Evaluating Numerical Stability in High-Accuracy Simulations: A Comparative Study of Time Discretization Methods for the Linear Convection Equation

Vibhanshu Dev GAUR*,[1], Mayur PATHAK[1], Anirudh SHANKAR[1], Nidhi SHARMA[1]

*Corresponding author
[1]Department of Aerospace Engineering,
Punjab Engineering College (Deemed to be University),
Sector 12, Chandigarh, 160012, India,
vibhanshudevgaur.mt22aero@pec.edu.in

DOI: 10.13111/2066-8201.2024.16.3.3

Received: 22 June 2024/ Accepted: 19 July 2024/ Published: September 2024

Abstract: *With the advent of technology, it has become possible to perform direct numerical simulations and the demand for high accuracy computing is increasing. Numerical simulations play an important part in understanding physics of the flow and instability mechanism in flows. For high accuracy, numerical schemes must be chosen that satisfy the physical dispersion relation, should not amplify or attenuate the solution and resolve all possible length and time scales. In the present paper, spectral stability analysis of linear convection equation is performed using first order forward difference (FD1) method and fourth order Runge Kutta (RK4) method, consisting of four stages, for time discretization and a second order central difference (CD2) method for evaluating spatial derivative. The results show that the presence of numerical instability for FD1 method is independent of the CFL number, consistent with the stability analysis which showed FD1 method to be unconditionally unstable. However, for RK4 method, the solution is found to be neutrally stable only for a particular range of CFL number, even stable solution introduced error by attenuating the computed or analytical solution.*

Key Words: *spectral stability, convection equation, RK4, FD1, CD2*

1. INTRODUCTION

In many applications of applied physics, it is crucial to learn about the evolution of errors that accompany the transmission of signals over a continuous medium. Error dynamics have been extensively studied using methods attributed to von Neumann analysis [1],[2]. The finite difference schemes generally use von Neumann analysis. Stable solutions remain bounded by perturbations to the input, while unstable solutions grow with time [2]. For a scheme to be stable, its computational domain must enclose the mathematical or analytical domain of dependence at every point in space and time; not meeting this criterion is equivalent to neglecting some of the time-marching data essential for an advancement in the solution. Otherwise, excessive data would have to be fed into the solution because there would be additional information that the model knows nothing about. The time step in such a discrete framework should not be greater than the time taken by a wave to travel in a uniform space between two neighboring points. According to Courant–Friedrichs–Lewy (CFL) condition,

this is the part of the grid cell through which a fluid wave goes over in convection during one time-step. Courant number imposed by the von Neumann analysis [3] determines possible mesh and temporal resolutions for stable solutions [1]. Gustafsson, Kreiss, and Sundstrom (GKS) stability theory criteria has a different meaning in the physical space which is expressed in relation to group velocity: for a finite difference model to be unstable, the basic condition says that the model along with its boundary conditions should support a set of waves at the boundaries while the group velocities should be pointing into the field [4]. While solving time dependent partial differential equations computationally by employing finite difference methods, a higher number of boundary condition are required than the problem's physics. This presents the demand of selecting additional numerical boundary conditions, where numerical stability is an extremely important factor to be taken into consideration. The stability problem for hyperbolic equations is mathematically solved by using the theory of Gustafsson, Kreiss, and Sundstrom's theory[5]. In the present work, the spectral stability analysis of a fundamental equation i.e. one dimensional linear convection equation is performed using two different explicit methods. Firstly, the time derivative is evaluated using first order forward–in time method and spatial derivatives are obtained by second order central difference (CD2)-scheme with Dirichlet type boundary condition [1]. The stability plots are then compared with a higher order method, four stage fourth order Runge-Kutta (RK4)-method for time integration and spatial scheme is kept as CD2. Since, the Navier-Stokes equation reduces to the linear convection equation in its basic form, hence, the inferences obtained here applies to the Navier-Stokes equation as well. The inferences obtained from the property charts are then used to solve the one-dimensional convection equation by the two different time-integration schemes.

