) :

$$2\nabla U_0\cdot\nabla\psi + U_0\nabla^2\psi = 0 \tag{86}$$

$$2\nabla U_1\cdot\nabla\psi + U_1\nabla^2\psi = -\nabla^2 U_0 \tag{87}$$

$$\cdots$$

Solutions of this problem can be expressed in a Lagrangian form, in a way similar to the characteristic solutions for compressible hydrodynamics and more generally for hyperbolic systems, in which Riemann invariants are advected along characteristic lines. For the sake of efficiency the Eikonal equation is recast in the following canonical Hamiltonian form:

$$\frac{d\mathbf{x}}{d\tau} = \mathbf{p}, \quad \frac{d\mathbf{p}}{d\tau} = \frac{1}{2}\nabla n^2 \tag{88}$$

where $\mathbf{p} = \nabla\psi$ is the pseudo-momentum of the ray and τ is a time-like parameter defined along the ray tied to its length l by $d\tau = dl/n$. The direction of propagation of the ray is $\mathbf{p}/\|\mathbf{p}\|$.

For a given initial condition $u^0 = U_0 e^{ik\psi^0}$, the solution is obtained by integration along the ray:

$$\psi = \psi^0 + \int_0^\tau n^2(\mathbf{x}(\tau))d\tau \tag{89}$$

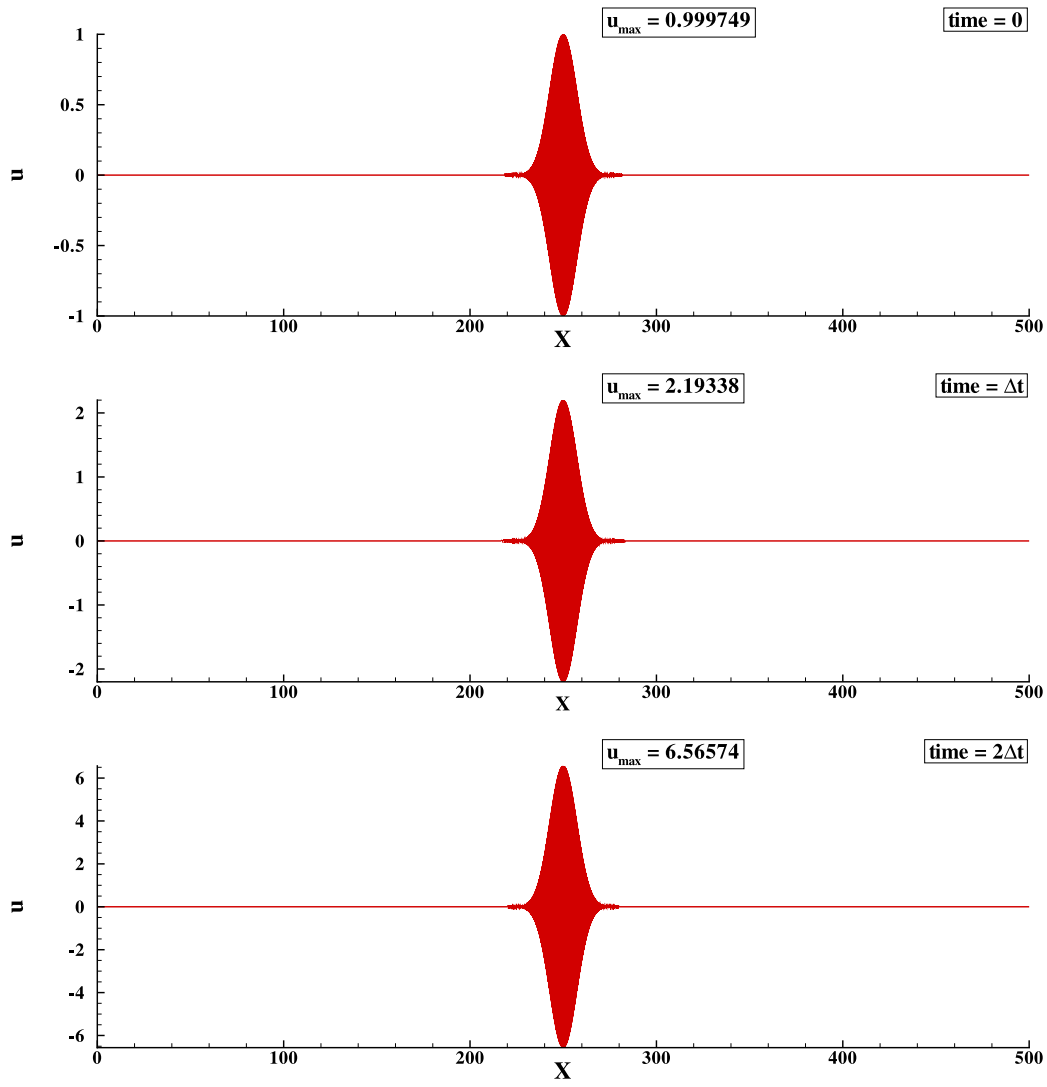

Fig. 17. Numerical solution of the 1D CE in the physical plane for $AB_2 - UD_3$ scheme at times (a) $t = 0$, (b) $t = \Delta t$ and (c) $t = 2\Delta t$. For the chosen parameters, numerical instability is mostly due to the numerical mode, while effect of physical mode is negligible.

Relations (88) emphasize the characteristic nature of rays, and show that the local instantaneous solution at $(\mathbf{x}, t)$ can be obtained from the initial solution by summing contributions from all rays reaching $(\mathbf{x}, t)$:

$$u(\mathbf{x}) = \sum_{j=1,N} U_0^{(j)} e^{ik\psi_j} \tag{90}$$

where $N, U_0^{(j)}$ and ψ_j denote the number of rays converging at point $\mathbf{x}$, the amplitude and the eikonal of the jth ray, respectively.

If the propagation speed c is uniform, rays are parallel straight lines, corresponding to a non-dispersive solution, while curved lines may exist in the case of dispersive solution associated to non-uniform propagation speed, i.e. space and/or time-varying $n(\mathbf{x}, t)$. Rays are lines along which energy is transported, i.e. they are energy trajectories. This is seen by introducing the vector of density of energy flux of u:

$$\mathbf{I} = \frac{1}{2ik}\left(u^* \nabla u - u \nabla u^*\right) \tag{91}$$

which is such that

$$\mathrm{div}(\mathbf{I}) = 0 \tag{92}$$

according to (80). In the geometrical approximation (84), the leading order expansion yields $\mathbf{I} \simeq \mathbf{p} U_0^2$, yielding

$$\mathrm{div}(\mathbf{p} U_0^2) = 0 \tag{93}$$

which is identical to (86) since $\mathbf{p} = \nabla\psi$. Therefore, the density of energy flux vector is parallel to $\mathbf{p}$, i.e. the energy flows along the rays.

Caustics (surfaces or lines) are mathematically defined as envelopes of the family of rays at which the field intensity increases sharply compared to the neighborhood. This phenomena has been analyzed using the catastrophe theory by many authors. This occurs when a large number of rays converge to the same location, resulting in a local concentration of a huge amount of energy. This phenomenon is illustrated in Fig. 18. This kind of singularity shares several features with shock waves in hydrodynamic, which occur at singular points in the space–time plane at which characteristic lines of the same family interact, leading to a ill-posed multi-valued problem. By analogy, the concentration of the error associated to the discretized version of Eq. (77) due to the dispersive nature of the numerical error can be understood as the result of the superimposition of the error energy carried by characteristic lines. In the case of a uniform base flow c, this can originate only in the scale-dependence of the numerical propagation speed c_N given by (13). Therefore, the linear focusing phenomenon is out of reach of the classical monochromatic plane-wave analysis, which must be replaced by a polychromatic analysis, as done in GSA.

In order to recover a local initial perturbation (or at least a perturbation with a compact support) and to mimic ray-like phenomena, wave

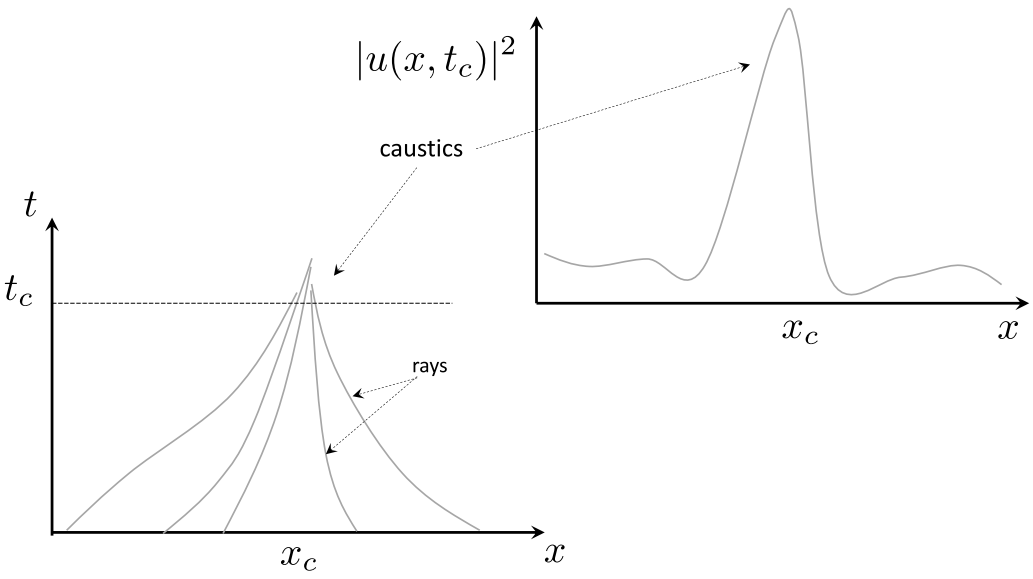

Fig. 18. Schematic view of the caustics phenomenon due to the focusing of rays.

packets should be considered, as done by Giles and Thompkins [86] who developed the concept of wavepacket particle to analyze the motion of wavepackets. These authors modeled wave packets as a solution of the form

$$u_j^n = A(k,j,n)e^{i\Psi(k,j,n)} \tag{94}$$

where $A(k,j,n)$ is assumed to be a slowly varying amplitude and $\Psi(k,j,n)$ is the phase, with $\omega_N = -\partial\Psi/\partial n$ being the frequency and $k = \partial\Psi/\partial j$ the wavenumber. An asymptotic analysis yields the following evolution equation for the amplitude:

$$\frac{\partial A}{\partial n} + V_{gN}\frac{\partial A}{\partial j} = \epsilon A \tag{95}$$

where V_{gN} is the discrete group velocity and ϵ is a scheme-dependent function of k, ω and j. Since V_{gN} depends on k, one can see that wave packets with different wave numbers will travel at different speeds, permitting the occurrence of the focusing phenomenon. This can be recast in a Lagrangian framework closer to the ray theory by defining the total derivative:

$$\frac{d}{dn} = \frac{\partial}{\partial n} + V_{gN}\frac{\partial}{\partial j} \tag{96}$$

yielding the Lagrangian formulation of the wave packet evolution:

$$\frac{dj}{dn} = V_{gN}, \quad \frac{dA}{dn} = \epsilon A, \quad \frac{dk}{dn} = V_{gN}\frac{\partial k}{\partial j} \tag{97}$$

These relations show that wavepackets travel along characteristic lines associated with the group velocity, along which both the amplitude of the envelope and the wavenumber may evolve. These theoretical predictions have been assessed by some numerical experiments, e.g. see Section 4.1.

Defining the energy of the wavepacket, $E(n)$ by integration over its compact support, one has

$$E(n) = \Delta x \sum_0^N |A(j,n)|^2 \tag{98}$$

along with the following discrete evolution equation (the asterisk denotes the complex conjugate)

$$\frac{dE}{dn} = \left(\epsilon + \epsilon^* + \frac{\partial V_{gN}}{\partial j}\right) E \tag{99}$$

These relations illustrate the similarities between the theory of wave propagation in dispersive media and the dynamics of discrete dispersive numerical solutions.

Numerical experiments were conducted on the linear scalar advection equation in [87] in which the authors identified local blow-up of the local error due to the pile-up of two wavepackets with different scales. The authors also proposed a criterion to identify schemes for which the rise of spurious caustics is favored. More precisely, spurious caustics are likely to occur if there exist a wavenumber k_c with $0 < k_c\Delta x \leq \pi$ such that the numerical group velocity exhibits an extremum, i.e.

$$\frac{dV_{gN}}{dk}(k_c) = 0 \tag{100}$$

where V_{gN} is given by the GSA relations (8) and (59). The GSA analysis of several cases is detailed in the following sections.

6.2. *Non-uniform base flow: collective interactions and side-band instabilities*

The previous developments are related to the focusing phenomenon, i.e. the concentration of the energy of the initial error due to the numerical dispersive error in the presence of a uniform base flow c.

Another linear error growth mechanism for polychromatic solutions has also been identified considering fluctuations about a sinusoidal base flow, e.g. [72,74,88–93]. In such a case, which can be interpreted as a model for error growth about a non-uniform discrete solution of the model nonlinear equation

$$\frac{\partial u}{\partial t} + (c+u)\frac{\partial u}{\partial x} = 0 \tag{101}$$

a linear growth of the error can be triggered by so-called sideband instabilities, which originate in resonance (coined as grid resonance by Cloot and Herbst [88]) between the base flow and the fluctuations. Such a mechanism is observed in the Benjamin–Feir instability in free-surface wave dynamics [94] or more general wavetrain instabilities [95].

This is now illustrated by considering a second-order centered finite difference scheme and a Leapfrog time integration, one obtains the following linearized equation for the fluctuations **u** around the base flow solution $(c + U_j^n)$:

$$\begin{aligned} &u_j^{n+1} - u_j^{n+1} + \alpha(u_{j+1}^n - u_{j-1}^n) + \gamma\theta(U_{j+1}^n u_{j+1}^n - U_{j-1}^n u_{j-1}^n) \\ &\quad + \gamma(1-\theta)[(U_{j+1}^n - U_{j-1}^n)u_j^n + U_j^n(u_{j+1}^n - u_{j-1}^n)] = 0 \end{aligned} \tag{102}$$

where $\gamma = \Delta t/\Delta x$ and $\sigma = c\Delta t/\Delta x$, and the θ parameter is a weighting coefficient used to mix the conservative and the quasi-linear formulation of the convection term. The base flow selected for the stability

analysis is made of the sum of a uniform base flow, c and a space–time dependent component :

$$U_j^n = \epsilon \left(\sum_{l=1,2} e^{i(kx_j - \omega_k^l t_n)} + c.c. \right) + O(\epsilon^2) \tag{103}$$

which is a so-called 2-modes stable solution of the nonlinear discrete equation with a small amplitude $\epsilon \ll 1$, and the fluctuating error field is taken equal to

$$u_j^n = \zeta \left(\sum_{m=0,N/2} \alpha_m^n e^{imx_j} + c.c. \right), \quad \zeta \ll \epsilon \tag{104}$$

where N is the number of grid points. It is worth noting that the key mechanisms at play for the growth of the error are related to wave resonance between a nonlinear solution and small superimposed disturbances, rather that the local pile-up of energy due to the convergence of characteristic lines due to the scale-dependence of the numerical phase speed. The splitting of the total solution as

$$v_j^n = c + U_j^n + u_j^n \tag{105}$$

can be seen as the result of a multiscale expansion of the full solution. Several theoretical approaches have been proposed in mathematical physics to perform such a decomposition that will not be discussed here for the sake of brevity, e.g. the Wentzel–Kramers–Brillouin (WKB) method . Frequencies of the two modes of the base flow are obtained considering dispersion relation associated with the linear discretized problem, i.e.

$$\sin(\omega_k \Delta t) = \sigma \sin(k\Delta x) \tag{106}$$

yielding

$$\omega_k^1 = \frac{1}{\Delta t} \arcsin(\sigma \sin(k\Delta x)), \quad \omega_k^2 = \frac{1}{\Delta t}(\pi - \Delta t \omega_k^1) \tag{107}$$

Inserting (103) and (104) into Eq. (102), one obtains discrete equations for the amplitude coefficients α_m^n:

$$\begin{aligned} &\frac{\alpha_s^{n+1} - \alpha_s^{n-1}}{2\Delta t} + i w_s \alpha_s^n = -i\epsilon C_s \left(\sum_{l=1,2} e^{-i\omega_k^l t_n} \right) (\alpha_{s-k}^n + \alpha_{s-k}^{n*}) \\ &- i\epsilon D_s \left(\sum_{l=1,2} e^{-i\omega_k^l t_n} \right) (\alpha_{s+k}^n + \alpha_{s+k}^{n*}) \end{aligned} \tag{108}$$

with $w_s = (U_0/\Delta x)\sin(s\Delta x)$ and

$$C_s = \frac{1}{\Delta x} (\theta \sin(s\Delta x) + (1-\theta)[\sin(k\Delta x) + \sin((s-k)\Delta x)]) \tag{109}$$

$$D_s = \frac{1}{\Delta x} (\theta \sin(s\Delta x) - (1-\theta)[\sin(k\Delta x) - \sin((s+k)\Delta x)]) \tag{110}$$

Considering that ϵ is very small, the right-hand-side is interpreted as a source term that must slightly perturb the solution of the homogeneous equation, whose solution is

$$\alpha_s^n \sim e^{-i\omega_s^l t_n}, \quad l = 1,2 \tag{111}$$

Substituting this solution in the right-hand-side, it is shown that instability will occur, i.e. α_s will grow linearly in time if one of the two resonance conditions is fulfilled:

$$\omega_s^m = \begin{cases} \omega_k^l + \omega_{s-k}^p \\ -\omega_k^l + \omega_{s+k}^p, \end{cases} \qquad m,l,p = 1,2 \tag{112}$$

In this case, the error fluctuations get in resonance with the base flow variations, leading to a constructive interaction (without feedback on the base flow solution). Looking at Eq. (107), the resonance conditions is fulfilled if:

$$s = \begin{cases} k & m = l = 1,2,\ p = 1 \\ N/2 & m = 2, l \neq p, \quad l,p = 1,2 \\ N/2 - k & m = 2, l = 1, p = 2 \end{cases} \tag{113}$$

showing that the solution is numerically unstable, since at least the two modes $s = N/2$ and $s = N/2 - k$ are unstable.

7. Explaining focusing using GSA: The mechanisms for the CE and focusing due to reflection of q-waves from NSE

In this section, focusing mechanism(s) for a wave-packet propagation following model linear CE as described using GSA in [57] is discussed and a mechanism is demonstrated for the solution of NSE. Focusing is a phenomenon which is observed during numerical solution of PDEs where the solution progresses smoothly for a long time accompanied by a quick solution break down due to a spectacular growth of numerical error at the fixed preferential node and the error-packet wavenumber. In the ensuing subsections, the focusing mechanisms for the linear CE are presented followed by the demonstration of a mechanism involving reflection of spurious q-waves at the boundary for the NSE.

7.1. Focusing mechanisms for linear CE

We recall that the model CE is given by Eq. (2). As discussed in the earlier section, GSA accurately incorporates the variation of numerical properties due to boundary closure schemes for explicit and implicit methods. It was also noted from the analyses that high accuracy methods suffer from numerical instability at the boundary and the near boundary nodes. Furthermore, for non-periodic problems, the instability when noted near the boundaries, is observed to be violently unstable compared to events in the interior of the domain. This prompted researchers in [72,74,83,84] to investigate the focusing phenomena as due to nonlinear mechanism(s).

Briggs et al. [72] proposed a nonlinear mechanism where the error gets focused at one point in the computational domain with respect to a nonlinear partial differential equation. They considered a periodic nonlinear problem which was quasilinearized and a three time level leap-frog method was used for time advancement, along with second order central difference scheme. As noted in an earlier section, the use of three time level method introduces a spurious computational mode, in addition to the physical mode. In their results, the spurious mode was found to be central for this nonlinear instability. For the employed leap-frog method for time advancement, Sloan & Mitchell [74] highlighted Fourier side-band instability in the context of envelope modulation. The same time discretization method was also employed in [83], for the nonlinear instability problem.

In [57], the instability is shown as due to a linear mechanism. This is enabled using GSA, where the violent instability is related to the nodal properties of the discretization, owing to the non-periodic nature of the problem. In demonstrating the results, the authors have employed a symmetrized OUCS3 scheme [67] and two time level RK4 method for time integration. The choice for the time integration method as noted by the authors, is to ensure that the instability is not due to computational mode.

In Figs. 19 and 20, the variation of numerical amplification, phase speed and group velocity contours with nodal locations is shown for the RK4-SOUCS3 scheme. In plotting these properties, 101 grid points are chosen for analysis and the results are plotted for near boundary nodes $j = 2,4,6,96,98,100$ and the interior point $j = 51$. It is noted that the interior point, which is also the middle point, is not influenced by the boundary stencils. As the symmetrization removes the directional bias of the upwind compact scheme, the nodal properties of points which are equidistant from the boundaries will remain the same. Comparison of numerical properties at near boundary nodes shows significant differences from the interior node. It is also noted that the interior node has a larger stability region compared to the nodes, $j = 2$ and 100. The comparison also shows large errors to be introduced at the near boundary nodes for numerical phase speed and numerical group velocity even for small values of N_c and $k\Delta x$. As a result, qualitatively different numerical solution is obtained at the interior and near boundary nodes for a propagating wave. Furthermore, in these plots, a dashed line along with a circle corresponding to $N_c =$

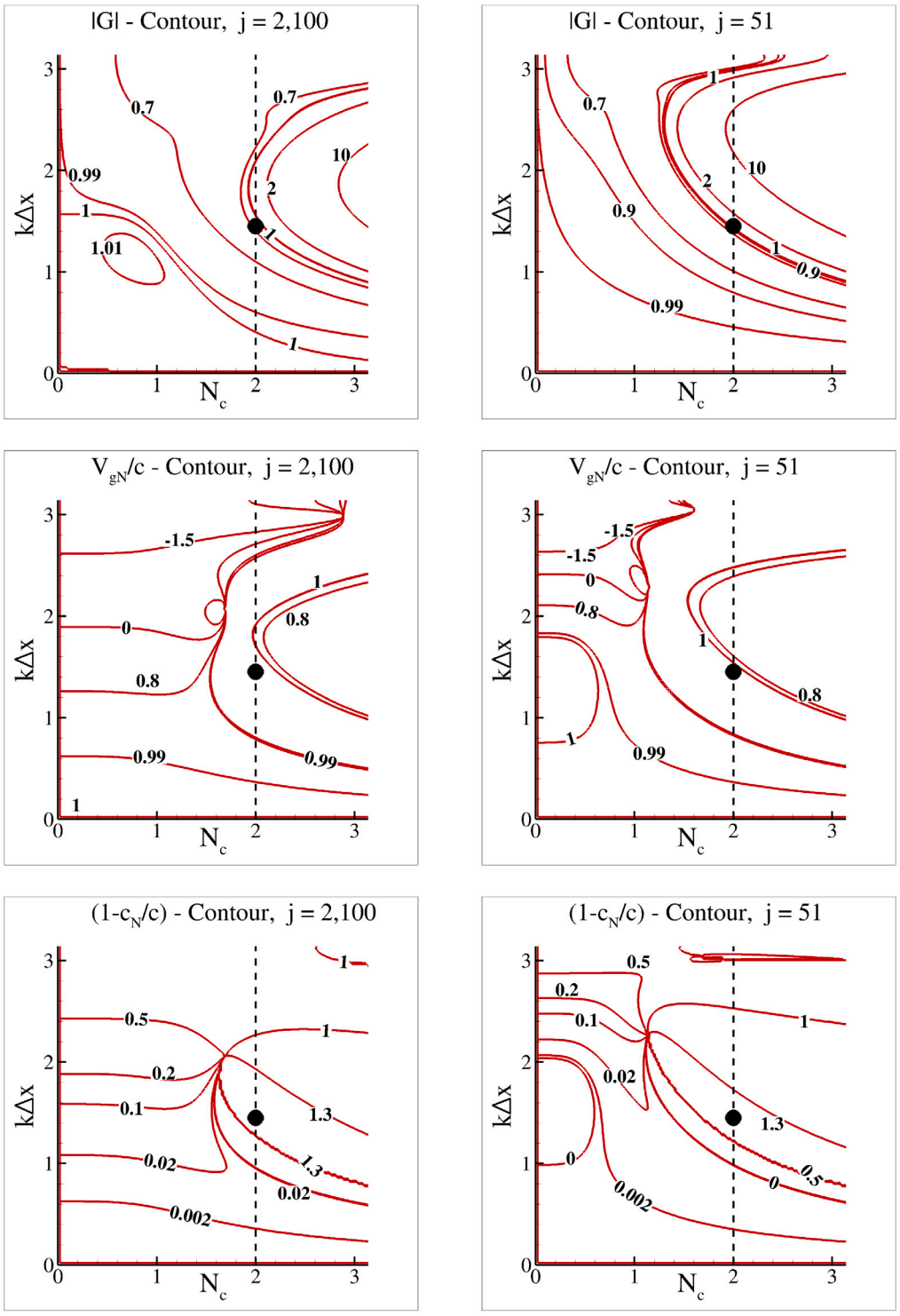

Fig. 19. Comparison of $|G_j|$, V_{gN}/c and $1-c_N/c$ contours for the near-boundary nodes $j = 2, 100$ (left column) and central node $j = 51$ (right column), using RK4-SOUCS3 scheme for the solution of 1D CE Eq. (2). The line at $N_c = 2$ corresponds to the cases computed and shown in Fig. 22.

2 is marked. This corresponds to the numerical test case chosen to demonstrate focusing, as described next.

Three different cases of focusing mechanism are shown in [57] by solving the propagation of wave-packet. The first is attributed to the instability arising at the near boundary nodes. The second case was shown as a consequence of discontinuity in the numerical solution. It was shown that focusing arising out of solution discontinuity starts at the location of discontinuity instead of boundary nodes. The third mode of focusing was attributed to the chosen numerical method of discretization. In this case, the error was always focused at a specific wavenumber scale which corresponded to numerical group velocity equal to zero, and was referred to as absolute instability. It should be noted this is not a necessary and sufficient condition of all spectacular error growth as the results correspond to the central spatial discretization schemes and does not hold for upwind spatial discretization stencil. In the next subsections, the three mechanisms are described.

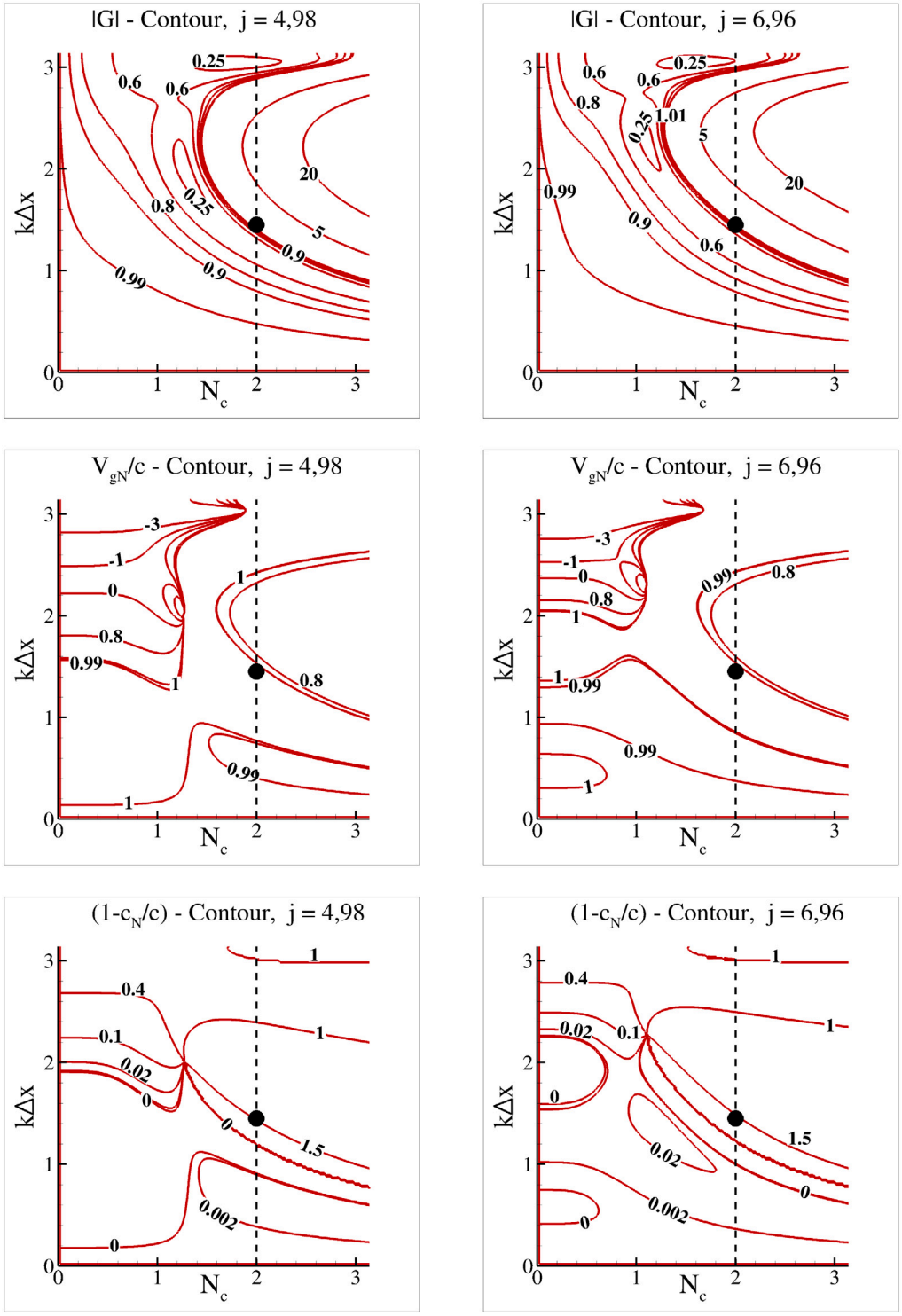

Fig. 20. Comparison of $|G_j|$, V_{gN}/c and $1-c_N/c$ contours for the near-boundary nodes $j = 4, 98$ and nodes $j = 6, 96$, using RK4-SOUCS3 scheme for the solution of 1D CE Eq. (2). The line at $N_c = 2$ corresponds to the cases computed and shown in Fig. 22.

7.1.1. Focusing phenomenon at a near-boundary node

Focusing is demonstrated using the propagation of a wave-packet whose initial solution is given by

$$u(x,0) = e^{-\alpha(x-x_0)^2}\sin(k_0 x) \tag{114}$$

where x_0 indicates the center of the packet, k_0 denotes the central wavenumber and α defines the spectral bandwidth of the packet. For the numerical simulation, the considered domain is $0 \le x \le 3$ with $x_0 =$ 1 and two different values of α equal to 10 and 24. The initial solutions are shown in panels (a) and (b) of Fig. 21 considering the packet to be centered at $k_0\Delta x = 1.45$. The computational domain employs 512 equi-spaced points. Although, both wave-packets look albeit with different width, major differences are noted for the tail signal of these wave-packets as shown by the enlarged views in the left panels of (a) and (b) in Fig. 21. The implications of the small differences are understood from the spectral band-width of the wave-packets as shown

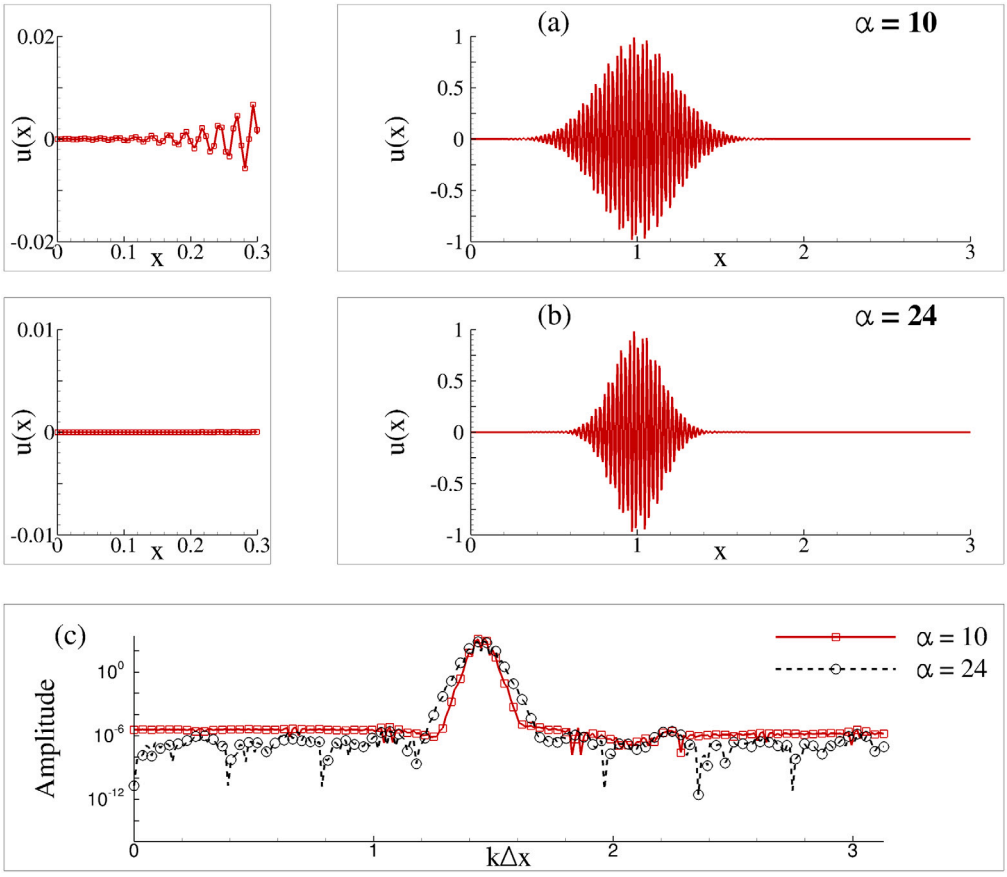

Fig. 21. Initial wave-packets given by Eq. (114) used for solving Eq. (2) are shown for (a) $\alpha = 10$ and (b) $\alpha = 24$. For both the cases $x_0 = 1$. In (a) and (b), the left panels are the enlarged view near the upstream boundary. Bottom frame (c) shows the FFT of the wave-packets in (a) and (b).

in Fig. 21(c). In the property charts in Fig. 19, the simulation conditions are indicated by the dotted line ($N_c = 2$) and the central wavenumber of the packet by a solid circle $k_0\Delta x = 1.45$. The differences of spectral band-widths of the packets and the values of the signal amplitude in the tail of the packets cause different numerical solutions as demonstrated next.

Figs. 22(a) and 22(b) show the evolution of numerical solution for $\alpha = 10$ and 24, respectively for the indicated times. The speed of propagation in these cases i.e. c, is equal to 0.01. For the chosen central wavenumber ($k_0\Delta x = 1.45$) and CFL number $N_c = 2$, the nodal amplification factor is noted as $|G_j| = 1.089$ from Fig. 19 for the interior node, i.e. $j = 51$. The unstable case is specifically chosen to demonstrate error growth and the focusing phenomenon where error localizes on a spatial grid as a high wavenumber instability. The numerical solutions in Figs. 22(a) and 22(b) display instability over the entire domain with the rates corroborating with the analyses in Figs. 19 and 20. In contrast, the computed solution using CD2 and leapfrog time marching scheme was identified as nonlinear growth in [72]. For the case of $\alpha = 24$ shown in Fig. 22(b), the wave-packet is noted to travel from left to right as the numerical group velocity, $\frac{V_{gN}(k_0\Delta x)}{c} > 0$, as shown in Fig. 19 for the central node $j = 51$. For the case of $\alpha = 10$ as shown in Fig. 22(a), one notes a rapidly growing wave-packet near the inflow boundary in addition to the original wave-packet. Following [57], this is attributed to (i) the location in the computational domain of the site where the error appears and (ii) the wavenumber selection procedure of this additional wave-packet. It was noted in [57] that such error was also numerically obtained in [75] but could not be explained satisfactorily.

Researchers have investigated error growth from different perspectives-as in [75,96,97]. In [96,97], phase error was calculated for different numerical methods for propagation of the monochromatic waves i.e. $c - c_n(k_0)$. However, the effect and quantification of phase error on signal error $u - u_N$ was not discussed. The authors also stated that the phase error could be reduced to very small values by refining the mesh alone for a fixed Nc. However, $1 - c_N/c$ contours shown in the property charts in Figs. 19 and 20 show that the error does not decay monotonically with wavenumber for any N_c. It should also be noted that according to the assertion in [75], the rapidly growing localized error originates from the decoupled solution at even and odd nodes implying that the error would correspond to $k\Delta x = \pi$. However, the present numerical solutions clearly show that is not the case in Fig. 22(a).

The FFT of the computed signals in Fig. 22(a) are shown in Fig. 22(c) to reveal the scale selection of errors. FFT results show the dominant error-packet corresponds to the central wavenumber $k_e\Delta x = 2.355$ for $t \geq 9\Delta t$. This is explained by noting the distinction between the cases of Figs. 22(a) and 22(b), regarding the wavenumber content and bandwidth of the initial signal. For $\alpha = 10$, the wavenumber bandwidth is less than $\alpha = 24$, as seen in Fig. 21(c). However the amplitude of the tail i.e. solution away from the packet center, is noted to be higher by a order of magnitude for the lower α case from Fig. 21(a) and (b). Hence, the larger signal for the tail is magnified more and as a result the error becomes visible near the left boundary as shown in Fig. 22(a). As noted before, the differences of instability between interior and near-boundary nodes manifest in focusing phenomenon.

From the FFT of the computed signal, the wavenumber corresponding to the maximum error growth is identified as $k_e\Delta x = 2.355$. In order to understand the node selection process for the focusing of error, the numerical properties for $N_c = 2$ and wavenumbers $k_0\Delta x$ and $k_e\Delta x$ are plotted for all the nodes in Fig. 22(d). From the figure, one notes that at few nodes, the numerical amplification factor $|G_j|$ is almost 9

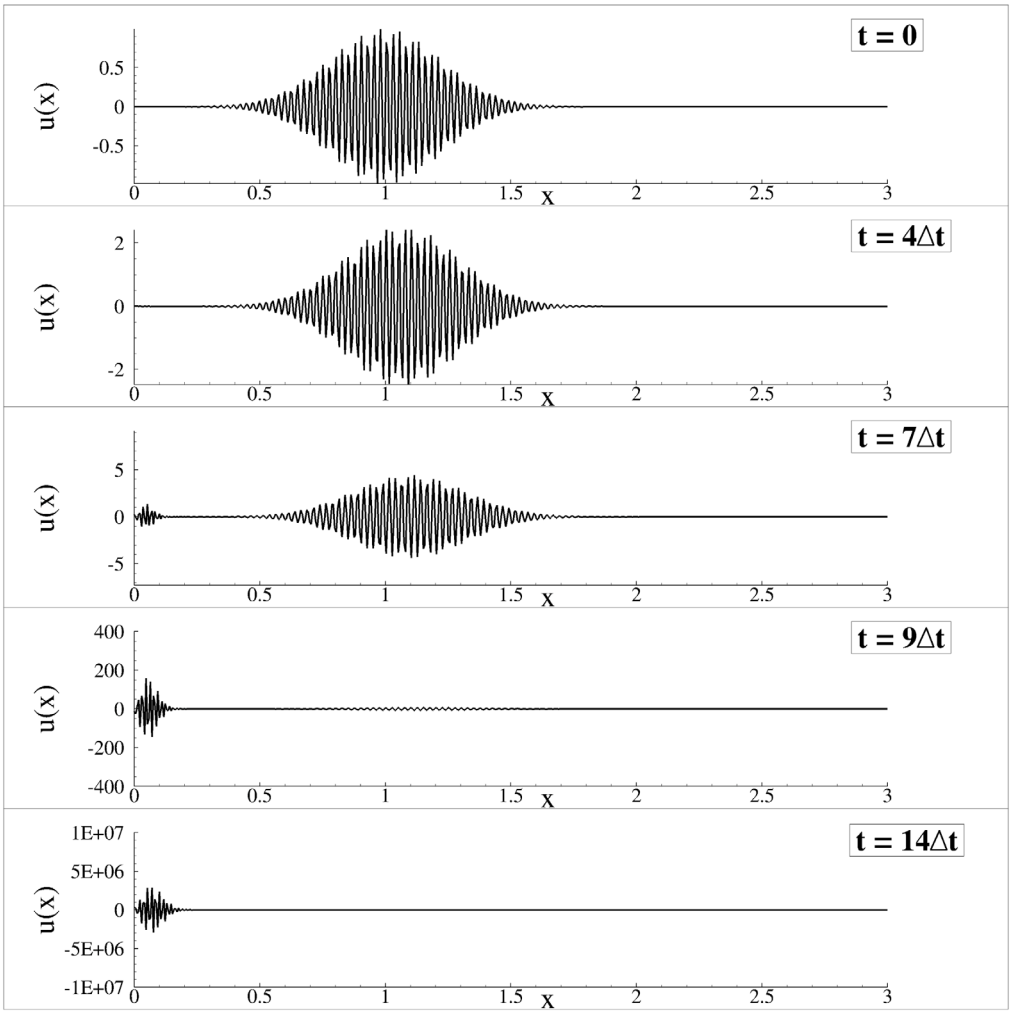

Fig. 22(a). Computed solution of Eq. (2), using RK4-SOUCS3 scheme, for the case of Fig. 21(a) with $\alpha = 10$ in Eq. (114). Focusing is evident from the appearance of the additional packet near the upstream boundary.

to 10 times larger for $k_e\Delta x = 2.355$. This is the reason for the rapid increase of amplitude of the error-packet as compared to the input wave-packet. One also notes from the figure that the sixth node has highest numerical amplification factor thus explaining the observed maximum growth of error at the node $j = 6$. This is again corroborated by plotting the numerical properties for the node at $j = 6$ across the entire wavenumber range for $N_c = 2$, in Fig. 22(e). In all the frames of the figure, wavenumbers corresponding to $k_0\Delta x = 1.45$ and $k_e\Delta x = 2.355$ are marked as A and B, respectively. This demonstrates the utility of GSA in explaining the mechanism by which the error is focused at the specific node and at the specific wavenumber for a given spatio-temporal discretization scheme. We also note that $|G|$ value at $k_0\Delta x$ is 1.022 which indicates mild instability compared to value of 13.22 at $k_e\Delta x$ implying violent instability. Furthermore, the contribution of numerical phase speed error $1 - c_N/c$ to the numerical error can be quantified by noting that it has a value of 1.55 at $k_0\Delta x$ and 1.059 at $k_e\Delta x$.

7.1.2. Focusing phenomenon due to solution discontinuity

In this section, focusing phenomenon is demonstrated at the interior nodes of the domain, with the help of a small discontinuity in the initial condition. The presence of the discontinuity in the numerical solution excites all the resolved wavenumbers and leads to focusing. The initial wave-packet solution with the small discontinuity is given by,

$$u(x,0) = e^{-32(x-x_0)^2}\sin(k_0x) + 0.1e^{-50000(x-x_1)^2} \tag{115}$$

where the first term on the right hand side of Eq. (115) is the regular wave-packet and is denoted by packet-A. The second term on the right hand side is the small discontinuity and is denoted as B. Results of the computations are shown in Fig. 23(a) with the top panel showing the initial solution. Packet-A is located at $x_0 = 2.5$ while disturbance B is at $x_1 = 1$ at $t = 0$. For the simulation, a domain of size $0 \le x \le 3$ with 512 equi-spaced points is chosen. The central wavenumber corresponding to packet-A is chosen to be $k_0\Delta x = 1.45$. As before, the convection speed c is chosen as 0.01. Time-step is chosen such that the CFL number $N_c = 2$. The results of the computation in Fig. 23(a) show a spectacular growth of the small disturbance B. Due to the small convection speed, the error-packet moves slowly towards the right side from its starting position $x_1 = 1$. In Fig. 23(b), the FFT of the corresponding numerical solution is shown for different time instants. One notes the wavenumber for the dominant error to be centered at $k_e\Delta x = 2.411$. For this wavenumber, the corresponding numerical group velocity is $\frac{V_{gN}}{c} = 0.8359$ for a central node as shown in Fig. 23(c). From this figure, one also notes the maximum G_j value to attain at $k_e\Delta x$ and thus leads to focusing.

7.2. Focusing due to reflections of q-waves from NSE

The focusing phenomenon is not only observed in a numerical solution of the model wave propagation problem but also observed in the solutions of NSE. A particular case is demonstrated in this section which is attributed to the reflection of spurious q-waves at

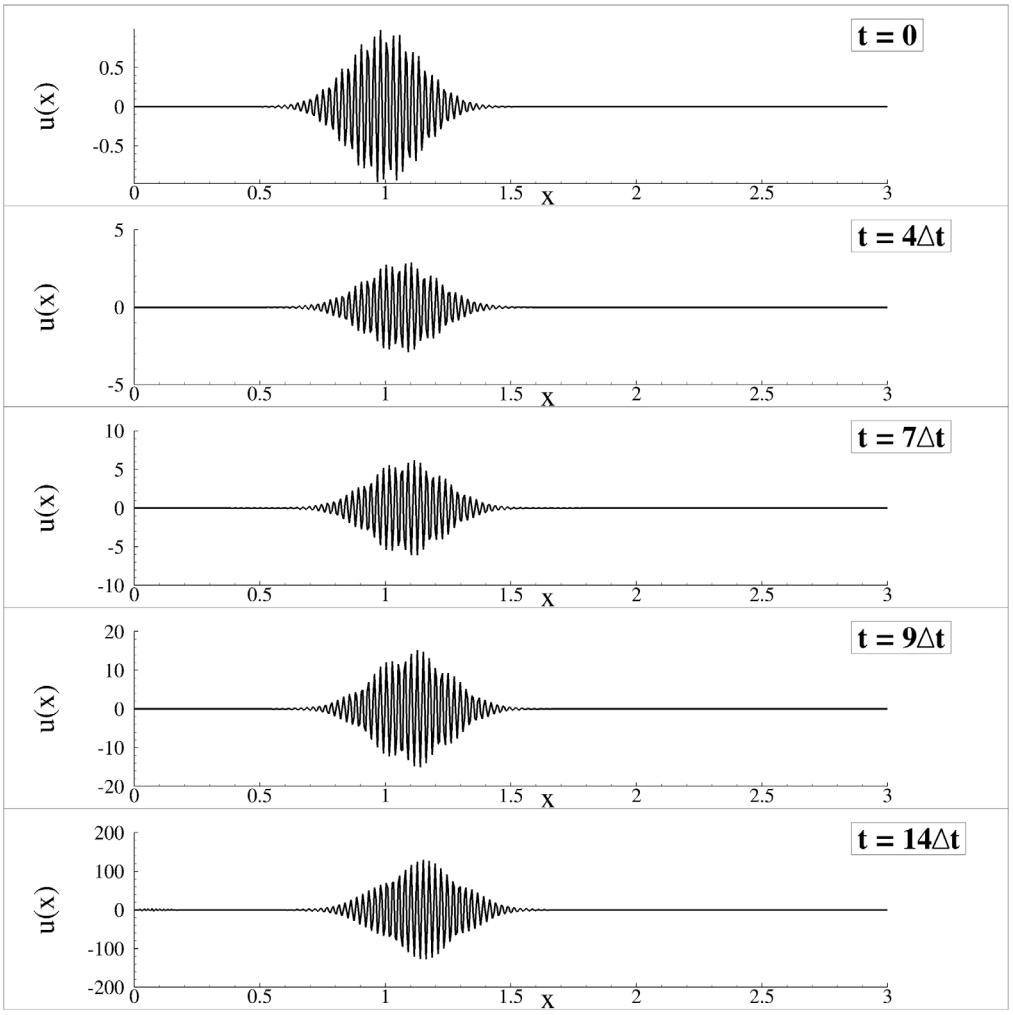

Fig. 22(b). Computed solution of Eq. (2), using RK4-SOUCS3 scheme, for the case of Fig. 21(b) with $\alpha = 24$ in Eq. (114). Here, the wave-packet growth is spectacular and the amplitude of the error-packet is very small as compared to Fig. 22(a).