2. SPECTRAL STABILITY ANALYSIS

The one-dimensional (1D) convection equation is given below:

$$\frac{\partial u}{\partial t} + c\frac{\partial u}{\partial x} = 0 \quad ; \quad c > 0 \tag{1}$$

where, c is the phase speed with which the solution travels to the right. The numerical solution of the wave equation is identified as;

$$u_i^n = u(x_m, t^n) = \int \hat{U}(k, t^n)\, e^{ikx_m} dk \tag{2}$$

where, quantity with a hat is the amplitude of the unknown function in spectral plane and x_m is the grid node given as $m*h$ with h as the grid spacing. Since, we are using explicit methods therefore, time dependence is explicit and space dependence is given by non-dimensional wavenumber *(kh)*. If, the unknown is defined as

$$u_i^n = u(x_m, t^n) = \hat{U}(k, \omega) e^{i(kx_m - \omega t)} dk\, d\omega \tag{3}$$

Then, the corresponding physical dispersion relation is given by, $\omega = ck$. While, c denotes phase speed, the group velocity *(V_g),* determines the speed at which energy propagates [6], [7], [8] and for Eq. (1), it is given as;

$$V_g = \frac{d\omega}{dk} = c \tag{4}$$

Hence, for the 1D convection equation which is non-dispersive in nature, the physical group velocity is equal to the physical phase speed. Therefore, to computationally solve Navier-Stokes equations, the chosen numerical methods must be stable, consistent and should

numerically satisfy the physical dispersion relation i.e. should preserve the correct space-time dependence simultaneously, called as the Dispersion Relation Preserving (DRP) schemes. The stability analysis for two different time-integration schemes is discussed in the following sections.

2.1 Forward-in-time and Centered-in-space (FTCS) method

The resultant discretized equation using FTCS method [4] for solving Eq. (1) is given as:

$$u_i^{n+1} = u_i^n + \frac{Nc}{2}(u_{i+1}^n - u_{i-1}^n) \tag{5}$$

where, 'i' is the grid node point, 'n' is the time level and N_c is the CFL number defined as $c\Delta t/\Delta x$, where Δt is the time-step and Δx is the uniform grid spacing given as, $x_{i+1} - x_i$. The presence and extent of numerical stability is measured by the numerical amplification factor [2] which is the ratio of amplitude of solution at the current time-step to the previous time-step, given as:

$$G(kh, Nc) = \frac{\widehat{U}(k, t + \Delta t)}{\widehat{U}(k, t)} \tag{6}$$

Substituting, Eq. (5) and (6) into Eq. (2), one gets;

$$|G| = \sqrt{1 + N_c^2 Sin^2 kh} \tag{7}$$

For this propagation problem, the general solution at any arbitrary time is:

$$u_i^n = u(x_m, t^n) = \int A_o(k)[G(k)]^n \, e^{i(kx_m - n\beta_i)} dk \tag{8}$$

where, $A_0(k)$ is the initial amplitude and β_i gives the measure of the phase speed of the numerical scheme that is given as:

$$c_N = \frac{\beta_i}{k\Delta t} \tag{9}$$

Using the physical dispersion relation ($\omega = ck$), the non-dimensional numerical phase speed is

$$\frac{c_N}{c} = \frac{\beta_i}{\omega \Delta t} \tag{10}$$

As, we can see from Eq. (7), the amplification factor is always greater than 1, i.e. |G| > 1 for FTCS scheme, hence the method is unconditionally unstable. Hence, whatever value of CFL number is chosen, the method will always be unstable. The amount of dispersion added by a numerical method is quantified with the help of numerical group velocity, (Vg_N). It is crucial to note that for a one-dimensional convection equation model, the difference between numerical group velocity and the physical group velocity must be approximately zero for a vast range of wavenumbers since the group velocity and disturbance energy moves together [4]. Thus, highly accurate method yields Vg_N/c and c_N/c as 1 and any deviation from 1 gives us a measure of dispersion and phase error, respectively. The non-dimensional numerical group velocity and phase speed are computed using the numerical dispersion relation, $\omega_{eq} = c_N k$ as:

$$\frac{Vg_N}{c} = \frac{1}{N_c}\frac{d\beta_i}{dkh}; \qquad \frac{c_N}{c} = \frac{\beta_i}{\omega\Delta t} = \frac{\beta_i}{N_c kh} \tag{11}$$

where, $\beta_i = tan^{-1}\left(-G_{imag}/G_{real}\right)$ is the phase angle. Figure 1 shows the numerical amplification factor, *(|G|)* in the top frame. It is noticed that the |*G*| values ranges from 1 to 3.295 i.e. greater than one everywhere, implying unstable numerical method. However, there exists a neutrally stable region for very small values of *Nc* resolving low and high wavenumbers, shown by the shaded area in grey. The non-dimensional numerical group velocity *(Vg_N/c)* and the non-dimensional numerical phase speed *(c_N/c)* are shown in the middle and bottom frames, respectively. It is noticed that for a narrow range of *kh*, the phase error is of the order of *$O(10^{-4})$* and the error increases with increase in *Nc* and *kh*. From the bottom frame, one notices that the dispersion error is minimum for a narrow range of small wavenumbers. However, for high wavenumbers, one notices q-waves i.e. negative group velocity, implying that the solution travels in the direction which is opposite to the physical solution.

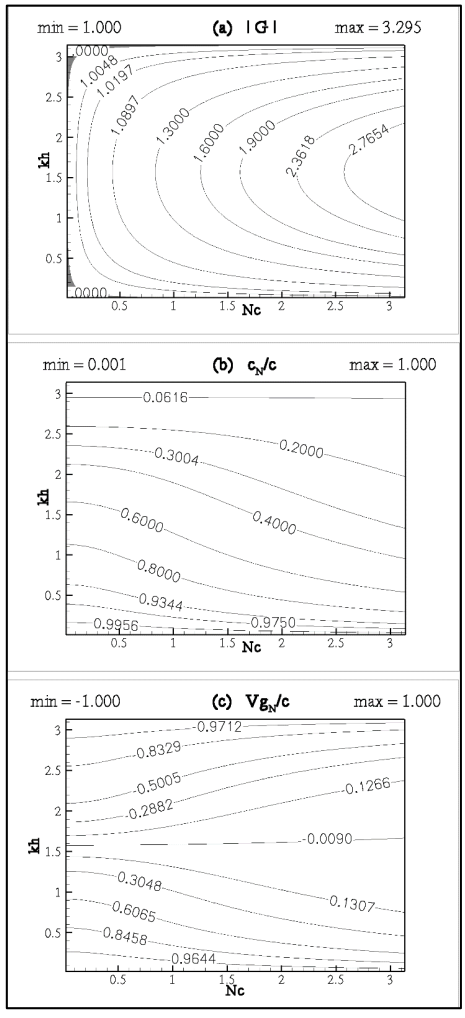

Figure1. Numerical property charts for 1D convection equation using FTCS scheme (a) numerical amplification factor, *(|G|),* (b) non-dimensional phase speed, *(c_N/c),* and (c) non-dimensional group velocity, *(Vg_N/c).* The shaded area in |G| shows the neutrally stable region *(|G|=1)*

2.2 Four-step 4th order Runge-Kutta (RK4) method with 2nd order central difference (CD2) scheme

The stencil for RK4 method (Trefethen, 1982) is given as:

$$\begin{aligned} &\textit{Step-1}\ \ u^{(1)} = u^{(n)} + \frac{\Delta t}{2} L[u^{(n)}] \\ &\textit{Step-2}\ \ u^{(2)} = u^{(n)} + \frac{\Delta t}{2} L[u^{(1)}] \\ &\textit{Step-3}\ \ u^{(3)} = u^{(n)} + \Delta t L[u^{(2)}] \\ &\textit{Step-4}\ \ u^{(n+1)} = u^{(n)} + \frac{\Delta t}{6}\begin{bmatrix} L\left[u^{(n)}\right] + 2L\left[u^{(1)}\right] + \\ 2L\left[u^{(2)}\right] + L[u^{(3)}] \end{bmatrix} \end{aligned} \tag{12}$$

The numerical amplification factor for RK4-CD2 given below is derived by substituting the spectral representation of the unknown, Eq. (2) and Eq. (6) into Eq. (12) and simplifying:

$$G = \left(1 - \frac{N_c^2\, sin^2\, k\, h}{2} + \frac{N_c^4\, sin^4\, k\, h}{24}\right) + \ i\left(-N_c\, sin\, k\, h + \frac{N_c^3\, sin^3\, k\, h}{6}\right) \tag{13}$$

Unlike, FTCS method, the numerical amplification factor of RK4-CD2 depends on the value of CFL number.