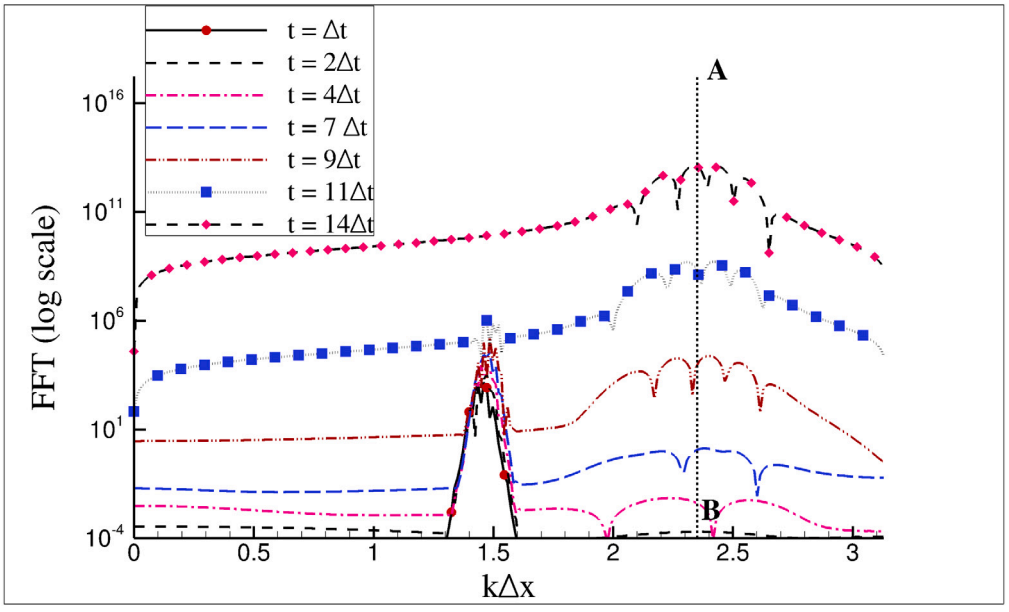

Fig. 22(c). FFT of the wave-packets are shown corresponding to wave-packets in Fig. 22(a), at the indicated time instants. A line AB at $k_e \Delta x = 2.355$ is drawn to represent the central wavenumber of the error-packet.

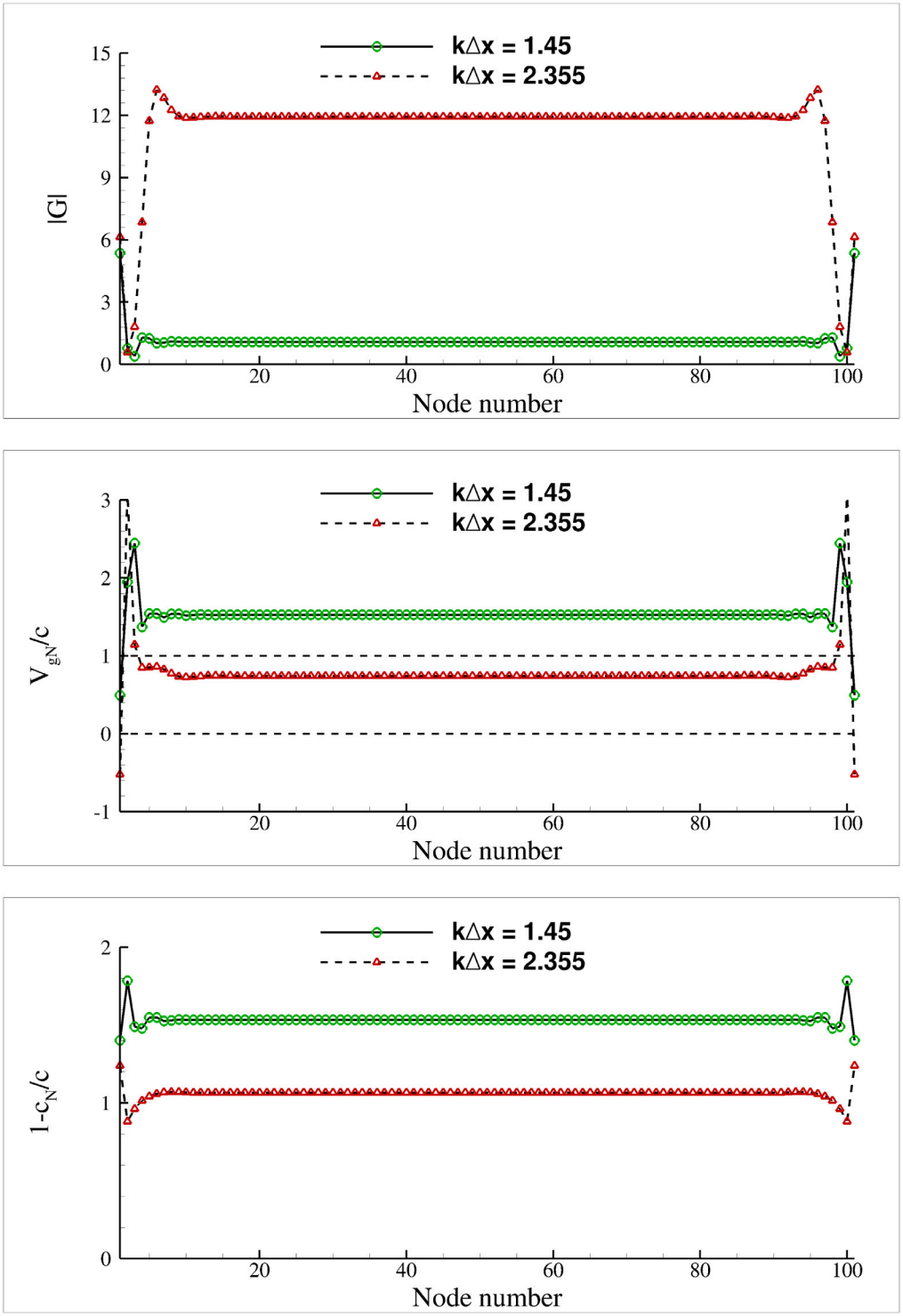

Fig. 22(d). Variations of $|G_j|$, V_{gN}/c and $1-c_N/c$ are shown, using RK4-SOUCS3 scheme, at different nodes of the domain corresponding to $N_c = 2$, for (i) the input wave-packet with $k_0\Delta x = 1.45$ (left column) and (ii) dominant response at $k_e\Delta x = 2.355$ (right column).

the boundaries. As noted earlier q-waves are spurious waves which have negative group velocity. It is noted that while these q-waves have been demonstrated in solution of model convection [36], convection–diffusion [49] and convection–diffusion–reaction equations [47], the present results demonstrate this from the direct solution of NSE. It is interesting to note that the observations and subsequent conclusions drawn from the solutions of model wave propagation equations provide important clues for explaining the focusing phenomenon in fluid flow simulations.

Here, we consider a propagation of a discrete shielded vortex along with the flow. Consider a two-dimensional rectangular domain $(-1 \le x \le 7)$, $(-1 \le y \le 1)$ containing equi-spaced grid points with grid spacing $\Delta x = \Delta y = 0.008$. A uniform flow enters the domain from the left boundary $(x = -1)$ and leaves the domain from the right hand

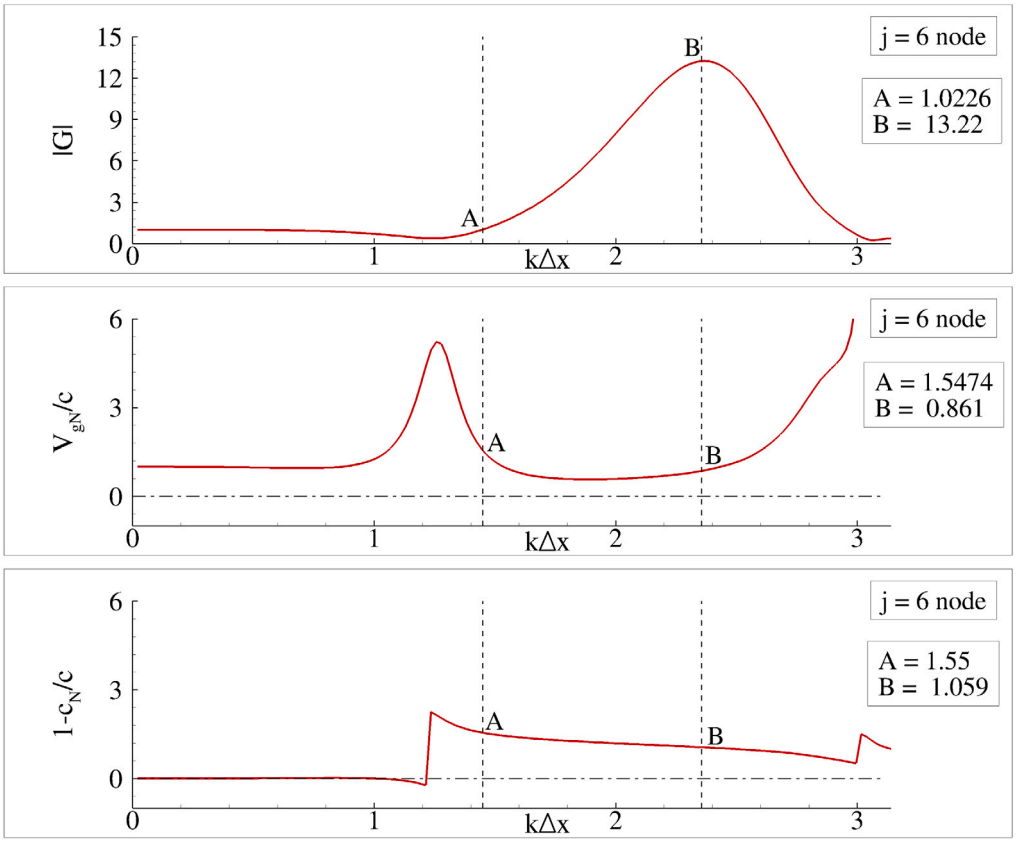

Fig. 22(e). Variations of $|G_j|$, V_{gN}/c and $1 - c_N/c$ are shown, using RK4-SOUCS3 scheme, for $j = 6$ node across the whole $k\Delta x$ range for $N_c = 2$.

side boundary. As an initial condition, we have considered a discrete shielded vortex centered at the origin $(0, 0)$ and superimposed on a uniform flow. If r denotes the distance between any point in the domain and the center of the vortex then the discrete shielded vortex can be prescribed as in [98,99],

$$\omega = k\,(1 - 100\,r^2)\,e^{-100\,r^2} \tag{116}$$

where, k denotes initial maximum vorticity centered at the origin and is prescribed as $k = 500$ in the present simulations. Propagation of the vortex along with the flow has been simulated using the NSE formulated in the stream function (ψ) - vorticity (ω) formulation which are given as [100],

$$\frac{\partial^2\psi}{\partial x^2} + \frac{\partial^2\psi}{\partial y^2} = -\omega \tag{117}$$

$$\frac{\partial\omega}{\partial t} + u\frac{\partial\omega}{\partial x} + v\frac{\partial\omega}{\partial y} = \frac{1}{Re}\left[\frac{\partial^2\omega}{\partial x^2} + \frac{\partial^2\omega}{\partial y^2}\right] \tag{118}$$

Eqs. (117) and (118) are non-dimensionalized using free-stream velocity U_∞ as a reference velocity and $1/k$ as the time scale. The Reynolds number is defined as $Re = \frac{U_\infty^2}{\nu k}$. Simulations have been performed for $Re = 10^5$. Here, we have used the $OUCS3$ scheme for discretization of the convective derivative terms while second order central discretization scheme has been used for the discretization of the second order derivative terms. Time integration has been performed using four stage, fourth order Runge–Kutta $(RK4)$ scheme. Calculations are performed using a time step of $\Delta t = 0.0007$. A uniform flow has been prescribed at the inflow boundary $x = -1$ while the stream function and vorticity values at the remaining boundaries have been updated using a convective outflow boundary condition.

Fig. 24 shows the propagation of the shielded vortex along with the flow using vorticity contours at the indicated instants. We have purposefully taken a steep vortex as an initial condition. One observes that the vortex moves with a unit non-dimensional velocity along the positive x-axis. However, one also observes nonphysical high wavenumber components propagating towards left hand side of the domain against the flow direction. These waves are identified as q-waves which have been generated due to the strong vorticity gradient associated with the shielded vortex. Solutions display q-waves as the numerical method is not able to preserve the physical dispersion relation numerically at high wavenumber region as observed before for the solution of the $1D$ wave equation. Although amplitude of these q-waves is small, these spurious waves can trigger numerical instability. Fig. 25 displays variation of vorticity on the line $y = 0$ at the indicated instants. One does not observe q-waves as the amplitude of these waves is small and are only observed in the zoomed view as displayed in Fig. 26. One observes small amplitude q-waves are propagating towards left boundary of the domain and their subsequent reflection from the left boundary. Although one does not observe any numerical instability in this example, if large amount of nonphysical q-waves are generated in the simulation, then one does observe numerical instability.

Authors in [100] reported numerical instability for a laminar flow past a rotary oscillating cylinder due to generation of spurious q-waves upon reflection of a convecting vortex from the outflow boundary. It was observed that if the convecting vortex has sufficient vorticity upon reaching the outflow boundary, then spurious high wavenumber oscillations are triggered at the outflow boundary which lead to numerical instability. This particular aspect has been highlighted here by reducing the domain size $(-1 \le x \le 1)$, $(-1 \le y \le 1)$ and carrying out simulations with the same initial condition. The grid spacing $\Delta x = \Delta y = 0.008$ and time step $\Delta t = 0.0007$ have been also kept same. As the outflow boundary has been brought from $x = 7$ to $x = 1$, it is expected that the vortex will create spurious high wavenumber oscillations at the outflow boundary which will lead to numerical instability.

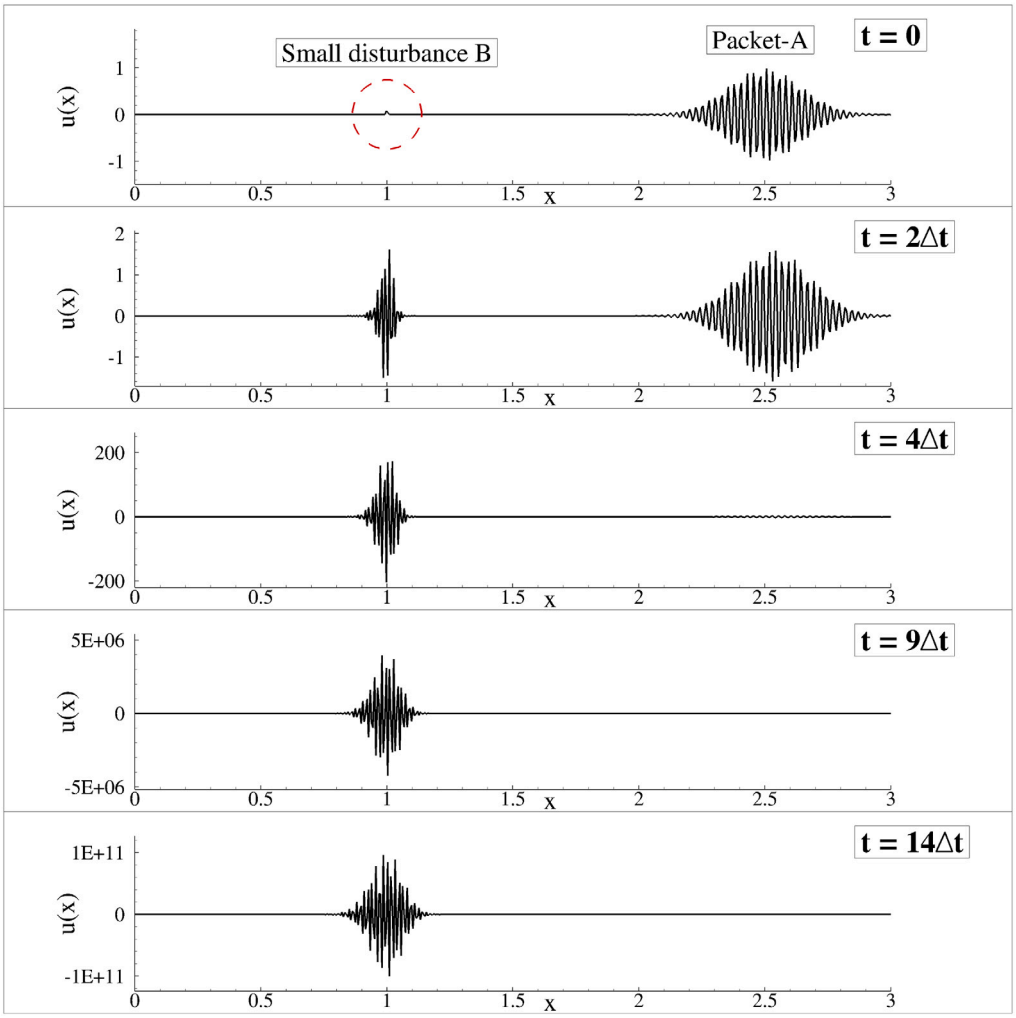

Fig. 23(a). Computed solution of Eq. (2) corresponding to the initial condition given in Eq. (115) using RK4-SOUCS3 scheme. Note the initial small disturbance at $x = 1$ undergoes spectacular amplification displaying focusing phenomenon.

Fig. 27 shows propagation of a shielded vortex in the reduced size domain at the indicated instants. The figure also displays spurious q-waves propagating towards the left boundary of the domain. It is observed that the simulation experienced numerical instability at $t = 0.93$ as the vortex reaches the outflow boundary. This is due to generation of large amount of spurious reflected waves at the domain outflow boundary. The numerical instability takes place over a very small duration as observed in the vorticity time variation at the locations close to outflow boundary as shown in Fig. 28. The vorticity close to outflow boundary ($x = 1$) starts picking up after $t = 0.9$ and magnitude increases by few orders over a small duration which indicates focusing phenomenon.

8. Explaining focusing using GSA: Focusing mechanisms for CDE and NSE

In this section, we analyze the 2D CDE using GSA in the context of calibrating numerical methods and understanding focusing phenomenon reported for numerical solution of nonlinear NSE. As discussed in the previous section, focusing is an instability reported for problems involving long time integration. Focusing, first reported by weather community, was observed during simulations using three time-level leapfrog method where the solution suddenly blew up after running successfully for a long time when corrective actions were not taken [101,102]. Phillips [102] hypothesized the blow-up/focusing to a non-linear computational instability and also proposed an adhoc rectification by filtering high wavenumber(k) components of the computed solution which yielded non-physical results [102]. Since then, different authors [72,74,83,84] have also led support to the nonlinear numerical instability hypothesis until recent studies [44,45,49,57,73] which showed the mechanism to be of linear origins.

In Section 7, a linear mechanism for focusing was discussed which is based on the analysis of CE. In this section, new linear mechanisms of focusing will be presented using the analysis of 2D CDE and simulation of NSE which establishes further evidence on the linear origins of the instability. This work is reported recently in [49,73] and is discussed in brief details. Furthermore, focusing for solution of NSE using the three time-level AB_2 method is demonstrated here for the first time.

8.1. GSA of 2D CDE and mechanisms for focusing

The model linear 2D CDE for a single dependent variable (u) is given in the Cartesian frame by,

$$\frac{\partial u}{\partial t} + c_x \frac{\partial u}{\partial x} + c_y \frac{\partial u}{\partial y} = \alpha \left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} \right) \tag{119}$$

where c_x, c_y denote the constant convection speeds in x- and y-directions, respectively, and α denotes the constant coefficient of diffusion. Generally, α is positive, indicating the stabilizing nature of diffusion. It should be noted that when α is a negative value, it is called anti-diffusion. Anti-diffusion can be physical, as in interface

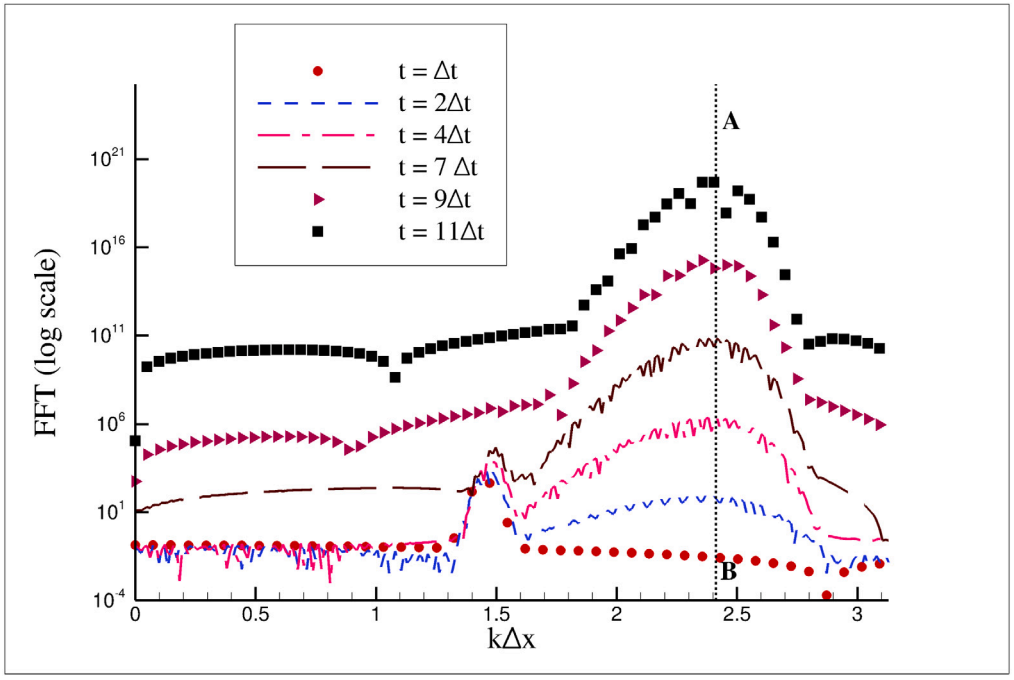

Fig. 23(b). FFT of the wave-packets are shown corresponding to wave-packets in Fig. 23(a), at the indicated time instants. A line AB at $k_e\Delta x = 2.411$ is drawn to represent the central wavenumber of the error-packet.

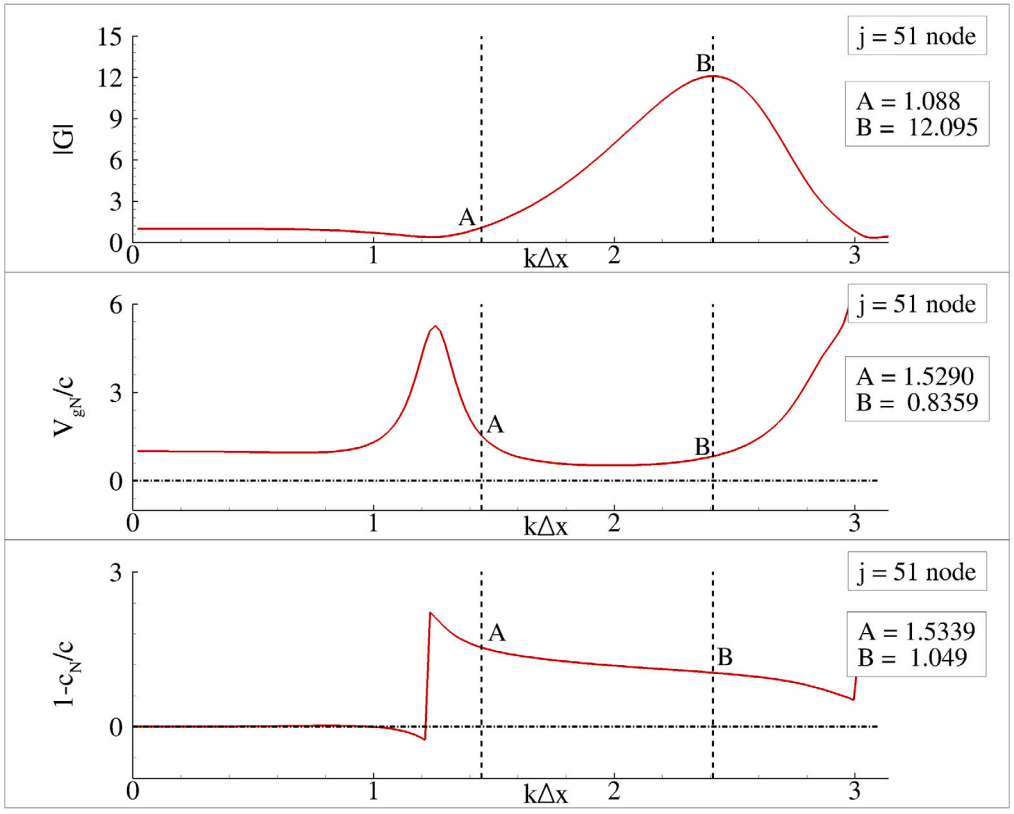

Fig. 23(c). Variations of $|G_j|$, V_{gN}/c and $1 - c_N/c$ are shown, using RK4-SOUCS3 scheme, for $j = 51$ node across the whole $k\Delta x$ range for $N_c = 2$.

steepening in two-phase incompressible flow [103], in granular transport, considering an interplay between drift and anti-diffusion [104], in solving geomagnetic storm problem at near-Earth [105], etc. In the context of coupling between heat flow and diffusion to create order in a chaotic system, anti-diffusion is discussed and is related to *a negative contribution to entropy production* [106]. In the context of image processing, a relation between deconvolving an image resulting from a Gaussian filter and integrating anti-diffusion equation is noted [107]. In interfacial flow instabilities, e.g. Rayleigh–Taylor instability, effects of anti-diffusion is related to negentropy [108,109].

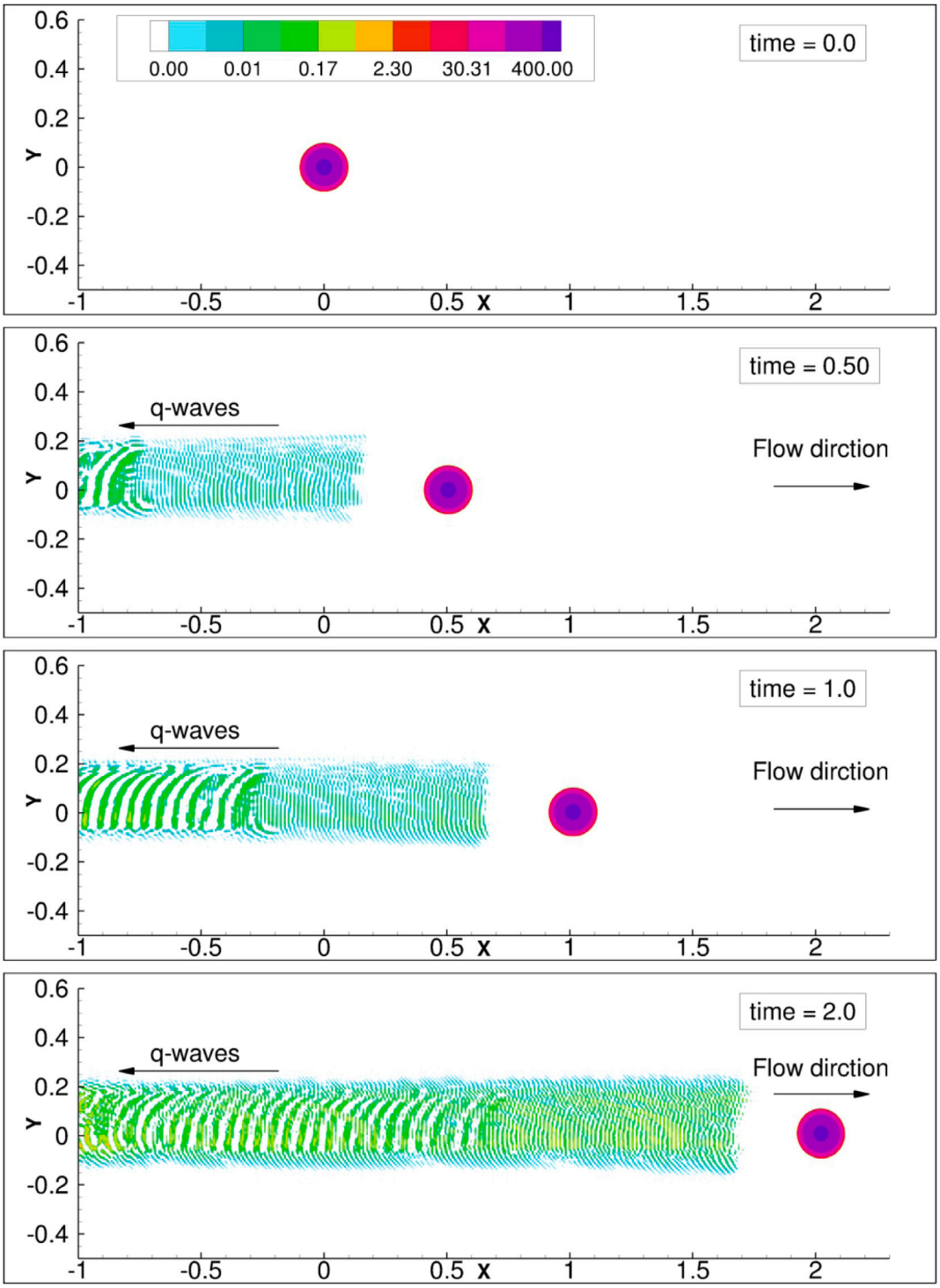

Fig. 24. Propagation of a shielded vortex in the domain at the indicated instants has been shown. Figure also displays spurious q-waves propagating towards left boundary of the domain.

In performing GSA of Eq. (119), $u(x, y, t)$ is expressed in the hybrid-spectral plane as,

$$u(x, y, t) = \iint \hat{U}(k_x, k_y, t) e^{i(k_x x + k_y y)} dk_x dk_y \tag{120}$$

where $\hat{U}$ is the Fourier–Laplace amplitude and k_x, k_y are the wavenumber components in the x- and y-directions, respectively. Substituting the expression for u in Eq. (119) gives the transformed equation as,

$$\frac{\partial \hat{U}}{\partial t} + ic_x k_x \hat{U} + ic_y k_y \hat{U} = -\alpha(k_x^2 + k_y^2)\hat{U} \tag{121}$$

Representing the initial condition as

$$u(x, y, 0) = f(x, y) = \iint U_0(k_x, k_y) e^{i(k_x x + k_y y)} dk_x dk_y$$

one obtains the exact solution of Eq. (119) as,

$$\hat{U}(k_x, k_y, t) = U_0(k_x, k_y)\, e^{-\alpha(k_x^2 + k_y^2)t} e^{-i(k_x c_x + k_y c_y)t} \tag{122}$$

Expressing u in the Fourier–Laplace transform as $u(x, y, t) = \iiint \hat{U}(k_x, k_y, \omega)\, e^{i(k_x x + k_y y - \omega t)} dk_x dk_y d\omega$, one obtains the physical dispersion

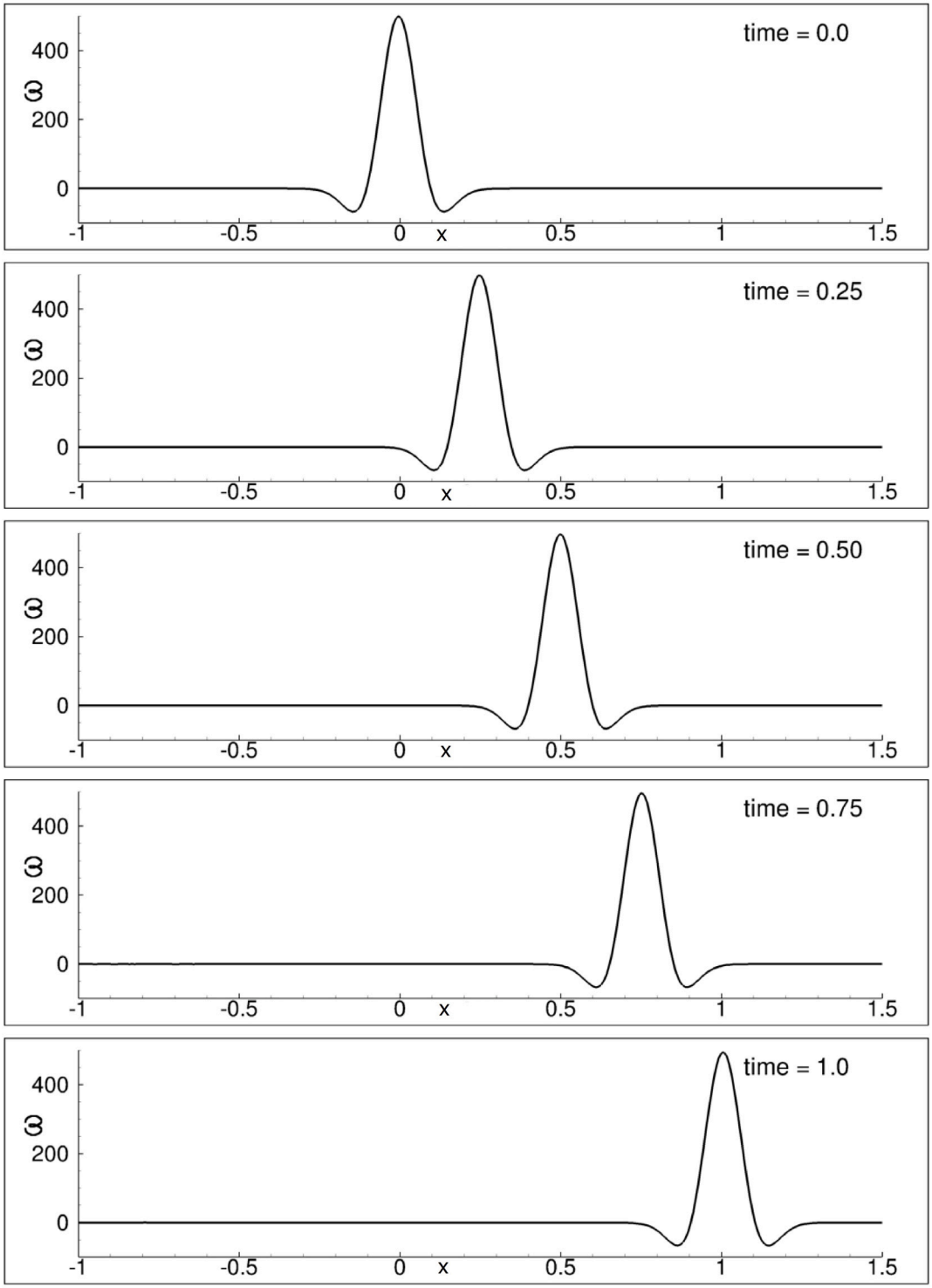

Fig. 25. Variation of vorticity on the line $y = 0$ has been shown at the indicated instants.

relation for the 2D CDE as

$$\omega = c_x k_x + c_y k_y - i\alpha(k_x^2 + k_y^2) \tag{123}$$

In solving the CDE accurately the numerical schemes must obey the dispersion relation to minimize phase and dispersion errors [1,31]. From the physical dispersion relation, one obtains the complex phase speed as given by,

$$c = \frac{\omega}{\sqrt{k_x^2 + k_y^2}} = \frac{c_x k_x + c_y k_y - i\alpha(k_x^2 + k_y^2)}{\sqrt{k_x^2 + k_y^2}} \tag{124}$$

Also, the physical group velocity components are obtained from Eq. (123) as,

$$V_{gx} = \frac{\partial \omega}{\partial k_x} = c_x - 2i\alpha k_x \tag{125}$$

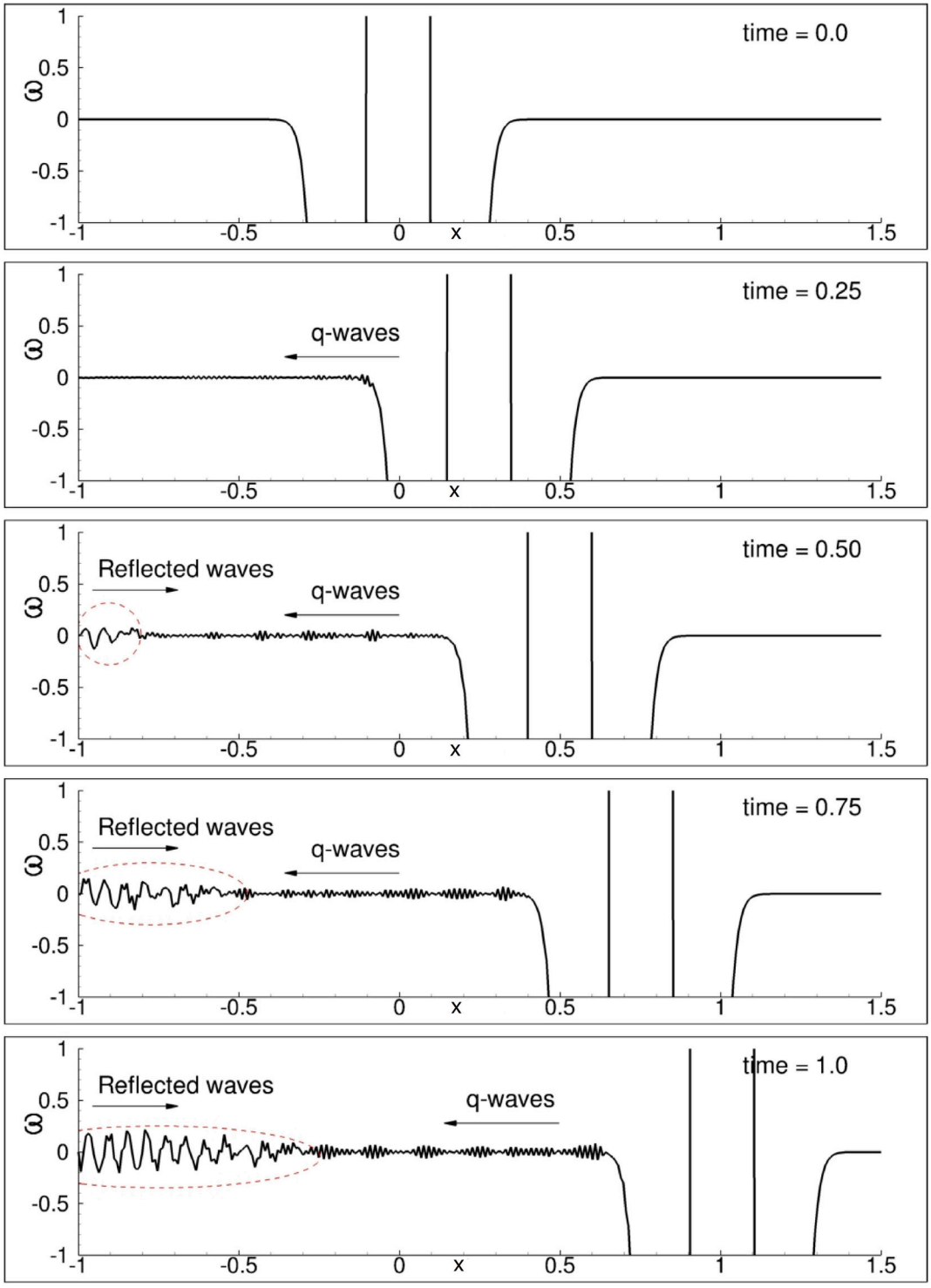

Fig. 26. Zoomed view of the variation of vorticity on the line $y = 0$ has been shown at the indicated instants. Figure shows small amplitude q-waves propagating towards left boundary of the domain and subsequent reflection.

$$V_{gy} = \frac{\partial \omega}{\partial k_y} = c_y - 2i\alpha k_y \tag{126}$$

From the exact solution of the CDE Eq. (121), the physical amplification factor is obtained as,

$$\begin{aligned} G &= \frac{\hat{U}(k_x, k_y, t + \Delta t)}{\hat{U}(k_x, k_y, t)} \\ &= e^{-\alpha(k_x^2 + k_y^2)\Delta t} e^{-i(k_x c_x + k_y c_y)\Delta t} \\ &= e^{-[Pe_x (k_x h_x)^2 + Pe_y (k_y h_y)^2]} e^{-i[N_{cx} k_x h_x + N_{cy} k_y h_y]} \end{aligned} \tag{127}$$

where Δt is the discrete time-step and h_x and h_y are the grid spacings in x- and y-directions, respectively. We also introduce the CFL and Peclet numbers in 2D as, $N_{cx} = \frac{c_x \Delta t}{h_x}$; $N_{cy} = \frac{c_y \Delta t}{h_y}$; $Pe_x = \frac{\alpha \Delta t}{h_x{}^2}$; $Pe_y = \frac{\alpha \Delta t}{h_y{}^2}$, which are the relevant parameters for analysis. The physical solution decays with time, as is noted from G and is in accordance with the physical nature of diffusion.

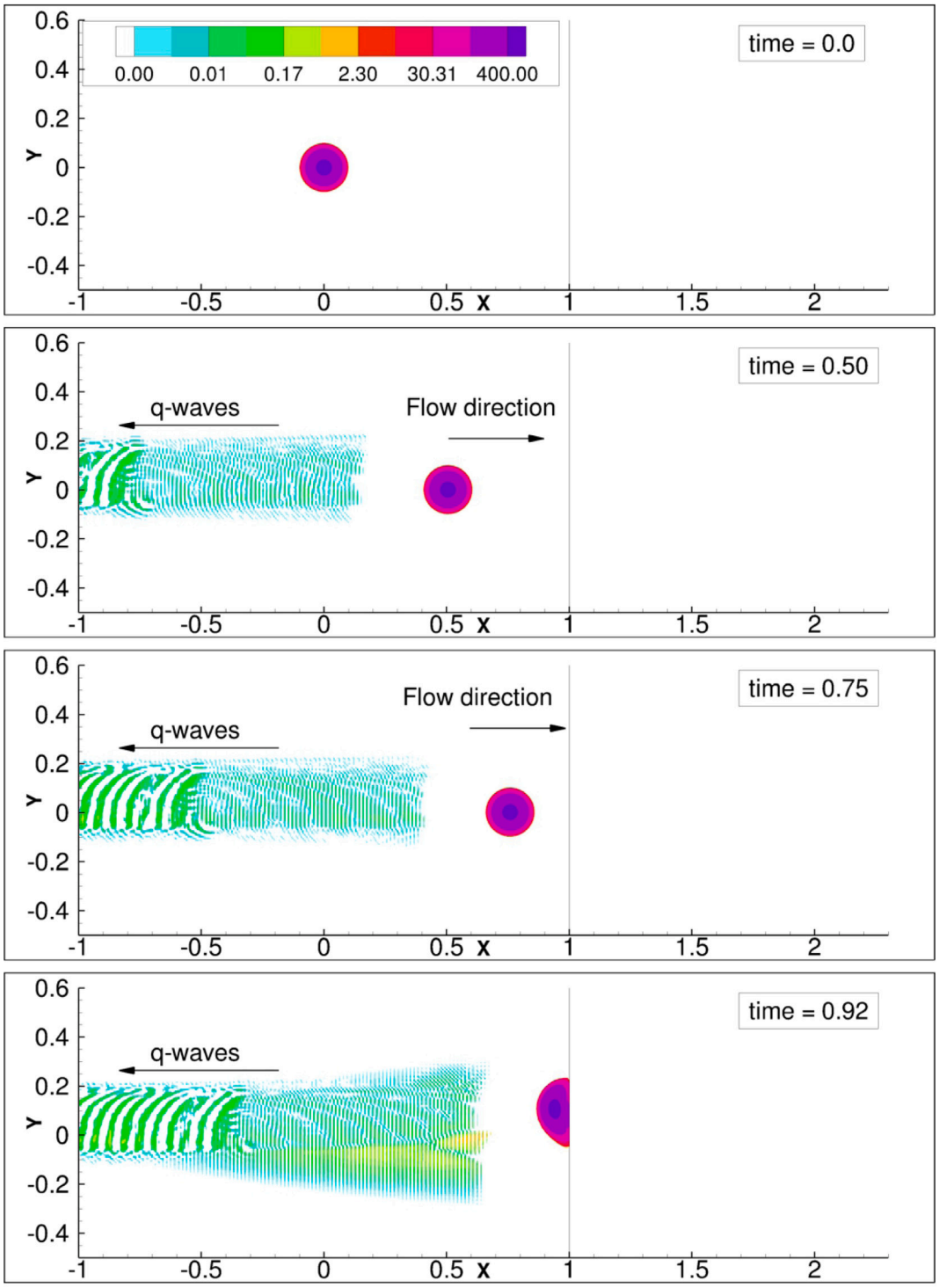

Fig. 27. Propagation of a shielded vortex in the reduced size domain at the indicated instants has been shown. Figure also displays spurious q-waves propagating towards left boundary of the domain. Simulation experiences numerical instability at $t = 0.93$ as the vortex reach outflow boundary.

The above analysis is exact for the CDE. However, as noted before, the solution of the CDE using numerical methods leads to the case where phase speed and α are non-constants. In order to study the performance of numerical schemes, we first obtain the numerical dispersion relation governing the evolution of solution. This relation is obtained by analogy from Eq. (123) as,

$$\omega_N = \left(\sqrt{{k_x}^2 + {k_y}^2}\right) c_N - i\alpha_N({k_x}^2 + {k_y}^2) \tag{128}$$

where c_N and α_N are not constants for any simulation. The numerical dispersion relation is generic and applies to any discretization. From the numerical dispersion relation, one obtains numerical amplification factor G_N as

$$G_N = e^{-i\omega_N \Delta t} = e^{-\alpha_N({k_x}^2+{k_y}^2)\Delta t}\ e^{-i\left(\sqrt{{k_x}^2+{k_y}^2}\right)c_N \Delta t} \tag{129}$$

It is clear that for accurate simulations, G_N should follow G as closely as possible, i.e. we require $G_N/G \approx 1$, with $|G_N| < 1$. From

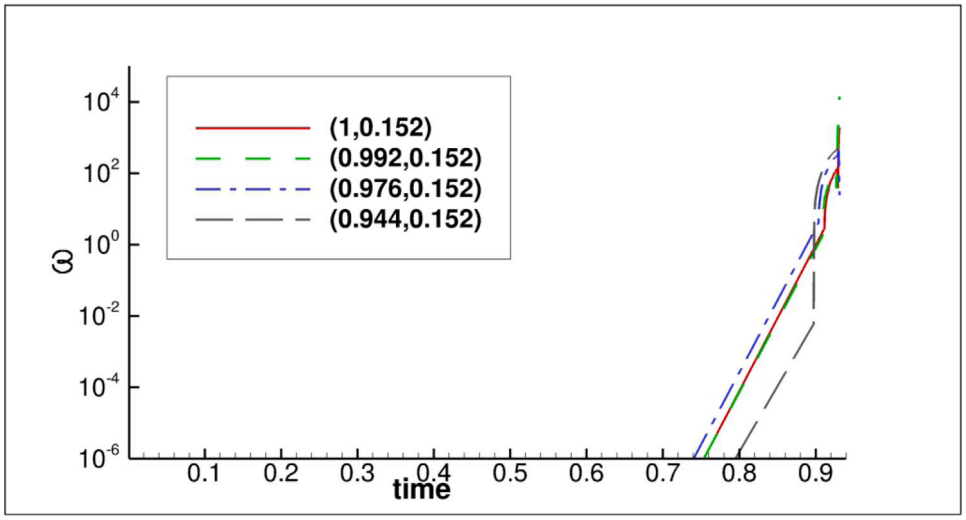

Fig. 28. Time variation of vorticity at the indicated locations close to outflow boundary. One observes the focusing phenomenon as the vorticity values shoot up in a few time steps close to $t = 0.93$.