While, the corresponding non-dimensional group velocity and phase speed are obtained by substituting the real part and imaginary part of $|G|$ from Eq. (13) in Eq. (11).

Similar to the graphs obtained from FTCS method, we have three graphs pertaining to numerical amplification factor *(|G|),* the non-dimensional group velocity *(Vg_N/c)* and the non-dimensional phase speed (c_N/c) for RK4-CD2 method as shown in Figure 2 in top, middle and bottom frames, respectively.

The values of $|G|$ in case of RK4 method is always positive and ranges from 0.5 to 2.023. For small values of N_c, one notices a neutrally-stable region ($|G|=1$), resolving the entire range of wavenumbers, as shown by shaded are in grey.

However, for higher values of CFL number, the solution will be stable with damped magnitude for intermediate wavenumber range.

From the middle frame of Figure 2, one notices that the phase speed in case of RK4 is a bit more random and not as linear as FTCS because RK4 is a fourth order method thus showing more uncertainty with phase shifts.

However, for small wavenumbers, the phase error is of the order $O\ (10^{-3})$ and keeps on increasing with wavenumber.

From the bottom frame of Figure 2, it is noticed that the Vg_N/c is of the order of $O\ (10^{-2})$ for small wavenumbers and keeps on increasing with *kh*.

Also, one notices the presence of q-waves i.e. unphysical group velocity for high wavenumber range.

With the inferences obtained in the present section, the 1D convection equation is solved using FTCS method and results are then compared with that of RK4-CD2 method. Additionally, the effect of numerical parameter such as CFL number is shown for the RK4-CD2 solution in the following section.

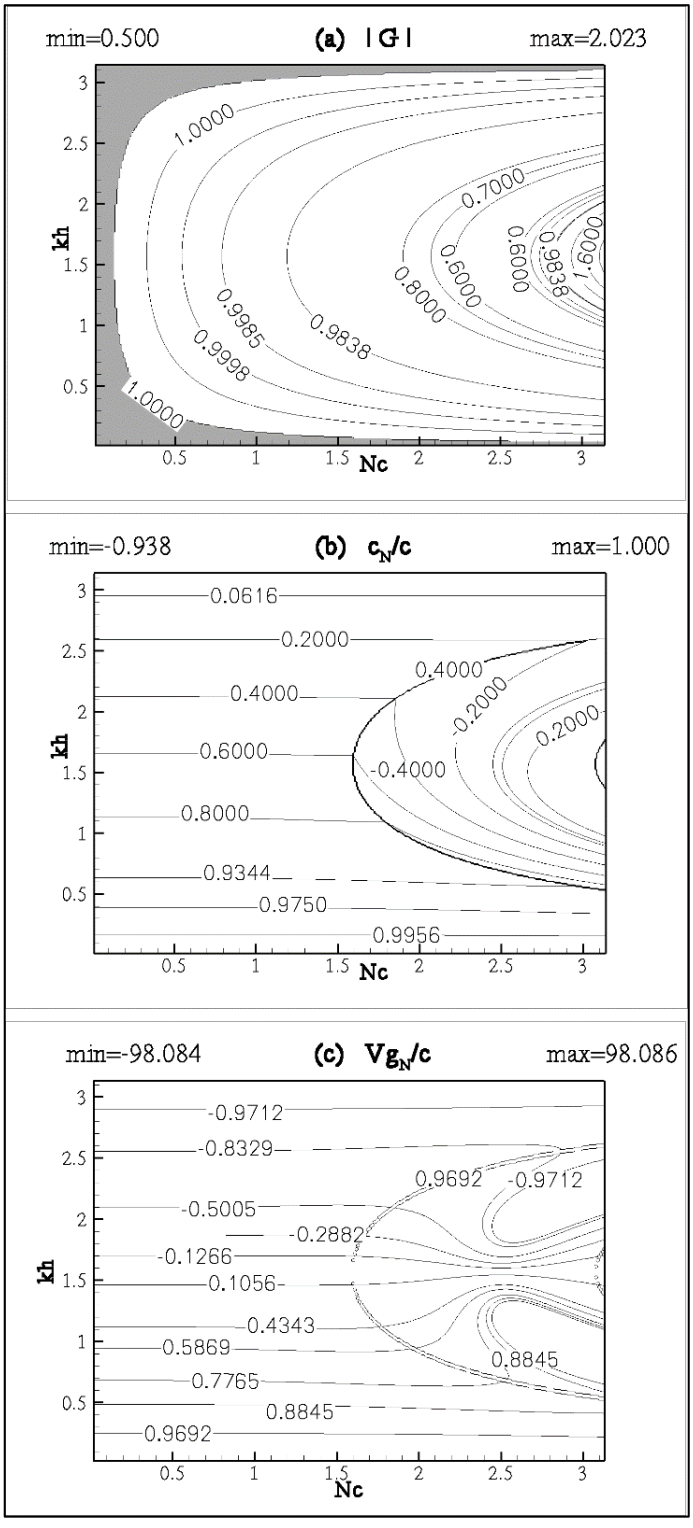

Figure 2. Numerical property charts for 1D convection equation using RK4 scheme (a) numerical amplification factor, (|G|), (b) phase velocity, (c_N/c), and (c) group Velocity, (Vg_N/c). The shaded area in |G| shows the neutrally stable region (|G|=1).

3. RESULTS AND DISCUSSIONS

The linear convection Eq. (1), given below, is solved here in one dimensional (1D) grid with 5000 grid points and a domain length varying from $0 < x < 10$.

$$\frac{\partial u}{\partial t} + c\frac{\partial u}{\partial x} = 0 \quad ; \quad c > 0$$

With the initial condition given as the Gaussian wave-packet:

$$u(x,0) = e^{-\alpha(x-x_0)^2}\ cos[k_0(x-x_0)]$$

The wave packet is centered at x_0=2.0 at t = 0 and the width size of the wave-packet is regulated by the parameter, α = -5 and initial wave number is k_0=50 with physical phase speed taken as c=0.1. The FTCS simulation is carried out for N_c = 0.09 while, the RK4-CD2 simulations are carried out for two different values of CFL number, N_c = 0.09 and N_c = 0.1 to see the effect of time-step on the accuracy of the solution. These particular values of N_c are chosen from the analysis of property charts discussed in section A, as these CFL number values lie in the neutrally stable region and correspond to minimum dispersion and phase error for both the numerical methods.

3.1 Solution of 1D convection Equation using FTCS

The analytically computed solutions are represented by solid lines in Figure 3 along with the exact or true solution by dashed lines for *Nc* = 0.09. From the property charts shown in Figure 1 for FTCS, the chosen N_c value lies in the neutrally stable region either for very small wavenumber range or for high wavenumbers. However, the majority of intermediate wave-numbers shows instability with $|G|>1$.