G_N the numerical phase shift ϕ_N per unit time step Δt is obtained as

$$\tan(\phi_N) = -\left(\frac{(G_N)_{Img}}{(G_N)_{Real}}\right) \quad \text{where} \quad \phi_N = \left(\sqrt{{k_x}^2 + {k_y}^2}\right) c_N \Delta t \tag{130}$$

The non-dimensional numerical phase speed is obtained as

$$\frac{c_N}{c} = -\left[\frac{1}{N_{cx}(k_x h_x) + N_{cy}(k_y h_y)}\right] \tan^{-1}\left(\frac{(G_N)_{Img}}{(G_N)_{Real}}\right) \tag{131}$$

where c is the physical phase speed in Eq. (124), i.e. $c = \frac{c_x k_x + c_y k_y}{\sqrt{k_x^2 + k_y^2}}$. The numerical group velocity components are obtained from their definitions and are given by

$$\frac{(V_{gx})_N}{c_x} = \frac{1}{N_{cx}} \frac{\partial \phi_N}{\partial (k_x h_x)} \tag{132}$$

$$\frac{(V_{gy})_N}{c_y} = \frac{1}{N_{cy}} \frac{\partial \phi_N}{\partial (k_y h_y)} \tag{133}$$

where c_x and c_y are the physical group velocity components in the x- and y-directions, respectively.

The numerical diffusion coefficient α_N is evaluated from G_N in non-dimensional form as

$$\frac{\alpha_N}{\alpha} = -\frac{\ln |G_N|}{[Pe_x (k_x h_x)^2 + Pe_y (k_y h_y)^2]} \tag{134}$$

The significance of numerical diffusion coefficient is explained here. If $\frac{\alpha_N}{\alpha} = 1$, then the numerical scheme models the physical diffusion exactly. If the ratio is greater than unity, then the numerical diffusion is higher than the physical diffusion, else we have lower numerical diffusion as compared to physical diffusion. Negative values of the ratio indicate anti-diffusion and hence, leads to numerical instability. Thus, numerical diffusion can also contribute to numerical instability. We also note that for accuracy of solution, all the quantities $\frac{\alpha_N}{\alpha}$, $\frac{c_N}{c}$, $\frac{(V_{gx})_N}{c_x}$ and $\frac{(V_{gy})_N}{c_y}$ should be equal to unity.

We analyze the RK_4-NCCD scheme and provide the performance metrics, $\frac{\alpha_N}{\alpha}$, $\frac{c_N}{c}$, $\frac{(V_{gx})_N}{c_x}$ and $\frac{(V_{gy})_N}{c_y}$. In [49], the authors have found this method to be the most accurate in solving 1D CDE among all the schemes analyzed. As noted earlier, NCCD scheme being a combined compact difference scheme, enables simultaneous evaluation of first and second derivatives. Expressing the simultaneous equations in a compact form, one obtains the equations for the scheme as [65]

$$[A]\{du\} = \{b\}$$

with the details of the matrix $[A]$ and the vectors $\{du\}, \{b\}$ as given in [64,65]. The simultaneous equations are solved for the derivatives using

$$\begin{aligned} \{u'\} &= \frac{1}{h}[D_1]\{u\} \\ \{u''\} &= \frac{1}{h^2}[D_2]\{u\} \end{aligned} \tag{135}$$

The analytical expressions for $[D_1]$ and $[D_2]$ are given in [64], with block tridiagonal matrix algorithm (TDMA) used to obtain these derivatives.

For the 2D CDE, the G_N for the RK_4-NCCD scheme is given by

$$(G_N)_{mn} = 1 - A_{mn} + \frac{{A_{mn}}^2}{2} - \frac{{A_{mn}}^3}{6} + \frac{{A_{mn}}^4}{24} \tag{136}$$

with

$$A_{mn} = -\frac{L(\hat{U})}{\hat{U}} \tag{137}$$

where, m, n are the nodal indices in x-and y-directions, respectively. The variable A_{mn} in Eq. (136) is determined as

$$\begin{aligned} A_{mn} = {} & N_{cx} \sum_{l=1}^{N_x} D_{1_{ml}} e^{ik(x_l - x_m)} - Pe_x \sum_{l=1}^{N_x} D_{2_{ml}} e^{ik(x_l - x_m)} \\ & + N_{cy} \sum_{l=1}^{N_y} D_{1_{nl}} e^{ik(y_l - y_n)} - Pe_y \sum_{l=1}^{N_y} D_{2_{nl}} e^{ik(y_l - y_n)} \end{aligned} \tag{138}$$

where N_x, N_y are the number of nodes in x- and y-directions, respectively. By substituting A_{mn} in Eq. (136), G_N is determined for the node (m, n). The other properties viz. $\frac{c_N}{c}$, $\frac{(V_{gx})_N}{c_x}$, $\frac{(V_{gy})_N}{c_y}$ and $\frac{\alpha_N}{\alpha}$ are determined from Eqs. (131), (132), (133) and (134), respectively.

For the 2D analysis, two more parameters: $AR = \frac{h_y}{h_x}$ and $\theta = \tan^{-1}(c_y/c_x)$ are introduced following the analysis of anisotropy for the 2D convection problem [110]. Analysis is performed only for the case of $AR = 1$, due to the choice of the grid employed for the simulations of NSE.

Fig. 29 shows the stable regions plotted for the middle stencil of the RK_4-NCCD scheme for different N_{cx}, N_{cy}, Pe combinations and wave propagation angle is considered as 45°. The iso-surface corresponding to $|G_N| = 1$ is shown in the figure, which demarcates the region of stability from instability. It is noted that the numerical scheme is stable when $|G_N| < 1$, while $|G_N| > 1$ leads to numerical instability. The horizontal plane $Pe = Pe_{cr}$ in the figure shows the onset of instability with respect to Pe. Below this plane the solution is noted to be stable.

Three distinct routes of numerical instability can be observed from the figure. One route corresponds to the case when instability occurs for any nonzero values of CFL numbers when Pe is above a critical value i.e. $Pe_{cr} = 0.1451$, as indicated in the panels (a), (b) and (c) of Fig. 29.

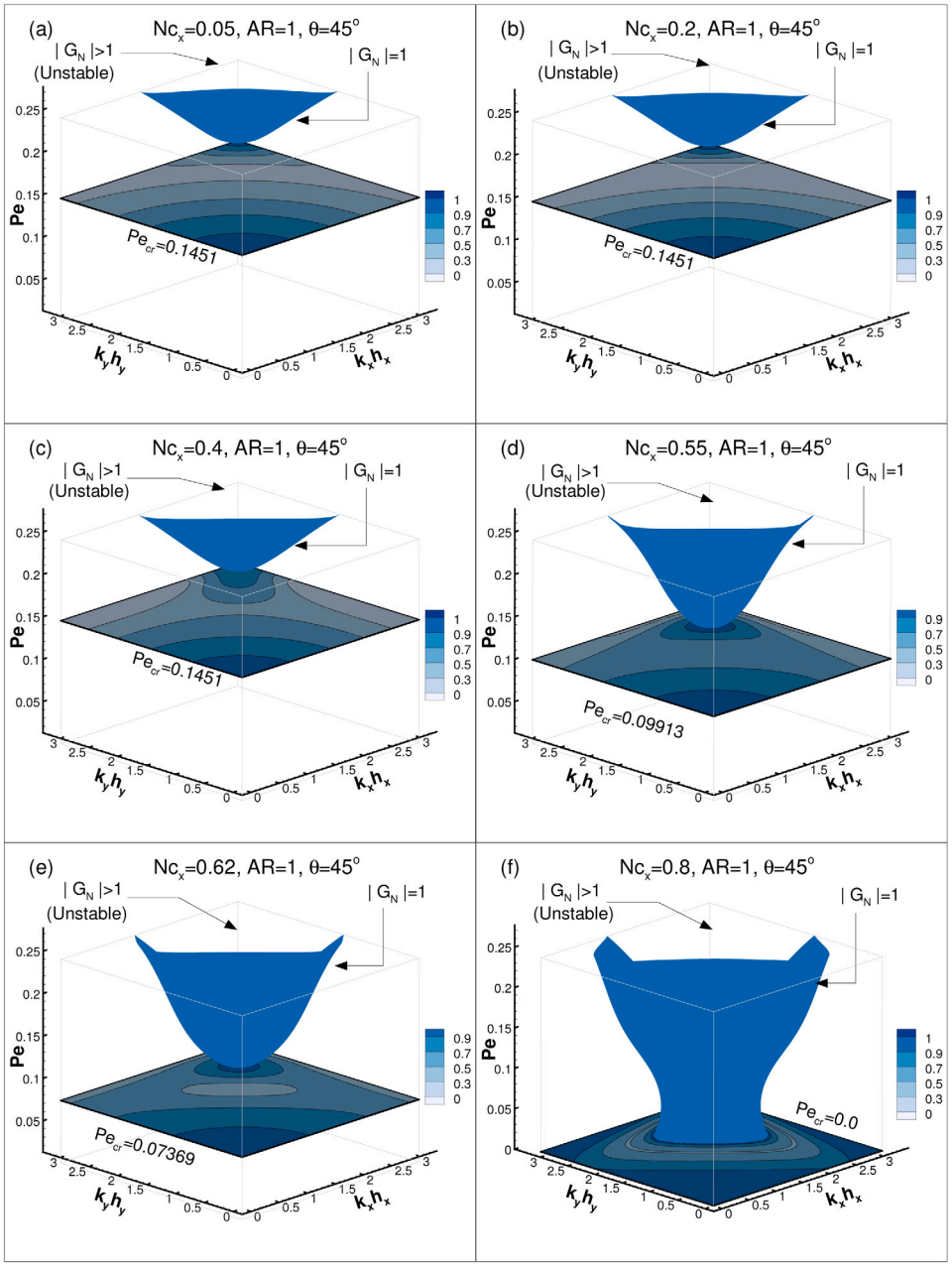

Fig. 29. Iso-contour of numerical amplification factor $|G_N| = 1$ for RK_4-NCCD scheme for grid aspect ratio $AR = 1$, wave propagation angle $\theta = 45°$ and CFL number $N_{cx} = 0.05, 0.2, 0.4, 0.55, 0.62, 0.8$ plotted in the (k_xh_x, k_yh_y, Pe)-plane. The plane $Pe = Pe_{cr}$ denotes the critical Peclet number above which numerical instability exists.

For Pe less than the critical value, instability arises at higher values of N_{cx} (and hence N_{cy}) as seen in panels (d) and (e), e.g., $Pe_{cr} = 0.09913$ and 0.07369 for $N_c = 0.55$ and 0.62, respectively. This is the second route of instability. The third route is noted when N_{cx} and/or N_{cy} values are increased further, say to $N_c = 0.8$, leading to instability for all Pe (i.e. $Pe_{cr} = 0.0$), as noted in the bottom right panel (f). These three cases of instability is understood by relating numerical errors arising from convective and diffusive terms. The first mode arises due to dominance of errors by diffusion term. The second mode exists due to combination of errors from convection and diffusion terms, while the last mode is due to dominance of convective errors only.

The scale selection for instability for the three modes are noted here. For the first mode, instability first appears when $k_xh_x = k_yh_y = \pi$ and percolates into lower wavenumber regions as Pe increases. For the other modes, errors are amplified first at moderately high wavenumbers and the spread increases in extent with increasing N_{cx} and N_{cy} values.

The previous discussion is for the case of $\theta = 45°$. In order to analyze the effects of θ on the numerical instability for $AR = 1$, in Fig. 30, the stable and unstable regions are marked for different θ values in the $Nc(= \sqrt{N_{cx}^2 + N_{cy}^2})$ and Pe plane. The plot shows that numerical instability does not depend on θ for the considered $AR = 1$, due to invariance of Eq. (138) on θ. As noted earlier, for lower Nc, $Pe_{cr} = 0.1451$ and this value decreases as Nc increases beyond a certain limit. From the plot, one attributes the instability in Route-1 to errors from diffusion discretization as instability occurs even for $N_{cx} = N_{cy} = 0$, when $Pe \geq 0.1451$.

The analysis indicates that iff $Pe \geq 0.1451$ numerical instability will arise causing solution blow-up in finite time. This is due to the appearance of anti-diffusion. It is noteworthy that the instability mimics absolute instability, as postulated in [111]. When Pe is close to and above the critical limit i.e. $Pe \to (0.1451 + \epsilon)$, as ϵ becomes small, error

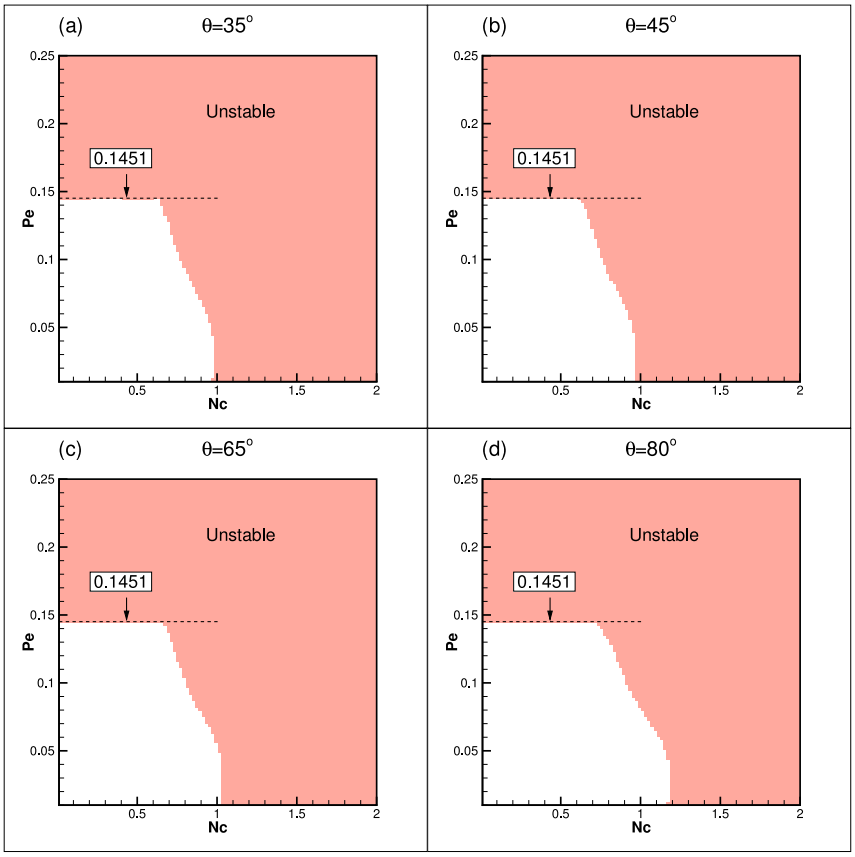

Fig. 30. Influence of wave propagation angle θ on the critical Peclet number Pe_{cr} for RK_4-NCCD scheme for the grid aspect ratio $AR = 1$ shown in the (Nc, Pe)-plane. Nc here is defined as $\sqrt{N_{cx}^2 + N_{cy}^2}$.

too grows slowly as $|G_N| \rightarrow (1 + \epsilon_1)$ with ϵ_1 also being a small positive quantity. This implies that blow-up occurs after a long time if epsilon is very small. The slow growth of error characterized by a selective scale of instability is known as focusing and the identified routes can explain its mechanism. For steady state simulations, one would obtain perfect steady solutions initially, matching in every respect with the non-focused solution, but the solution will blow-up after a long time due to focusing. The background omnipresent disturbances corresponding to specific length scales for focusing are amplified with the solution exploding eventually. For the computations involving unsteady dynamics, focusing can occur early. This is because of the reduced time step size (Δt) needed for unsteady computations (as compared to steady flow cases); which means more number of computations n $(= \frac{t}{\Delta t})$ is required for the unsteady case to reach the same value of time, t.

In the previous paragraphs, GSA was performed for RK_4-NCCD scheme whose middle stencil is a central scheme. Next, GSA for the upwind discretization of the convective terms in the governing CDE is performed and the effect of its inherent numerical dissipation on stability is assessed. For the present study, we adopt a third order upwind (UD_3) scheme pioneered by Kuwahara [112]. Diffusion terms are discretized by a second order central differencing (CD_2) scheme and time integration is performed by RK_4 method.

For this numerical scheme, the discretized right hand side term for the 2D CDE is given by

$$L(u_{lm}^n) = -c_x \frac{u_{l+2m}^n - 2u_{l+1m}^n + 9u_{lm}^n - 10u_{l-1m}^n + 2u_{l-2m}^n}{6h_x} - c_y \frac{u_{lm+2}^n - 2u_{lm+1}^n + 9u_{lm}^n - 10u_{lm-1}^n + 2u_{lm-2}^n}{6h_y} + \alpha \left[\frac{u_{l+1m}^n - 2u_{lm}^n + u_{l-1m}^n}{h_x^2} + \frac{u_{lm+1}^n - 2u_{lm}^n + u_{lm-1}^n}{h_y^2} \right] \tag{139}$$

where n indicates the time index, l, m are the indices in x- and y-directions, respectively. A_{lm}, given below, is used to determine G_N and the properties $\frac{c_N}{c}$, $\frac{(V_{gx})_N}{c_x}$, $\frac{(V_{gy})_N}{c_y}$ and $\frac{\alpha_N}{\alpha}$.

$$\begin{aligned} A_{lm} =& N_{cx} \left[\frac{e^{2ik_xh_x} - 2e^{ik_xh_x} + 9 - 10e^{-ik_xh_x} + 2e^{-2ik_xh_x}}{6} \right] \\ &+ N_{cy} \left[\frac{e^{2ik_yh_y} - 2e^{ik_yh_y} + 9 - 10e^{-ik_yh_y} + 2e^{-2ik_yh_y}}{6} \right] \\ &+ 2\,Pe_x \left[1 - \cos(k_xh_x)\right] + 2\,Pe_y \left[1 - \cos(k_yh_y)\right] \end{aligned} \tag{140}$$

Fig. 31, shows the stability region for $\theta = 35°$, $45°$, $65°$ and $80°$ for aspect ratio $AR = 1$ in the $(Nc - Pe)$-plane. Numerical instability is indicated by the colored region. Comparing with Fig. 30 for central scheme, one immediately notes the absence of critical Peclet number value (Pe_{cr}) independent of CFL number. Therefore, numerical instability for the upwind scheme is caused due to combinations of errors from both diffusion and convective discretizations. Further, one observes that as Nc increases, Pe_{cr} decreases, linearly. Numerical stability is therefore constrained to a triangular region enclosing the origin. The maximum Pe_{cr} value for the upwind scheme is noted to be higher than its corresponding central counterpart of the same order (not shown here) due to the inherent diffusion contained in the former which increases the strength of the overall numerical diffusion. However, this may not be beneficial for computing as a slight increase in Nc reduces Pe_{cr} value.

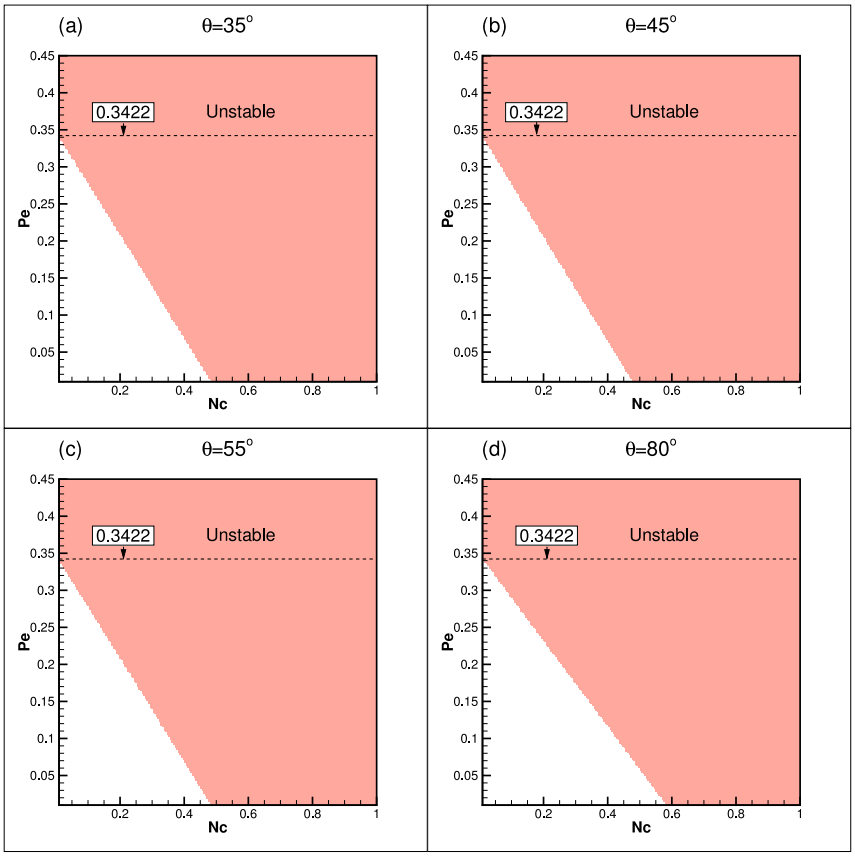

Fig. 31. Stability region for RK_4-UD_3-CD_2 scheme for grid aspect ratio $AR = 1$ plotted for wave propagation angles $\theta = 35°$, $45°$, $65°$ and $80°$ in the (Nc, Pe)-plane. Nc here is defined as $\sqrt{N_{cx}^2 + N_{cy}^2}$.

8.2. Focusing of 2D incompressible NSE

Focusing is demonstrated by solving 2D incompressible NSE for the canonical flow inside a square lid driven cavity (LDC). The governing equations are solved in streamfunction (ψ) - vorticity (ω) formulation which ensures direct satisfaction of mass conservation in the computational domain. In this approach, one solves a Poisson equation for ψ and a transport equation for ω which resembles the 2D CDE, given by

$$\nabla^2\psi = -\omega \tag{141}$$

$$\frac{\partial\omega}{\partial t} + (\vec{V}\cdot\vec{\nabla})\omega = \frac{1}{Re}\nabla^2\omega \tag{142}$$

where Re is the Reynolds number resulting from the non-dimensionalization of the equations using the side of the LDC and the speed of the upper lid. The velocity vector $\vec{V}$ is given by $\vec{V} = u\hat{i} + v\hat{j}$ and its components are related to the stream-function by $\vec{V} = \vec{\nabla}\times\vec{\psi}$, where $\vec{\psi} = \langle 0\ 0\ \psi\rangle^T$.

The stream function equation, Eq. (141) is discretized using CD_2 scheme and then solved using the BiCGSTAB iterative method [113]. In solving the vorticity transport equation, Eq. (142), the NCCD scheme is used for spatial derivatives of ω. For the numerical solution using upwind scheme, convective terms are discretized by UD_3 and CD_2 scheme is used for the viscous terms. Time integration is performed using the RK_4 method.

Focusing in the solution of incompressible NSE is demonstrated using three cases. In all the cases considered here the flow is unsteady as a super-critical Reynolds number $Re = 10{,}000$ is chosen. The first case shows focusing due to errors from diffusion term discretization while the other two cases highlight the role of errors due to both convection and diffusion terms discretizations. It should be noted that for the first two cases RK_4-NCCD scheme is employed and the last case considers the upwind scheme RK_4-UD_3.

For the first case, two computations are performed using identical, uniform grids of size (4001×4001) in the (x, y)-plane but with different time step sizes. The first computation employs a time step of $\Delta t_1 = 9.06625 \times 10^{-5}$ and for the other computation a slightly higher value of $\Delta t_2 = 9.06875 \times 10^{-5}$ is used. The chosen time steps and the grid size result in Peclet numbers of $Pe_1 = 0.14506$ and $Pe_2 = 0.1451$, respectively, considering $\alpha = \frac{1}{Re}$. Maximum N_c in the domain is evaluated as $N_{cx} = N_{cy} = 0.375$ by considering the velocity to be equal to the velocity scale i.e. $u = v = 1$.

Property charts of $\frac{\alpha_N}{\alpha}$ based on 2D linear CDE and employing the simulation parameters for the two computations are shown in Fig. 32. The Peclet number corresponding to Pe_2 shows numerical instability due to anti-diffusion ($\alpha_N < 0$) occurring in a small region close to $((k_x h_x, k_y h_y) = (\pi, \pi))$. This is indicated by a red arrow in the top right panel of the figure. This indicates that the focusing is due to errors from diffusion discretization. The scale selection and focusing of errors will happen at these wavenumbers showing up as grid-scale oscillations in the computed solutions. The other simulation does not display focusing as anti-diffusion is absent.

Results are presented for the two unsteady NSE computations in Fig. 33. In the figure, the left panels represent the simulation with $Pe = 0.14506$, and the right panel represents the second simulation with $Pe = 0.1451$, respectively. The results for $Pe = 0.1451$ show grid-scale oscillations at an early time $t = 2$, and the solution blows-up at $t \simeq 2.8$ (not shown in the figure). This is due to a combination of the reasons: (i) the small time-step size leading to an increase in the number of iterations n $(= \frac{t}{\Delta t})$ and hence, early blow-up as amplification factor becomes $|G_N|^n$; (ii) non-negligible FFT amplitudes of initial ω

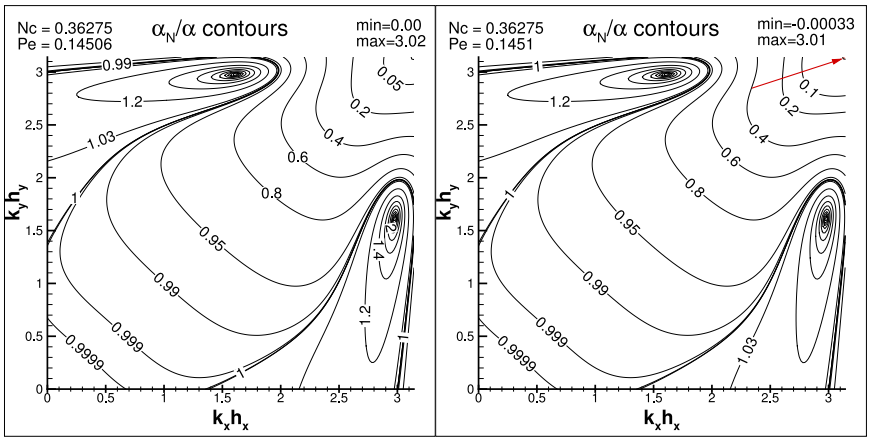

Fig. 32. Numerical diffusion ($\frac{\alpha_N}{\alpha}$) contours for the RK_4-NCCD scheme for the 2D CDE. The left panel corresponds to the space time properties $Pe_x = 0.14506$, $N_{cx} = 0.375$, $AR = 1$, $\theta = 45°$ and the right panel is for the parameters $Pe_x = 0.1451$, $N_{cx} = 0.375$, $AR = 1$, $\theta = 45°$, respectively.

field ($\mathcal{O}(10^{-5})$) at Nyquist limits arising due to the higher boundary vorticities (ω_b) for the finer grid ($\omega_b \propto \frac{1}{h_x \text{ or } h_y}$) and (iii) smaller convection time scale compared to diffusion due to high Re. Exactly opposite effects are noted for the steady flow case at low Re. The results in the left panel do not suffer from any instability for the computed times. This corroborates very well with the analysis of linear 2D CDE in Fig. 32, where anti-diffusion is seen for higher Peclet number whereas for the lower value case it is absent.

The scale selection for error growth is determined by the FFT plots of the numerical solution shown in Fig. 34. For the computation with $Pe = 0.1451$, the omnipresent background errors (due to round-off, truncation errors) are amplified in a region close to $((k_x h_x, k_y h_y) = (\pi, \pi))$ and corresponds exactly with the observations of the property chart. The presence of high wavenumbers with non-negligible amplitudes is the reason for the grid-scale oscillations noted in Fig. 33.

In the next case discussed here, focusing due to errors from both convection and diffusion term discretizations is demonstrated. This is done using a single simulation employing a uniform grid of size (525×525) in the (x, y)-plane. A time step of $\Delta t = 2.8626 \times 10^{-3}$ is chosen and it corresponds to a fixed Peclet number of $Pe = 0.0786$. The CFL number based on reference freestream speed is $Nc = 1.5$.

Results are presented in Fig. 35 with the vorticity contours in the left panel and its corresponding FFT contours in the right panel, respectively. The times $t = 249.7615$ and 317.2901 correspond to an intermediate state and just before solution blow-up, respectively. Wave packet like oscillations in vorticity contours develop and remain just below the top lid (location indicated by the red arrow) until the solution blows up. An instantaneous snapshot of these oscillations is displayed in the bottom panel. These oscillations are unphysical (absent in non-focused solution) and their amplitudes grow with time eventually leading to blow-up. This is attributed to the focusing of error due to anti-diffusion for the stencils at the observed locations and is shown next.

It is noted that the location of the non-physical wave packet corresponds to the grid line just below the top wall. Hence, property charts of numerical diffusion $\frac{\alpha_N}{\alpha}$ are plotted in Fig. 36 for the stencil at a representative nodal location $(263, 524)$, which is the mid point of line. Property charts show anti-diffusion appearing for $Nc \simeq 1.3$. The instability appears near the top right and it progresses vertically towards the bottom increasing in strength, as Nc increases. For the NSE results, the location(s) with $Nc > 1.3$ in the computational domain is shown to be near the top boundary as indicated by a red arrow in Fig. 37. This is expected as the CFL numbers are higher near the moving lid. It is noted that the location with $Nc > 1.3$ coincides with the observed unphysical wave packet thus establishing anti-diffusion to the genesis of the unphysical wave packet and solution break down.

The property charts indicate the scale selection for error to originate from the region $k_x h_x \in (2.2, 2.5)$. This is confirmed from the FFT plots in Fig. 35 where the instability is in the same region as shown by the property charts. The scales are to be contrasted with the previous case where focusing is seen to occur at wavenumbers close to the Nyquist limit. It should be noted that focusing occurs for this case as there is a sustained value of $Nc > 1.3$. However, it is not clear if this situation is guaranteed for all unsteady systems. Nevertheless, the present results indicate that for finite difference solution of any unsteady system, the use of appropriate property charts will always determine focusing.

In the previous two cases, focusing was shown for the NCCD scheme whose middle stencil is a central scheme. In the final case, focusing is also demonstrated for upwind discretization of the convection terms. A uniform grid of size (2001×2001) is chosen and a time step of $\Delta t = 2.65 \times 10^{-4}$ is employed. These parameters fix the Pe and Nc values to be 0.106 and 0.5, respectively.

From the analysis results in Fig. 31, the chosen parameters can cause focusing. This is confirmed from the vorticity contour plots (not shown). To illustrate the scale selection for error amplification, property charts of $\frac{\alpha_N}{\alpha}$ are plotted for two values of Nc. For $Nc = 0.55$, we note anti-diffusion occurring near the top left corner. The scale selection of error for NSE results can be confirmed from the contour plots of FFT of computed vorticity shown in Fig. 39. The simulation eventually blows up at $t = 1.69653$ due to focusing.

These results conclusively demonstrate a one-to-one correspondence of the solution of the NSE with the predictions from the GSA of the 2D linear CDE. This unequivocally proves that the focusing phenomenon for NSE, is due to a linear instability and not due to nonlinear instability mechanism as conjectured in [102,114]. Such an analysis based on a linear CDE to explain the error dynamics of numerical methods for the nonlinear NSE has not been reported before. This demonstrates the accuracy and utility of the GSA.

8.3. Focusing due to time integration with three-time level methods

In this section, we will demonstrate focusing phenomenon using a three-time level time integration method for the solution of the NSE. We will use the well-established $NCCD$ scheme for spatial discretization to solve the flow inside a square lid-driven cavity for sub-critical and super-critical Reynolds numbers. Contrary to the previously reported solutions with polygonal vortices [70,115], here the solution breaks down after a finite time due to focusing of the physical and numerical modes of the AB_2 method. Preliminary investigation of this focusing for three time-level methods based on the 1D CE is reported in [45]. Here, we will extend the exercise to the 2D NSE based on GSA of the 2D CDE.

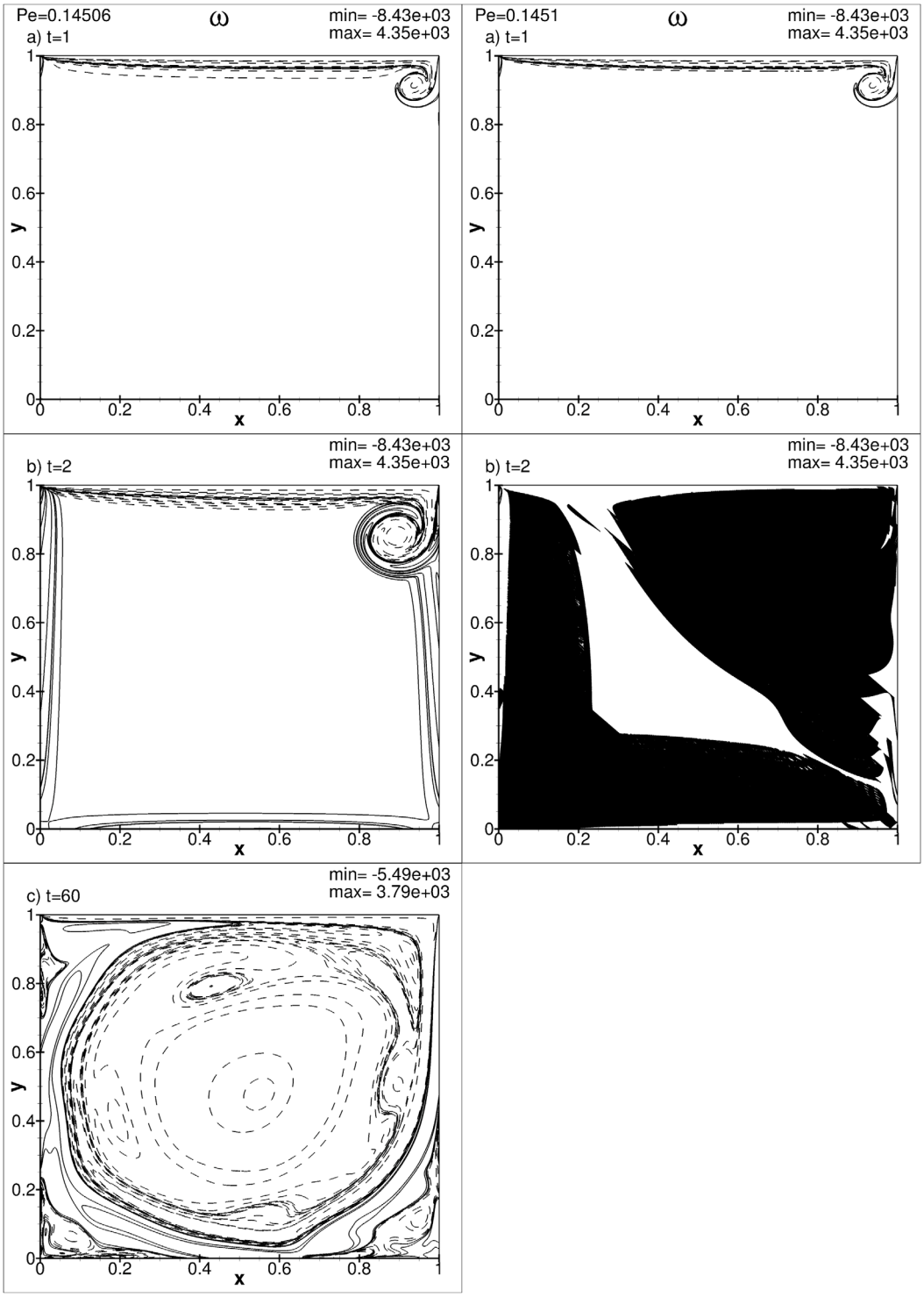

Fig. 33. Comparison of vorticity contours for the two simulations of LDC problem for Re = 10,000 at the indicated times. (a) Left panels show contours for the case with $Pe = 0.14506$ and (b) Right panels display contours for the simulation with $Pe = 0.1451$, respectively.

In Fig. 40, the unstable regions in the (Nc, Pe)-plane are marked by the dark patches for the $AB_2 - NCCD$ scheme by solving the 2D CDE with different wave propagation angles θ. In the left frames, these are shown for the physical mode of AB_2. Contrary to the $RK_4 - NCCD$ scheme in Fig. 30, there is no clear demarcation on the basis of a critical Pe for the physical mode. In the right frames, the stability map is shown for the numerical mode of AB_2. Here, a critical $Pe_{cr} = 0.05$ can be clearly seen for the different values of θ considered. We will use this figure to identify two sets of numerical parameters: (i) when the physical mode is unstable and the numerical mode is stable $(Nc = 0.4, Pe = 0.01, G_{phys} = 1.74, G_{num} = 0.42)$ (ii) when the physical mode is stable and the numerical mode is unstable $(Nc = 0.2, Pe = 0.055, G_{phys} = 1, G_{num} = 1.075)$. Next, we will solve the 2D NSE for the lid-driven cavity using these (Nc, Pe) combinations to demonstrate: (i) the direct utility of the GSA of 2D CDE for the 2D NSE, (ii) focusing phenomena in three time-level methods either due to an unstable

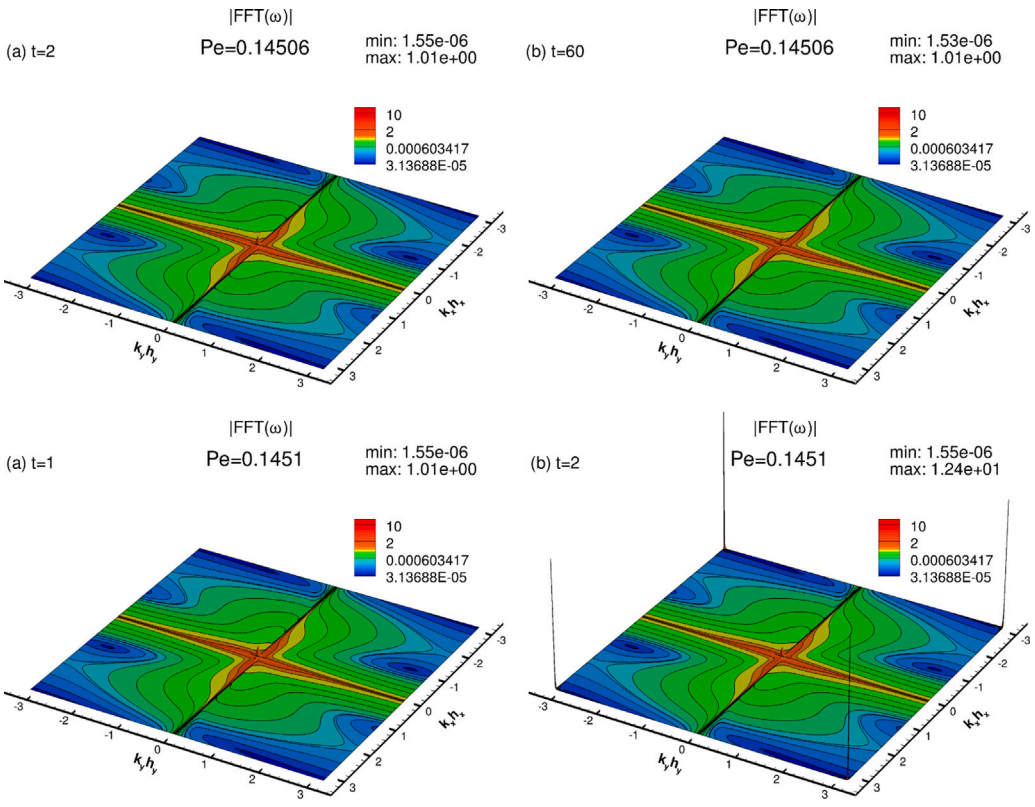

Fig. 34. Comparison of 2D FFT amplitude contours of vorticity for the two simulations of LDC problem for $Re = 10{,}000$ at the indicated times. Top panels display contours for the simulation with $Pe = 0.14506$ and bottom panels display contours for the simulation with $Pe = 0.1451$.

physical or numerical mode (iii) earlier solution breakdown due to unstable numerical mode than physical one.

In Fig. 41, solution of flow inside lid-driven cavity for super-critical $Re = 12{,}000$ using $AB_2 - NCCD$ shows the focusing mechanism due to the unstable physical mode of AB_2. In frames (a) to (d), $(Nc = 0.4, Pe = 0.01)$ are chosen such that the physical mode of AB_2 is unstable while the numerical mode is stable. From the vorticity contours in frames (a) to (d) and the time-series, it is clear that the solution undergoes focusing of error leading to catastrophic breakdown beyond $t = 6$. For the vorticity contours in frames (e) to (h) and the corresponding time-series, $(Nc = 0.025, Pe = 0.01)$ are chosen such that both modes of AB_2 are numerically stable. Using this configuration, the solution continues to be indefinitely stable. At $t = 1000$, coherent vortical structures, representative of such super-critical Re, are visualized. It should be noted that for computing the diffusion term of the NSE here, $NCCD$ is used twice instead of the traditional route of using CD_2 [70]. This has been done to ensure an efficient comparison with the stability criterion shown in Fig. 40.

In Fig. 42, solution of flow inside lid-driven cavity for sub-critical $Re = 5000$ using $AB_2 - NCCD$ shows the focusing mechanism due to the unstable numerical mode of AB_2. In frames (a) to (d), $(Nc = 0.2, Pe = 0.055)$ are chosen such that the numerical mode of AB_2 is unstable while the physical mode is stable. From the vorticity contours in frames (a) to (d) and the time-series, it is clear that the solution undergoes focusing of error leading to catastrophic breakdown at early time of $t = 0.05$. Although the numerical amplification factor corresponding to the unstable numerical mode ($G_{num} = 1.075$) is lesser than that for the physical mode in Fig. 41 ($G_{phys} = 1.74$), the solution undergoes earlier breakdown here. This suggests that for the AB_2 method, the unstable numerical mode has a more dire consequence on the stability of solution than the physical mode. Vorticity contours in frames (e) to (h) and corresponding time series are evaluated using $(Nc = 0.2, Pe = 0.05)$ for which both physical and numerical modes of AB_2 are stable. Here, the solution can be computed indefinitely and flow features captured match well with prior computations at this sub-critical Re. Thus, the localized error growth in both Figs. 41 and 42 eventually contaminates the entire flow field and leads to solution breakdown, as predicted by the numerical property chart in Fig. 40.

8.4. Remedy for focusing

The numerical instability arising due to anti-diffusion for the incompressible NSE from the solution of flow inside a LDC has been demonstrated using NCCD and Kuwahara's schemes. A one-to-one correspondence is noted between the analysis and the results of NSE for the scale selection of errors for numerical instability. If all the scales displaying instability are attenuated/removed then focusing can be eliminated. Hence, filtering is an ideal solution to cure focusing.

In computing, filtering has been used to perform multiple functions ranging from numerical stabilization at high wavenumbers arising due to highly stretched meshes and boundary conditions [116–118] to alleviating nonlinear instabilities due to aliasing error [1] and performing LES without the need for a sub-grid scale (SGS) model [119,120]. Filtering is investigated here as a cure for the focusing phenomenon arising due to anti-diffusion. The presented simulations of NSE establish focusing phenomenon as a linear instability. Therefore, by studying the effects of filtering on the model linear 2D CDE will help address the question concerning its utility in removal of focusing. The methodology is firmly established by performing simulations of NSE with filtering.

8.4.1. GSA of 2D CDE with filtering

Here, we perform GSA using the concepts and methodology introduced in [120] for analysis and development of new filters for LES and detached eddy simulation (DES). As 2D filtering is implemented as a post-processing operation, it alters the original numerical amplification

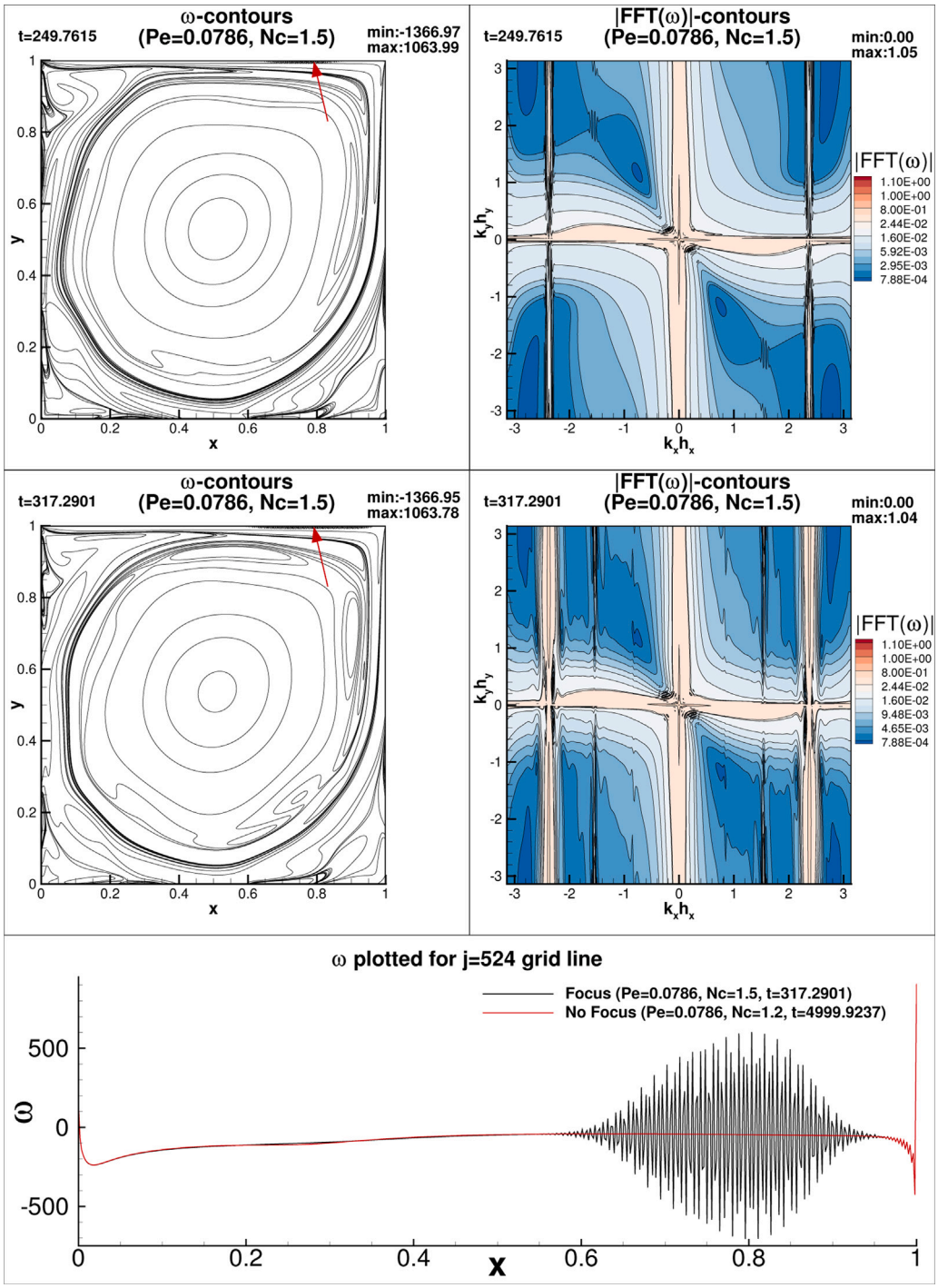

Fig. 35. Plot showing the vorticity contours (left) and FFT of vorticity contours (right) at the indicated time instants for the simulation of LDC problem for $Re = 10{,}000$. The red arrow indicates the location of the unphysical wave packet. Bottom panel shows the vorticity distribution plotted for the grid line immediately below the top lid for the cases of with and without focusing. Note the unphysical wave packet like appearance of the distribution.

factor as

$$\hat{G}_N(kh) = G_N(kh) \times TF(kh) \tag{143}$$

where $\hat{G}$ is the numerical amplification factor incorporating the effects of filtering, TF is filter's transfer function and right hand side is a simple multiplication. Knowing the TF of the filter, the numerical amplification factor, and all other numerical properties, viz. $\frac{\hat{\alpha}_N}{\alpha}$, $\frac{\hat{c}_N}{c}$, $\frac{(\hat{V}_{gx})_N}{c_x}$ and $\frac{(\hat{V}_{gy})_N}{c_y}$ can be obtained. The hat superscript denotes quantities with the inclusion of filtering. Brief details regarding the determination of TF and its interpretation readers are present in [1,120].