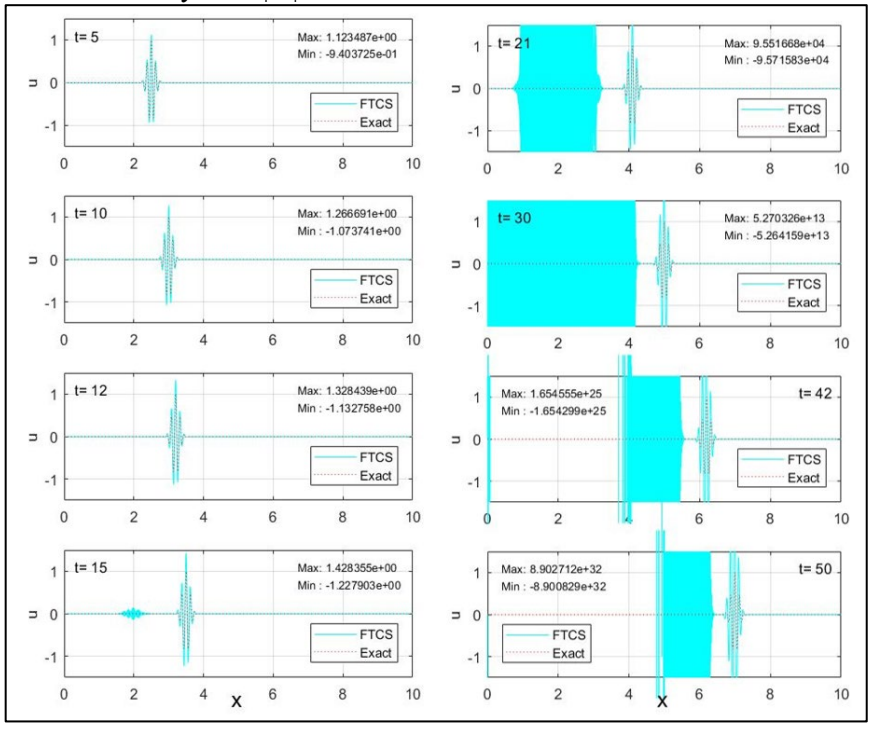

Figure. 3 Results of FTCS method for N_c = 0.09. The analytical solution is compared with exact or real solution at t = 5, 10, 12, 15, 21, 30, 42, 50

Hence, the computed solution also shows amplification in the magnitude from t = 5. Additionally, since the FTCS method is unconditionally unstable, one notices the rise of static instability at initial wave-packet location i.e. x_0 = 2.0 at t = 15, which keeps on amplifying with time. Furthermore, as V_{gN}/c and c_N/c shows contamination in the third/ second decimal place (Figure. 1), thus, one notices in Figure 3b, both dispersion and dissipation errors at later times of t = 30, which increases with time.

3.2 Solution of 1D convection Equation using RK4-CD2

To examine the impact of utilizing a higher order time discretization scheme on the analytical solution, a fourth order (RK4) method with four stages is used while keeping the spatial discretization method same as CD2 for the same CFL number of *0.09.* Figure 4 below shows the analytical solution with solid lines and exact solution by dashed lines at indicated times of t = 5, 10, 12 and 15. Here, N_c = 0.09 lies in the neutrally stable region (from Figure 2, $|G|$=1) for the entire wavenumber range. Hence, the computed solution magnitude matches completely with that of the exact solution. However, the computed solution shows amplification in the maximum magnitude with respect to time with $(|G|_{max})$ as 1.018 at t = 5 and $(|G|_{max})$ as 1.057 at t = 15. Figure 4 also shows the computed and exact solution by solid and dashed lines respectively at later times of t = 21, 30, 42 and 50. Here also, the amplitude keeps on increasing with $(|G|_{max})$ = 1.197 at t = 50. However, one notices the dissipation errors and dispersion errors in the computed solution from t = 21, which keeps on increasing with respect to time. This is due to the contamination by V_{gN}/c and c_N/c in the third/ second decimal as shown in Figure 2 for the RK4-CD2 method. Though, the computed solution does not show instability for any moment of the simulation carried out up to t = 50.