For the current demonstration, a second order, 2D filter stencil developed in [100] is adopted and is given by,

$$\hat{u}_{m,n} + \gamma(\hat{u}_{m-1,n} + \hat{u}_{m+1,n} + \hat{u}_{m,n-1} + \hat{u}_{m,n+1}) = \sum_{i=0}^{1} \frac{a_i}{2}(u_{m\pm i,n} + u_{m,n\pm i}) \tag{144}$$

where m, n are the nodal indices, quantities with the hat ($\hat{}$) superscript denote the filtered quantities, γ denotes the strength of filtering in the range $\gamma \in [-0.25, 0.25)$, and $a_0 = \frac{1}{2} + 2\gamma$, $a_1 = \frac{1}{4} + \gamma$.

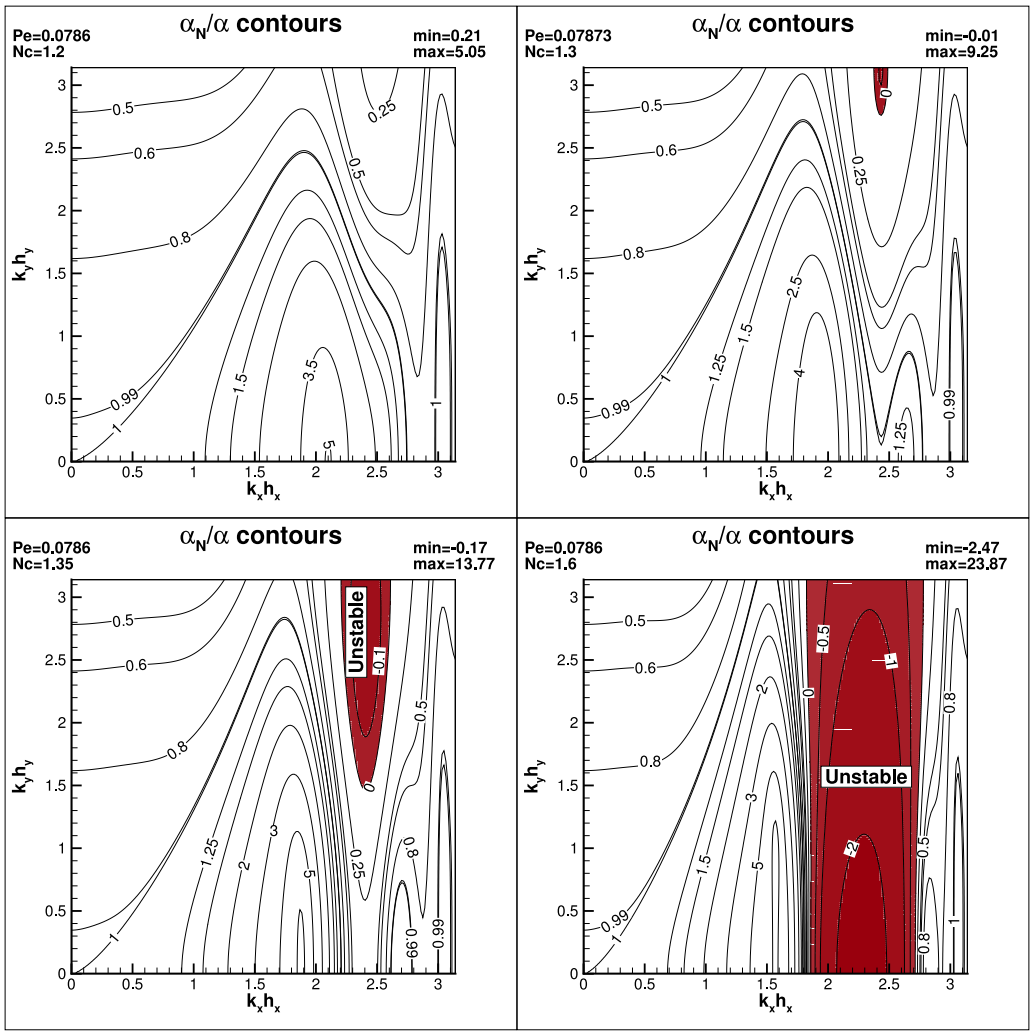

Fig. 36. Numerical diffusion ($\frac{\alpha_N}{\alpha}$) contours for the RK_4-NCCD scheme for the 2D CDE. The contours are plotted for the location which corresponds to a mid node in the x-direction and the nearest boundary node in the y-direction. The space time properties are indicated by the CFL (Nc) and Pe values with $AR = 1$, $\theta = 0°$, respectively.

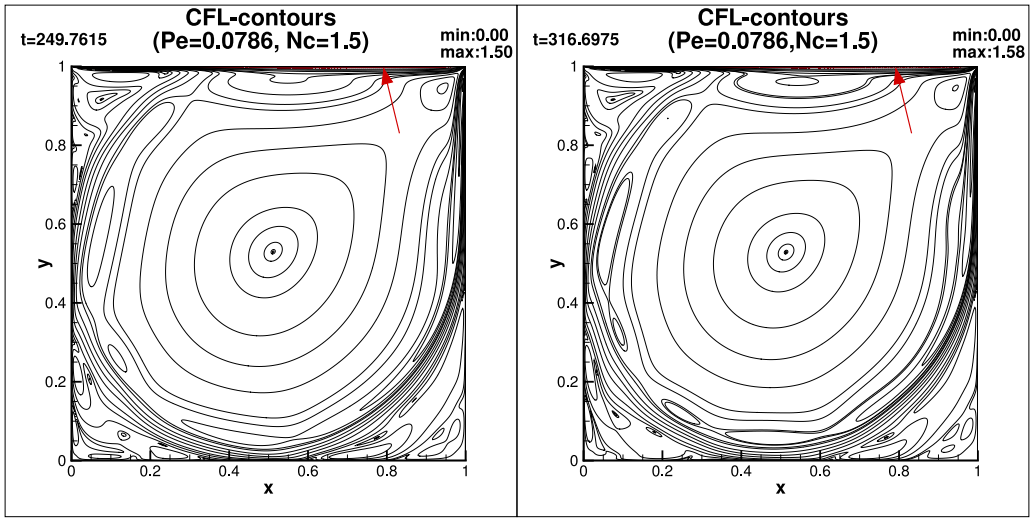

Fig. 37. Local CFL values in the computational domain at the indicated times for the simulation of LDC problem for $Re = 10{,}000$. The non-dimensional Pe and Nc values for this simulation are 0.0786 and 1.5, respectively. Red arrow shows the location of the maximum CFL in the computational domain.

The filtering operation for the full domain can be expressed in the form $[A_f]\{\hat{u}\} = [B_f]\{u\}$ where $[A_f]$, $[B_f]$ are matrices as determined by the filter stencil. It should be noted that the variables at boundaries are not filtered. From this relation, the filtered quantity can be determined

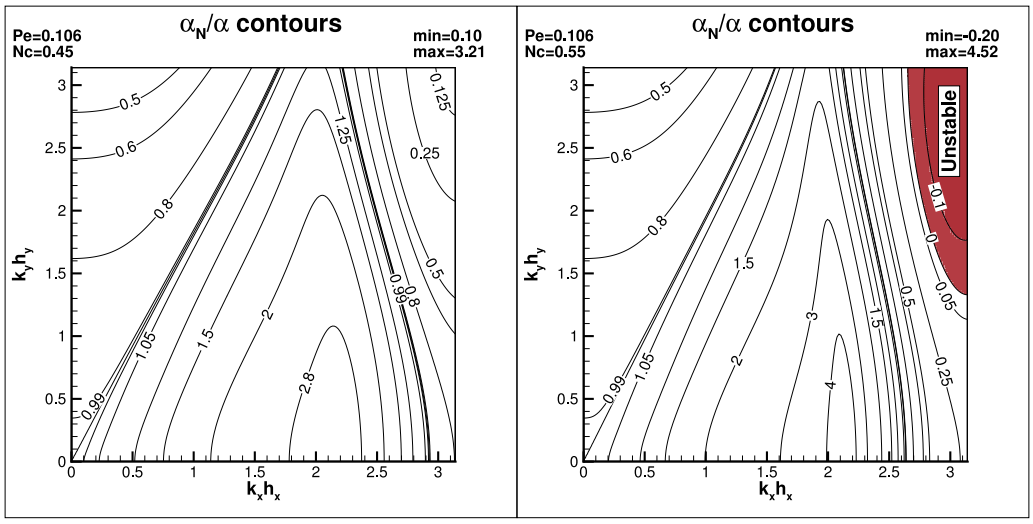

Fig. 38. Numerical diffusion-$\frac{\alpha_N}{\alpha}$ contours for the RK_4-UD_3-CD_2 scheme for the 2D CDE. The space time properties are indicated by the CFL (Nc) and Pe values with $AR = 1$, $\theta = 0°$, respectively.

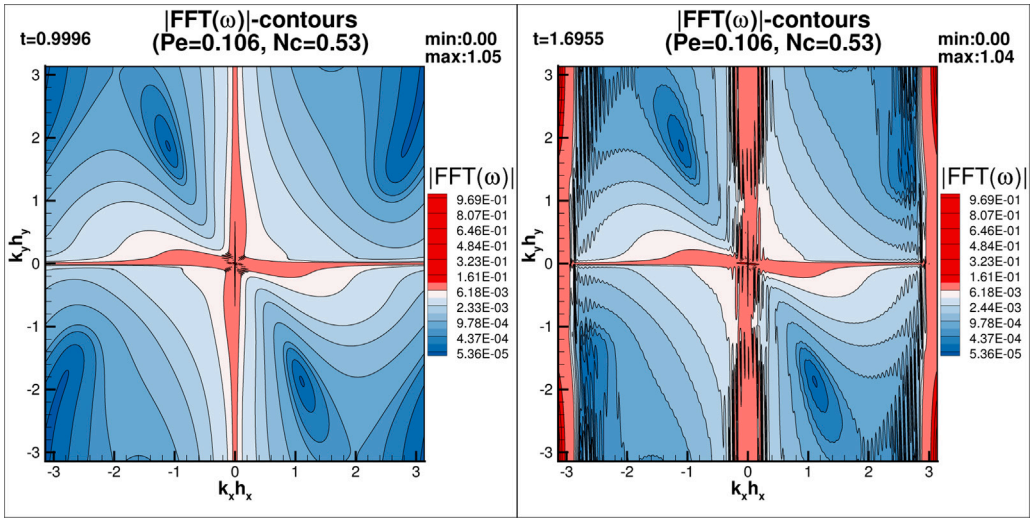

Fig. 39. Plot showing the FFT of vorticity contours at the indicated time instants for the simulation of LDC problem for $Re = 10{,}000$ using RK_4-UD_3-CD_2 scheme.

as $\{\hat{u}\} = [C_f]\{u\}$ with $[C_f] = [A_f]^{-1}[B_f]$ and therefore, the transfer function at a node (m, n) can be determined as

$$TF_{m,n}(k_x h_x, k_y h_y) = C_{f\,N_y(n-1)+m,j}\; e^{i((m_x-m)k_x h_x + (n_y-n)k_y h_y)} \tag{145}$$

where N_x and N_y are the number of grid points in x- and y-directions, respectively. The integer variables m_x and n_y are the grid indices of the point whose corresponding variable is u_j i.e., $n_y = \frac{j}{N_y} + 1$ and $m_x = j - ((n_y - 1) \times N_y)$. These are determined by noting that the vector u is stored as $u = [u_{1,1} u_{2,1} u_{3,1} \ldots u_{N_x,N_y}]^T$.

With the transfer function determined, the numerical properties for the filtered 2D CDE are evaluated using the equations given below.

$$\tan(\hat{\phi}_N) = -\left(\frac{(\hat{G}_N)_{Img}}{(\hat{G}_N)_{Real}}\right) \quad \text{where} \quad \hat{\phi}_N = \left(\sqrt{{k_x}^2 + {k_y}^2}\right) \hat{c}_N \Delta t$$

$$\frac{\hat{c}_N}{c} = -\left[\frac{\hat{\phi}_N}{N_{cx}(k_x h_x) + N_{cy}(k_y h_y)}\right]$$

$$\frac{(\hat{V}_{gx})_N}{c_x} = \frac{1}{N_{cx}} \frac{\partial \hat{\phi}_N}{\partial (k_x h_x)}$$

$$\frac{(\hat{V}_{gy})_N}{c_y} = \frac{1}{N_{cy}} \frac{\partial \hat{\phi}_N}{\partial (k_y h_y)}$$

$$\frac{\hat{\alpha}_N}{\alpha} = -\frac{\ln |\hat{G}_N|}{[Pe_x (k_x h_x)^2 + Pe_y (k_y h_y)^2]} \tag{146}$$

Due to the central nature of the filter stencil (the transfer function is real), the group velocities are unaffected. Therefore, the present filtering operation affects only α_N, while preserving the numerical dispersion relation of the unfiltered case.

The property charts are shown in Fig. 43, comparing the unfiltered and filtered cases for the critical value of $Pe_{cr} = 0.1451$ for the RK_4-NCCD scheme. The unfiltered case is shown in the left panel and the 2D second order filter with $\gamma = 0.248$ with filtering performed at every $25\Delta t$ is shown in the right panel. The figure clearly shows the advantage of filtering by noting that it removes the numerical instability associated with anti-diffusion, as seen in the contour plots of $\frac{\hat{\alpha}_N}{\alpha}$. This is due to the fundamental nature of filtering operation, which is the attenuation/removal of high wavenumber components from the numerical solution.

To establish whether filtering eliminates focusing or not, the simulation case of flow inside LDC for $Re = 10{,}000$ with $Pe = 0.1451$ is computed with the filtering implemented. This case displayed absolute numerical instability as noted in the earlier subsection. The second order 2D filter stencil with $\gamma = 0.248$ is employed and ω field is filtered

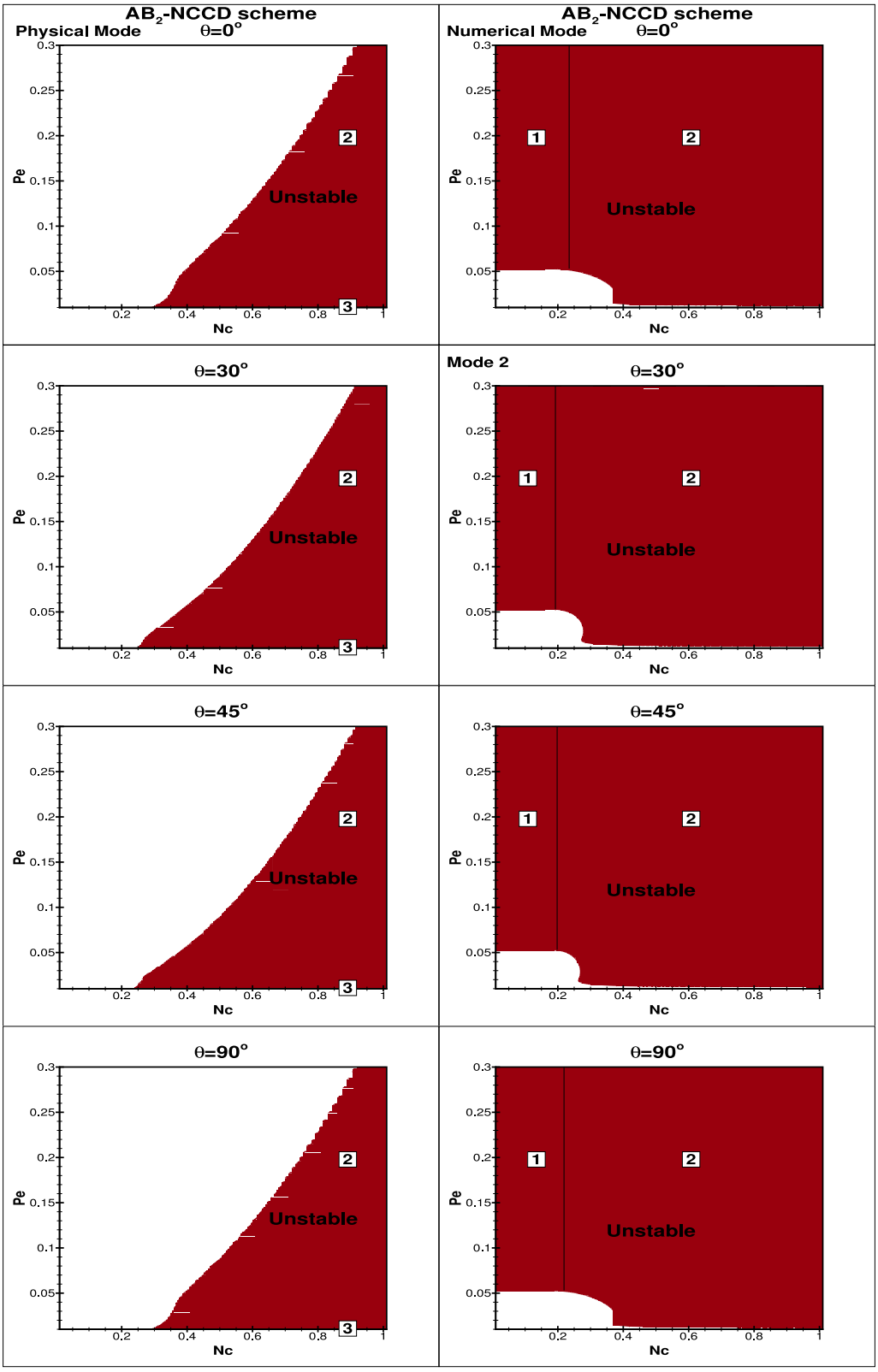

Fig. 40. Influence of wave propagation angle θ of the 2D CDE on the critical Peclet number for $AB_2 - NCCD$ scheme with grid aspect ratio AR = 1. Unstable regions are indicated by the dark patches in the (Nc, Pe)-plane. Left frames correspond to the physical mode while right frames correspond to the numerical mode of AB_2.

after every $25\Delta t$. Filtering is implemented here as a post-processing operation i.e. the governing equations are not filtered. This methodology has minimal impact on the speed of computations as one does not require solution of Eq. (144) using expensive iterative schemes. For the current demonstration, a point Jacobi method is employed although one can use the same Bi-CGSTAB method employed for solution of stream function equation, Eq. (141). It is noted that for the point Jacobi method, converged solutions are obtained within 5 iterations.

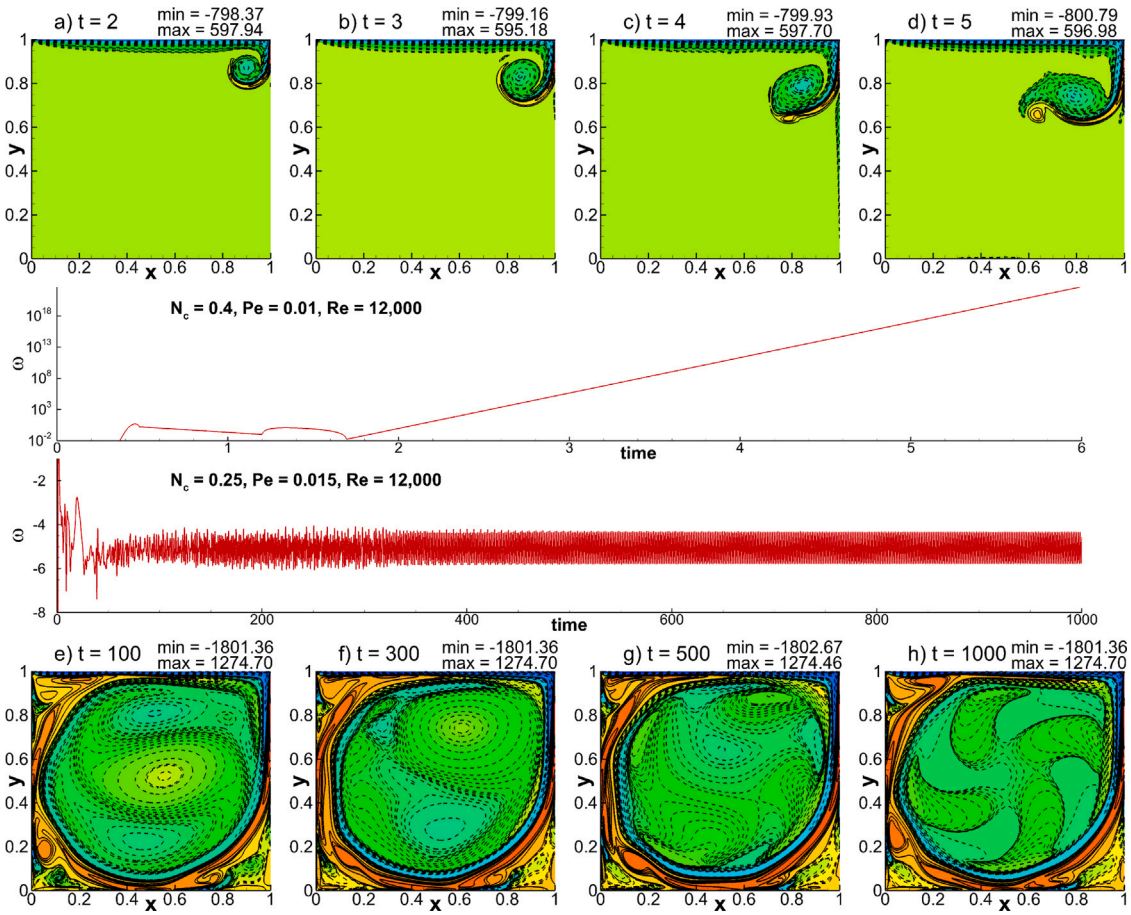

Fig. 41. Vorticity contours for the lid-driven cavity problem solved using $AB_2 - NCCD$ for $Re = 12{,}000$ with $Nc = 0.4$ and $Pe = 0.01$ in frames (a) to (d) and with $Nc = 0.25$ and $Pe = 0.015$ in frames (e) to (h). The first set of (Nc, Pe) corresponds to the unstable physical mode of AB_2 in Fig. 40 while the second set are for a numerically stable set up. Focusing phenomenon is displayed via the time-series of vorticity for the unstable (Nc, Pe) beyond $t = 6$.

In Fig. 44, the vorticity contours are shown for the filtered case with $Pe = 0.1451$ in the right side panels. The benefit of filtering can be confirmed by noting the suppression of filtering thereby validating the GSA approach. Without filtering the solution blows up as $Pe \geq Pe_{cr}$. To evaluate the quality of solution with filtering, its results are compared with the unfocused case ($Pe = 0.14506$) with the latter serving as a reference. Apart from a phase shift between the vortical structures one notes the filtered solution to display similar structures.

The effect of increasing filtering frequency on the accuracy of numerical solution is noted from the results in Fig. 45. Comparison with the reference unfocused case at $t = 6$ shows very good match. This is due to reduced numerical diffusion/damping of the filtering on the solution. For all the computed cases with filtering, the frequencies are chosen using GSA such that focusing is just eliminated. The presented results are for cases with weak instabilities ($|G_N| \simeq 1 + \epsilon$) i.e. near the critical values for focusing. However, filtering should also function effectively for stronger instabilities as we have noted the removal of instability at Peclet numbers higher than the critical value (not shown here). In such cases, filtering has to be performed at frequent intervals (lower frequencies) as a consequence of the stronger instability.

In this discussion, a simple filtering strategy for eliminating focusing is presented. One can explore other strategies of filtering, particularly the adaptive filtering [121], which will yield a significant reduction in computational efforts in addition to eliminating the focusing problem.

9. Linearized rotating shallow water equations

The linearized rotating shallow water equations (LRSWE) based on single-layer approximation [122–124] are extensively used for the numerical modeling of atmosphere and ocean dynamics. The LRSWE representing a dispersive system of hyperbolic conservation laws are given as

$$\frac{\partial u}{\partial t} - f v + g\frac{\partial \eta}{\partial x} = 0 \tag{147}$$

$$\frac{\partial v}{\partial t} + f u + g\frac{\partial \eta}{\partial y} = 0 \tag{148}$$

$$\frac{\partial \eta}{\partial t} + H\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) = 0 \tag{149}$$

where, u and v are the velocity components in x- and y-directions, respectively; η is the time dependent surface-elevation from the mean level at $z = 0$; f is the Coriolis frequency; g is the acceleration of gravity and H is the mean depth. The system of first-order Eqs. (147)–(149) can also be reduced to a single equation by eliminating u and v, as

$$\frac{\partial}{\partial t}\left[\frac{\partial}{\partial t^2} + f^2 - gH\nabla^2\right]\eta = 0 \tag{150}$$

where, $\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$, is the two-dimensional Laplacian operator. The dispersion relation based on bilateral Fourier–Laplace transform for Eqs. (147)–(149) or Eq. (150) is given as [125],

$$\omega\left(\omega^2 - c^2\left|\vec{K}\right|^2 - f^2\right) = 0 \tag{151}$$

where, $\vec{K} = \hat{i}k_x + \hat{j}k_y$ and $c = \sqrt{gH}$, with k_x and k_y denoting wavenumber components in x- and y- directions, respectively. Roots of the Eq. (151) are given as

$$\omega_1 = 0\,, \quad \omega_{2,3} = \pm\sqrt{f^2 + c^2\left|\vec{K}\right|^2} \tag{152}$$

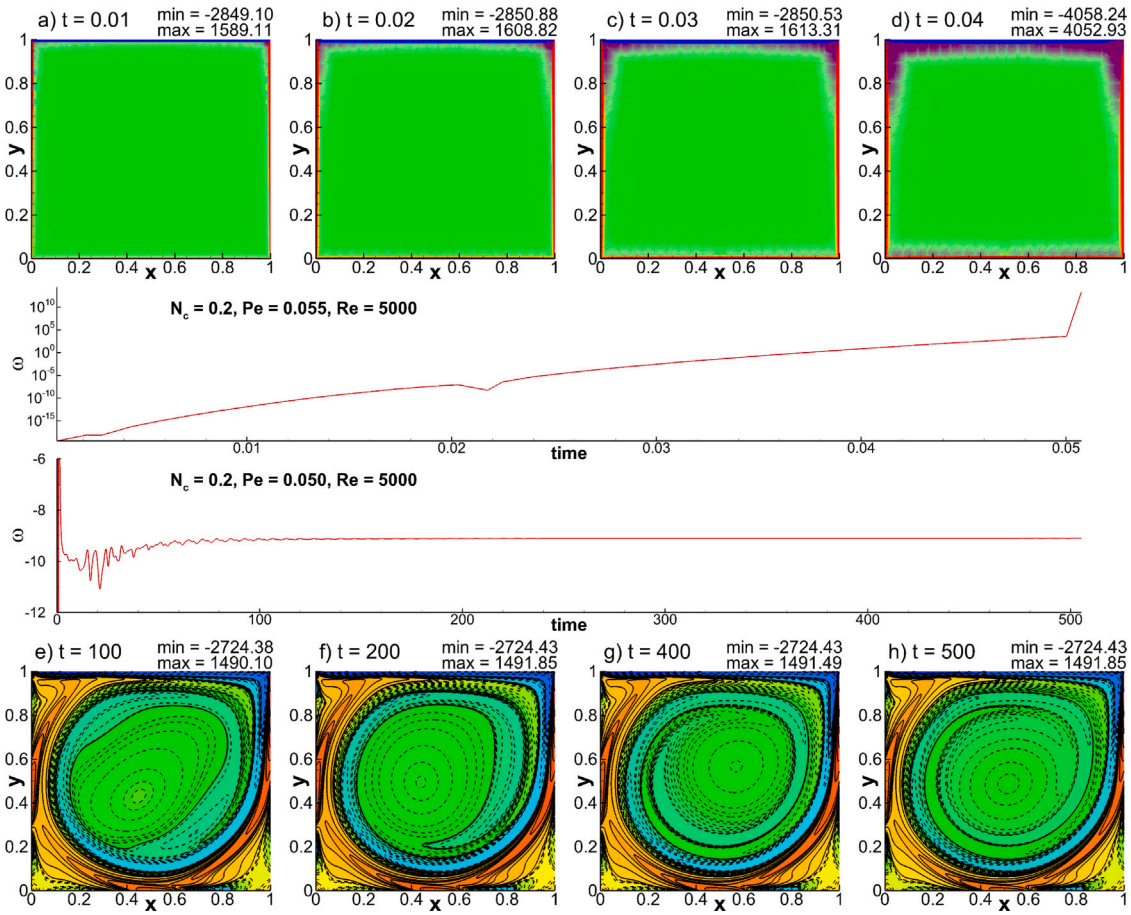

Fig. 42. Vorticity contours for the lid-driven cavity problem solved using $AB_2 - NCCD$ for $Re = 5000$ with $Nc = 0.2$ and $Pe = 0.055$ in frames (a) to (d) and with $Nc = 0.2$ and $Pe = 0.05$ in frames (e) to (h). The first set of (Nc, Pe) corresponds to the unstable numerical mode of AB_2 in Fig. 40 while the second set are for a numerically stable set up. Focusing phenomenon is displayed via the time-series of vorticity for the unstable (Nc, Pe) beyond $t = 0.05$.

where, ω_1 corresponds to the geostrophic mode, while $\omega_{2,3}$ correspond to the inertia-gravity modes [125]. Expressions for the group velocity components in x- and y-directions and phase speeds in x- and y-directions for the inertia-gravity modes are obtained from Eq. (152) as [36,110,125],

$$(V_{gx})_{2,3} = \frac{\partial \omega_{2,3}}{\partial k_x} = \pm \frac{k_x c^2}{\sqrt{f^2 + c^2 |\vec{K}|^2}}, (V_{gy})_{2,3} = \frac{\partial \omega_{2,3}}{\partial k_y} = \pm \frac{k_y c^2}{\sqrt{f^2 + c^2 |\vec{K}|^2}} \tag{153}$$

$$(c_{ex})_{2,3} = \frac{\omega_{2,3}}{k_x} = \pm \frac{\sqrt{f^2 + c^2 |\vec{K}|^2}}{k_x}, \ (c_{ey})_{2,3} = \frac{\omega_{2,3}}{k_y} = \pm \frac{\sqrt{f^2 + c^2 |\vec{K}|^2}}{k_y} \tag{154}$$

Moreover, resultant group velocity for the inertia-gravity modes are obtained from Eq. (154), as $V_g = \sqrt{V_{gx}^2 + V_{gy}^2} = (c^2 |\vec{K}|) / \sqrt{f^2 + c^2 |\vec{K}|^2}$, which makes an angle, $\theta_{ex} = \tan^{-1}(V_{gy}/V_{gx})$ with the positive x-axis.

9.1. Dispersion analysis of the linearized rotating shallow water equations

Here, the dispersion analysis of LRSWE on Arakawa grids is performed by considering space–time discretizations together, as in [30,31,110,126]. The analysis presented in this section is valid for an explicit two-level time integration method. Using vector notations $Z = [u, v, \eta]^T$, Eqs. (147)–(149) can be rewritten as

$$\frac{\partial Z}{\partial t} + [A] Z + [B] \frac{\partial Z}{\partial x} + [C] \frac{\partial Z}{\partial y} = 0 \tag{155}$$

with $A = \begin{bmatrix} 0 & -f & 0 \\ f & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$, $B = \begin{bmatrix} 0 & 0 & g \\ 0 & 0 & 0 \\ H & 0 & 0 \end{bmatrix}$ and $C = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & g \\ 0 & H & 0 \end{bmatrix}$

Similar to non-dispersive case, here also using Fourier–Laplace transform vector of unknowns Z can be represented as

$$Z = \int \int \hat{Z} \, e^{i (k_x x + k_y y)} \, dk_x \, dk_y$$

where, $\hat{Z} = [\hat{U}, \ \hat{V}, \ \hat{E}]^T$, where $\hat{U}$, $\hat{V}$ and $\hat{E}$ denote the bilateral Fourier–Laplace transforms [26,127] of u, v and η, respectively. Furthermore, spatial discretization schemes for the first-order derivatives can also be represented as

$$\begin{aligned} \frac{\partial Z}{\partial x} &= \int \int i \, (k_x)_{eq} \, \hat{Z} e^{i (k_x x + k_y y)} \, dk_x \, dk_y \\ \frac{\partial Z}{\partial y} &= \int \int i \, (k_y)_{eq} \, \hat{Z} e^{i (k_x x + k_y y)} \, dk_x \, dk_y \end{aligned} \tag{156}$$

where $(k_x)_{eq}$ and $(k_y)_{eq}$ are equivalent wavenumbers in obtaining the first-order derivatives in x- and y-directions, respectively. Except one, all grids proposed in Mesinger & Arakawa [128] are of staggered type. Thus, to evaluate relevant Coriolis component one has to interpolate the grid variables to the location where it is required. Interpolation of unknowns effectively changes the value of the Coriolis parameter from f to f_{eq}. Thus, upon using spatial discretization and interpolation schemes in Eq. (155), we have the following

$$\frac{\partial \hat{Z}}{\partial t} + [D] \, \hat{Z} = 0 \tag{157}$$

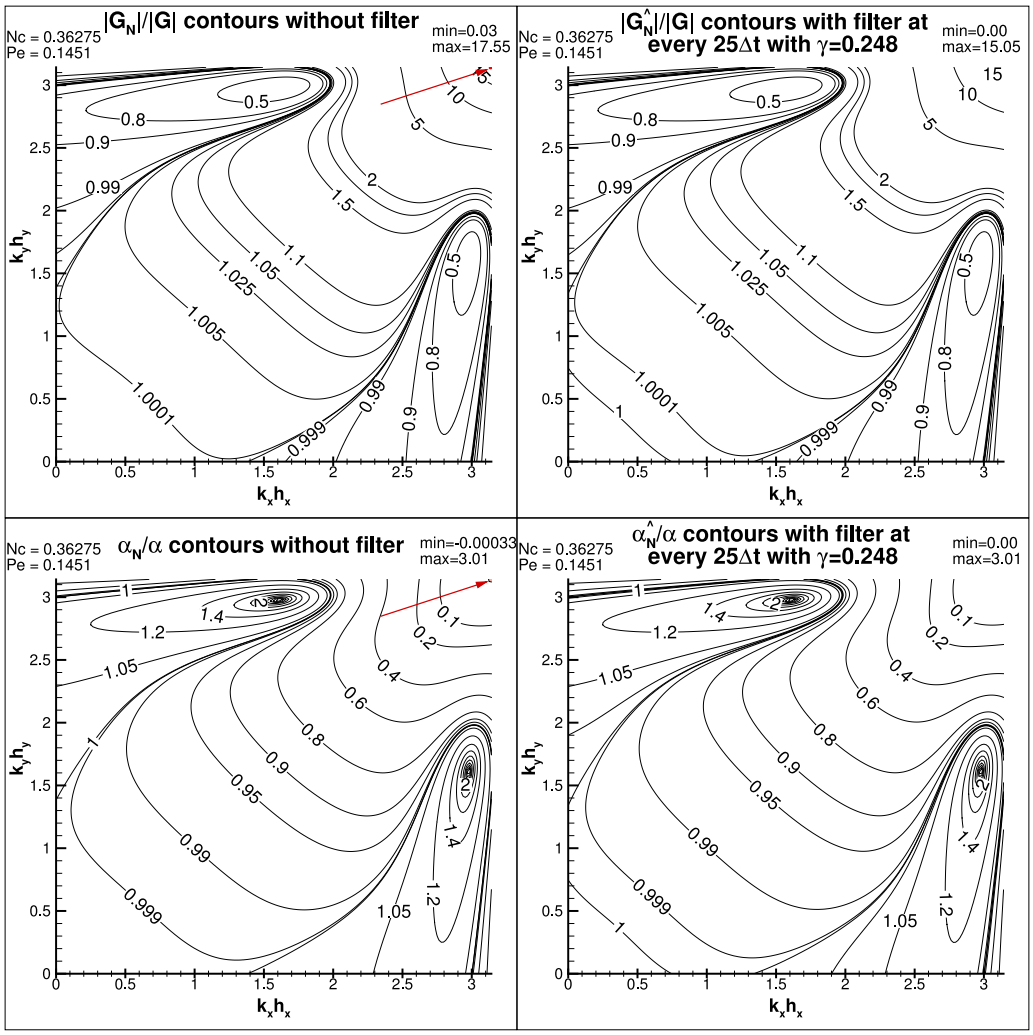

Fig. 43. Normalized numerical amplification factor ($\frac{|G_N|}{|G|}$) and numerical diffusion coefficient ($\frac{\alpha_N}{\alpha}$) shown for the RK_4-NCCD scheme for the 2D CDE. Left panels are for unfiltered case with $Pe_x = 0.1451$, $Nc_x = 0.375$, $AR = 1$, $\theta = 45°$ and right panels are for the unfiltered case with 2D filter with $\gamma = 0.248$ applied every $25\Delta t$, respectively.

$$\text{with } [D] = \begin{bmatrix} 0 & -f_{eq} & i(k_x)_{eq}\, g \\ f_{eq} & 0 & i(k_y)_{eq}\, g \\ i(k_x)_{eq}\, H & i(k_y)_{eq}\, H & 0 \end{bmatrix}$$

In particular, using RK_4 time integration method in Eq. (157) we can relate the value of $\hat{Z}$ at $(n+1)$th time-level with nth time-level value as

$$\hat{Z}^{n+1} = [P_{RK_4}]\, \hat{Z}^n \tag{158}$$

where evolution matrix is obtained as, $P_{\mathrm{RK}_4} = I - \Delta t\ D + \frac{\Delta t^2}{2!}\ D^2 - \frac{\Delta t^3}{3!}\ D^3 + \frac{\Delta t^4}{4!}\ D^4$. The modal-amplification factors, $G = [G_1, G_2, G_3]^T$ are the eigenvalues of $[P_{RK_4}]$. Also, equivalent wavenumbers for the spatial discretization of the first-order spatial derivatives can be represented as

$$(k_x)_{eq} = \frac{1}{\Delta x}\zeta(k_x\Delta x)\,, \quad (k_y)_{eq} = \frac{1}{\Delta y}\psi(k_y\Delta y)$$

where, ζ and ψ are functions of $k_x\Delta x$ and $k_y\Delta y$, with Δx and Δy denoting the mesh-widths in x- and y-directions, respectively. Denoting $a \equiv Nc_x\ \zeta(k_x\Delta x)$ and $b \equiv Nc_y\ \psi(k_y\Delta y)$, where $Nc_x = \frac{c\,\Delta t}{\Delta x}$ and $Nc_y = \frac{c\,\Delta t}{\Delta y}$ denote the CFL numbers based on mesh-widths in x- and y-directions, respectively. Using symbolic toolbox, eigenvalues of P_{RK_4} are obtained as [125],

$$G_1 = 1, \quad G_{2,3} = \delta \mp i\ \epsilon \tag{159}$$

where, $\delta = 1 - \frac{1}{2}\ \gamma^2 + \frac{1}{24}\gamma^4$; $\epsilon = \gamma - \frac{1}{6}\gamma^3$, with $\gamma = \sqrt{(a^2 + b^2 + p^2)}$ and $p = f_{eq}\Delta t$. Here, G_1 represents the geostrophic mode and $G_{2,3}$ correspond to the inertia-gravity modes. In general, $f_{eq} = fTF_x(k_x\Delta x)TF_y(k_y\Delta y)$, where, TF_x and TF_y are the Fourier transfer functions associated with the chosen interpolation scheme in x- and y-directions, respectively.

Numerical circular frequency ω_N, for a mode is calculated from,

$$\omega_N = -\frac{1}{\Delta t}\tan^{-1}\left(G_{\mathrm{imag}}/G_{\mathrm{real}}\right)$$

where, G_{imag} and G_{real} denote the imaginary and real parts of the modal-amplification factor, respectively. Using Eq. (159), numerical circular frequencies for geostrophic and inertia-gravity modes are obtained as

$$(\omega_N)_1 = 0\,, \quad (\omega_N)_{2,3} = \pm\frac{1}{\Delta t}\tan^{-1}\left(\frac{\epsilon}{\delta}\right) \tag{160}$$

using numerical circular frequency ω_N, numerical group-velocity components in x- and y-directions, V_{gNx} and V_{gNy}, for the inertia-gravity modes [125] are given as

$$V_{gNx} = \frac{\partial\omega_N}{\partial k_x}\,, \quad V_{gNy} = \frac{\partial\omega_N}{\partial k_y} \tag{161}$$

and expressions for the numerical phase speeds in x- and y-directions, c_{Nx} and c_{Ny}, are also obtained as

$$c_{Nx} = \frac{\omega_N}{k_x}\,, \quad c_{Ny} = \frac{\omega_N}{k_y} \tag{162}$$

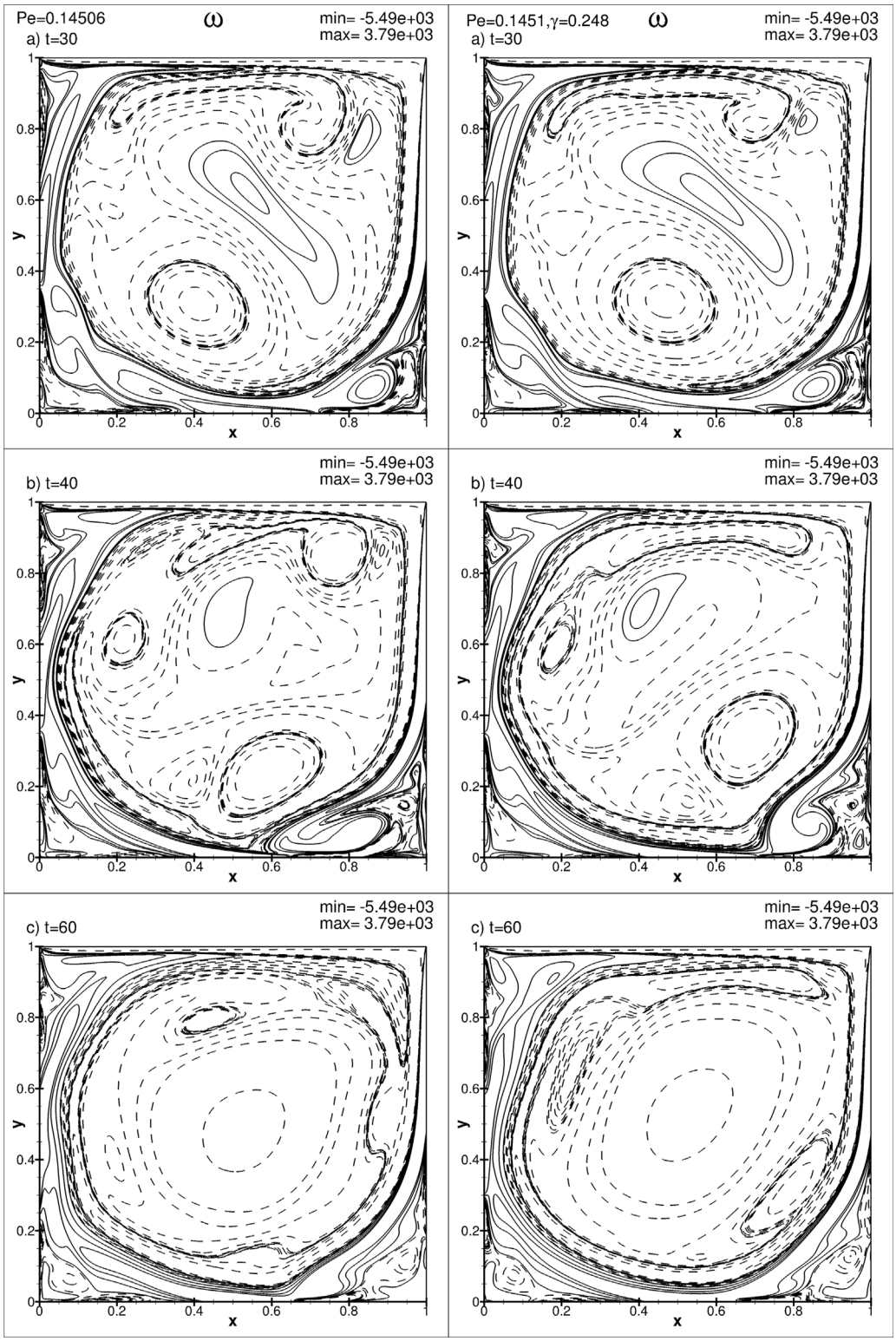

Fig. 44. Comparison of vorticity contours for $Re = 10{,}000$. The left panels are for the unfiltered sub-critical case ($Pe = 0.14506$), and, the right panels are for the case with 2D, second order filter with $\gamma = 0.248$ applied every $25\Delta t$ for the critical case $Pe = 0.1451$.

Using Eqs. (160) and (161), expressions for the numerical group velocity components in x- and y-directions for the inertia-gravity modes are obtained as

$$\left(V_{gNx}\right)_{2,3} = \pm\, c\, \frac{1}{\gamma}\, \frac{d\xi}{d\gamma} \left[a \frac{d\zeta}{d(k_x \Delta x)} + p \frac{\Delta x}{L_r} TF_y \frac{d(TF_x)}{d(k_x \Delta x)} \right]$$

$$\left(V_{gNy}\right)_{2,3} = \pm\, c\, \frac{1}{\gamma}\, \frac{d\xi}{d\gamma} \left[b \frac{d\psi}{d(k_y \Delta y)} + p \frac{\Delta y}{L_r} TF_x \frac{d(TF_y)}{d(k_y \Delta y)} \right] \tag{163}$$

where, $\xi = \tan^{-1}\left(G_{\text{imag}}/G_{\text{real}}\right)$ and $L_r = \sqrt{(gH)}/f$ is the Rossby radius. Moreover, using Eq. (163), absolute numerical group velocity V_{gN} for

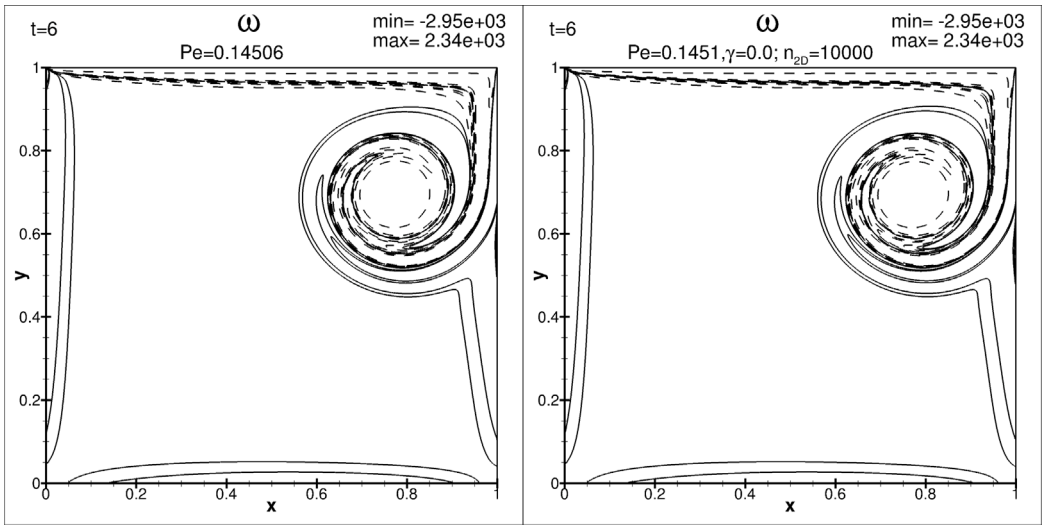

Fig. 45. Comparison of vorticity contours of the LDC problem for $Re = 10{,}000$ at $t = 6$. The left panel shows the contours for the unfiltered case with $Pe = 0.14506$. The right panel represents the solution with 2D, second order filter with $\gamma = 0$ and filtering after every $10{,}000\Delta t$ for $Pe = 0.1451$.

inertia-gravity modes and its angle of propagation θ_N are obtained as

$$V_{gN} = \sqrt{\left(V_{gNx}\right)^2 + \left(V_{gNy}\right)^2}, \quad \theta_N = \tan^{-1}\left(\frac{V_{gNy}}{V_{gNx}}\right) \tag{164}$$

Similarly, from Eqs. (160)–(162) expressions for the numerical phase speed components are obtained as

$$\left(c_{Nx}\right)_{2,3} = \pm\, c\, \frac{\zeta}{N c_x k_x \Delta x}, \quad \left(c_{Ny}\right)_{2,3} = \pm\, c\, \frac{\zeta}{N c_y k_y \Delta y} \tag{165}$$

It can be noticed that the geostrophic mode has zero phase speed and group velocity. Also, for the inertia-gravity modes, modulus of amplification factors is the same, as second inertia-gravity mode is the complex conjugate of the first inertia-gravity mode. Therefore, numerical group velocity components and phase speeds in x- and y-directions for the second inertia-gravity mode are same in magnitude, as for the first inertia-gravity mode, with change in sign.

9.2. Compact schemes for the first-order derivatives and interpolation on Arakawa meshes

For collocated A-grid, the OUCS3 scheme given in Eqs. (48) to (50) is employed. On staggered meshes, optimized staggered compact scheme (OSCS) with optimized staggered interpolation scheme for spatial discretization [125] are used. Staggered compact scheme for evaluating the first-order spatial derivative at jth node is given as [129]

$$\alpha_1 u'_{i-1} + u'_i + \alpha_1 u'_{i+1} = b_1 \frac{u_{i+3/2} - u_{i-3/2}}{3\Delta x} + a_1 \frac{u_{i+1/2} - u_{i-1/2}}{\Delta x} \tag{166}$$

Eq. (166) represents a single parameter (α_1) family of fourth-order schemes with $a_1 = \frac{3}{8}(3 - 2\alpha_1)$ and $b_1 = \frac{1}{8}(-1 + 22\alpha_1)$. For the present case we have chosen $\alpha_1 = 0.18$, as given in [125]. Moreover, use of staggered schemes also requires mid-point interpolation of unknowns. Optimized compact scheme for mid-point interpolation [125] is given as

$$\alpha_2 \hat{u}'_{i-1} + \hat{u}'_i + \alpha_2 \hat{u}'_{i+1} = b_2 \frac{u_{i+3/2} + u_{i-3/2}}{2} + a_2 \frac{u_{i+1/2} + u_{i-1/2}}{2} \tag{167}$$

where $\hat{u}_j$ denotes the interpolated values of the unknown u at the jth node. This approximation is of fourth-order accuracy if $a_2 = \frac{1}{8}(9 + 10\alpha_2)$ and $b_2 = \frac{1}{8}(-1 + 6\alpha_2)$. In the present case, the optimized value $\alpha_2 = 0.35$ [125] is used in the computations. Boundary closures for the use of Eqs. (166)–(167) for non-periodic problems are discussed next.

9.3. Boundary closures for the optimized staggered compact scheme

Boundary stencils for non-periodic problems using staggered compact scheme are derived by following the similar approach as discussed in [29] for the OUCS3 scheme. Boundary closures at full (integer) locations are obtained by using function values at half (non-integer) locations.

9.3.1. Boundary closures or full locations

For $j = 1$ node, boundary stencil is obtained as

$$u'_1 = \frac{1}{h}\left[-\frac{71}{24}u_{3/2} + \frac{47}{8}u_{5/2} - \frac{31}{8}u_{7/2} + \frac{23}{24}u_{9/2}\right] \tag{168}$$

Furthermore, for $j = 2$ node we have considered the combination of dispersive and dissipative boundary stencils, obtained as

$$u'_2 = \frac{1}{h}\left[-\frac{71}{24}u_{5/2} + \frac{47}{8}u_{7/2} - \frac{31}{8}u_{9/2} + \frac{23}{24}u_{11/2}\right] \tag{169}$$

$$u'_2 = \frac{u_{5/2} - u_{3/2}}{h} + h^3 \beta_f \left.\frac{\partial u_2^4}{\partial x^4}\right|_{\mathrm{CD}_2} \tag{170}$$

As discussed in [29], here also by taking $(1 : 2)$ blending of Eqs. (169)–(170) we have obtained the boundary stencil at $j = 2$, as

$$\begin{aligned} u'_2 = \frac{1}{3h}\Bigg[&(2\beta_f - 2)u_{3/2} - \left(\frac{23}{24} + 8\beta_f\right)u_{5/2} + \left(\frac{47}{8} + 12\beta_f\right)u_{7/2} \\ &- \left(\frac{31}{8} + 8\beta_f\right)u_{9/2} + \left(\frac{23}{24} + 2\beta_f\right)u_{11/2}\Bigg] \end{aligned} \tag{171}$$

Similarly, boundary closures at $j = N$ and $N - 1$ are obtained as

$$u'_N = -\frac{1}{h}\left[-\frac{71}{24}u_{N-1/2} + \frac{47}{8}u_{N-3/2} - \frac{31}{8}u_{N-5/2} + \frac{23}{24}u_{N-7/2}\right] \tag{172}$$

$$\begin{aligned} u'_{N-1} = -\frac{1}{3h}\Bigg[&(2\beta_{fN} - 2)u_{N-1/2} - \left(\frac{23}{24} + 8\beta_{fN}\right)u_{N-3/2} \\ &+ \left(\frac{47}{8} + 12\beta_{fN}\right)u_{N-5/2} \\ &- \left(\frac{31}{8} + 8\beta_{fN}\right)u_{N-7/2} + \left(\frac{23}{24} + 2\beta_{fN}\right)u_{N-9/2}\Bigg] \end{aligned} \tag{173}$$

As in [29], here also for the global accuracy and numerical stability we have chosen $\beta_f = -0.025$ for $j = 2$ node and $\beta_{fN} = 0.09$ for $j = N-1$ node. Boundary stencils at half (non-integer) locations are obtained using function-values at full (integer) locations, and are discussed next.

9.3.2. Boundary closures for half locations

Similar to full locations, boundary closure at $j = 3/2$ node is obtained as

$$u'_{3/2} = \frac{1}{h}\left[-\frac{23}{24}u_1 + \frac{7}{8}u_2 + \frac{1}{8}u_3 - \frac{1}{24}u_4\right] \tag{174}$$

Again for $j = 5/2$ node, we have considered the combination of dispersive and dissipative boundary stencils obtained as

$$u'_{5/2} = \frac{1}{h}\left[-\frac{23}{24}u_2 + \frac{7}{8}u_3 + \frac{1}{8}u_4 - \frac{1}{24}u_5\right] \tag{175}$$

$$u'_{5/2} = \frac{u_3 - u_2}{h} + h^3\beta_h \frac{\partial u^4_{5/2}}{\partial x^4}\bigg|_{\mathrm{CD}_2} \tag{176}$$

here also by considering $(1 : 2)$ blending of Eqs. (175)–(176), near-boundary stencil at $j = 5/2$ node is obtained as

$$u'_{5/2} = \frac{1}{3h}\left[2\beta_h u_1 + \left(-\frac{71}{24} - 8\beta_h\right)u_2 + \left(\frac{23}{8} + 12\beta_h\right)u_3 + \left(\frac{1}{8} - 8\beta_h\right)u_4 + \left(-\frac{1}{24} + 2\beta_h\right)u_5\right] \tag{177}$$

Also, boundary closures at the right boundary (downstream) are obtained as

$$u'_{N-1/2} = -\frac{1}{h}\left[-\frac{23}{24}u_N + \frac{7}{8}u_{N-1} + \frac{1}{8}u_{N-2} - \frac{1}{24}u_{N-3}\right] \tag{178}$$

$$u'_{N-3/2} = -\frac{1}{3h}\left[2\beta_{hN}u_N + \left(-\frac{71}{24} - 8\beta_{hN}\right)u_{N-1} + \left(\frac{23}{8} + 12\beta_{hN}\right)u_{N-2} + \left(\frac{1}{8} - 8\beta_{hN}\right)u_{N-3} + \left(-\frac{1}{24} + 2\beta_{hN}\right)u_{N-4}\right] \tag{179}$$

here also for the global accuracy and numerical stability we have chosen $\beta_h = -0.025$ for $j = 5/2$ node and $\beta_{hN} = 0.09$ for $j = N - 3/2$ node. Boundary stencils for the optimized compact interpolation scheme are discussed next.

9.4. Boundary closures for the optimized compact interpolation scheme

Next, boundary closures for the compact interpolation scheme are derived at full and half locations, respectively.

9.4.1. Boundary closures for full locations

Boundary schemes for full locations $(j = 1, 2)$ using three-points (third-order) are obtained as

$$u_1 = \frac{15}{8}u_{3/2} - \frac{5}{4}u_{5/2} + \frac{3}{8}u_{7/2}$$
$$u_2 = \frac{15}{8}u_{5/2} - \frac{5}{4}u_{7/2} + \frac{3}{8}u_{9/2} \tag{180}$$

and using four-points (fourth-order), boundary schemes are obtained as

$$u_1 = \frac{35}{16}u_{3/2} - \frac{35}{16}u_{5/2} + \frac{21}{16}u_{7/2} - \frac{5}{16}u_{9/2}$$
$$u_2 = \frac{35}{16}u_{5/2} - \frac{35}{16}u_{7/2} + \frac{21}{16}u_{9/2} - \frac{5}{16}u_{11/2} \tag{181}$$

similarly, for the right boundary $(j = N, N-1)$, third- and fourth-order accurate stencils are obtained as

$$u_N = \frac{15}{8}u_{N-1/2} - \frac{5}{4}u_{N-3/2} + \frac{3}{8}u_{N-5/2}$$
$$u_{N-1} = \frac{15}{8}u_{N-3/2} - \frac{5}{4}u_{N-5/2} + \frac{3}{8}u_{N-7/2} \tag{182}$$

$$u_N = \frac{35}{16}u_{N-1/2} - \frac{35}{16}u_{N-3/2} + \frac{21}{16}u_{N-5/2} - \frac{5}{16}u_{N-7/2}$$
$$u_{N-1} = \frac{35}{16}u_{N-3/2} - \frac{35}{16}u_{N-5/2} + \frac{21}{16}u_{N-7/2} - \frac{5}{16}u_{N-9/2} \tag{183}$$

9.4.2. Boundary closures for half locations

Boundary stencils for half locations $(j = 3/2, 5/2)$ nodes using three-points are obtained as

$$u_{3/2} = \frac{3}{8}u_1 + \frac{3}{4}u_2 - \frac{1}{8}u_3$$
$$u_{5/2} = \frac{3}{8}u_2 + \frac{3}{4}u_3 - \frac{1}{8}u_4 \tag{184}$$

and boundary stencils based on four-points are obtained as

$$u_{3/2} = \frac{5}{16}u_1 + \frac{15}{16}u_2 - \frac{5}{16}u_3 + \frac{1}{16}u_4$$
$$u_{5/2} = \frac{5}{16}u_2 + \frac{15}{16}u_3 - \frac{5}{16}u_4 + \frac{1}{16}u_5 \tag{185}$$

Similarly, boundary stencils at $(j = N-1/2, N-3/2)$ nodes using three- and four-points are obtained as

$$u_{N-1/2} = \frac{3}{8}u_N + \frac{3}{4}u_{N-1} - \frac{1}{8}u_{N-2}$$
$$u_{N-3/2} = \frac{3}{8}u_{N-1} + \frac{3}{4}u_{N-2} - \frac{1}{8}u_{N-3} \tag{186}$$

$$u_{N-1/2} = \frac{5}{16}u_N + \frac{15}{16}u_{N-1} - \frac{5}{16}u_{N-2} + \frac{1}{16}u_{N-3}$$
$$u_{N-3/2} = \frac{5}{16}u_{N-1} + \frac{15}{16}u_{N-2} - \frac{5}{16}u_{N-3} + \frac{1}{16}u_{N-4} \tag{187}$$

For the present computations, three-point boundary interpolation stencils are used, as the higher-order interpolation generates oscillations in discrete computing.

9.5. Focusing for two-dimensional dispersive LRSWE

Focusing phenomena for dispersive LRSWE on Arakawa grids is discussed next. For collocated A-grid, focusing phenomena as given in [57] is briefly discussed next. To exhibit the focusing mechanism, propagation of 2D wave-packet following LRSWE is considered, as

$$u(x, y, t = 0) = 0, \quad v(x, y, t = 0) = 0$$
$$\eta(x, y, t = 0) = 5e^{-\alpha\left[(x-x_0)^2 + (y-y_0)^2\right]}\sin(k_x x + k_y y) \tag{188}$$

For A-grid, RK_4 – SOUCS3 scheme [67] is used for the space–time discretization of the LRSWE. Numerical properties of the LRSWE using global spectral analysis (GSA) [1] are shown in Figs. 46(a)–46(b). Moreover, comparison of numerical properties of LRSWE using RK_4 – SOUCS3 and RK_4 – OUCS3 schemes are also shown in Fig. 46(a). It is evident from Fig. 46(a) that the directional bias in the OUCS3 introduced by the corresponding boundary closures is rectified using SOUCS3 schemes. Next, numerical solutions to the propagation problem following LRSWE is considered. Here, computational domain of size (120×120) is considered with uniform mesh-width, where propagation angle, $\theta = 45°$ and grid-aspect ratio of $\lambda = \Delta y/\Delta x = 1$. Other parameters are chosen as, $g = 10$ m/s^2, $H = 2.5$ m, $\Delta x = \Delta y = 0.3$ m, $f\Delta t = 2 \times 10^{-7}$ and $Nc_x = Nc_y = 0.2$. The wave-packet given by Eq. (188) is centered at $(k_0\Delta x, k_0\Delta y) = (1.40, 1.40)$. The schematic of the grids A-E used in the computations are shown in Fig. 47.

As in [57], numerical results for collocated A-grid using RK_4 – SOUCS3 are discussed next. Numerical solution to LRSWE for collocated A-grid are shown in Figs. 48–49 at indicated time instants with $\alpha = 0.05$ and $\alpha = 1.0$, respectively. Numerical solutions shown in Fig. 49 are also plotted against z-axis in Fig. 50 to demonstrate the spectacular view of focusing phenomena at the corner nodes of the domain for $\alpha = 1.0$. Moreover, variation of nodal numerical amplification factor for the LRSWE on A-grid with nodes along the diagonal of the computational domain is shown in Fig. 51. The focusing phenomena as shown in Figs. 49–50 for $\alpha = 1.0$ can be explained from the property contours shown in Figs. 46(a)–46(b). The focusing phenomena is observed for the relatively higher value of $\alpha = 1.0$ for which the wave-packet admits upstream propagating higher wavenumber components (q-waves). These q-waves get amplified by the boundary nodes as shown in Fig. 50,

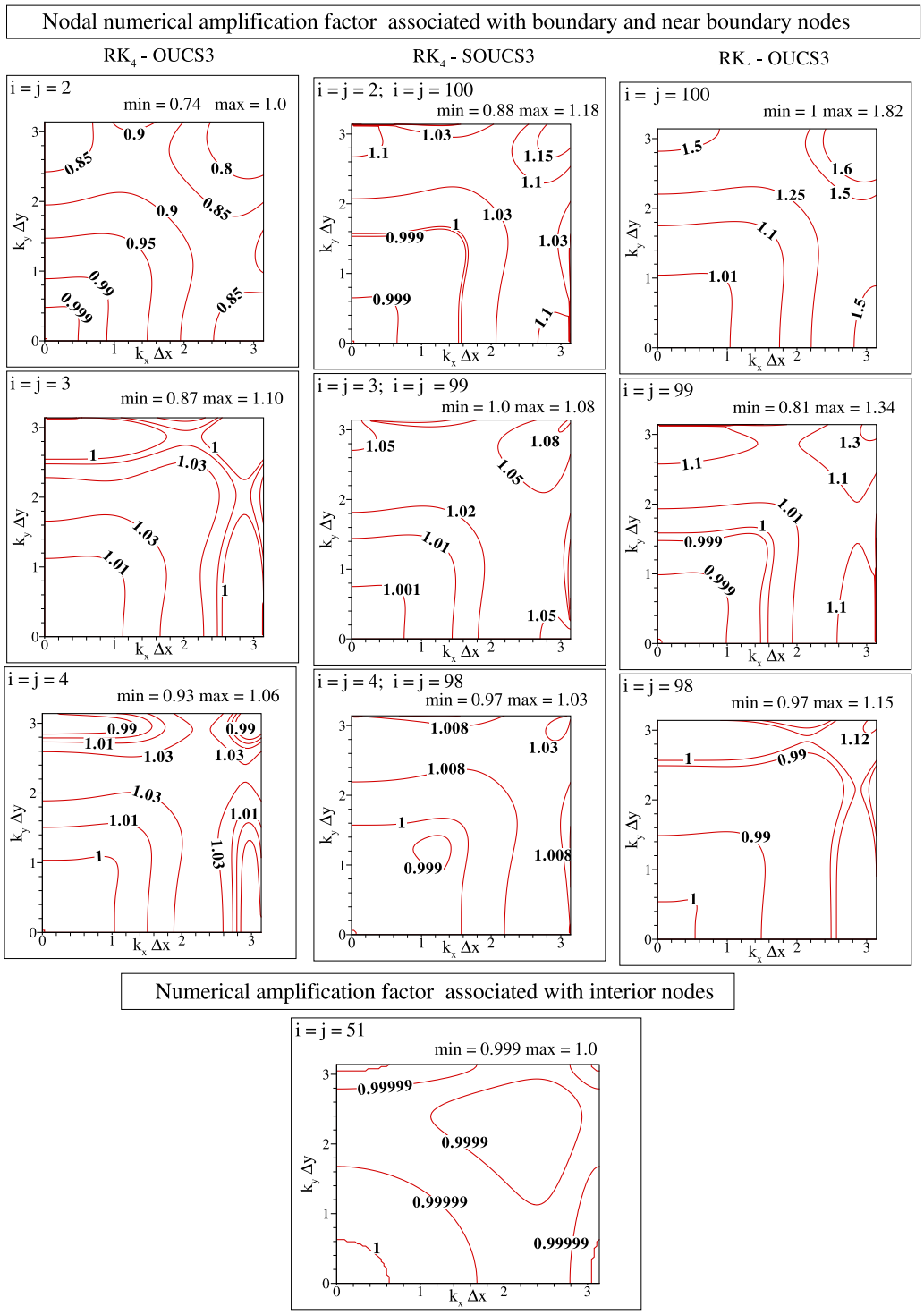

Fig. 46(a). Nodal amplification factor for the LRSWE on *A*-grid using RK_4 – OUCS3 and RK_4 – SOUCS3 schemes.

which is evident from variation of nodal-amplification factors as shown in Fig. 51.

Next, numerical solutions to LRSWE on Arakawa's staggered grids are discussed. For the full-domain analysis on staggered grids, boundary closures discussed in Sections 9.3 and 9.4 are used. For the staggered grids also, we have chosen the same set of parameters as in Figs. 48 and 49. For staggered B-grid, numerical results are shown in Figs. 52 and 53 for values of $\alpha = 0.05$ and $\alpha = 1.0$, respectively. As in [125] (Figs. 2 and 4), it can be noticed that q-wave regions are not present for staggered *B*-grid. Due to absence of q-waves, no focusing phenomena is observed even for $\alpha = 1.0$, as shown in Fig. 53, however, dispersion error can be seen in the steep packet case ($\alpha = 1.0$). Similar trend was noticed for *C*-grid, as shown in Figs. 54 and 55 for $\alpha = 0.05$ and $\alpha = 1.0$, respectively. Here also, due to absence of q-waves no focusing phenomena is observed, however, dispersion errors are noticed for the steep packet ($\alpha = 1.0$) case. As evident from the numerical properties discussed in [125], dispersion errors for *C*-grid are lesser than *B*-grid.

Numerical solutions corresponding to *D*-grid are shown plotted in Figs. 56 and 57 for $\alpha = 0.05$ and $\alpha = 1.0$, respectively. *D*-grid has poorer numerical properties as compared to other Arakawa grids [125],

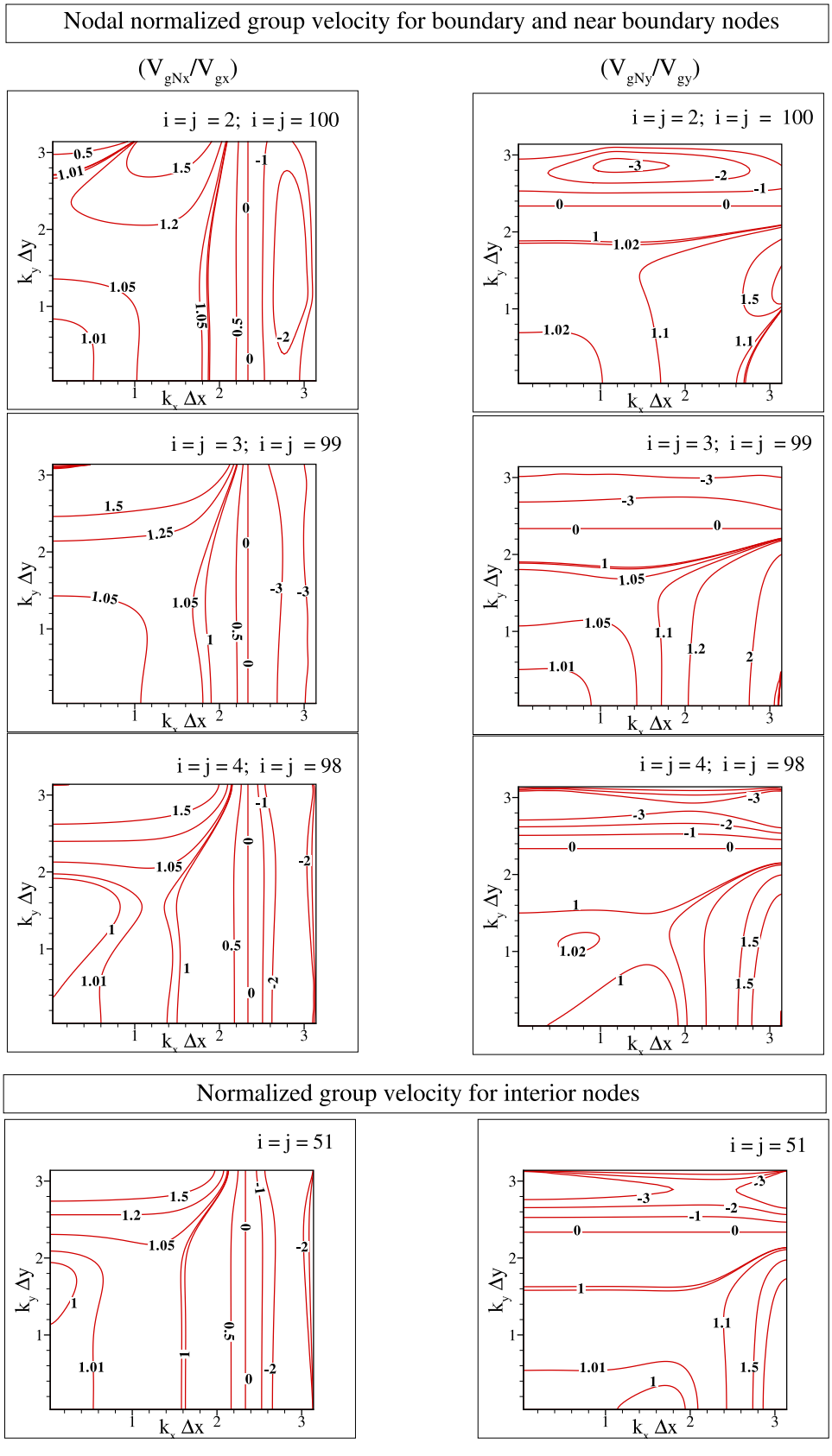

Fig. 46(b). Nodal normalized group velocity contour for the LRSWE on *A*-grid using RK_4 – SOUCS3 scheme.

however q-waves limit for *D*-grid is about $k\Delta = 2.8$ which is higher than *A*-grid (2.39). Thus, huge dispersion error can be noticed for both the cases (with $\alpha = 0.05$ and $\alpha = 1.0$). Numerical solutions on *E*-grid (which is 45^o rotated *B*-grid) are shown in Figs. 58 and 59 for $\alpha = 0.05$ and $\alpha = 1.0$, respectively. For the case of *E*-grid, similar trend, as for *B*- and *C*-grids, is observed. Here also due to absence of q-waves no

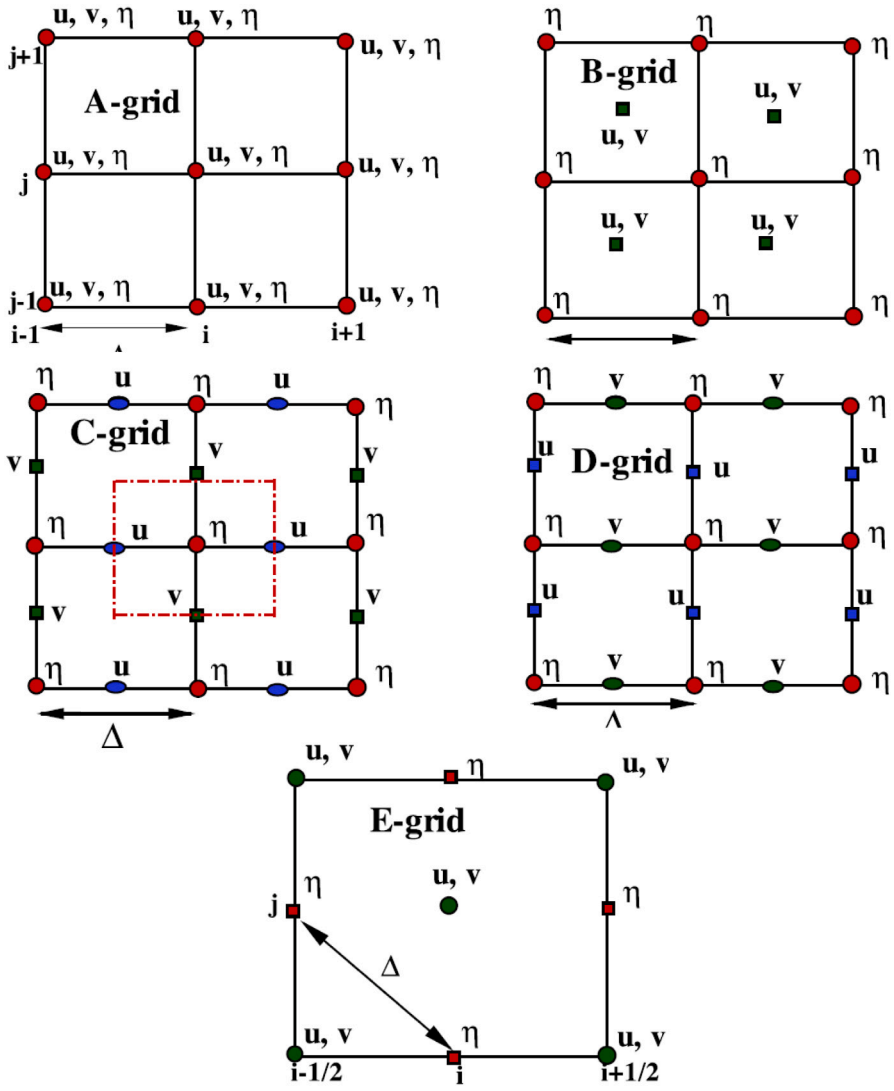

Fig. 47. Schematic of grid types for solving LRSWE with Arakawa's A- to E-grids. C-grid with dash-dotted lines indicates the marker and cell (MAC) method.

focusing phenomena is observed, however, dispersion errors are there. Furthermore, as *C*-grid displays better numerical properties [125], therefore the dispersion errors for $\alpha = 1.0$ case are lower for *C*-grid, as compared to other grids.

Finally, from numerical solutions for the LRSWE on *B*- to *E*-grids as shown in Figs. 52–59 for $\alpha = 0.05$ and 1.0 it is evident that presence of q-waves is necessary for triggering the focusing phenomena. As staggering of unknowns alters the dispersion relation, it in turn removes/lowers the q-waves barrier [125]. Moreover, staggering of unknowns also introduces numerical dissipation, which is another reason for the absence of focusing phenomena for the LRSWE on staggered *B*- to *E*-grids. For the present computations, it is checked that focusing phenomena was absent even after $t = 40$ for $\alpha = 1.0$.

10. Recent developments related to GSA in HPC

In this section, we demonstrate applications of GSA to explain how GSA can explain past activities termed as DNS, as well as, present better methods for ongoing high performance computing using high accuracy compact schemes.

10.1. Evaluation of DNS of homogeneous isotropic turbulence

Ever since the appearance of GSA as an analysis tool, it has been used to study many numerical methods which were not understood well before. Not surprisingly, this also includes the pseudo-spectral method used for DNS of homogeneous isotropic turbulence (HIT) reported for the first time by Orszag and Patterson [130], who used two-stage, second order Runge–Kutta (RK2) method for time integration. A more detailed account of the same was reported by Rogallo [131], and this particular version of the code has been used in many dissertations and part of the results reported in [132–134]. The same methodology continues to be used in many other researches, as reported in [135]. For HIT problem being periodic, application of Fourier spectral method is supposed to give the best spatial resolution. The results are reported with respect to a Reynolds number based on the Taylor's microscale (λ), which was given by $R_\lambda = 35$ in [130] who used a 32^3 computational periodic box, and in [135], the authors have used 12288^3 points for $R_\lambda = 1300$. In all these cited references in this section, RK2 time integration has been used.

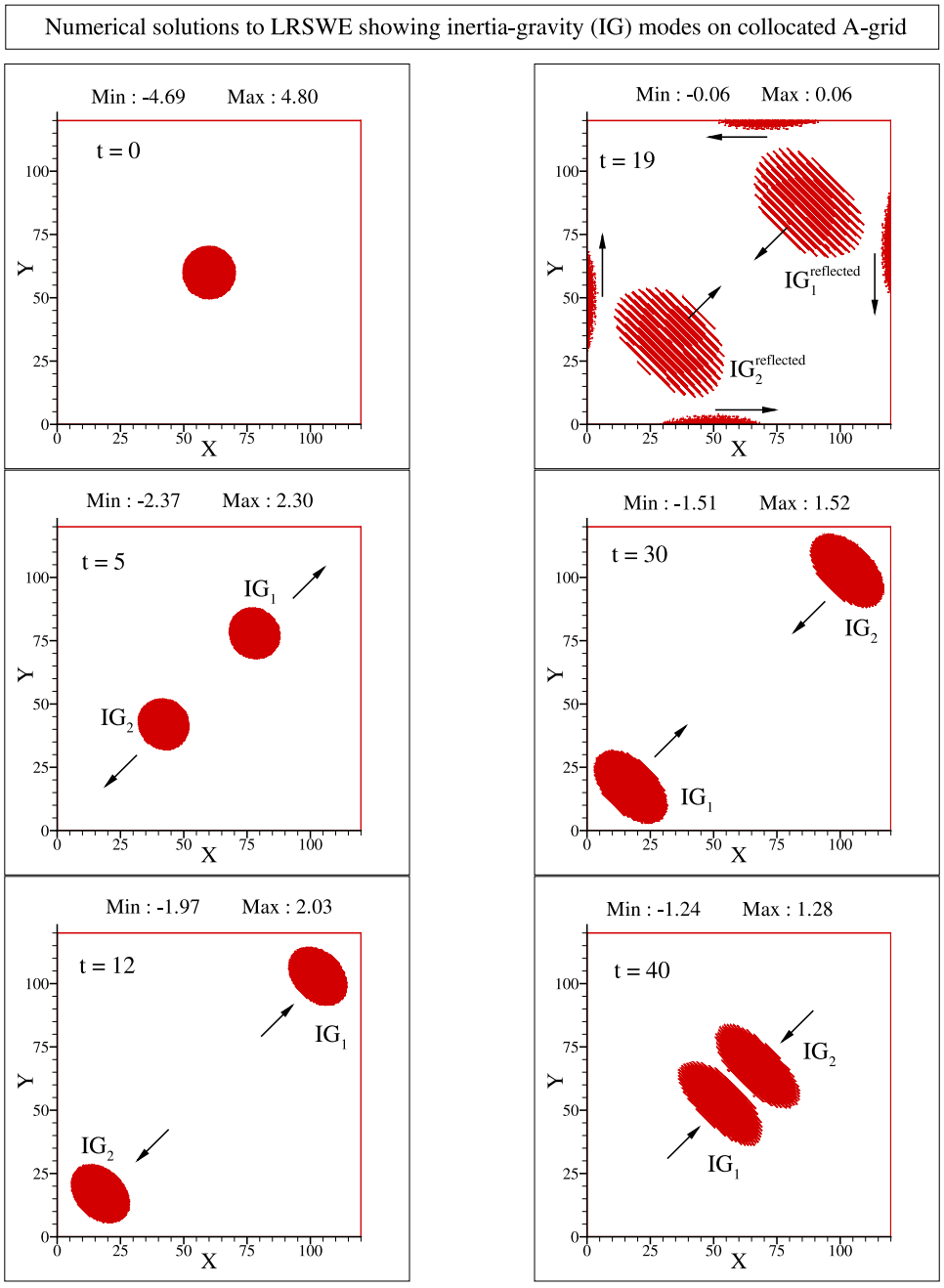

Fig. 48. Numerical solutions to LRSWE on A-grid at indicated time instants using $RK_4 - SOUCS3$ scheme with $N_c = 0.2$, $k_o h = 1.40$ and $\alpha = 0.05$.

In solving the HIT problem, one may like to solve the incompressible NSE (INSE) with a generic form given by,

$$\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot \frac{\partial \mathbf{u}}{\partial x} = -\nabla P/\rho + \nu \nabla^2 \mathbf{u} + \mathbf{f}, \tag{189}$$

where $\mathbf{u}$ is the solenoidal velocity field, P is the pressure, ρ is the density of the fluid, ν is the kinematic viscosity, and $\mathbf{f}$ is a forcing term imposed at a large length scale added to the INSE to weakly justify to ensure statistical stationarity of the computed turbulent signal [132] by solving HIT by the RK2-Fourier spectral method, as in [135]. This method was originally used by Rogallo [131] with the added forcing in INSE.

It is noted that such DNS often show solution "blow-up" in finite time (specifically in the limit, $\nu \to 0$), which some researchers [135] have interestingly conjectured to correspond to turbulent solutions of the INSE. Lamorgese et al. [136] have noted that "experimental measurements in homogeneous turbulence at high Reynolds numbers show the unequivocal presence of a "bottleneck" effect", and the authors introduced hyperviscosity in the INSE as,

$$\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot \frac{\partial \mathbf{u}}{\partial x} = -\nabla P/\rho + (-1)^{h+1} \nu_h \nabla^{2h} \mathbf{u} + \mathbf{f}, \tag{190}$$

where ν_h is the specified constant hyperviscosity coefficient, and $\mathbf{f}$ is the forcing function. The formulation for the DNS is recovered with $h = 1$ and $\mathbf{f} = 0$. The authors [136] have noted that the "(u)se of $\mathbf{f}$ is unnatural (as are most other ways of forcing turbulence). However, we are primarily concerned with bottleneck effects on energy spectra, i.e., we investigate one particular characteristic of small-scale turbulence. In this case, use of a large-scale forcing (in order to analyze statistically stationary rather than decaying turbulence) is justifiable

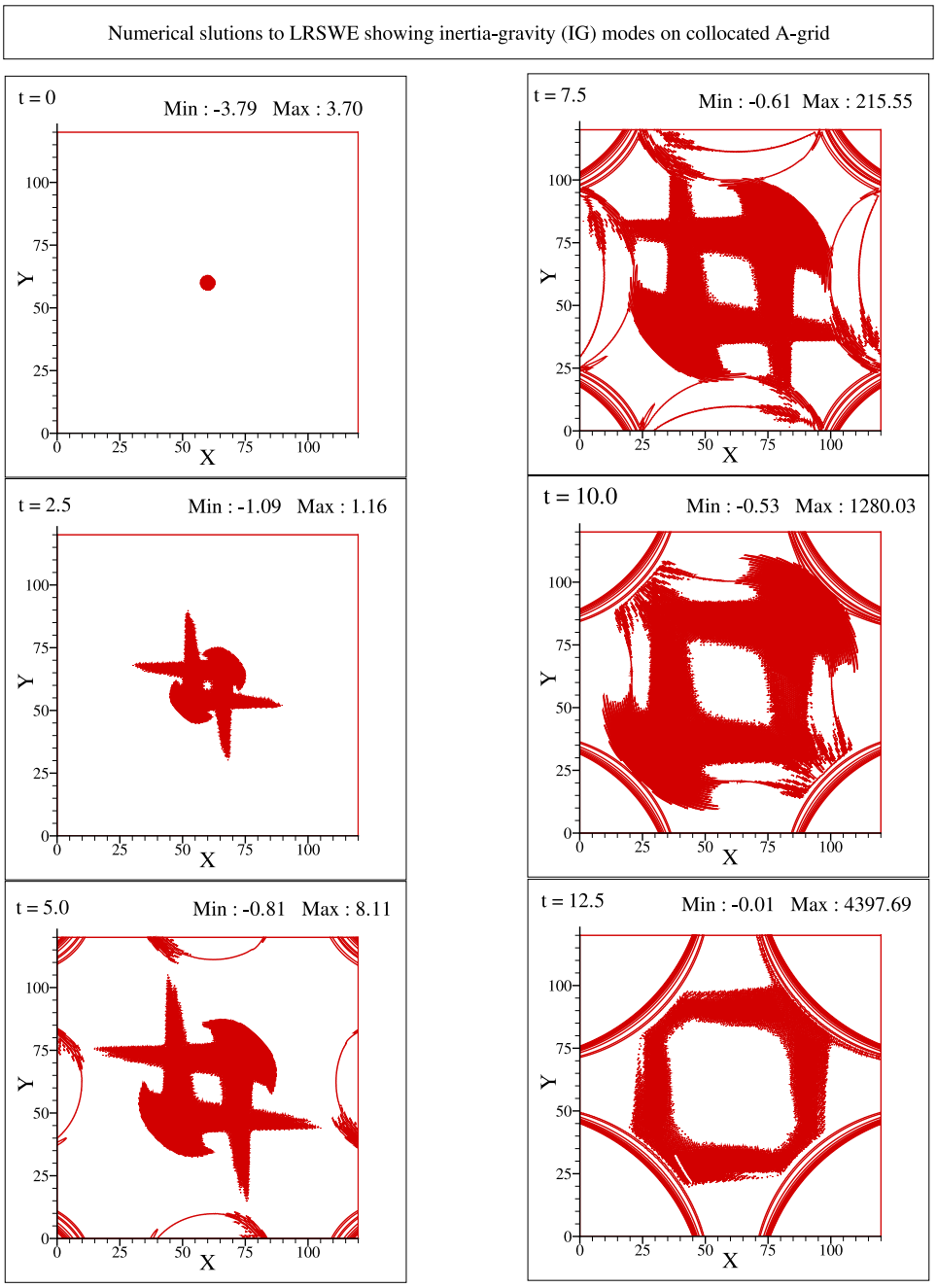

Fig. 49. Numerical solutions to LRSWE on A-grid at indicated time instants using RK_4 – SOUCS3 scheme with $N_c = 0.2$, $k_o h = 1.40$ and $\alpha = 1.0$.

on the grounds that the details of the forcing have little effect on the small-scale statistics".

To avoid making conjectures, one can instead analyze the CE and CDE where multi-stage RK methods of time integration is used in conjunction with spatial discretization of convection and diffusion terms given by, $\frac{\partial u}{\partial x} = \int ik\hat{U}e^{ikx}dk$ and $\frac{\partial^2 u}{\partial x^2} = \int -k^2\hat{U}e^{ikx}dk$. These derivatives are evaluated using fast Fourier transform (FFT) and the difference equation provides the complex G as function of $k\Delta x$ and N_c for the CE and $k\Delta x$, N_c, Pe for the CDE. With $|G|$ and ϕ obtained, one can readily obtain c_N/c and $V_{g,N}/c$ for the CE, and for CDE one finds ν_N/ν additionally, as the non-dimensional dispersive coefficient of diffusion. Here we discuss typical results involving the normalized numerical amplification factor to highlight the aspect of the claim made for DNS by Fourier spectral and RK2 method in the past.

In Fig. 60, the magnitude of non-dimensional numerical amplification factor ($|G|$) of RK2-Fourier spectral method is considered for the CE and CDE. For the 1D CE, $|G|$-contours are shown plotted in the $(N_c, k\Delta x)$-plane on the left hand side of Fig. 60. It is noted that except a very small region close to the origin, the method is unconditionally unstable, as reported earlier in [22]. Even when one chooses a very small CFL number in this near-origin region, there will always be some resolved wavenumbers within the Nyquist region, for which the method will be unstable, with the higher wavenumbers more unstable. It is to be noted that the round-off error provides the seed for instability and the unstable high wavenumbers interact to create many wave-packets for the periodic problem. For non-periodic problems, outflow boundary conditions would allow the possibility of error-packets leaving the computational domain. To highlight the growth of error inherent to numerical methods, a periodic problem is more suited, as the signal is

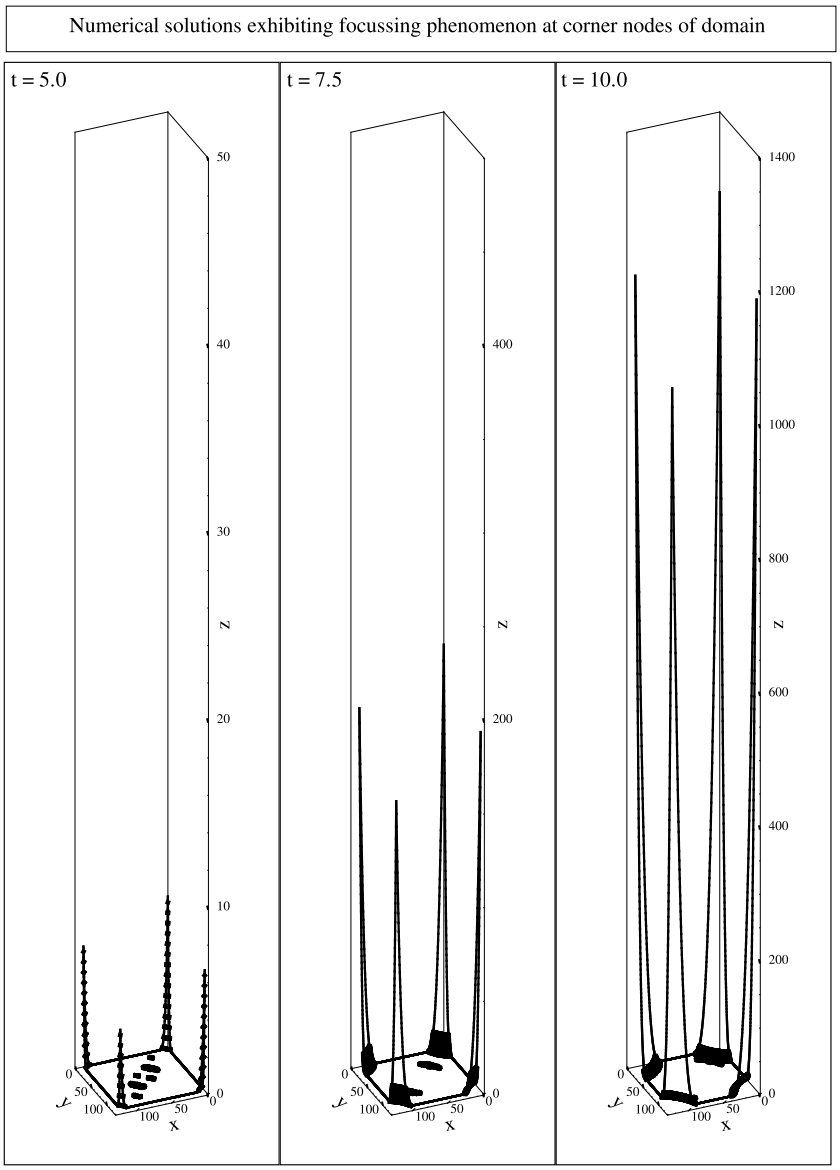

Fig. 50. Numerical solution to LRSWE on A-grid corresponding to case shown in Fig. 49. Solutions plotted against Z-axis shows the focusing phenomena at the corner nodes of the domain.

trapped in the domain, in addition to evanescent error continuously. The property chart for the CE clearly explains why the solution of Euler's equation always indicates a finite time solution "blow up" [137]. Because the property chart of CE is similar to that one would expect the solution of linearized Euler's equation to display solution behavior. For the CDE, the error dynamics is also affected via the term involving ν_N/ν and is given by [138,139],

$$
\begin{aligned}
e_t + ce_x - \nu e_{xx} = & \int_{-k_{max}}^{k_{max}} (\nu_N - \nu)\, k^2\, e^{-\nu_N k^2 n\Delta t}\, U_0(k)\, e^{ik(x-c_N t^n)}\, dk \\
& + ikc_N\, e^{-\nu_N k^2 n\Delta t}\, U_0(k)\, e^{ik(x-c_N t^n)} \Big|_{-k_{max}}^{k_{max}} \\
& - \int_{-k_{max}}^{k_{max}} \left(\frac{V_{gN} - c_N}{k} \right) \\
& \times \left\{ \int_{-k_{max}}^{k} ik'\, e^{-\nu_N k'^2 n\Delta t}\, U_0(k')\, e^{ik'(x-c_N t^n)}\, dk' \right\} dk \\
& - \int_{-k_{max}}^{k_{max}} ikc\, e^{-\nu_N k^2 n\Delta t}\, U_0(k)\, e^{ik(x-c_N t^n)}\, dk \qquad (191)
\end{aligned}
$$

Once again, the error does not follow the same dynamics as the signal, as was demonstrated for the CE [31]. This is a universal feature of all scientific computing, even for linear systems. For the error dynamics of CDE, the sources of error for CE are also noted. To focus on the error due to numerical amplification/attenuation in analyzing numerical errors by GSA, the contours of $|G_{Num}|/|G_{Phys}|$ are shown in this figure. On the right hand side frames of Fig. 60, the contours of $|G_{Num}|/|G_{Phys}|$ are shown in $(N_c, k\Delta x)$-plane for $Pe = 0.05$ and 0.2. One notes a finite range of N_c for which there is no instability over the complete resolved scales, for which $|G_{Num}|$ is not greater than one, for the case of $Pe = 0.05$.