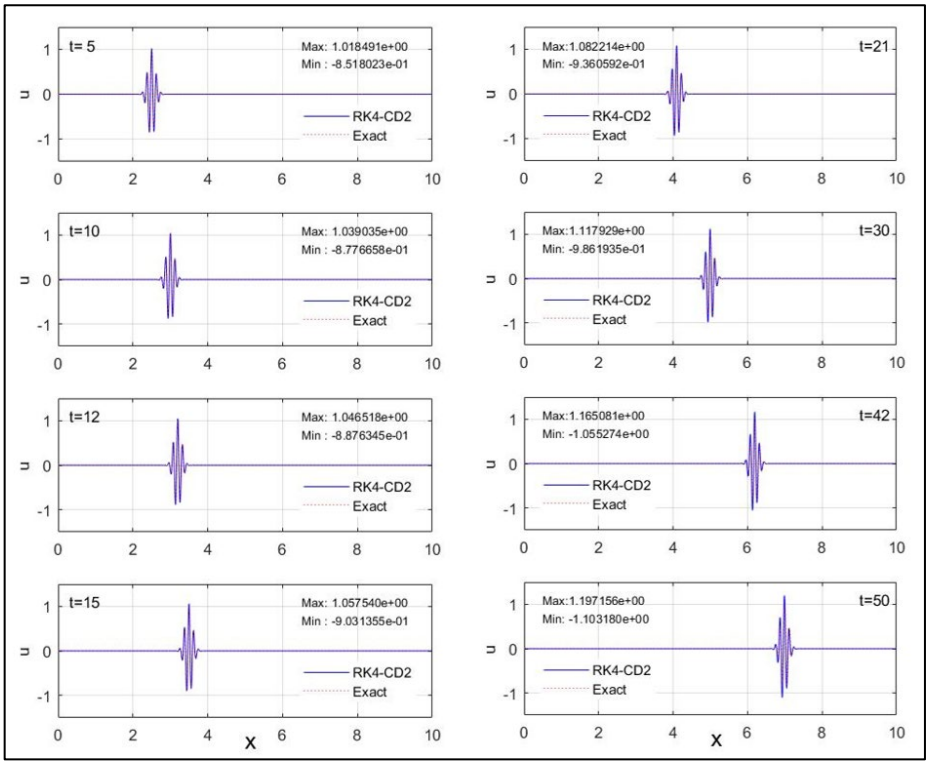

Figure. 4 Results of RK4-CD2 method for N_c = *0.09*. The analytical solution is compared with the exact or true solution at t = 5, 10, 12, 15, 21, 30, 42, 50

In order to examine the effect of CFL number on the stability of the computed solution, the RK4-CD2 method is computed for a higher N_c value of 0.1 as discussed in the following section. For the same grid size and physical phase speed, higher value of CFL number offers the advantage of computing faster as one can use a higher value of time-step

3.3 Comparison

Figure 5 shows the comparison of the error = $u_{exact} - u_{computed}$, in the computed solution at the same time instant between (a) FTCS method for N_c = 0.09, (b) RK4-CD2 scheme for N_c = 0.09 and (c) RK4-CD2 method for N_c = 0.1. The error plotted for the FTCS scheme for N_c = 0.09 in frame (a) shows the generation and evolution of static and local instability at the initial wave packet location of x_0 = 2.0 at t = 2. Also, the error magnitude is of the order of 0.2 at t = 8 and is increasing with time with a value of 0.4 at t = 15, consistent with the observation from the spectral stability analysis which showed the method to be unstable. However, for the RK4-CD2 method, the error is plotted for two different values of CFL number, N_c = 0.09 and for N_c = 0.1 represented in frames (b) and (c), respectively. From frame (b), it is noticed that the error magnitude for the RK4-CD2 method is lower in comparison to that for the FTCS method for the same CFL number of 0.09, owing the higher order discretization RK4 method. On comparing, frames (b) and (c), one notices that on increasing the value of N_c from 0.09 to 0.1 for the RK4-CD2 method, the error magnitude increases from 0.15 to 0.25 (approx.) at a time instant of t = 15. Although, from frames (a) and (c), it is observed that the error is still lower in comparison to FTCS method where the error magnitude is of the order of *0.4* at t = 15. Additionally, the RK4-CD2 solution for both the values of N_c is free from numerical instability.

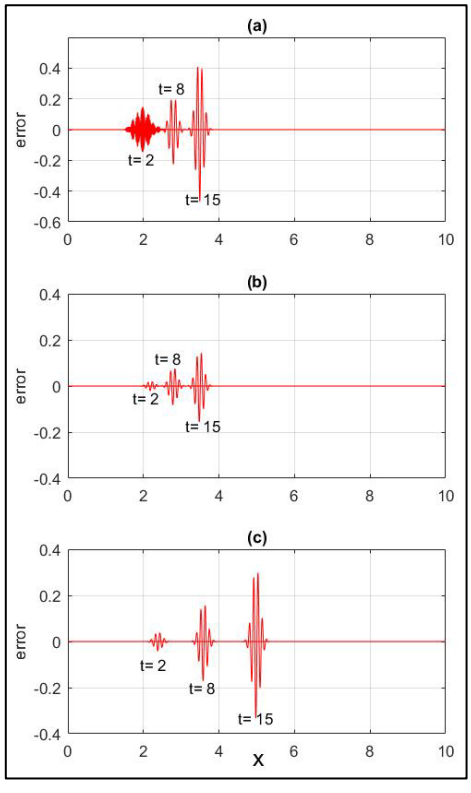

Figure 5. Variation of error for (a) FTCS scheme for N_c = *0.09*, (b) RK4 - CD2 scheme for N_c = *0.09*, and (c) RK4-CD2 scheme for N_c = *0.1*.

4. SUMMARY AND CONCLUSIONS

In order to numerically solve the Navier-Stokes equation accurately, the chosen numerical methods should capture the physics of the flow without contaminating the solution with errors. Thus, the efficacy of the discretization scheme for each and every term in the governing equation is investigated. The present research work is one such investigation where a fundamental one-dimensional (1D) convection equation is solved as it governs many convection dominated flows, using different numerical methods. The 1D convection equation is non-dispersive in nature and is used for spectral stability analysis of various space-time discretization methods.