However in [138,139], the ratio of $|G_{Num}|/|G_{Phys}|$ have been investigated for the values of $Pe = 0.01$ and 0.1, and it is seen that for both the Pe, there are contour values in the Nyquist limit which are stable. Results for $Pe = 0.01$ indicated $|G_{Num}|/|G_{Phys}|$ almost equal to 1 for $N_c = 0.2$, for the entire Nyquist limit [138,139]. It was also noted that for $Pe = 0.1$, one cannot choose any value of N_c for which all the resolved scales in the range, $0 \le kh \le \pi$, show the ideal attribute of $|G_{Num}|/|G_{Phys}|$ to be very close to one [138,139]. The property charts for these two values of Pe in [139] seem to indicate that increased Pe

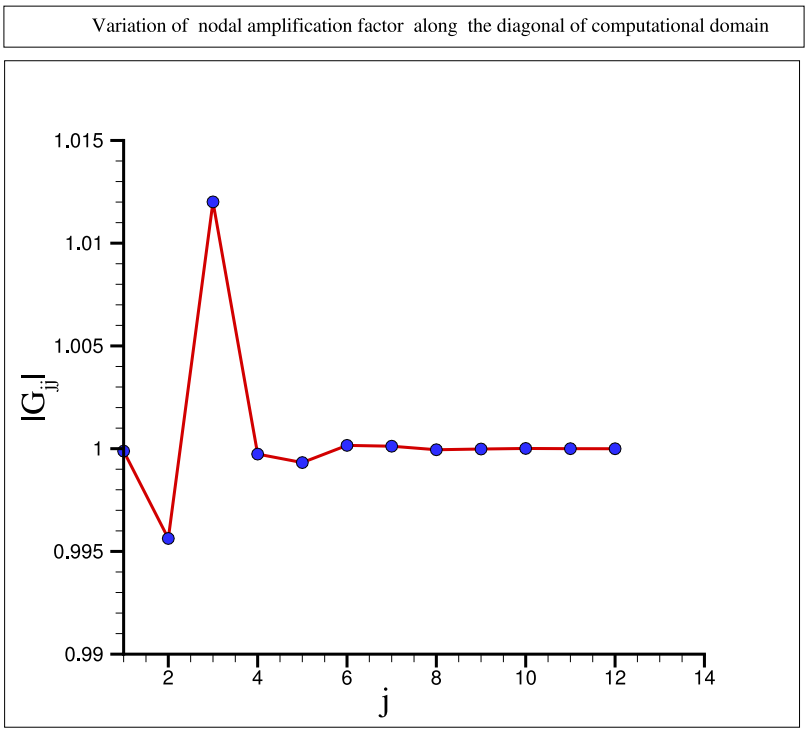

Fig. 51. Variation of nodal numerical amplification factor for LRSWE on A-grid with nodes along the diagonal of the computational domain.

shrinks the region of unconditionally unstable region ($|G_{Num}| > 1$). In Fig. 60, the property shown for the case of $Pe = 0.2$, shows that the range of unconditionally unstable region again increases, and there are no N_c values for which the ratio $|G_{Num}|/|G_{Phys}|$ remains very close to 1 for the entire Nyquist limit. Present results and those in [139] clearly indicates that there is no possibility of performing DNS by considering only this single consideration of the ratio $|G_{Num}|/|G_{Phys}|$. This is due to the fact that in solving INSE, one cannot simply consider only single value of N_c, and instead of CE/CDE, one may like to study instead the one-dimensional viscous Burgers' equation, as the canonical equation. Presented results and the discussion of the property chart clearly explains why the Fourier spectral and RK2 methods can never provide DNS results, by this simple application of GSA. All such previous claims [130–133,135] are to be correctly evaluated in the light of the presented results here and in [139]. In the name of DNS, all these cited references altered the governing INSE for DNS to the altered equation given in Eq. (190) via the forcing and hyperviscosity terms.

10.2. Application of GSA for high performance computing using compact schemes

In describing the compact schemes in Section 3, it has been noted that due to its implicit nature, one needs extra care in closing the system of equations for evaluating the derivatives at the interior of the domain. Even when the stencil size is same, as in an explicit scheme, this constitutes the typical boundary closure problem for compact scheme. For this reason, the compact schemes in [33] have been very sparingly used by researchers, as in [66,117] with limited success even for sequential computing. The issues have been identified in proposing the GSA by the authors in [1,29,34] due to numerically adding anti-diffusion for the near boundary points. The authors in proposing a series of upwind compact schemes solved the boundary closure problems for sequential computing with limited flow distortions localized near the inflow and outflow. These attributes of localized nature of boundary closure problem was used in proposing a parallel algorithm for compact scheme employing Schwarz domain decomposition in [140], which also requires additional filtering. This necessitated having overlapping points in the subdomain boundaries, making the use of additional resources overall, and still some flow distortions across the domain near the subdomain boundaries remain. This caused problems of accuracy for flow instability problems. To remove such distributed flow distortion near subdomain boundaries, a fresh approach of parallelization has been used in [70,141] to study problems of bifurcation and instabilities. The details of the developed method for compact scheme, without any overlap of points at subdomain boundaries and without any error caused by parallelization have been described in [142,143] and is described in the following. The method is found to be very robust and easy to implement, so that the same has been used for many other fluid dynamic problems in [144–148].

The design of this new parallel algorithm is based on GSA, with a simple observation that every implicit method for evaluating derivatives have their equivalent explicit algorithm. A first derivative is indicated by a prime in GSA by writing it as, $[A]\{f'\} = \frac{1}{h}[B]\{f\}$, with entries near the top and bottom rows of $[A]$ and $[B]$ contain boundary closure schemes, and h is the uniform spacing of the grid points. In this notation, $[A]$ is the identity matrix for an explicit method. For compact schemes to evaluate $\{f'\}$, one can alternately write the above linear algebraic equation as, $\{f'\} = \frac{1}{h}[C]\{f\}$, with $[C]$ matrix given by, $[C] = [A]^{-1}[B]$. The noticeable feature of compact scheme is that $[A]$ and $[B]$ matrices are strictly band-limited by design, while $[C]$ matrix is a wide-band matrix. It has been noted [29,142] that the entries of the $[C]$ depends upon the row in which they occur for a non-periodic problem, but are independent of the rows for a periodic problem. This observation helps in using this resolution property for subdomain boundary closure by treating the boundary points to be at the interior, by using the entries of the middle row of the $[C]$ matrix of a sequential computing exercise with sufficient number of points. Once the entries of the subdomain boundary points are used to calculate the derivatives there, the derivatives in the interior of each subdomain are obtained by using Thomas algorithm [1]. The entries of the middle row of the $[C]$ matrix are listed in the appendix in [143] for the OUCS3 scheme, which require 48 points on either side of the boundary from the neighboring subdomain to maintain 16 digit accuracy for

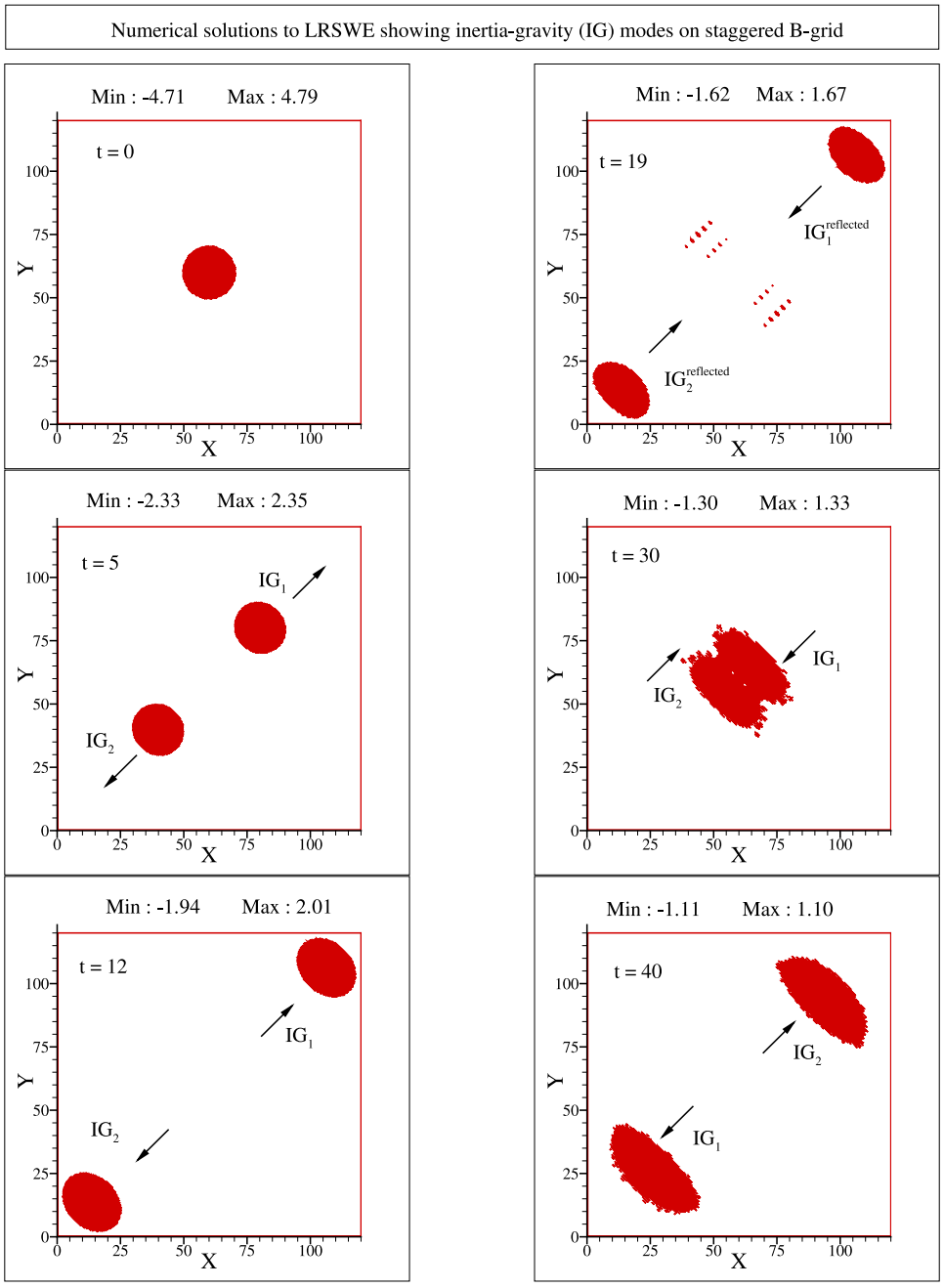

Fig. 52. Numerical solutions to LRSWE on B-grid at indicated time instants using $RK_4 - OSCS$ and optimized interpolation schemes with $N_c = 0.2$, $k_o h = 1.40$ and $\alpha = 0.05$.

real numbers in the computations. This non-overlapping high accuracy parallel (NOHAP) scheme has the unique feature of removing any error up to machine precision due to parallelization and widely reported, as noted above.

10.2.1. Implementation of NOHAP scheme

A schematic of uniformly spaced grid points in the direction along which derivative is to obtained is shown in Fig. 61. The computational domain is decomposed without overlapping points and equally distributed to available processors, as shown in the bottom frame of the figure. While i represents an arbitrary grid point in both sequential and parallel computing, the $\bar{i}$ and $\bar{j}$ represent the first and last point of the sub-domain distributed to the $(p+1)$th processor. The derivative at $\bar{i}$ and $\bar{j}$ is computed by the equivalent explicit scheme of the compact scheme developed for the interior nodes, as explained next.

The NOHAP scheme can be implemented in two stages [142]. While the first stage pre-processes the compact scheme to get the equivalent explicit scheme, the derivatives at the sub-domain boundaries are evaluated in the second stage using the equivalent explicit scheme. The stage-1 is performed only once to get an equivalent explicit scheme of the chosen compact scheme. In contrast, Stage-2 is executed to evaluate the derivatives in the time-accurate simulations at the sub-domain boundaries to eliminate the parallelization error up to 16th decimal place.

Stage-1:

1. A sufficiently large circulant $[A]$ and $[B]$ matrices are formed for the chosen compact scheme. In the present exercise, the dimensions of the matrices are 251×251.

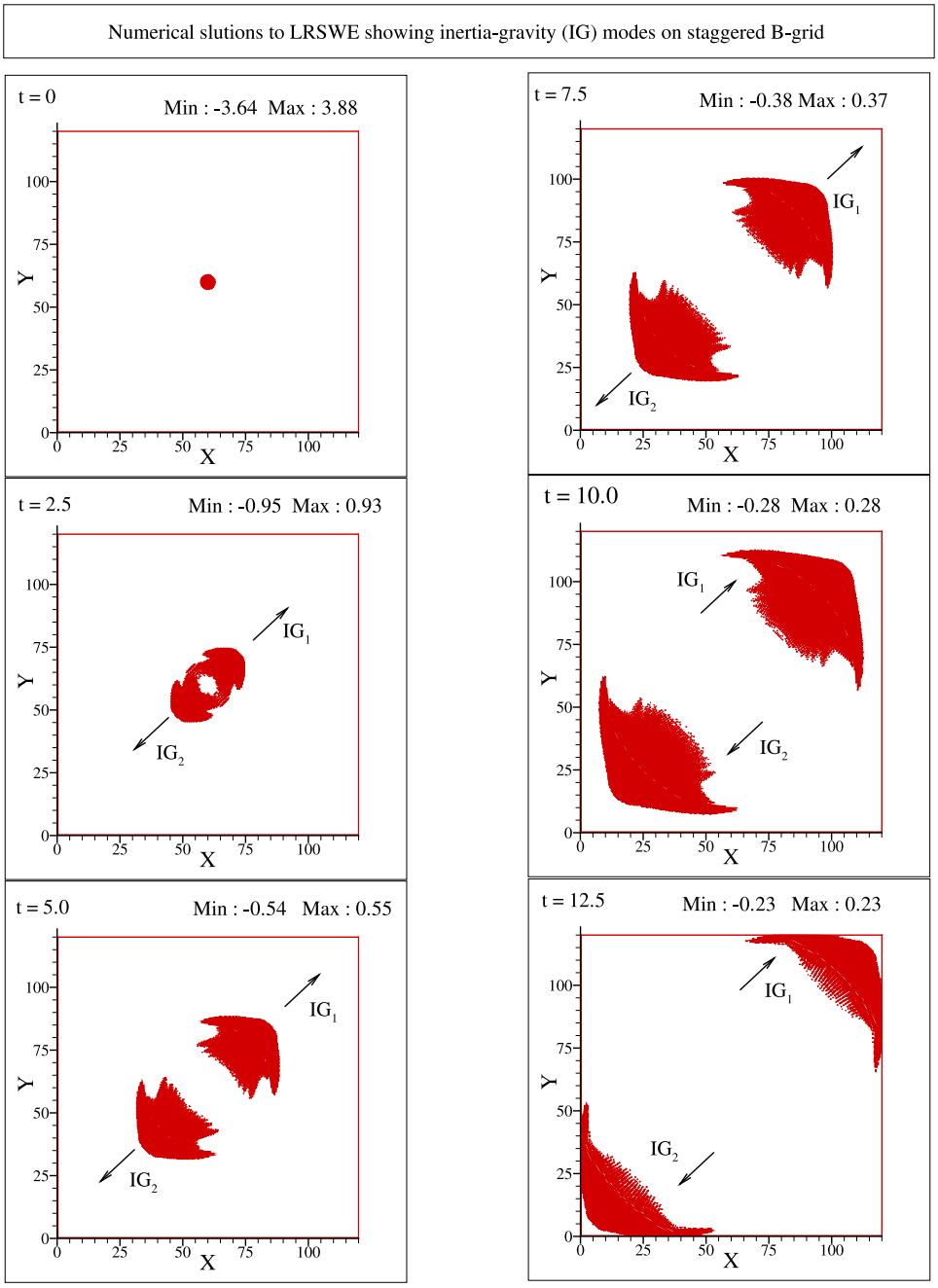

Fig. 53. Numerical solutions to LRSWE on B-grid at indicated time instants using $RK_4 - OSCS$ and optimized interpolation schemes with $N_c = 0.2$, $k_o h = 1.40$ and $\alpha = 1.0$.

2. $[C] = [A]^{-1}[B]$ matrix is obtained by using Thomas' algorithm. The magnitude of the coefficients of the $[C]$ matrix exponentially decay with respect to the diagonal [142,143].
3. The coefficients with magnitude less than 10^{-16} are replaced to zero, resulting in a banded $[C]$ matrix.
4. The non-zero coefficients of the mid-row of the $[C]$ matrix forms the equivalent explicit scheme of the chosen compact scheme, which is stored in an array (*gamma*) whose index varies from $-n_b$ to $+n_b$.
5. The derivative by the compact scheme at any grid point can be obtained by equivalent explicit stencil given by Eq. (192), which has the spectral resolution equivalent to the chosen compact scheme.

$$f'_i = \frac{1}{h} \sum_{j=-n_b}^{n_b} \gamma_j f_{i+j} \tag{192}$$

Stage-2:

1. Decompose the computational domain in each direction without overlapping points, as shown in Fig. 61(b).
2. Derivative at the first grid point of the $(p+1)$th-processor (marked as $\bar{I}+1$ in the figure) is obtained by Eq. (192). The

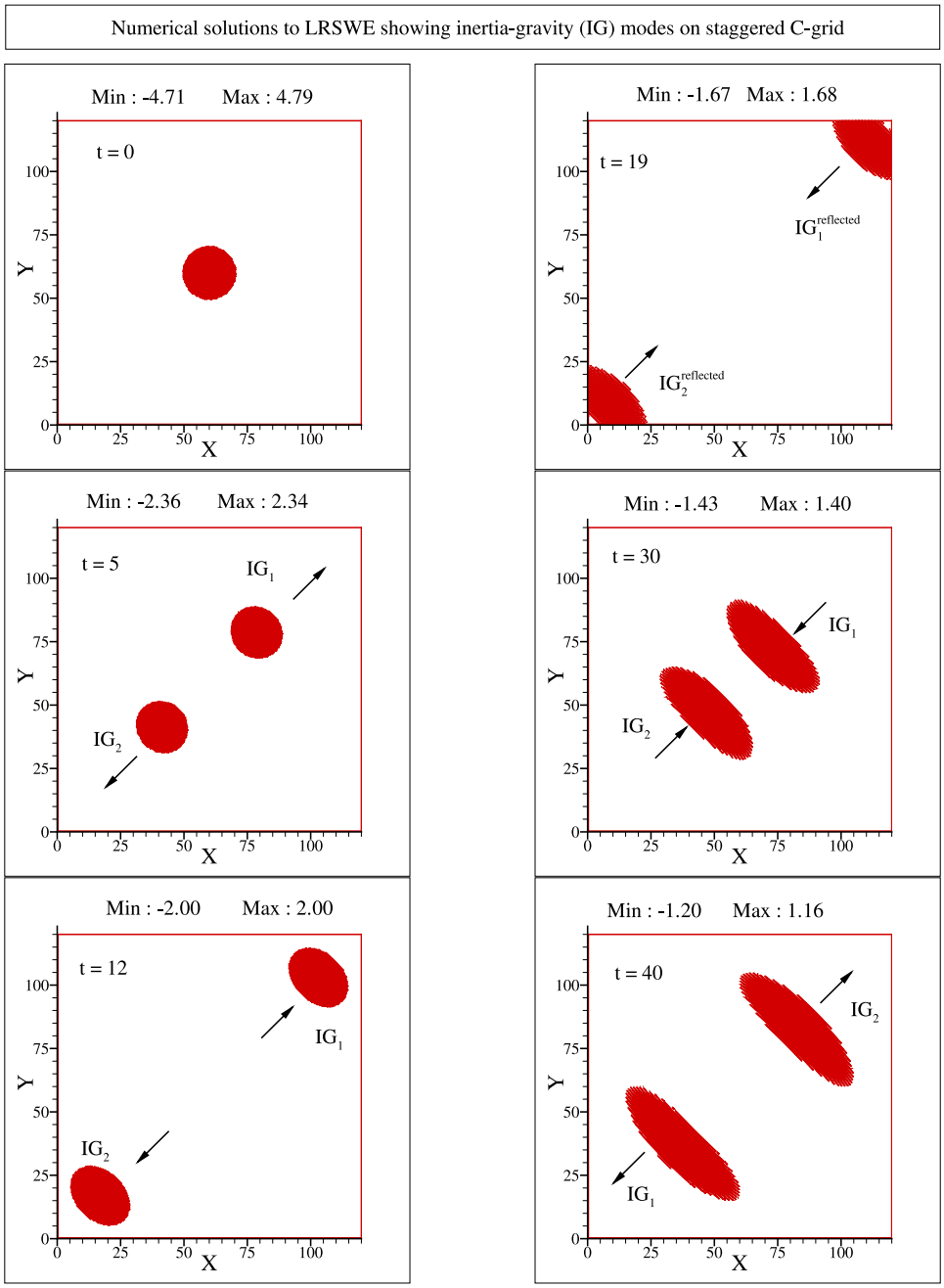

Fig. 54. Numerical solutions to LRSWE on C-grid at indicated time instants using $RK_4 - OSCS$ and optimized interpolation schemes with $N_c = 0.2$, $k_o h = 1.40$ and $\alpha = 0.05$.

Eq. (192) has been split as,

$$f'_{\bar{i}+1} = \frac{1}{h}\sum_{j=-n_b}^{-1} \gamma_j f_{\bar{i}+1+j} + \frac{1}{h}\sum_{j=0}^{n_b} \gamma_j f_{\bar{i}+1+j} \tag{193}$$

3. The pth-processor evaluates the first term of the above equation, and the $(p+1)$th-processor computes the second term.
4. The (p)th-processor transfers the partial sum obtained in the previous step to $(p+1)$th-processor to evaluate $f'_{\bar{i}+1}$ following Eq. (193). A similar procedure is followed for computing the derivative at the last grid point ($f'_{\bar{j}}$) of the $(p+1)$th-processor.
5. The formation of $[A]$ and $[B]$ matrix for the interior points in the $(p+1)$th-processor with Lele's scheme is demonstrated here. Eq. (45) for the second grid point $(\bar{i}+2)$ in the $(p+1)$th-processor is written as,

$$\alpha_6 f'_{\bar{i}+1} + f'_{\bar{i}+2} + \alpha_6 f'_{\bar{i}+3} = \frac{a_6}{2h}(f_{\bar{i}+3} - f_{\bar{i}+1}) + \frac{b_6}{4h}(f_{\bar{i}+4} - f_{\bar{i}})$$

Since $f'_{\bar{i}+1}$ is known from the previous step, the equation can be modified as,

$$f'_{\bar{i}+2} + \alpha_6 f'_{\bar{i}+3} = \frac{a_6}{2h}(f_{\bar{i}+3} - f_{\bar{i}+1}) + \frac{b_6}{4h}(f_{\bar{i}+4} - f_{\bar{i}}) - \alpha_6 f'_{\bar{i}+1}$$

The function value, $f_{\bar{i}}$ is the last grid point of the (p)th-processor, and hence, one more communication is needed to form the $[B]$ matrix.
6. The same procedure is followed at the second last point $(\bar{j}-1)$

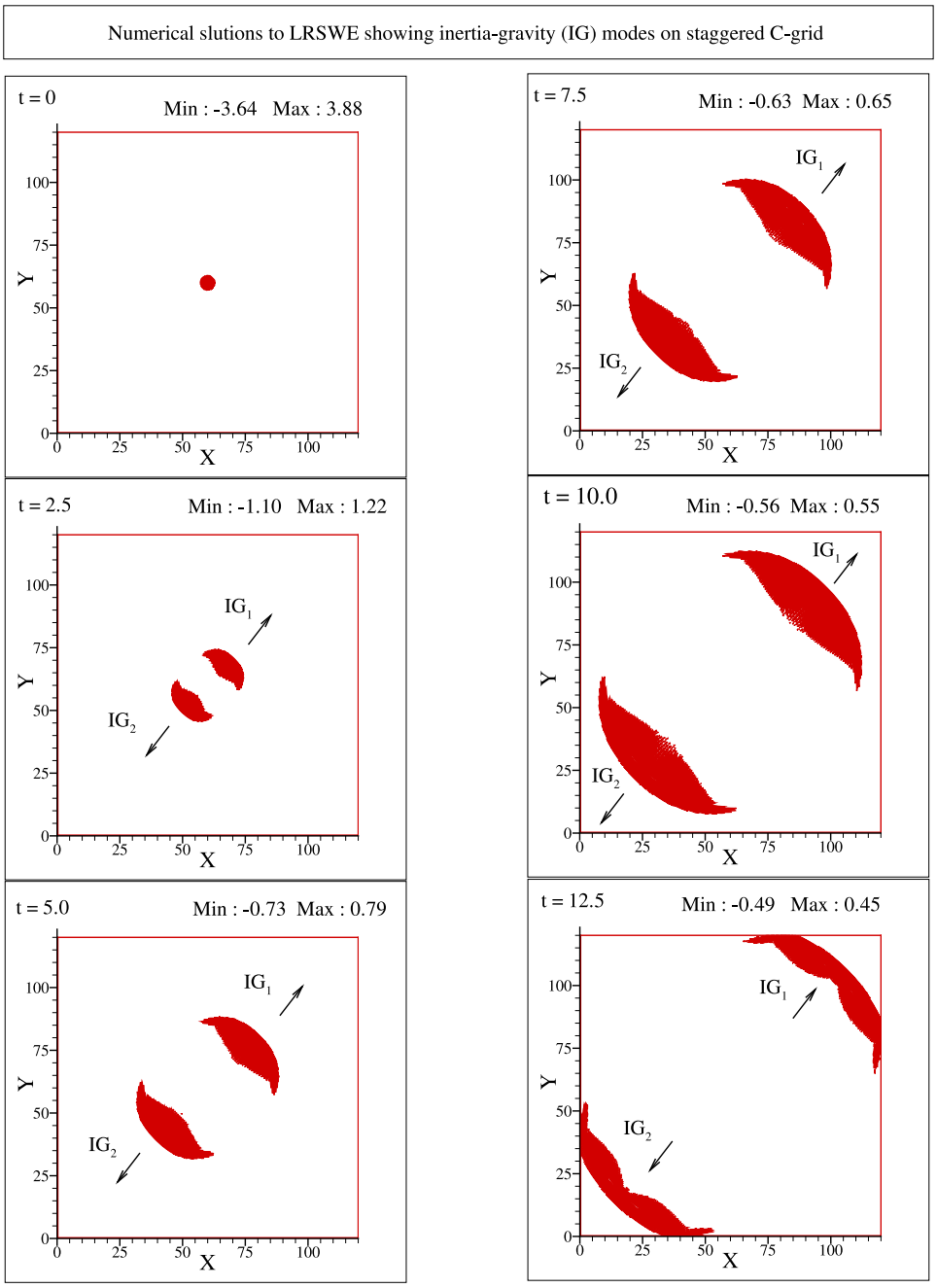

Fig. 55. Numerical solutions to LRSWE on C-grid at indicated time instants using RK_4 – OSCS and optimized interpolation schemes with $N_c = 0.2$, $k_o h = 1.40$ and $\alpha = 1.0$.

of the $(p+1)$th-processor. The system of equations solved by $(p+1)$th-processor is in the form of Eq. (194).

7. This stage is parallelly executed by all the processors to decouple the system of equations and solve them independently.

$$\begin{bmatrix} 1 & \alpha_6 & 0 & \dots & \dots & 0 & 0 \\ \alpha_6 & 1 & \alpha_6 & 0 & \ddots & \ddots & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \alpha_6 & 1 & \alpha_6 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \ddots & \ddots & 0 & \alpha_6 & 1 & \alpha_6 \\ 0 & 0 & \dots & \dots & 0 & \alpha_6 & 1 \end{bmatrix} \begin{Bmatrix} f'_{\bar{i}+2} \\ f'_{\bar{i}+3} \\ \vdots \\ f'_{i} \\ \vdots \\ f'_{\bar{j}-2} \\ f'_{\bar{j}-1} \end{Bmatrix}$$

$$= \begin{Bmatrix} \frac{a_6}{2h}(f_{\bar{i}+3} - f_{\bar{i}+1}) + \frac{b_6}{4h}(f_{\bar{i}+4} - f_{\bar{i}+1}) - \alpha_6 f'_{\bar{i}+1} \\ \frac{a_6}{2h}(f_{\bar{i}+4} - f_{\bar{i}+2}) + \frac{b_6}{4h}(f_{\bar{i}+5} - f_{\bar{i}+1}) \\ \vdots \\ \frac{a_6}{2h}(f_{i+1} - f_{i-1}) + \frac{b_6}{4h}(f_{i+2} - f_{i-2}) \\ \vdots \\ \frac{a_6}{2h}(f_{\bar{j}-1} - f_{\bar{j}-3}) + \frac{b_6}{4h}(f_{\bar{j}} - f_{\bar{j}-4}) \\ \frac{a_6}{2h}(f_{\bar{j}} - f_{\bar{j}-2}) + \frac{b_6}{4h}(f_{\bar{j}+1} - f_{\bar{j}-3}) \\ -\alpha_6 f'_{\bar{j}} \end{Bmatrix} \tag{194}$$

It is to be noted that Eq. (192) requires $2n_b+1$ operations per point to compute the derivative. Whereas computation of derivative from the solution of Eq. (194) by Thomas' algorithm requires only 5 operations per point [1]. Thus, although Eq. (192) is valid for any interior points

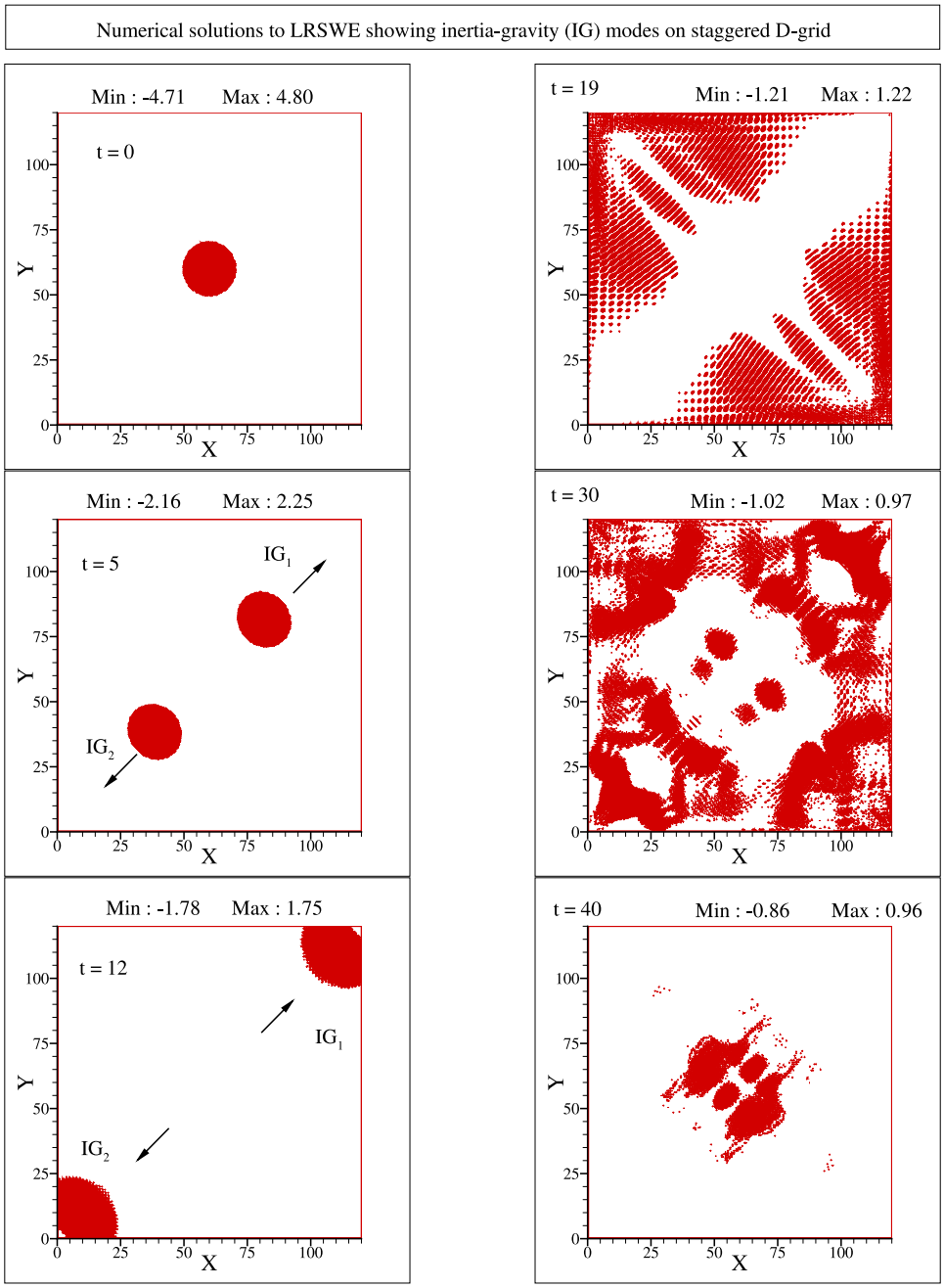

Fig. 56. Numerical solutions to LRSWE on D-grid at indicated time instants using RK_4 – OSCS and optimized interpolation schemes with $N_c = 0.2$, $k_o h = 1.40$ and $\alpha = 0.05$.

with $i > n_b$, it is only used at the sub-domain boundaries due to an immense requirement of arithmetic operations. Also, Eq. (193) assumes that each processor contains a minimum of n_b points in the direction in which the derivative is evaluated. When processors contain less than n_b points, partial summations are performed and sent by the second neighbor and beyond, which increases the communication overhead and may result in performance degradation.

Other efforts have been made to develop parallel computing with compact schemes which do not require large overlap points. The authors in [60,61] computed the NSE by the sixth order compact scheme [33] for the interior points with eighth order central difference (CD8) scheme for boundary closure. The strategy required one ghost point on either ends of the subdomains, and the author claimed superior performance with the unbiased subdomain boundary closure. However, the resultant discontinuity at the subdomain boundaries has been shown as the source of spurious disturbances in [142]. Fang et al. [149] reported parallel computing using compact scheme for the interior, without any overlap point, and explicit subdomain boundary closure of the same order.

10.2.2. Comparison of NOHAP with CD8 subdomain closures for parallel computing

Here, the canonical 2D CE is solved using four stage Runge–Kutta scheme for time advancement, with OUCS3 scheme in the interior [1]. The subdomain boundary closures are performed by the GSA based NOHAP scheme [142,143] with the CD8 closure, as used in [60,61]. The governing equation for the 2D CE is given by,

$$\frac{\partial u}{\partial t} + c\cos\theta\frac{\partial u}{\partial x} + c\sin\theta\frac{\partial u}{\partial y} = 0; \quad \text{with;} \quad c > 0 \tag{195}$$

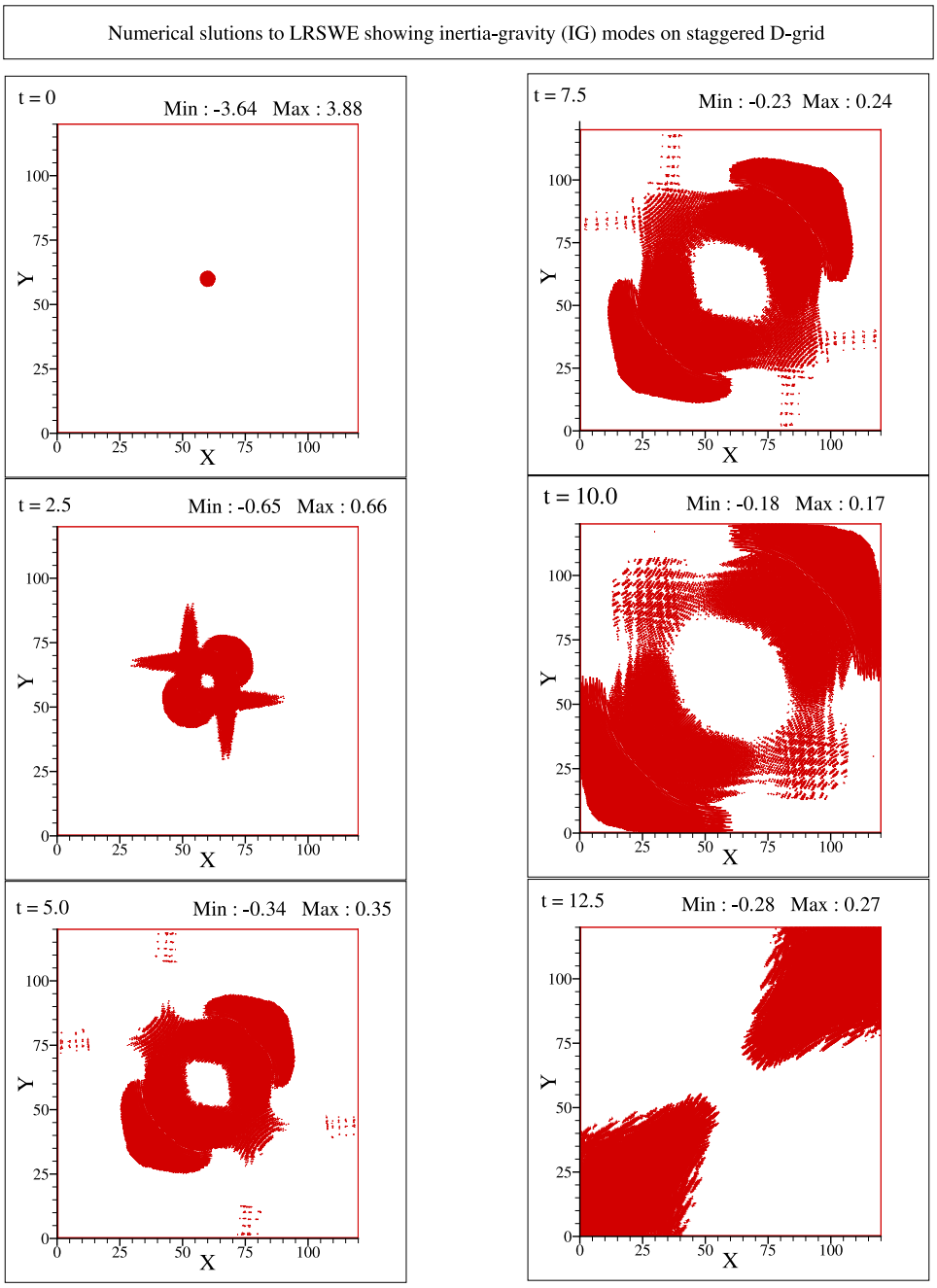

Fig. 57. Numerical solutions to LRSWE on D-grid at indicated time instants using RK_4 – OSCS and optimized interpolation schemes with $N_c = 0.2$, $k_o h = 1.40$ and $\alpha = 1.0$.

where c is the phase speed of the signal traveling at an angle θ, with respect to the x-axis. As usual in GSA, the unknown is represented by,

$$u(x, y, t) = \int\int\int \hat{U}(\omega_0, k_x, k_y) e^{i(k_x x + k_y y) - i\omega_0 t} dk_x dk_y d\omega_0$$

with the physical dispersion relation given by, $\omega_0 = k_x c \cos\theta + k_y c \sin\theta$, and the physical group velocity given by, $\vec{V}_{phys} = \hat{\imath} d\omega_0/dk_x + \hat{\jmath} d\omega_0/dk_y$.

For the sake of comparison, a double-periodic 2D domain is considered: $-1 \le x \le 19$ and $-1 \le y \le +1$; with 2001×201 equidistant points. To keep the analysis of the results easily tractable, we consider the case of $c = 0.1$ and $\theta = 0$, so that the physical phase speed and group velocity will only have the x-component. A time step of $\Delta t = 10^{-04}$ is chosen to compute the wave system with $N_c = 0.001$ to be small enough for the accuracy of computing by both the subdomain closure schemes. It is noted in [1,36] that CD8 scheme creates upstream propagating q-waves for wavenumbers in excess of $kh = 0.6507\pi$, while the OUCS3 schemes exhibits the same for $kh \ge 0.7664\pi$. To study the wave propagation, consider a Gaussian wave-packet as the initial condition given by,

$$u(x, y, t = 0) = e^{-25(x^2 + y^2)}$$

This initial wave-packet is symmetrically placed about zero wavenumber. The computed circular wave-packet is shown in Fig. 62.

In the top frames of Fig. 62, the propagation of the Gaussian packet is traced, when the subdomain closure is performed using CD8 scheme. The displayed results show the contours in the (x, y)-plane spanning the range from 10^{-8} to 10^{0}. In the top most frame, one notes the initial wave-packet, that travels to the right with minor distortion of the main packet, along with some q-waves originating when the signal traverses through the subdomain boundaries, caused by solution discontinuity due to mismatch between the interior (obtained by OUCS3 method of

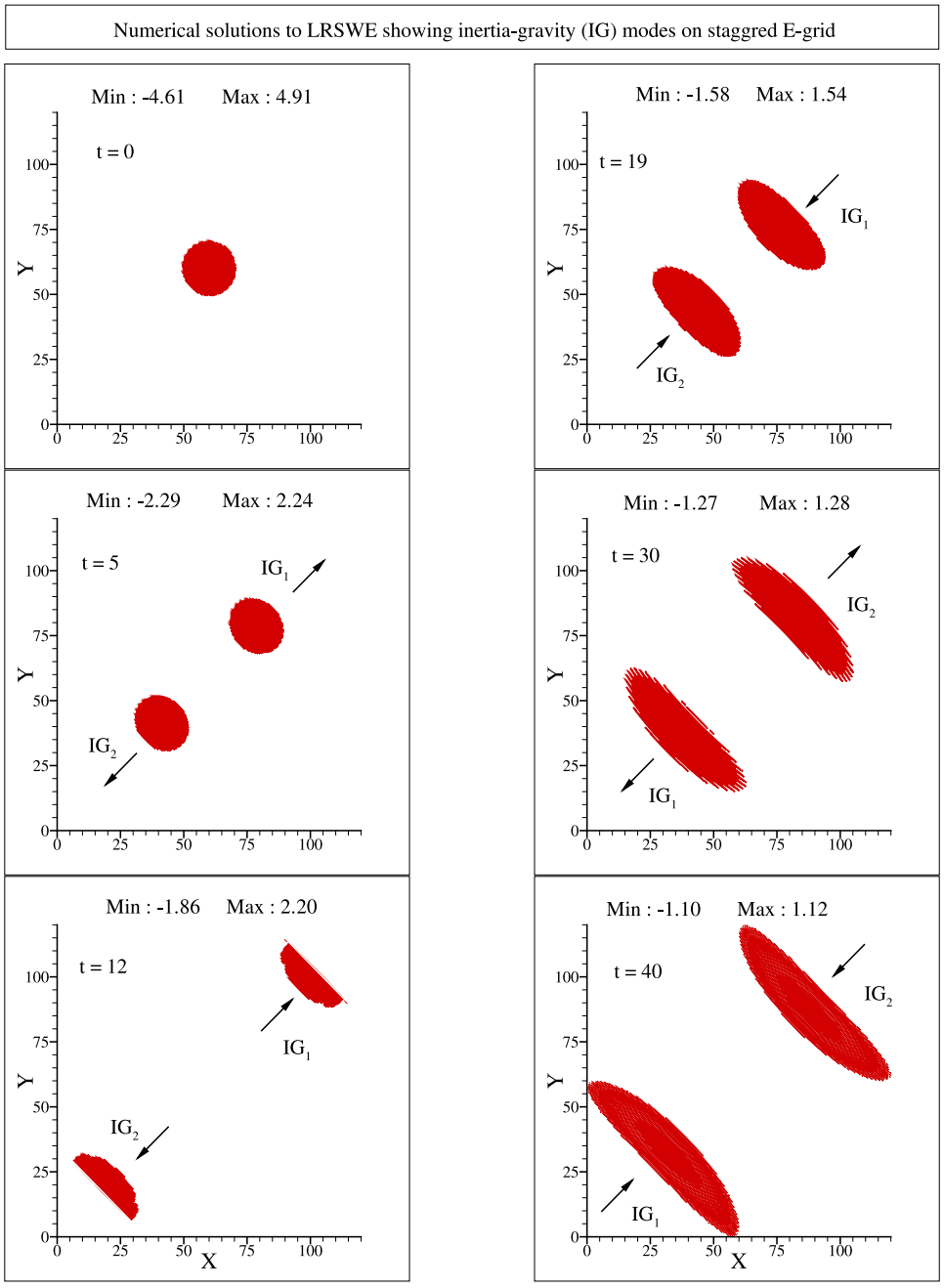

Fig. 58. Numerical solutions to LRSWE on E-grid at indicated time instants using RK_4 – OSCS and optimized interpolation schemes with $N_c = 0.2$, $k_o h = 1.40$ and $\alpha = 0.05$.

discretization) and the subdomain boundaries (obtained by the CD8 method). Such spurious waves created at very high wavenumbers are of very small magnitudes (of the order of 10^{-8}), and these remain trapped within the computational domain for this periodic problem. There are numerous such streaks noted over the full domain due to creation of the q-waves and their reflections and refractions at subdomain boundaries for the displayed solution at $t = 250$ and 300.

For the bottom two frames Eq. (195) is solved in the domains with subdomain boundary conditions obtained by using the GSA for the OUCS3 schemes as, $\{f'\} = \frac{1}{h}[C]\{f\}$, with the $[C]$-matrix entries given in [143]. However, the results shown in the bottom two frames of Fig. 62 for $t = 250$ and 300, one does not notice any q-waves for this subdomain closure strategy, which does not cause any solution discontinuity at the subdomain boundaries. Such discontinuities create Gibbs' phenomenon [150], which in turn creates the q-waves.

10.3. Analysis of 1D linear convection–diffusion–reaction equation with a realistic reaction term

Previous application focused on the linear convection–diffusion–reaction equation (LCDRE) with a constant source term has been performed for few commonly used numerical schemes which provided useful understanding of the problem [47]. It is however not representative of a flame problem. To be fully applicable and useful a necessary step is to recast a flame problem in a LCDRE type that is amenable to such a numerical analysis. The first objective of the following is to identify such a model. While using this model, here the popularly used Lax–Wendroff scheme using CD2 spatial discretization scheme is analyzed. This scheme has been used to study the linear CDE in detail for both 1D and 2D [151].