Here, the one-dimensional convection equation was assessed and a stability analysis was performed using the first order forward-in time and second order centered in space (FTCS) methods. To illustrate the effect of using a higher order time integration scheme, the four-stage fourth order (RK4) method along with CD2 for spatial discretization is also analyzed in the spectral plane.

It is found that the FTCS method is unconditionally unstable while the RK4 method provides neutrally stable solution for particular value of numerical parameter (CFL number, N_c). With the observations obtained from the spectral property charts, the model 1D convection equation is solved using FTCS and RK4-methods for particular values of N_c corresponding to near-neutral stability.

The FTCS results plotted in Figure 3 for N_c = 0.09, shows unstable solution with increasing magnitude from the beginning of the computation. Additionally, at some intermediate time of t = 15 onwards, one notices the generation of local instability where the wave-packet is placed initially at t = 0. This instability keeps on amplifying with respect to time, consistent with |G|-contours greater than 1 for any value of N_c, plotted in Figure 1. Furthermore, one can notice the dispersion and dissipation errors at later times of t = 30, due to the deviation in the V_{gN}/c and c_N/c values from unity (Figure. 1), respectively.

Unlike the FTCS method, the results of RK4-CD2 method plotted in Figure 4, did not show instability for any time instant for Nc = 0.09. The computed solution overlaps with the original or exact solution at examined times of t = 5, 10, 12, and 15. However, at later times, one notices the presence of errors known as dissipation errors and dispersion errors increasing with increasing time as depicted in Figure 2.

The contamination by V_{gN}/c and c_N/c is of the order of 10^{-2} and 10^{-3}, respectively. A bigger value of CFL number of 0.1 is used to compute the solution using the RK4-CD2 method with the objective to evaluate the variation resulting from a higher time-step value on the accuracy of the solution and to save computing time.

Observations to be made from Figure 5, which shows the error comparison between the FTCS (top frame) and RK4-CD2 (middle frame) methods for the same N_c = 0.09, are that the error is larger for the FTCS method.

At the same time, static instability is observed for FTCS, which is absent for RK4-CD2. The bottom frame shows the RK4-CD2 result for a larger value of N_c = 0.1, which has a larger error than that of RK4-CD2 for N_c = 0.09 but still smaller than that of the FTCS method.

Additionally, even for larger N_c value the RK4-CD2 method does not show any sign of numerical instability.

In the future, the present study can also be performed for two- and three-dimensional flows, such as benchmark problem of 2D Lid-driven cavity and flows for which an analytical solution for the Taylor-Green instantaneous vortex exists, to analyze the efficacy of space-time dependent numerical methods in uniform and non-uniform structured grids.

REFERENCES

[1] T. K. Sengupta, *High Accuracy Computing Methods*, Cambridge University Press, 2004.
[2] T. K. Sengupta, A. Dipankar, and P. Sagaut, Error dynamics: Beyond von Neumann analysis, *Journal of Computational Physics*, vol. **226**, pp. 1211-1218, 2007.
[3] B. Gustafsson, H.-O. Kreiss & A. Sundström, Stability theory of difference approximations for mixed initial boundary value problems. *Mathematics of Computation,* **26**(119), 649-686, 1972.
[4] L. N. Trefethen, Group velocity interpretation of the stability theory of GKS, *Journal of Computational Physics*, vol. **49**, pp. 199-217, 1983.
[5] L. N. Trefethen, Group velocity in finite difference schemes, *SIAM Review*, vol. **24**, no. 2, pp. 113-136, 1982.
[6] Lord Rayleigh, Scientific Papers, vol. **1**, Cambridge University Press, 1889.
[7] Lord Rayleigh, Scientific Papers, vol. **2**, Cambridge University Press, 1889.
[8] V. K. Suman, T. K. Sengupta, C. J. Durga Prasac, K. Surya, and D. Sanwalia, Spectral analysis of finite difference schemes for convection diffusion equation, *Computers & Fluids*, **150**, C, pp 95-114, DOI 10.1016/j.compfluid.2017.04.00, 2017.