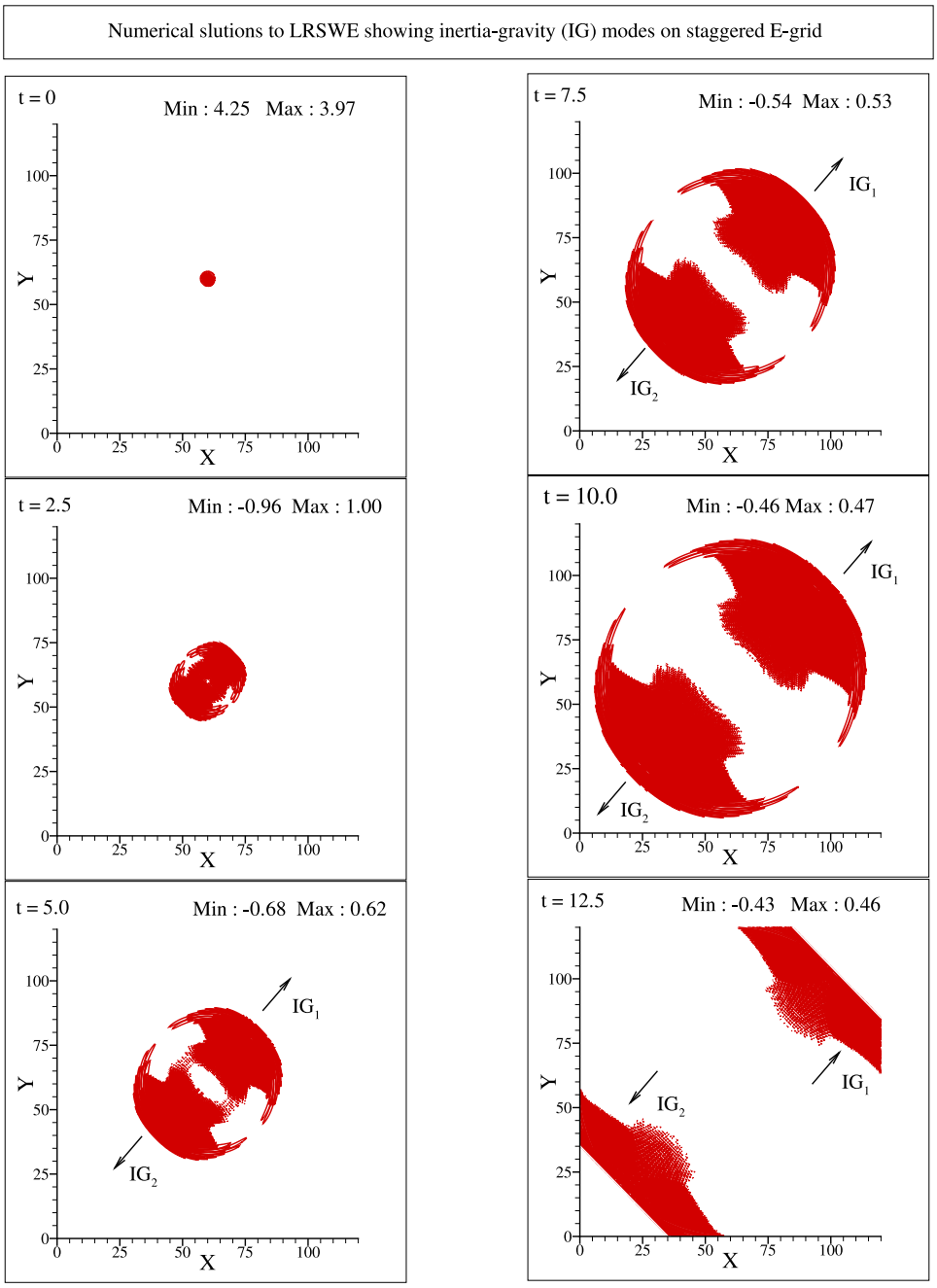

Fig. 59. Numerical solutions to LRSWE on E-grid at indicated time instants using RK_4 – OSCS and optimized interpolation schemes with $N_c = 0.2$, $k_o h = 1.40$ and $\alpha = 1.0$.

First, let us consider a 1D fully premixed flame front as the target for theoretical studies [71,152]; with adiabatic conditions; unity Lewis number for all species and a constant diffusivity, D. The transport equation describing such a problem can be reduced to the progress variable equation which reads [71]:

$$\frac{\partial \theta}{\partial t} + v\frac{\partial \theta}{\partial x} = D\frac{\partial^2 \theta}{\partial x^2} + \dot{\omega}_\theta \,, \tag{196}$$

where x, v are the spatial coordinate, the flame velocity, respectively, and $\dot{\omega}_\theta$ is the progress variable source term.

The progress variable source term is commonly modeled using an Arrhenius formulation, depending on the temperature field T and an activation temperature T_a: $\dot{\omega}_\theta \propto (1-\theta)\exp(-T_a/T)$ [71]. Another way to model such a reaction is to find analytical functions for θ and $\dot{\omega}_\theta$ which are solution of Eq. (196). In this spirit, Pfitzner et al. [153,154] proposed the progress variable source term as given by,

$$\dot{\omega}_\theta = \frac{(\rho_u s_L)^2}{\rho D}(m+1)(1-\theta^m)\theta^{m+1} \,, \tag{197}$$

where s_L stands for the premixed laminar flame speed and m is a model coefficient that can be tuned to match a reference Arrhenius-like chemistry model, and which will affect the laminar flame thickness. For any m value however, the laminar flame will always propagate at the laminar flame speed s_L specified in Eq. (197). Unlike the linear reaction source term of CDR equation used before in [47], this formulation vanishes for $\theta = 0$ as well as for $\theta = 1$ and peaks inside a thin reaction zone, thus behaving like a realistic combustion source term. However, such a model still renders impossible the linearization and therefore precludes from performing a straightforward GSA analysis for such a LCDRE.

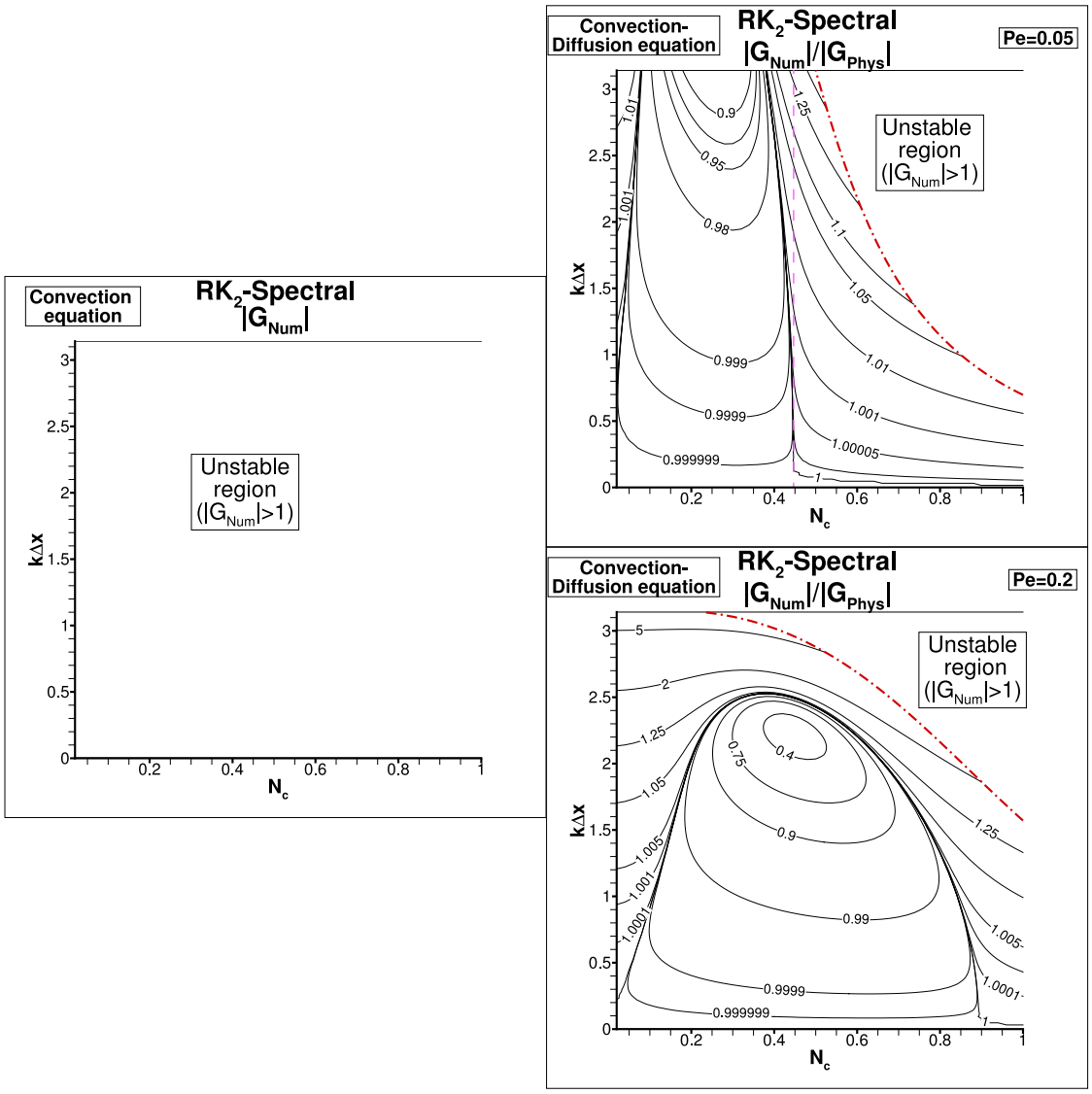

Fig. 60. Numerical amplification factor and ratio of numerical amplification factor to the physical amplification factor for the RK2-Fourier spectral scheme plotted in the $(N_c, k\Delta x)$-plane for linear convection and CDEs, respectively. For the latter, contours are plotted for two representative Peclet numbers- $Pe = 0.02$ and 0.2.

To circumvent this difficulty, while allowing a GSA analysis, expressions for the Fourier transform of $\widehat{\dot{\omega}}_\theta$ as a function of $\widehat{\theta}$ is needed. To do so, an approximate reaction term R is proposed. In doing so, it further is presumed that R is also the result of the convolution of a high-pass filter B with the progress variable front θ. That is:

$$R(x) = B(x) \star \theta(x)\,, \tag{198}$$

so that, $\widehat{R}$ is simply the product of $\widehat{B}$ with $\widehat{\theta}$ (their Fourier transport counterparts),

$$\widehat{R}(k) = \widehat{B}(k)\widehat{\theta}(k)\,. \tag{199}$$

Note that high-pass filtering of θ, results in a R profile, which goes to 0 in the fully burnt, as well as, in unburnt states and that peaks inside the flame front. Note also that if B is chosen so that R closely matches $\widehat{\dot{\omega}}_\theta$, their spectral behavior are expected to be the same and GSA can be performed by replacing $\widehat{\dot{\omega}}_\theta$ with the expression of $\hat{R}$ in Eq. (199).

In the following, for simplicity B is chosen to be a first-order Butterworth filter whose frequency response reads,

$$\widehat{B}(k) = \frac{B_0}{\sqrt{1 + \left(\frac{k_c}{k}\right)^2}}\,, \tag{200}$$

where k_c is the cutoff wavenumber and B_0 its gain for $k \to +\infty$. Note that, k_c and B_0 control the width and the amplitude of the approximated reaction peak and they can be tuned to match a given Pfitzner source term, as illustrated in Fig. 63(a) or more complex expressions if needed. For the specific case considered in Fig. 63, the Pfitzner source term for m = 0.2, peaks inside the flame front and decays to 0 in the fresh (left of the figure) and burnt (right) states, as expected. Furthermore, having $k_c = 265\ \mathrm{m}^{-1}$ and $B_0 = 3136\ \mathrm{s}^{-1}$, R clearly approaches $\dot{\omega}_\theta$ quite accurately. The computation non-dimensional Damkohler number, Da, for this problem is given by $Da = \frac{B_0 * h}{c}$. The Damkohler number is directly proportional to the amplitude of the approximated reaction peak, B_0. The largest errors arise near the cold boundary of the flame front and overall amount to less than 5% of the maximum source term value shown in frame (b) of Fig. 63.

Using this reaction term, the expressions resulting from the use of the Lax–Wendroff scheme using CD2 spatial discretization yield,

$$|G_{num}| = 1 - iN_c sin(kh) + 2[\xi cos(kh) - 1] + Da * Nc * \tau \tag{201}$$

where, $\tau = \frac{kh}{\sqrt{(kh)^2 + (k_c h)^2}}$. Here, and in comparison to the previous case with a linear reaction source term, an additional parameter τ expresses

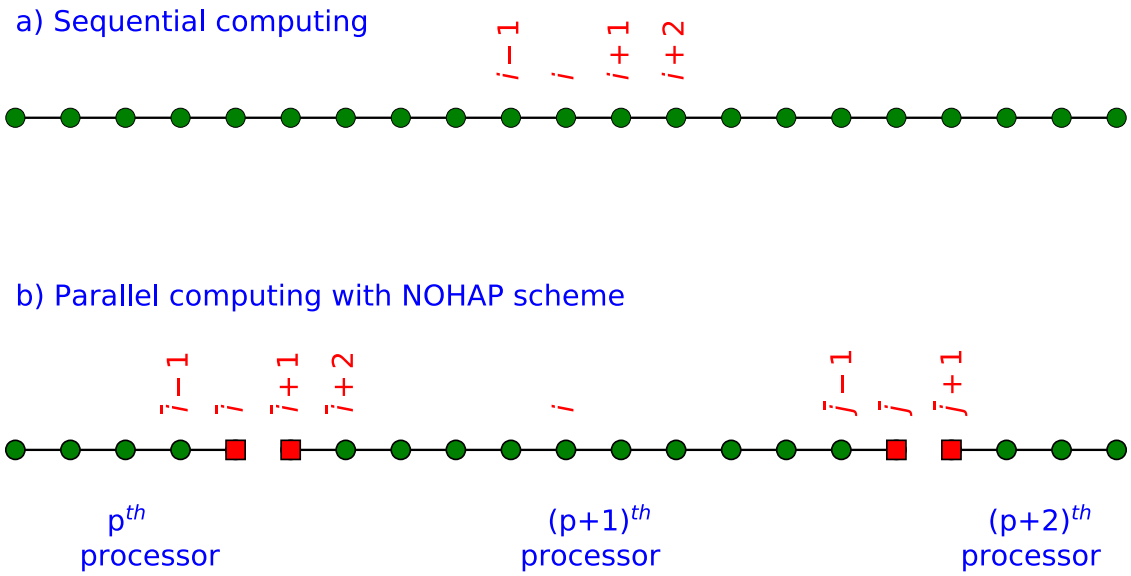

Fig. 61. A general representation of grid points distribution in the direction along which derivative is obtained by (a) sequential computing and (b) parallel computing with NOHAP scheme.

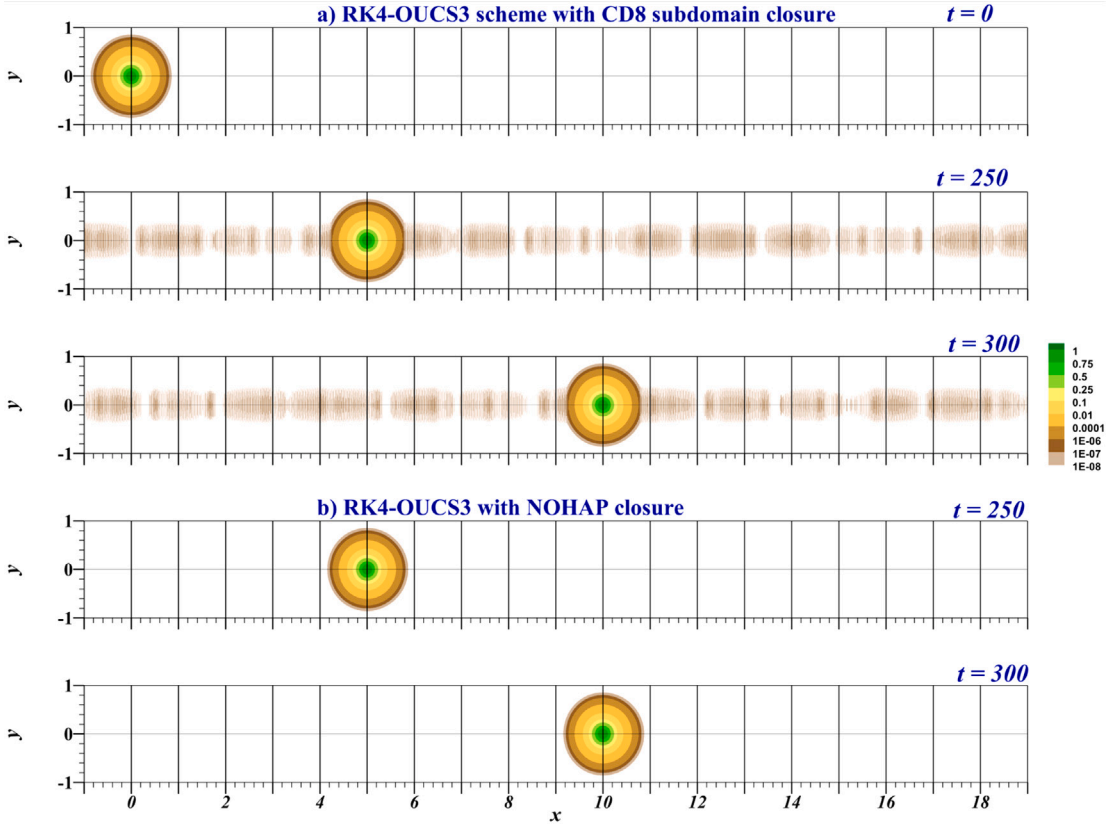

Fig. 62. Solution of Eq. (195) using RK4-OUCS3 scheme.

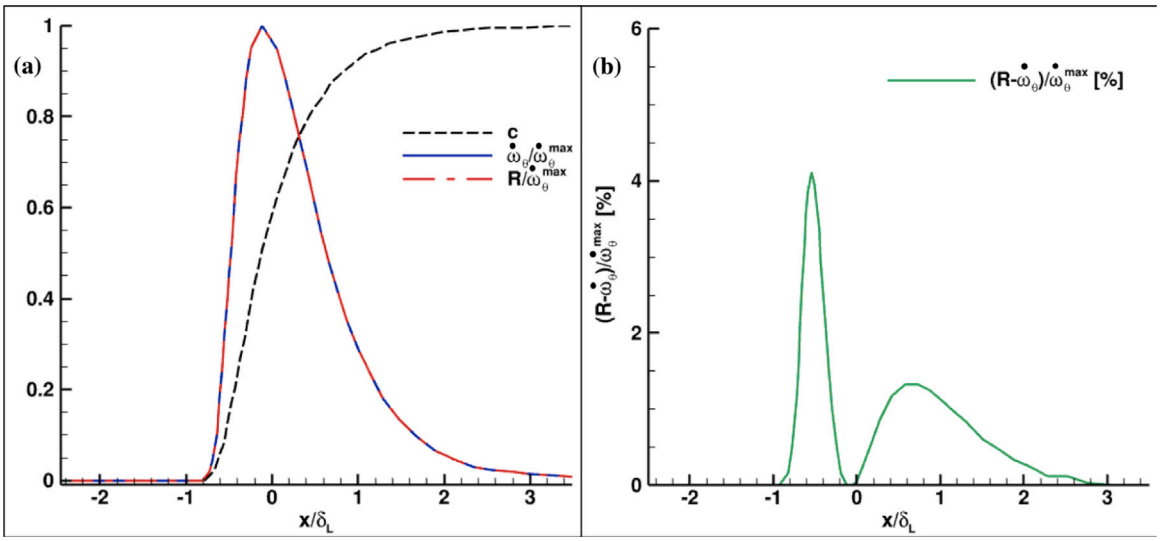

Fig. 63. Profiles of (a) progress variable θ (black, long dash), Pfitzner source term $\dot{\omega}_\theta$ (blue,solid), filtered approximation R (red, dash dot), and (b) error of the approximation (green, solid) from the simulation of a 1D propagating premixed flame. For visualization purposes, the last three quantities are normalized by the maximum value of $\dot{\omega}_\theta$ in the flame.

the wavenumber, k, grid spacing, h, and dependency on the cutoff wavenumber, k_c.

A direct consequence is that the non-dimensional phase speed reads,

$$\frac{c_{num}}{c} = -\frac{1}{khN_c} tan^{-1} \frac{sin(kh)N_c}{1 + 2\xi(cos(kh) - 1) + DaN_c\tau}, \tag{202}$$

while the non-dimensional group velocity can be expressed as functions of the real and imaginary parts of numerical amplification factor (G_r and G_i), as well as the derivatives with respect to (kh) noted (G_r' and G_i' respectively), yields:

$$\frac{V_{g,num}}{V_g} = -\frac{1}{N_c} \frac{G_r G_i' - G_i G_r'}{(G_r^2 + G_i^2)}. \tag{203}$$

Introducing the non-dimensional parameters (Pe, N_c &Da) in Eq. (201), and further simplifying, obtains

$$ln|G_{num}| = -\frac{\alpha_{num}}{\alpha}(kh)^2 Pe - \frac{B_{0_{num}}}{B_0}\tau * N_c * Da. \tag{204}$$

The non-dimensional numerical reaction coefficient can then be estimated by evaluating the above expression for $Pe = 0$,

$$\frac{B_{0_{num}}}{B_0} = -\left(\frac{ln|G_{num}|_{Pe=0}}{N_c\tau Da}\right). \tag{205}$$

The numerical amplification factor is noted to be the function of all the four non-dimensional parameters, kh, N_c, Pe and Da while the non-dimensional numerical reaction coefficient is determined by N_c and Da. By substituting the numerical reaction coefficient in Eq. (204), one obtains the expression for the non-dimensional numerical diffusion coefficient as,

$$\frac{\alpha_{num}}{\alpha} = \frac{ln|G_{num}|_{Pe=0} - ln|G_{num}|}{(kh)^2 Pe}. \tag{206}$$

Looking at the expression for the numerical amplification factor for this model with the Lax–Wendroff scheme using CD2 spatial discretization, the effect of the reaction source term can be directly observed by looking at Eq. (204). The reaction source term plays an additional role that is similar to a diffusion term observed before. The addition is here due to the dependency of τ, which is a function of k and k_c. Compared to the corresponding expressions for linear reaction source term, this addition is expected to affect the stability limits differently. Now, using these expressions the property charts can be made to analyze the numerical scheme and eventually be used to solve a reacting NSE problem.

The variable parameters in this model can be tweaked and changed to match other theoretical or practical combustion models and thus provide a gateway into analyzing realistic reacting problems using GSA.

10.4. Analysis and determination of simulation parameters for prescribed accuracy for the Lax–Wendroff method

GSA has been used recently to analyze the well known Lax–Wendroff scheme which can be considered as a DRP scheme. Soumyo et al. [151] have analyzed the Lax–Wendroff central difference scheme employed in the ABVP code, which is developed at CERFACS, with respect to the model convection–diffusion equation with an aim to understand the scheme's stability in solving the Navier–Stokes equations. Using GSA, the authors obtained property charts from which acceptable range(s) of simulation parameters are determined that satisfy numerical stability as well as having good dispersive properties. Further, the property charts are calibrated by solving the 2D Navier–Stokes equations for Taylor–Green vortex problem and the numerical behaviors are explained.

One of the recent developments involving GSA is its application in determining simulation parameters for performing numerical simulations of fluid flows with a prescribed accuracy for the explicit CD_2 based Lax–Wendroff (LW-CD_2) method [155]. The researchers developed a framework using GSA, for analyzing numerical schemes for their suitability for high fidelity simulations such as LES/DNS. This was achieved by first analyzing the LW-CD_2 method for 2D CDE using GSA. The optimal parameters for prescribed accuracy were then determined by minimizing the contribution to the diffusion error. In addition to this, the researchers derived and assessed two variants of the LW method for the CDE- (i) full scheme and (ii) applied to convection only terms and established the efficacy of the latter for the simulations.

The full LW scheme for the 1D CDE is given by [155]

$$u_j^{n+1} = u_j^n - \frac{N_c}{2}(u_{j+1}^n - u_{j-1}^n) + (Pe + \frac{N_c^2}{2})(u_{j+1}^n - 2u_j^n + u_{j-1}^n) - PeN_c D^3 u_j^n + \frac{Pe^2}{2} D^4 u_j^n \tag{207}$$

where N_c, Pe are the CFL and Peclet numbers, D^3 and D^4 are the third and fourth order derivative terms. From the full LW scheme, one notes added diffusive terms arising from the pure convection and diffusion terms. It is interesting to note a third order dispersive term due to the interaction between convection and diffusion terms and a fourth order

diffusion term which reduces the excessive numerical dissipation at high wavenumbers [155]. The second variant of the method is obtained by setting D^3 and D^4 terms to 0 in Eq. (207). Using 1D GSA, the authors noted minimal differences between the two variants and established the efficacy of the second variant for computing.

In order to determine the optimal parameters the LW method applied to convection terms was analyzed using GSA. The scheme for the 2D CDE is given by

$$u(t+\Delta t) = u(t) - \Delta t\left(c_x\frac{\partial u}{\partial x} + c_y\frac{\partial u}{\partial y}\right) + \alpha\Delta t\left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}\right) + \frac{(\Delta t)^2}{2}\left(c_x^2\frac{\partial^2 u}{\partial x^2} + 2c_xc_y\frac{\partial^2 u}{\partial x\partial y} + c_y^2\frac{\partial^2 u}{\partial y^2}\right) \tag{208}$$

where CD_2 schemes are used for the derivatives appearing above. Comparing this scheme with its 1D equivalent given in Eq. (207), the authors noted the additional presence of a mixed derivative term which leads to asymmetrical numerical properties.

Optimal parameters were determined by finding out the combination resulting in a prescribed diffusion error i.e. by finding region(s) where $|1 - |G_{num}/G_{phys}\ || < \epsilon$. ϵ is a chosen tolerance parameter with a lower value denoting higher accuracy and vice versa. The authors identified two values of ϵ namely 10^{-4} and 10^{-6} with the latter being a representative of finer simulations such as LES or unresolved DNS.

Finally, the optimal parameters thus determined are established by the authors by solving the 2D flow inside a square LDC at a Reynolds number $Re = 10{,}000$. These simulation demonstrates an excellent agreement with the benchmark results by capturing the evolution of a transient triangular core vortex.

Due to the generic nature of the framework, this process can be employed for any other numerical schemes resulting in their efficient application for fluid flow simulations.

11. Beyond GSA: nonlinear instabilities and N-mode analysis

Previous sections are devoted to the analysis of the discrete linearized solution via GSA. Thanks to its completeness (it accounts for space discretization, time integration and boundary conditions) and versatility, it has been shown that GSA is a very powerful tool that is able to recover the results of the classical single-monochromatic-plane-wave disturbance analysis, but also to account for polychromatic collective effects, such as local focusing due to dispersive effects and side-band instabilities due to grid resonance between large and small disturbances. A common feature of these collective polychromatic effects is that they can lead to a very large local growth of the energy of the solution. In such a case, the cornerstone hypothesis (78) is no longer valid, and nonlinear mechanisms may become dominant when nonlinear problem are addressed. Therefore, numerical schemes that are stable according to the linearized analysis may be observed to be unstable.

In order to investigate the nonlinear stability properties of numerical schemes, several approaches have been proposed, the most popular being the N-mode analysis, that will be discussed below.

The method, that originates in the discrete dynamical system theory, consists of finding finite discrete sets of waves that span a closed exact solution of the non-linear discrete problem. The associated amplitude evolution equations generate a nonlinear dynamical system, whose stability is investigated numerically in practice. Since it accounts for resonant wave interactions among a closed set of modes, this method can also be interpreted as a kind of discrete wave turbulence-weak turbulence theory (see e.g. [156]) for numerical scheme analysis. This approach was pioneered in the late 1970s and 1980s, with application to a broad range of problems and numerical methods, as illustrated in Table 2. It is now illustrated considering the model nonlinear advection–diffusion equation, that is an extension of the classical

Table 2
Survey of investigations related to numerical nonlinear instabilities. Kdv: Korteweg de Vries equation; VdP: Van der Pol equation.

Ref.	Model	Type	Envelope/side-band analysis
Briggs et al. [72]	1D advection	1-2-3-mode	✓
Sloan [90]	1D KdV	1-mode	✓
Aoyagi [93]	1D advection	3-mode	✓
Cai et al. [157]	VdP	2-mode	
Fornberg [83]	1D advection	1-mode	
Herbst [158]	1D Schrödinger	1-2-mode	
Hsia et al. [89]	1D Burgers	2-mode	✓
Newell [84]	1D heat, 1D KdV	2-mode	✓
Sloan et al. [74]	1D advection	1-2-mode	✓
Vadillo et al. [159]	1D advection	2-mode	✓

Burgers equation:

$$\frac{\partial u}{\partial t} + \frac{\theta}{2}\frac{\partial u^2}{\partial t} + \{(1-\theta)u + c\}\frac{\partial u}{\partial x} = \nu\frac{\partial^2 u}{\partial x^2} \tag{209}$$

where c is related to the base flow and the coefficient θ is a weighting parameter between the conservative and the quasilinear expressions of the nonlinear convection term. The general form of the corresponding semi-discrete equation found considering finite difference schemes with a centered stencil with $(2M+1)$ grid points for convection along with the 3-point second order accurate scheme for second-order space derivative is

$$\dot{u_l} + \frac{1}{\Delta x}\left(\frac{\theta}{2}\sum_{j=-M}^{M} a_j u_{l+j}^2 + \{(1-\theta)u_l + c\}\sum_{j=-M}^{M} a_j u_{l+j}\right) - \frac{\nu}{\Delta x^2}\sum_{j=-1}^{1} b_j u_{l+j} = 0 \tag{210}$$

where $u_l(t)$ denotes the semi-discrete solution at the lth grid point. The coefficients a_j and b_j are related to the discretization of the first-order and second-order spatial derivatives, respectively. Eqs. (209) and (210) exhibit a quadratic nonlinearity, therefore closed set of modes found considering the adequate complex roots of unity. As an illustration, the 1-Mode, 2-Mode and 3-Mode solutions are defined as:

$$u_l(t) = \begin{cases} A(t)\exp(2\pi il/3) + A^*(t)\exp(-2\pi il/3) & \text{1-mode} \\ A(t)\exp(\pi il/2) + A^*(t)\exp(-\pi il/2) + B(t)\exp(\pi il) & \text{2-mode} \\ A(t)\exp(\pi il/3) + B(t)\exp(2\pi il/3) + C(t)\exp(\pi il) & \\ \quad + A^*(t)\exp(-\pi il/3) + B^*(t)\exp(-2\pi il/3) + C^*(t)\exp(-\pi il) & \text{3-mode} \end{cases} \tag{211}$$

where $A(t), B(t)$ and $C(t)$ are the complex amplitudes of the modes under consideration, and the *asterisk* denotes the complex conjugate. The time-continuous evolution equations are recovered by inserting one of the solutions given in (211) into (210). For the sake of illustration, taking $M = 3$, the equations for the 1-mode solution are

$$\dot{A}(t) + \left(i\sqrt{3}(a_1 - a_2)\frac{c}{\Delta x} + \frac{3\nu}{\Delta x^2}\right)A(t) - \frac{i\sqrt{3}}{2\Delta x}(a_1 - a_2)(2-3\theta)A^{*2}(t) = 0 \tag{212}$$

$$\dot{A}^*(t) + \left(-i\sqrt{3}(a_1 - a_2)\frac{c}{\Delta x} + \frac{3\nu}{\Delta x^2}\right)A^*(t) + \frac{i\sqrt{3}}{2(\Delta x)}(a_1 - a_2)(2-3\theta)A^2(t) = 0 \tag{213}$$

while the 2-mode solution yields

$$\dot{A}(t) + \left(2i(a_1 - a_3)\frac{c}{\Delta x} + \frac{2\nu}{\Delta x^2}\right)A(t) - \frac{2i}{\Delta x}(a_1 - a_3)(1-2\theta)A^*(t)B(t) = 0 \tag{214}$$

$$\dot{A}^*(t) + \left(-2i(a_1 - a_3)\frac{c}{\Delta x} + \frac{2\nu}{\Delta x^2}\right)A^*(t) + \frac{2i}{\Delta x}(a_1 - a_3)(1-2\theta)A(t)B(t) = 0 \tag{215}$$

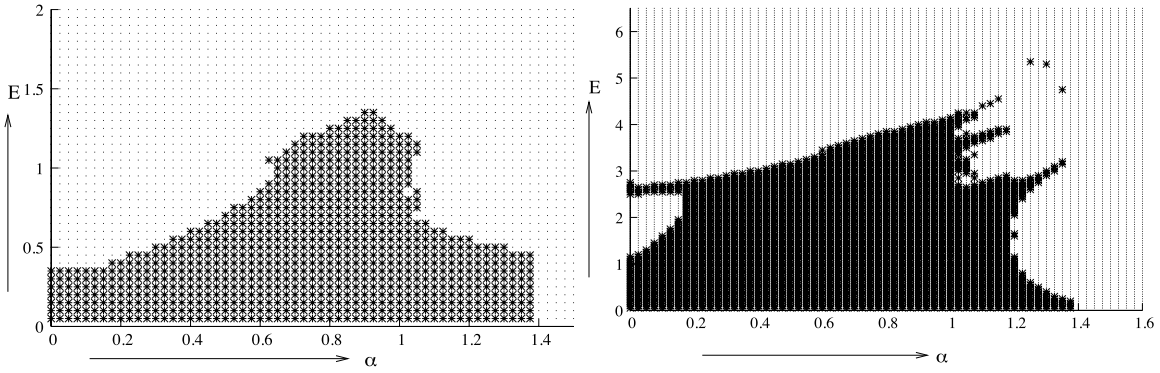

Fig. 64. Stability regions using the RK3 time integration scheme in the (E, α) for the 1-mode analysis, for different values of the weighting parameters θ. Left: $\theta = 0$, Right: $\theta = 1$. Gray area corresponds to stable solutions. Courtesy of Dr. S. Pandit.

$$\dot{B}(t) + \frac{4\nu}{\Delta x^2} B(t) + \frac{2i}{\Delta x}(a_1 - a_3)(1-\theta)\left(A^2(t) - A^{*2}(t)\right) = 0 \tag{216}$$

The last step consists of applying the selected time integration schemes to these evolution equations to obtain a fully discretized problem. It is worth noting that this approach does not incorporate the boundary conditions.

Due to the genuinely nonlinear character of the resulting dynamical systems, the possibility to perform a purely analytical analysis of the stability is restricted to a few particular cases. In practice, the stability is investigated by prescribing an initial condition, i.e. giving a initial value for each mode, and then integrating numerically the system over a arbitrary number of time steps to check if the energy of the solution has grown over an arbitrary threshold or not. If the energy is larger than this arbitrary value, the system is said to be unstable. In the opposite case, it is considered as stable. Using the Vaschy–Buckingham theorem, one can see that the stability depends on several parameters, i.e. the initial energy E, the Courant number $\alpha = c\Delta t/\Delta x$ and the cell Reynolds number $R_c = c\Delta x/\nu$. It is important noting that E is an important parameter, since the non-linear analysis is based on finite amplitude disturbances, while the usual linearized one considers asymptotically small perturbations.

Some typical results are illustrated in Figs. 64–66, which display the stability region in the (E, α) plane obtained using Tam's DRP scheme to discretize the convection term coupled to several time integration schemes and different numbers of modes. A first observation is that the topology stability regions is much more complicated than those observed in the linearized cases, and that non-simply convex stable regions are present. The second point is that the stability is directly tied to the initial energy of the disturbances. This is the reason why linear polychromatic mechanisms discussed above that may lead to a local rise of the energy are very important: they can trigger nonlinear instability due to finite-amplitude wave resonance. One can see that there is an initial energy threshold E_{max} above which the solution is linearly stable but nonlinearly unstable. The global picture of the stability analysis is schematized in Fig. 67.

12. Summary and conclusions

A review of the GSA of numerical methods has been presented using linear and non-linear equations, and is compared with analysis by von Neumann and semi-discrete analyzes. Von Neumann analysis is restrictive as it provides stability characteristics without the accuracy of the methods, as demonstrated with the CE which is neutrally stable. Furthermore, this does not provide the phase/propagation speed of the numerical solution – a vital parameter of propagation problems encountered in fluid dynamics and many disciplines of science. GSA, on the other hand, is an analysis tool revealing the behavior of numerical methods by correctly identifying numerical dispersion relation, handling non-periodic problems due to its generalized approach.

Wave propagation problems involve spatial and temporal terms governed by the fundamental properties of the physical dispersion relation and the physical amplification factor. These depend on the governing equation and the boundary conditions, dictating the evolution of the solution. For accurate solutions, the adopted methods having numerical dispersion relation and amplification factor, must follow the corresponding physical properties. This is the basis of dispersion relation preserving (DRP) schemes. The correct DRP properties are obtained when both temporal and spatial discretizations are analyzed together as opposed to the semi-discrete analysis where only spatial discretization is considered. The former is termed as the Γ-form and the latter as the Π-form analysis.

The GSA performs Γ-form analysis using Fourier–Laplace transform valid also for non-periodic problems. The critical distinction between Γ and Π-forms is demonstrated using leap-frog and CD_2 scheme for the 1D CE. The numerical phase speeds (c_N) computed from the Γ-form analysis incorporates correctly the spatial and temporal discretizations enabling comparison between the numerical and physical modes. In contrast, the Π-form analysis based on spatial discretization obtains an incorrect c_N.

The global spectral resolution of classical explicit and high accuracy compact schemes are represented for first and second order spatial derivatives in Section 3. For first derivative, the effectiveness is obtained as k_{eq}/k with k as the wavenumber and the numerical derivative is determined by k_{eq}, and k_{eq}/k is in general, a complex quantity. The real part of k_{eq}/k represents the accuracy of the numerical method in obtaining the spatial derivative. The imaginary part signifies added numerical dissipation or anti-diffusion when its sign is either negative or positive, respectively. Schemes showing anti-diffusion are undesirable for computing as this leads to solution blowup.

From Fig. 2, one infers that increased order increases resolution of the scheme. However, near-spectral accuracy can be achieved by lower order, optimized implicit schemes, as in case of the OUCS3 scheme. The effects of anti-diffusion/dissipation are noted for implicit schemes in Fig. 3 for non-periodic problem which necessitates one sided boundary closures. Considering solution propagation from left to right, analysis show anti-diffusion to be present near the inflow boundary, whereas dissipation is noted near the outflow boundary. It is noted that anti-diffusion cannot be eliminated for one-sided schemes used at the inflow. However, a careful design of near-boundary closure schemes can drastically reduce it, as shown in Fig. 3 for the OUCS3 and NCCD schemes. The effectiveness of second derivative discretization for various schemes are summarized in Fig. 4. Unlike the first derivative, the effectiveness for second derivative does not become zero at the

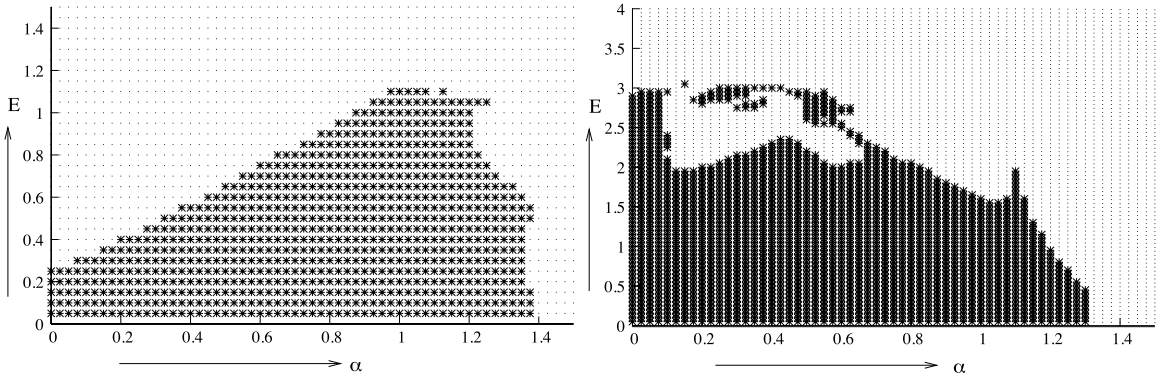

Fig. 65. Stability regions using the RK3 time integration scheme in the (E, α) for the 3-mode analysis, for different values of the weighting parameters θ. Left: $\theta = 0$, Right: $\theta = 1$. Gray area corresponds to stable solutions. Courtesy of Dr. S. Pandit.

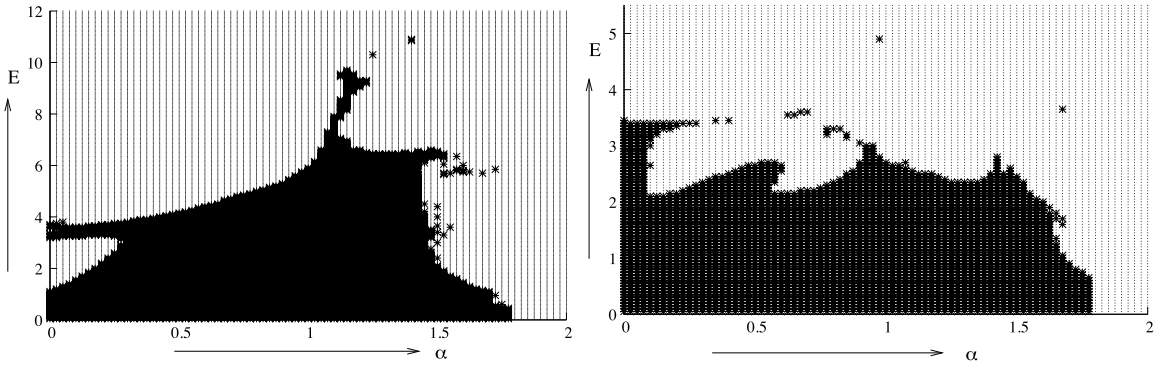

Fig. 66. Stability regions using the RK4 time integration scheme in the (E, α) for the 1-mode analysis (left) and the 3-mode analysis (right), for $\theta = 1$. Gray area corresponds to stable solutions. Courtesy of Dr. S. Pandit.

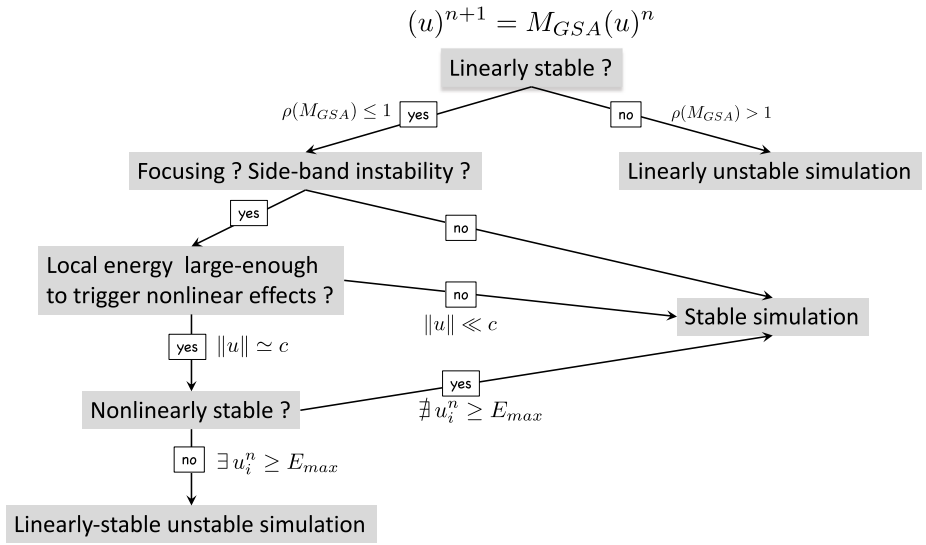

Fig. 67. Schematic view of the full stability analysis for numerical scheme.

Nyquist limit, implying diffusion to be present for higher wavenumbers at a reduced value.

The importance of the Γ-form analysis is shown using the 1D CE in Section 4. This model equation enables assessment of the accuracy of numerical schemes, as the exact solution is non-dispersive and non-dissipative. For any numerical scheme, one associates a numerical amplification factor (G) and a numerical dispersion relation. The former describes whether the solution is dissipative/growing, while the latter indicates dispersion effects. Using GSA, a governing equation for error is developed in Eq. (60), which provides the metrics driving the error dynamics. Unlike the von Neumann analysis indicating only the stability/instability for periodic problems, it is worth noting that GSA provides dispersion and dissipation errors for all resolved scales.

To illustrate the space–time analysis, the two time-level, four stage Runge–Kutta method (RK_4) is chosen for time integration along with different explicit central and compact difference schemes for the spatial derivative discretizations. In Figs. 6 to 8, G, c_N and V_{gN} are plotted for different schemes using GSA and an analysis employing incorrect

dispersion relation. To ascertain the correctness of the dispersion relation obtained by GSA, numerical tests have been performed which validates it in Fig. 9. Furthermore, the property charts signify the typical aspect of computing that c_N is a function of k. One notes that among the compared schemes, the OUCS3 scheme provides higher accuracy, which can be achieved with lower order scheme designed optimally.

In Section 5, the roles of multi time-level methods, in terms of three time-level schemes, are discussed using GSA. A typical attribute of these methods is the presence of at least one spurious numerical mode. Property charts are presented for explicit central and compact difference schemes for the Adams–Bashforth AB_2 scheme in Figs. 12 to 14. From these charts it is noted that the spurious mode is always present in such computations and one never obtains higher accuracy for any computational parameters. In Fig. 17, the evidence of numerical mode is established via an experiment using solution of the CE, validating the findings of GSA.

In Section 6, the aspect of nonlinear numerical instability (not accounted by the linear analysis) is discussed in the context of a nonlinear advection equation as the governing equation. It is hypothesized that the origins of non-linear instability are due to either focusing of energy due to dispersion errors or collective interactions and side band instabilities due to non-uniform base flow. The former is explained using caustics theory for geometric optics where a local concentration of huge amount of energy may occur. This can be also explained using focusing from GSA by noting that c_N is dispersive for the linear CE prompting one to replace classical monochromatic plane wave analysis with GSA. Furthermore, numerical experiments conducted for the linear CE [87] correctly identified the local blow-up of error from GSA due to focusing of dispersive errors. Also, a criterion for identifying likely cause of spurious caustics based on the extrema of V_{g_N} is proposed. Another linear error growth mechanism for polychromatic solutions exists when the base flow is sinusoidal and it is attributed to the side-band instabilities due to resonance between base flow and fluctuations. Such mechanisms have been observed in the Benjamin–Feir instability in free surface wave dynamics [94] and wave-train instabilities [95].

While the study of nonlinear instabilities is vital for understanding the overall error dynamics, the role of linear analysis cannot be underestimated during the early stages of disturbance evolution. This is true in the context of physical instabilities, as recent high accuracy simulations [160–164] of transition induced by deterministic wall and free stream excitation for zero-pressure gradient boundary layer shows the spatio-temporal wave front (STWF) to cause transition. It is to be noted that although STWF originates due to a linear process, the action of nonlinearity causes the flow to transition and eventually become turbulent. It has also been noted in the other canonical problem of Rayleigh–Taylor instability the STWF is caused due to pressure pulses [108,165,166]. Maddipati et al. [167] have also shown from a linear analysis that STWF plays a major role in 2D flows, as compared to 3D flows — analogous to Squires' theorem stated for normal mode analysis.

In Sections 7 and 8, focusing mechanisms for the nonlinear 2D NSE are explained using GSA following the analysis of linear CE and CDE. Focusing is the violent concentration of energy for selective wavenumbers which leads to abrupt breakdown of the numerical solution. This phenomenon has been reported by Phillips and other researchers in weather forecasting, who observed the smoothly progressing simulation to blow up suddenly without any indication. Prevailing theories attempted to explain focusing due to nonlinear instabilities. However, numerical results show focusing due to also a linear mechanism explained by GSA [49,57].

Three different mechanisms have been identified for focusing of the 1D linear CE: due to instability at near boundary nodes; solution discontinuity and chosen numerical discretization. The first cause is demonstrated in Figs. 22(a) and 22(b), where one notes instability in the entire domain with Fig. 22(a) showing the error at the inflow boundary. The corresponding property charts are displayed in Figs. 19 and 20. The scale selection of error is confirmed from Fig. 22(c), which corroborates well with the spectral bandwidth of the initial solution and GSA. The focusing due to solution discontinuity is demonstrated in Fig. 23(a), which corresponds to the property charts in Fig. 23(c).

A mechanism of focusing is shown for the nonlinear NSE due to reflection of q-waves, explained by GSA. These spurious upstream propagating waves occur at higher k, revealed by GSA. Results in Figs. 24 and 27, show the q-waves traveling in the upstream direction during the simulation of convection of a shielded vortex. It is noted that the computations in Fig. 24 does not display focusing, whereas the results in Fig. 27 does. It is attributed to the generation of large amount of q-waves which leads to numerical instability.

Another focusing mechanism for the NSE has been identified by GSA of CDE as the model equation [49]. Here, focusing is due to the creation of anti-diffusion $\alpha_N < 0$, which indicates spurious concentration of energy. It is noted that the analysis shows focusing as either due to the error due to diffusion discretization or errors due to combined convective–diffusive discretizations. Numerical results based on the solution of the NSE for a high Reynolds number flow inside a square lid driven cavity and associated GSA analysis presented in Figs. 32–34 and Figs. 35–37 demonstrate two mechanisms of focusing. These results corroborate very well with the GSA, and shows the focusing in this case to be governed by a linear mechanism. The general nature of focusing is further established by demonstrating it for cases involving upwind convective term discretization and three time-level methods in Figs. 38, 39 and Figs. 40–42, respectively. In addition to demonstrating focusing, a remedy to cure focusing is also proposed by filtering, and results shown in Figs. 43–45, with the magnitude of filtering decided by GSA. The one-to-one correspondence noted between the solution of 2D NSE and GSA of linear 2D CDE further establishes the utility of GSA.

Another problem of focusing is studied using GSA for the linearized rotating shallow water equations in Section 9 for different grid strategies. Results establish the necessity of q-waves for triggering focusing. The results shown in Figs. 48, 49 for the collocated Arakawa grid shows focusing, whereas staggered grid cases (Arakawa, B-E) do not display focusing, as confirmed from GSA. It is noted that staggering alters numerical dispersion relation which lowers/removes the q-waves. Staggering also introduces numerical dissipation which helps remove the focusing.

Recent developments arising out of GSA are described in Section 10, on the aspects of analysis of numerical methods and design of peta- and exa-scale HPC by using compact schemes. In the former, the much touted DNS of homogeneous isotropic turbulence by Fourier spectral — RK2 method is analyzed. There are a continuous stream of papers in the literature, with the common feature that all of these modify the NSE by adding explicit forcing term, in addition to hyperviscosity (some even add hypoviscosity) terms. Presented analysis clearly demonstrate that such claims of performing DNS is not only exaggerated, but are completely misplaced. This has been made possible by the GSA reported earlier for the CE [22] and a detailed analysis here for CDE. The utility of GSA is demonstrated via the design of subdomain boundary closure for the high accuracy compact schemes for HPC. The parallelization error caused in other methods like that is used in Schwarz domain decomposition method [140] has been completely eliminated, up to machine precision, by the design of the closure scheme by the GSA. Another extension of GSA has been demonstrated for the problem of combustion and flame propagation via the nonlinear model for the source as a convolution term for the CDR equation.

While GSA is a powerful analysis tool which accounts for polychromatic effects such as local focusing due to dispersion and side-band instabilities, a large growth of local error is a common feature of such cases thereby invalidating the linearity hypothesis. This necessitates the investigation of nonlinear stability properties of schemes and this is

emphasized in Section 11. A popular analysis method, namely the N-mode analysis, whose origins are based on dynamical systems theory is discussed, and its application is illustrated with the example of nonlinear viscous Burgers' equation. The fundamental operating principle of this approach is the prescription of an initial condition for each mode with the numerical system then being integrated over an arbitrary number of steps to check if the energy of the solution grows above an arbitrary threshold value. This determines whether the system is nonlinearly stable or unstable. Typical results are shown in Figs. 63–65, which show the complex topology of the stability regions with that non-simple convex stable regions present. Also, it is noted that the stability is directly tied to the initial energy of the disturbances which can trigger nonlinear instability by finite-amplitude wave resonance. The global picture of the stability analysis is schematically summarized in Fig. 67. Dynamical system approach to receptivity is very recently presented comprehensively in the book by Sengupta [163] for fluid flows, which has a direct analog to numerical simulation.

13. Perspectives of GSA

The discussion in the present work concerns about uniform grids, GSA has also been extended to non-uniform grids where high accuracy compact schemes have been designed and analyzed [168,169]. The GSA can be further extended to non-uniform grids by studying problems such as focusing in near future.

The different topics and tools discussed in the present review can be supplemented by another ways to scrutinize the features of numerical methods in fluid mechanics and to characterize numerical errors, Which are still rare in the literature. A first example consists of performing a symmetry analysis of the discrete system, the key point being that symmetries of the original continuous equations should be preserved by the numerical methods. An important point is that these preservation properties must be satisfied at the discrete level at finite Δx and Δt. Therefore this approach is not tied to the order of accuracy of the scheme. The issue of deriving numerical schemes that preserve the Lie group of one-parameter symmetries of the continuous equations was pioneered in the 1990's by Dorodnitsyn and other Russian researchers [170,171]. Some results have been obtained since the publication of first seminal works for several physical models, including the heat equation, shallow-water equations, wave equations and the NSE, e.g. see [172–179]. This topic was renewed in the field of CFD by Verstappen and Veldman [180,181] followed by many researchers, e.g. see [182–188], whose researches aim at deriving stable and accurate schemes that will preserve inviscid invariants of the INSE, namely the kinetic energy supplemented by the vorticity in 2D and helicity in 3D.

It is observed that the preservation of these quantities is associated to the skew-symmetry of the convective and pressure-dependent terms in the continuous equations, so that skew-symmetric schemes must be designed. It is observed that the preservation of some key symmetries leads to dramatic improvement of the efficiency of the schemes, by preventing the occurrence of spurious source terms in the discrete evolution equations of the physical invariant quantities. As a matter of fact, preserving the energy and the vorticity or the helicity amounts to controlling the L_2 norm of the solution and the rotational part of its gradient, leading to improved stability properties, without relying on artificial viscosity or other dissipative techniques.

This can be understood by invoking extensions of Noether theorem to discretized equations [189–191], which bridges between symmetries and the existence of conserved quantities. Some extensions have been proposed for compressible flows, e.g. [188,192–198], with the additional problem of searching for schemes that preserve linear and quadratic inviscid invariants and the second law of thermodynamics at the same time. It is worth noting that most of the results dealing with energy-preserving schemes are only related to semi-discrete analyses with continuous time-integration. The definition of energy-preserving time integration methods is more difficult, since the NSE do not have a friendly Hamiltonian formulation for practical CFD applications, and therefore efficient strictly symplectic time-integration methods have not been designed for them. Several energy-preserving schemes and pseudo-symplectic schemes have been proposed for incompressible flows, e.g. [199–201], but preservation of other inviscid invariants is not guaranteed. Kinetic energy-preservation being a property of continuous incompressible Euler equations, it is also possible to recast the discrete energy-preservation requirement as a time-reversibility property of the numerical method [199,202]. A striking result is that, at present time, no existing numerical scheme has been observed to be able to preserve all symmetries and all linear and non-linear (inviscid) invariants of the NSE (looking at the Lie-group of continuous one-parameter symmetries, which includes different scaling and generalized Galilean invariance).

The second example is related to the possible unphysical appearance of some specific events in the discrete solutions [203,204], which is sometimes referred to as structural stability of the numerical method, since the discrete solution admits solutions that are not supported by the continuous equations, but which mimic physical phenomena observed in fluid mechanics or other fields of physics. Within the discrete dynamical system framework, these spurious states can be described as spurious numerical attractors, that can be either steady or unsteady, stable or unstable. The case of the rise of spurious caustics due to numerical dispersive errors is an example, but more phenomena have been studied in the past, among which:

- The existence of spurious steady states, e.g. [205–209]. In the general framework of GSA, a steady state is derived from Eq. (26) as

$$(I_d - M_{GSA})\mathbf{u}_{steady} = (I_d - M_{GSA})\tilde{\mathbf{s}} = (I_d - M_{GSA})A^{-1}\mathbf{s} \qquad (217)$$

  and one can see that spurious discrete steady states can appear due the structure of the matrix A^{-1} which is tied to spatial discretization, or to an inaccurate inversion of the matrix $(I_d - M_{GSA})$ by the iterative method or the time marching method used.
- the existence of energy-bounded periodic [210] or chaotic solutions, e.g. [203,211–213]
- the existence of spurious waves [214], including very slowly decaying or self-sustained spurious solitons, e.g. [215,216] or unphysical shock wave propagation [217]

The last example deals with a change in the nature of physical instabilities due to numerical errors. This issue was raised in [111], where it is shown that numerical errors may lead to a change in the absolute/convective nature of hydrodynamic instabilities [218] in numerical simulations. A key change in the analysis compared to GSA is here that the solution is physically unstable and that its energy is algebraically or exponentially growing in both continuous and discrete systems, but the instability mechanisms may exhibit different features.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

[1] Sengupta TK. High accuracy computing methods: Fluid flows and wave phenomena. New York, USA: Cambridge Univ. Press; 2013.

[2] Richardson LF. Weather prediction by numerical process. U.K.: Cambridge University Press; 1922.

[3] Richardson LF. The approximate arithmetical solution by finite differences of physical problems involving differential equations, with an application to the stresses in a masonry dam. Philos Trans R Soc Lond A 1910;210:307–57.
[4] von Neumann J, Richtmyer RD. On the numerical solution of partial differential equations of parabolic type. Los Alamos Rept. Series A LA-657, 1947, p. 1–17.
[5] Charney JG, Fjørtøft R, von Neumann J. Numerical integration of the barotropic vorticity equation. Tellus 1950;2(4):237–54.
[6] Morton KW, Mayers DF. Numerical solution of partial differential equations. 2nd ed.. U.K.: Cambridge Univ. Press; 2005.
[7] Zingg DW. Comparison of high-accuracy finite-difference schemes for linear wave propagation. SIAM J Sci Comput 2000;22(2):476–502.
[8] Zingg DW, Lomax H, Jurgens H. High accuracy finite difference schemes for linear wave propagation. SIAM J Sci Comput 1996;17(2):328–46.
[9] Warming RF, Hyett BJ. The modified equation approach to the stability and accuracy analysis of finite difference methods. J Comput Phys 1974;14:159–79.
[10] Chang SC. A critical analysis of the modified equation technique of Warming and Hyett. J Comput Phys 1990;86:107–26.
[11] Li J, Yang Z. The von Neumann analysis and modified equation approach for finite difference schemes. Appl Math Comput 2013;225:610–20.
[12] Garabedian PR. Estimation of the relaxation factor of small mesh sizes. Math Tables Aids Comput 1956;10. 1983–185.
[13] Harten A, Hyman JM, Lax PD. On the finite difference approximations and entropy conditions for shocks. Comm Pure Appl Math 1976;29:297–322.
[14] Griffiths DF, Sanz-Serna JM. On the scope of the method of modified equations. SIAM J Sci Stat Comput 1986;7:994–1008.
[15] Milne WE. Numerical solution of differential equations. New York, USA: Wiley; 1953.
[16] Du Fort EC, Frankel SP. Stability conditions in the numerical treatment of parabolic differential equations, Vol. 7B5. Washington: Math. Table NRC; 1953, p. 135–53.
[17] Ames WF. Numerical methods for partial differential equations. 2nd ed.. New York, USA: Academic Press; 1977.
[18] Lax PD, Wendroff B. System of conservations laws. Comm Pure Appl Math 1960;13:217–37.
[19] Winnicki I, Jasinski J, Pietrek S. New approach to Lax–Wendroff modified differential equation for linear and nonlinear advection. Numer Methods Partial Differential Equations 2019;1–30. http://dx.doi.org/10.1002/num.22412.
[20] Shokin YI. The method of differential approximation. Berlin, Germany: Springer–Verlag; 1983.
[21] Yanenko NN, Fedotova ZI, Tusheva LA, Shokin Yu I. Classification of difference schemes of gas dynamics by the method of differential approximation–I. Comput & Fluids 1983;11(3):187–206.
[22] Sengupta TK. A critical assessment of simulations for transitional and turbulent flows. In: Sengupta TK, Lele SK, Sreenivasan KR, Davidson PA, editors. IUTAM symp. proc. advances in computation, modeling and control of transitional and turbulent flows. Singapore: World Sci. Publ. Co.; 2016, p. 491–532.
[23] Lomax H, Pulliam TH, Zingg DW. Fundamentals of CFD. Berlin, Germany: Springer–Verlag; 2002.
[24] Lighthill MJ. Waves in fluids. UK: Cambridge Univ. Press; 1978.
[25] Drazin PG, Reid WH. Hydrodynamic stability. UK: Cambridge Univ. Press; 1981.
[26] Sengupta TK. Instabilities of fluid flows and transition to turbulence. Florida, USA: CRC Press; 2012.
[27] Yanenko NN. The method of fractional steps: The solution of problems of mathematical physics in several variables. New York, USA: Springer-Verlag; 1971,
[28] Schiesser WE, Griffiths GW. A compendium of partial differential equation models: Method of lines analysis with Matlab. U.K.: Cambridge University Press; 2009.
[29] Sengupta TK, Ganeriwal G, De S. Analysis of central and upwind compact schemes. J Comput Phys 2003;192(2):677–94.
[30] Sengupta TK, Dipankar A. A comparative study of time advancement methods for solving Navier–Stokes equation. J Sci Comput 2004;21(2):225–50.
[31] Sengupta TK, Dipankar A, Sagaut P. Error dynamics: beyond von Neumann analysis. J Comput Phys 2007;226:1211–8.
[32] Vichnevetsky R, Bowles JB. Fourier analysis of numerical approximations of hyperbolic equations. SIAM stud. app. math., vol. 5, Philadelphia, USA; 1982.
[33] Lele SK. Compact finite difference schemes with spectral like resolution. J Comput Phys 1992;103:16–42.
[34] Sengupta TK. Fundamentals of computational fluid dynamics. Hyderabad, India: Universities Press; 2004.
[35] Sengupta TK, Sengupta R. Flow past an impulsively started circular cylinder at high Reynolds number. Comput Mech 1994;14(4):298–310.
[36] Sengupta TK, Bhumkar YG, Rajpoot MK, Suman VK, Saurabh S. Spurious waves in discrete computation of wave phenomena and flow problems. Appl Math Comput 2012;218:9035–65.
[37] Carpenter MH, Gottlieb D, Abarbanel S. The stability of numerical boundary treatments for compact high-order finite difference schemes. J Comput Phys 1993;108:272–95.
[38] Hu FQ, Hussaini MY, Manthey JL. Low-dissipation and low-dispersion Runge–Kutta schemes for computational acoustics. J Comput Phys 1996;124:177–91.
[39] Zhong X. High-order finite difference schemes for numerical simulation of hypersonic boundary-layer transition. J Comput Phys 1998;144:622–709.
[40] Trefethen LN. Group velocity in finite difference schemes. SIAM Rev 1982;24(2):113–36.
[41] LeVeque RJ. Finite difference methods for ordinary and partial differential equations: Steady state and time-dependent problems. Philadelphia, USA: SIAM; 2007.
[42] Strikwerda JC. Finite difference schemes and partial differential equations. 2nd ed.. Philadelphia, USA: SIAM; 2004.
[43] Tam CKW, Webb JC. Dispersion-relation-preserving finite difference schemes for computational acoustics. J Comput Phys 1993;107:262–81.
[44] Sengupta TK, Sengupta A, Saurabh K. Global spectral analysis of multi-level time integration schemes: Numerical properties for error analysis. Appl Math Comput 2017;304:41–5.
[45] Sengupta TK, Sagaut P, Sengupta A, Saurabh K. Global spectral analysis of three-time level integration schemes: Focusing phenomenon. Comput & Fluids 2017;157:182–95.
[46] Haltiner GJ, Williams RT. Numerical prediction and dynamic meteorology. 2nd ed.. New York, USA: John Wiley & Sons; 1980.
[47] Sengupta S, Sengupta TK, Puttam JK, Suman VK. Global spectral analysis for convection–diffusion-reaction equation in one- and two-dimensions: Effects of numerical anti-diffusion and dispersion. J Comput Phys 2020;408:109310.
[48] Sengupta TK, Bhole A. Error dynamics of diffusion equation: Effects of numerical diffusion and dispersive diffusion. J Comput Phys 2014;266:240–51.
[49] Suman VK, Sengupta TK, Durga Prasad CJ, Mohan KS, Sanwalia D. Spectral analysis of finite difference schemes for convection diffusion equation. Comput & Fluids 2017;150:95–114.
[50] Baker AB. Essentials of Padé approximants. Academic Press; 1975.
[51] Price HS, Varga RS, Warren JE. Application of oscillation matrices to diffusion-convection equations. J Math Phys 1966;45:301–11.
[52] Siemieniuch JL, Gladwell I. Analysis of explicit difference methods for a diffusion-convection equation. Internat J Numer Methods Engrg 1978;12:899–916.
[53] Griffiths DF, Christie I, Mitchell AR. Analysis of error growth for explicit difference schemes in conduction-convection problems. Internat J Numer Methods Engrg 1980;15:1075–981.
[54] Varga RS. Matrix iterative analysis. 2nd revised and expanded ed.. Springer; 2009.
[55] Gantmacher FR. Applications of the theory of matrices. New York, USA: Interscience Publishers, Inc.; 1959.
[56] Gustafsson B, Kreiss HA, Sundström A. Stability theory for difference approximations of mixed initial boundary value problems II. Math Comp 1972;26:649–86.
[57] Bhumkar YG, Rajpoot MK, Sengupta TK. A linear focusing mechanism for dispersive and non-dispersive wave problem. J Comput Phys 2011;230(4):1652–75.
[58] Suman VK, Sengupta TK, Mathur JS. Effects of numerical anti-diffusion in closed unsteady flows governed by two-dimensional Navier–Stokes equation. Comput & Fluids 2020;201:104479.
[59] Haras Z, Ta'asan S. Finite difference scheme for long time integration. J Comput Phys 1994;114:265–79.
[60] Keller MA, Kloker MJ. Direct numerical simulations of film cooling in a supersonic boundary-layer flow on massively-parallel supercomputers. In: Resch MM, Bez W, Focht E, Kobayashi H, Kovalenko Y, editors. Proc. Jt. workshop on sustained simulation performance. Cham, Switzerland: Univ. Stuttgart/ Tohoku Univ., Springer; 2013.
[61] Keller MA, Kloker MJ. DNS of effusion cooling in a supersonic boundary layer flow: Influence of turbulence. In: The 44th AIAA thermophysics conf.. AIAA-2013–2897, 2013.
[62] Bhumkar YG, Sheu TWH, Sengupta TK. A dispersion relation preserving optimized upwind compact difference scheme for high accuracy flow simulations. J Comput Phys 2014;278:378–99.
[63] Chu PC, Fan C. A three-point combined compact difference scheme. J Comput Phys 1998;140:370–99.
[64] Sengupta TK, Lakshmanan V, Vjay VVSN. A new combined stable and dispersion relation preserving compact scheme for non-periodic problems. J Comput Phys 2009;228(8):3048–71.
[65] Sengupta TK, Vijay VVSN, Bhaumik S. Further improvement and analysis of CCD scheme: Dissipation dicretization and de-aliasing properties. J Comput Phys 2009;228(17):6150–68.
[66] Adams NA, Shariff KA. High-resolution hybrid compact-ENO scheme for shock-turbulence interaction problem. J Comput Phys 1996;127:27–51.
[67] Dipankar A, Sengupta TK. Symmetrized compact scheme for receptivity study of 2D transitional channel flow. J Comput Phys 2006;215(1):245–73.
[68] Sengupta TK, Sircar SK, Dipankar A. High accuracy schemes for DNS and acoustics. J Sci Comput 2006;26(2):151–93.
[69] Ashwin VM, Saurabh K, Sriramkrishnan M, Bagade PM, Parvathi MK, Sengupta TK. KdV equation and computations of solitons: Nonlinear error dynamics. J Sci Comput 2015;62(3):693–717.

[70] Suman VK, Siva VS, Tekriwal MK, Bhaumik S, Sengupta TK. Grid sensitivity and role of error in computing a lid-driven cavity problem. Phys Rev E 2019;99:013305.

[71] Poinsot T, Veynante D. Theoretical and numerical combustion. 2nd ed.. Philadelphia, USA: R.T. Edwards Inc.; 2005.

[72] Briggs WL, Newell AC, Sarie T. Focusing: A mechanism for instability of nonlinear finite difference equations. J Comput Phys 1983;51:83–106.

[73] Sengupta TK, Suman VK. In: Pirozzoli S, Sengupta TK, editors. Focusing phenomenon in numerical solution of two-dimensional Navier–Stokes equation. CISM monograph: High-performance computing of big data for turbulence and combustion, Switzerland: Springer Nature; 2019.

[74] Sloan DM, Mitchell AR. On nonlinear instabilities in leap-frog finite difference schemes. J Comput Phys 1986;67:372–95.

[75] Vichnevetsky R, Peiffer B. Advances in computer methods for partial differential equations. Vol. 53. Ghent, Belgium: AICA; 1975.

[76] Baum JD, Levine JN. Numerical techniques for solving nonlinear instability problems in solid rocket motors. AIAA J 1982;20:955–61.

[77] Baum M, Poinsot TJ, Thevenin D. Accurate boundary conditions for multicomponent reactive flows. J Comput Phys 1994;116:247–61.

[78] Bashforth F, Adams JC. An attempt to test the theories of capillary action by computing the theoretical and measured forms of drops of fluid. UK: Cambridge University Press; 1883.

[79] Lilly DK. On the computational stability of numerical solutions of time-dependent non-linear geophysical fluid dynamics problems. Mon Weather Rev 1965;138:11.

[80] Durran DR. Numerical methods for wave equations in geophysical fluid dynamics. New York, USA: Springer Verlag; 1999.

[81] Bosshard C, Bouffanais R, Deville M, Gruber R, Latt J. Computational performance of a parallelized three-dimensional high-order spectral element toolbox. Comput & Fluids 2011;44:1–8.

[82] Karniadakis GE, Israeli M, Orszag SA. High-order splitting methods for incompressible Navier–Stokes equations. J Comput Phys 1991;97:414–43.

[83] Fornberg B. On the instability of leap-frog and Crank–Nicolson approximations of a nonlinear partial differential equation. Math Comp 1973;27:45–57.

[84] Newell AC. Finite amplitude instabilities of partial difference equations. SIAM J Appl Math 1977;32:133–60.

[85] Krastov YuA, Orlov YuI. Caustics, catastrophes and wave fields. Springer-Verlag; 2005.

[86] Giles MB, Thompkins WT. Propagation and stability of wavelike solutions of finite difference equations with variable coefficients. J Comput Phys 1985;58:349–60.

[87] David C, Sagaut P, Sengupta TK. A linear dispersive mechanism for numerical error growth: spurious caustics. Eur J Mech B Fluids 2009;28:146–51.

[88] Cloot A, Herbst BM. Grid resonances, focusing and Benjamin-Feir instabilities in Leapfrog time discretizations. J Comput Phys 1988;75:31–53.

[89] Hsia HM, Jeng YN. The weak nonlinear instability of Euler explicit scheme for the convective equation. J Comput Phys 1987;68:251–61.

[90] Sloan DM. On modulational instabilities in discretisations of the Korteweg–de Vries equation. J Comput Phys 1988;79:167–83.

[91] Aoyagi A, Abe K. Parametric excitation of computational modes inherent to leapfrog scheme applied to the Korteweg–de Vries equation. J Comput Phys 1989;83:447–62.

[92] Stuart A. Nonlinear instability in dissipative finite difference schemes. SIAM Rev 1989;31:191–220.

[93] Aoyagi A. Nonlinear Leapfrog instability for Fornberg's pattern. J Comput Phys 1995;120:316–22.

[94] Stuart JT, DiPrima RC. The Eckhaus and Benjamin-Feir resonance mechanisms. Proc R Soc Lond Ser A Math Phys Eng Sci 1978;362:27–41.

[95] Fornberg B, Witham GB. A numerical and theoretical study of certain non-linear wave phenomena. Philos Trans R Soc Lond Ser A Math Phys Sci 1978;289:373–404.

[96] Kreiss H, Oliger J. Comparison of accurate methods for the integration of hyperbolic equations. Tellus 1972;24:199–215.

[97] Swartz B, Wendroff B. The relative efficiency of finite difference and finite element methods. I: Hyperbolic problems and splines. SIAM J Numer Anal 1974;11:979–93.

[98] Green SI. Fluid vortices : Fluid mechanics and its applications. Springer; 1995.

[99] Sengupta TK, Bhumkar YG, Sengupta S. Dynamics and instability of a shielded vortex in close proximity of a wall. Comput & Fluids 2012;70:166–75.

[100] Sengupta TK, Bhumkar YG. New explicit two-dimensional higher order filters. Comput & Fluids 2010;39:1848–63.

[101] Smagorinsky J. Some historical remarks on the use of nonlinear viscosities. In: Galperin B, Orszag SA, editors. Large eddy simulation of complex engineering and geophysical flows. USA: Cambridge Univ. Press; 1993.

[102] Phillips NA. An example of non-linear computational instability. In: Bolin B, editor. The atmosphere and the sea in motion. USA: Rockfeller Inst. Press; 1959.

[103] So KK, Hu XY, Adams NA. Anti-diffusion method for interphase steepening in two-phase incompressible flow. J Comput Phys 2011;230(13):5155–77.

[104] Kanellopoulos G, Weele Kvan der. Critical flow and clustering in a model of granular transport: The interplay between drift and antidiffusion. Phys Rev E 2012;85:061303.

[105] Lee LC, Zhang L, Otto A, Choe GS, Cai HJ. Entropy antidiffusion instability and formation of a thin current sheet during geomagnetic substorms. J GeoPhys Res 1998;103(A12):29419–28.

[106] Prigogine I, Stengers I. Order out of chaos. New York, USA: Bantam Book; 1988.

[107] Konstantopoulos C, Mittag L, Sandri G. Deconvolution of Gaussian filters and antidiffusion. J Appl Phys 1990;68(4):1415–20.

[108] Sengupta TK, Sengupta A, Sharma N, Sengupta S, Bhole A, Shruti KS. Roles of bulk viscosity on Rayleigh–Taylor instability: Non-equilibrium thermodynamics due to spatio-temporal pressure fronts. Phys Fluids 2016;28(9):094102.

[109] Sengupta TK, Sengupta A, Shruti KS, Sengupta S, Bhole A. Non-equilibrium thermodynamics of Rayleigh–Taylor instability. J Phys Conf Ser 2016;759(1):012079.

[110] Sengupta TK, Rajpoot MK, Saurabh S, Vijay VVSN. Analysis of anisotropy of numerical wave solutions by high accuracy finite difference methods. J Comput Phys 2011;230(1):27–60.

[111] Cossu C, Loiseleux T. On the convective and absolute nature of instabilities in finite difference numerical simulations of open flows. J Comput Phys 1998;144(1):98–108.

[112] Kawamura T, Takami H, Kuwahara K. A new higher-order upwind scheme for incompressible Navier–Stokes equations. Fluid Dyn Res 1985;1(1):145–62.

[113] Van der Vorst HA. Bi-CGSTAB: A fast and smoothly converging variant of Bi-CG for the solution of non-symmetric linear systems. SIAM J Sci Stat Comput 1992;12:631–44.

[114] Adam Y. Nonlinear instability in advection-diffusion numerical models. Appl Math Model 1985;9(6):434–40.

[115] Lestandi L, Bhaumik S, Avatar GRKC, Azaiez M, Sengupta TK. Multiple Hopf bifurcations and flow dynamics inside a 2D singular lid driven cavity. Comput & Fluids 2018;166:86–103.

[116] Gaitonde DV, Shang JS, Young JL. Practical aspects of higher-order numerical schemes for wave propagation phenomena. Int J Numer Methods Eng 1999;45(12):1849–69.

[117] Gaitonde D, Visbal M. Further development of a Navier–Stokes solution procedure based on higher-order formulas. In: Aerospace sciences meetings. American Institute of Aeronautics and Astronautics; 1999.

[118] Visbal MR, Gaitonde DV. High-order-accurate methods for complex unsteady subsonic flows. AIAA J 1999;37(10):1231–9.

[119] Rizzetta DP, Visbal MR, Blaisdell GA. A time–implicit high-order compact differencing and filtering scheme for large–eddy simulation. Int J Numer Methods Fluids 2003;42(6):665–93.

[120] Sengupta TK, Bhumkar YG, Lakshmanan V. Design and analysis of a new filter for LES and DES. Comput Struct 2009;87(11):735–50.

[121] Bhumkar YG, Sengupta TK. Adaptive multi-dimensional filters. Comput & Fluids 2011;49(1):128–40.

[122] Pedlosky J. Geophysical fluid dynamics. Springer Verlag; 1979.

[123] Gill AE. Atmosphere-ocean dynamics. International geophysics series, New York: Academic Press; 1982.

[124] Vallis GK. Atmosphere and ocean fluid dynamics: Fundamentals and large scale circulations. Cambridge Univ. Press; 2006.

[125] Rajpoot MK, Bhaumik S, Sengupta TK. Solution of linearized rotating shallow water equations by compact schemes with different grid-staggering strategies. J Comput Phys 2012;231:2300–27.

[126] Rajpoot MK, Sengupta TK, Dutt PK. Optimal time advancing dispersion relation preserving schemes. J Comput Phys 2010;229:3623–51.

[127] Pol Bvan der, Bremmer H. Operational calculus based on the two-sided laplace transform. Cambridge Univ. Press; 2008.

[128] Mesinger F, Arakawa A. Numerical methods used in atmospheric models. 1. GARP publ. ser., no. 17, Geneva: WMO; 1976.

[129] Nagarajan S, Lele SK, Ferziger JH. A robust high-order compact method for large eddy simulation. J Comput Phys 2003;19:392–419.

[130] Orszag SA, Patterson G. Numerical simulation of three-dimensional homogeneous isotropic turbulence. Phys Rev Lett 1972;28(2):76–9.

[131] Rogallo RS. Numerical experiments in homogeneous turbulence. NASA Tech. Mem. 81315, 1981.

[132] Eswaran V, Pope SB. An examination of forcing in direct numerical simulations of turbulence. Comput & Fluids 1988;16(3):257–78.

[133] Yeung PK, Donzis DA, Sreenivasan KR. Dissipation, enstrophy and pressure statistics in turbulence simulations at high Reynolds numbers. J Fluid Mech 2012;700:5–15.

[134] Ranjan A, Davidson PA. DNS of a buoyant turbulent cloud under rapid rotation. In: Sengupta TK, Lele SK, Sreenivasan KR, Davidson PA, editors. IUTAM symp. proc. advances in computation, modeling and control of transitional and turbulent flows. Singapore: World Sci. Publ. Co.; 2016, p. 491–532.

[135] Buaria D, Pumir A, Bodenschatz E. Self-attenuation of extreme events in Navier–Stokes turbulence. Nature Commun 2020;11:5852.

[136] Lamorgese AG, Caughey DA, Pope SB. Direct numerical simulation of homogeneous turbulence with hyperviscosity. Phys Fluids 2005;17(1):015106.

[137] Beale JT, Kato T, Majda A. Remarks on the breakdown of smooth solutions for the 3-D Euler equations. Comm Math Phys 1984;94:61–6.

[138] Sengupta TK, Suman VK, Sundaram P, Sengupta A. Analysis of pseudo-spectral methods used for numerical simulation of turbulence. 2021, http://dx.doi.org/10.48550/arXiv.2109.00255, arXiv, arXiv:2109.00255.

[139] Sengupta TK, Suman VK, Sundaram P, Sengupta A. Analysis of pseudo-spectral methods used for numerical simulation of turbulence. WSEAS Trans Comput Res 2022;10:9–24.

[140] Sengupta TK, Dipankar A, Kameswara Rao A. A new compact scheme for parallel computing using domain decomposition. J Comput Phys 2007;220:654–77.

[141] Sengupta A, Sundaram P, Suman VK, Sengupta TK. Three-dimensional direct numerical simulation of Rayleigh–Taylor instability triggered by acoustic excitation. Phys Fluids 2022;34(5):054108.

[142] Sengupta TK, Sundaram P, Suman VK, Bhaumik S. A high accuracy preserving parallel algorithm for compact schemes for DNS. ACM Trans Parallel Comput 2020;7(4):21, 1–32.

[143] Sundaram P, Sengupta A, Sengupta TK. A non-overlapping high accuracy parallel subdomain closure for compact scheme: Onset of Rayleigh–Taylor instability by ultrasonic waves. J Comput Phys 2022;470:111593.

[144] Sundaram P, Sengupta TK, Sengupta A, Suman VK. Multi-scale instabilities of Magnus-Robins effect for compressible flow past rotating cylinder. Phys Fluids 2021;33(3):034129.

[145] Sengupta TK, Ghosh Roy A, Chakraborty A, Sengupta A, Sundaram P. Thermal control of transonic shock-boundary layer interaction over a natural laminar flow airfoil. Phys Fluids 2021;33:126110.

[146] Sengupta TK, Chakraborty A, Ghosh Roy A, Sengupta A, Sundaram P. Comparative study of transonic shock-boundary layer interactions due to surface heating and cooling on an airfoil. Phys Fluids 2022;34(4):046110.

[147] Chakraborty A, Ghosh Roy A, Sengupta A, Sundaram P, Sengupta TK. Controlling transonic shock-boundary layer interactions over a natural laminar flow airfoil by vortical and thermal excitation. Phys Fluids 2022;34:085124.

[148] Sundaram P, Sengupta S, Suman VK, Sengupta TK, Bhumkar YG, Mathpal RK. Flow control using single dielectric barrier discharge plasma actuator for flow over airfoil. Phys Fluids 2022;34.

[149] Fang J, Gao F, Moulinec C, Emerson DR. An improved parallel compact scheme for domain-decoupled simulation of turbulence. Internat J Numer Methods Fluids 2019;90(10):479–500.

[150] Sengupta TK, Ganeriwal G, Dipankar A. High accuracy compact schemes and Gibbs' phenomenon. J Sci Comput 2004;21(3):253–68.

[151] Sengupta S, Sreejith NA, Mohanamuraly P, Staffelbach G, Gicquel L. Global spectral analysis of the Lax–Wendroff-central difference scheme applied to Convection–Diffusion equation. Comput & Fluids 2022;242:1105508.

[152] Ferziger JH, Echekki T. A simplified reaction rate model and its application to the analysis of premixed flames. Combust Sci Technol 1993;89:293–315.

[153] Pfitzner M. A new analytic pdf for simulations of premixed turbulent combustion. Flow Turbul Combust 2021;106:1213–39.

[154] Pfitzner M, Breda P. An analytic probability density function for partially premixed flames with detailed chemistry. Phys Fluids 2021;33:1–16.

[155] Sengupta TK, Suman VK, Sengupta S, Sundaram P. Quantifying parameter ranges for high fidelity simulations for prescribed accuracy by Lax–Wendroff method. Comput & Fluids 2023;254.

[156] Nazarenko S. Wave turbulence. Springer-Verlag; 2011.

[157] Cai D, Aoyagi A, Abe K. Parametric excitation of computational mode of the leapfrog scheme applied to the Van der Pol equation. J Comput Phys 1993;107:146–51.

[158] Herbst BM, Mitchell AR, Weideman JAC. On the stability of the nonlinear Schrödinger equation. J Comput Phys 1985;60:263–81.

[159] Vadillo F, Sanz-Serna JM. Studies in numerical nonlinear instability. II. A new look at $u_t + uu_x = 0$. J Comput Phys 1986;66:225–38.

[160] Sengupta TK, Ballav M, Nijhawan S. Generation of Tollmien-Schlichting waves by harmonic excitation. Phys Fluids A 1994;6(3):1213–22.

[161] Sundaram P, Sengupta TK, Sengupta S. Is Tollmien-Schlichting wave necessary for transition of zero pressure gradient boundary layer flow? Phys Fluids 2019;31:031701.

[162] Sengupta A, Sundaram P, Sengupta TK. Nonmodal nonlinear route of transition to two-dimensional turbulence. Phys Rev Res 2020;2:012033.

[163] Sengupta TK. Transition to turbulence: A dynamical system approach to receptivity. Cambridge, U.K.: Cambridge Univ. Press; 2021.

[164] Sundaram P, Suman VK, Sengupta A, Sengupta TK. Effects of free stream excitation on the boundary layer over a semi-infinite flat plate. Phys Fluids 2020;32:094110.

[165] Sengupta A, Samuel RJ, Sundaram P, Sengupta TK. Role of Non-zero bulk viscosity in three-dimensional Rayleigh–Taylor instability: Beyond Stokes' hypothesis. Comput & Fluids 2021;225:104995.

[166] Sengupta A, Sundaram P, Suman VK, Sengupta TK. Three-dimensional direct numerical simulation of Rayleigh–Taylor instability triggered by acoustic excitation. Phys Fluids 2022;34(5):054108.

[167] Maddipati R, Sengupta TK, Sundaram P. Relevance of two- and three-dimensional disturbance field explained with linear stability analysis of Orr–Sommerfeld equation by compound matrix method. Comput & Fluids 2021;225:104965.

[168] Sengupta TK, Sengupta A. A new alternating Bi-diagonal compact scheme for non-uniform grids. J Comput Phys 2016;310:1–25.

[169] Sharma N, Sengupta A, Rajpoot M, Samuel RJ, Sengupta TK. Hybrid sixth order spatial discretization scheme for non-uniform cartesian grids. Comput & Fluids 2017;157(3):208–31.

[170] Dorodnitsyn V. Finite difference models entirely inheriting continuous symmetries of original differential equations. Internat J Modern Phys C 1994;5(4):723–34.

[171] Dorodnitsyn V. Applications of Lie groups to difference equations. CRC Press; 2011.

[172] Chhay M, Hamdouni A. On the accuracy of invariant numerical schemes. Commun Pure Appl Anal 2011;10:761–83.

[173] Chhay M, Hoarau E, Hamdouni A, Sagaut P. Comparison of some Lie-symmetry-based integrators. J Comput Phys 2011;230:2174–88.

[174] Razafindralandy D, Hamdouni A, Al Sayed N. Lie-symmetry group and modeling in non-isothermal fluid mechanics. Physica A 2012;391:4624–36.

[175] Bihlo A, Popovych RO. Invariant discretization schemes for the shallow-water equations. SIAM J Sci Comput 2012;34:B810–39.

[176] Ozbenli E, Vedula P. High order accurate finite difference schemes based on symmetry preservation. J Comput Phys 2017;349:376–98.

[177] Ozbenli E, Vedula P. Construction of invariant compact finite-difference schemes. Phys Rev E 2020;101:023303.

[178] Dorodnitsyn V, Kapstov EI. Shallow water equations in Lagrangian coordinates: Symmetries, conservation laws and its preservation i, difference models. Commun Nonlinear Sci Numer Simul 2020;89:105343.

[179] Cheviakov AF, Dorodnitsyn V, Kapstov EI. Invariant conservation law-preserving discretizations of linear and non-linear wave equations. J Math Phys 2020;61:081504.

[180] Verstappen RWCP, Veldman AEP. A spectro-consistent discretization of Navier–Stokes: a challenge to RANS and LES. J Engrg Math 1998;34:163–79.

[181] Verstappen RWCP, Veldman AEP. Symmetry-preserving discretization of turbulent flows. J Comput Phys 2003;187:343–68.

[182] Reiss J, Sesterhenn J. A conservative, skew-symmetric finite difference scheme for the compressible Navier–Stokes equations. Comput & Fluids 2014;101:208–19.

[183] Trias FX, Lehmkuhl O, Oliva A, Perez-Segarra CD, Verstappen RWCP. Symmetry-preserving discretization of Navier–Stokes equations on collocated unstructured grids. J Comput Phys 2014;258:246–67.

[184] Capuano F, Coppola G, Balarac G, de Luca L. Energy preserving turbulent simulations at a reduced computational cost. J Comput Phys 2015;298:480–94.

[185] Capuano F, Vallefuoco D. Effects of discrete energy and helicity conservation in numerical simulations of helical turbulence. Flow Turbul Combust 2018. http://dx.doi.org/10.1007/s10494-018-9939-x.

[186] Coppola G, Capuano F, de Luca L. Discrete energy-conservation properties in the numerical simulation of the Navier–Stokes equations. Appl Mech Rev 2019;71:010803.

[187] Veldman AEP. A general condition for kinetic-energy preserving discretization of flow transport equations. J Comput Phys 2019;398:108894.

[188] Rozema W, Verstappen RWCP, Veldman AEP, Kok JC. Low-dissipation simulation methods and models for turbulent subsonic flows. Arch Comput Methods Eng 2020;27:299–330.

[189] Dorodnitsyn V. Noether-type theorems for difference equations. Appl Numer Math 2001;39:307–21.

[190] Fu JL, Chen LQ, Chen BY. Noether-type theorem for discrete nonconservative dynamical systems with nonregular lattices. Sci China 2010;53:545–54.

[191] Dorodnitsyn V, Ibragimov NH. An extension of the Noether theorem: Accompanying equations possessing conservation laws. Commun Nonlinear Sci Numer Simul 2014;19:328–36.

[192] Honein AE, Moin P. Higher entropy conservation and numerical stability of compressible turbulence simulations. J Comput Phys 2004;201:531–45.

[193] Morinishi Y. Skew-symmetric form of convective terms and fully conservative finite difference schemes for variable density low-Mach number flows. J Comput Phys 2010;229:276–300.

[194] Pirozzoli S. Generalized conservative approximations of split convective derivative operators. J Comput Phys 2010;229:7180–90.

[195] Brouwer J, Reiss J, Sesterhenn J. Conservative time integrators of arbitrary order for skew-symmetric finite-difference discretizations of compressible flow. J Comput Phys 2014;100:1–12.

[196] van't Hof B, Vuik MJ. Symmetry-preserving finite-difference discretizations of arbitrary order on structured curvilinear staggered grids. J Comput Sci 2019;36:101008.

[197] Coppola G, Capuano F, Pirozzoli S, de Luca L. Numerically stable formulations of convective terms for turbulent compressible flows. J Comput Phys 2019;382:86–104.

[198] Sjögreen B, Yee HC, Kotov D. Skew-symmetric splitting and stability of high order central schemes. J Phys: Conf Ser 2019;837:012019.

[199] Sanderse B. Energy-conserving Runge–Kutta methods for the incompressible Navier–Stokes equations. J Comput Phys 2013;233:100–31.
[200] Capuano F, Coppola G, de Luca L. An efficient time advancing strategy for energy-preserving simulations. J Comput Phys 2015;295:209–29.
[201] Capuano F, Coppola G, Randez L, de Luca L. Explicit Runge–Kutta schemes for incompressible flow with improved energy-conservation properties. J Comput Phys 2017;328:86–94.
[202] Duponcheel M, Orlandi P, Winckelmans G. Time-reversibility of the Euler equation as a benchmark for energy-preserving schemes. J Comput Phys 2008;227:8736–52.
[203] Iserles A. Stability and dynamics of numerical methods for nonlinear ordinary differential equations. IMA J Numer Anal 1990;10:1–30.
[204] Yee HC, Sweby PK. Dynamics of numerics and spurious behaviours in CFD computations. NASA RIACS technical report, No 97.06, 1997.
[205] Yee HC, Sweby PK, Griffiths DF. Dynamical approach study of spurious steady-state numerical solutions of nonlinear differential equations. I. The dynamics of time discretization and its implications for the algorithm development in computational fluid dynamics. J Comput Phys 1991;97:259–310.
[206] Yee HC, Sweby PK, Griffiths DF. Dynamical approach of spurious steady-state numerical solutions of nonlinear differential equations. NAS Applied Research Technical Report, RNR-92-008, 1992.
[207] Yee HC, Sweby PK. Dynamical approach study of spurious steady-state numerical solutions of nonlinear differential equations. II. Global asymptotic behaviour of time discretizations. Comput Fluid Dyn 1995;4:219–83.
[208] Lafon A, Yee HC. Dynamical approach study of spurious steady-state numerical solutions of nonlinear differential equations. III. The effects of nonlinear source terms in reaction-convection equations. Int J Comput Fluid Dyn 1996;6:1–36.
[209] Yee HC, Sweby PK. On spurious behaviour of super-stable implicit methods. IJCFD 1997;8:265–86.
[210] Sleeman BD, Griffiths DF, Mitchell AR, Smith PD. Stable periodic solutions in nonlinear difference equations. SIAM J Sci Stat Comput 1988;9:543–57.
[211] Griffiths DF, Mitchell AR. Stable periodic bifurcations of an explicit discretization of a nonlinear partial differential equation in reaction diffusion. IMA J Numer Anal 1988;8:435–54.
[212] Griffiths DF, Sweby PK, Yee HC. On spurious asymptotic numerical solutions of explicit Runge–Kutta methods. IMA J Numer Anal 1992;12:319–38.
[213] Hataue I. Mathematical and numerical analyses of dynamical structure of numerical solutions of two-dimensional fluid equations. J Phys Soc Japan 1998;67:1895–911.
[214] Griffiths DF, Stuart AM, Yee HC. Numerical wave propagation in an advection equation with a nonlinear source term. SIAM J Numer Anal 1992;29:1244–60.
[215] David C, Sagaut P. Spurious solitons and structural stability of finite-diference schemes for non-linear wave equations. Chaos Solitons Fractals 2009;41:655–60.
[216] David C, Sagaut P. Structural stability of finite dispersion-relation preserving schemes. Chaos Solitons Fractals 2009;41:2193–9.
[217] Yee HC, Kotov DV, Wang W, Shu CW. Spurious behaviour of shock-capturing methods by the fractional step approach: problems containing stiff source terms and discontinuities. J Comput Phys 2013;241:266–91.
[218] Huerre P, Monkewitz PA. Local and global instabilities in spatially developing flows. Annu Rev Fluid Mech 1990;22. 473–457.