Implementation of Genetic Algorithm for the optimization of boundary characteristics of compact finite difference schemes

Sina Haeri, Jae Wook Kim

Aerodynamics and Flight Mechanics Research Group, University of Southampton, Southampton, SO17 1BJ, UK

Abstract

This paper presents an advanced strategy for the optimization of compact finite difference schemes with particular emphasis on boundary closure schemes. The present work is based on fourth-order pentadiagonal schemes on seven-point stencils. In contrast to the previous optimization strategies that required trial-and-error routines, the present paper introduces a fully comprehensive optimization platform that is systematically formulated without the need of ad-hoc procedures. This work employs a *Genetic Algorithm* to efficiently deal with a large-scale optimization problem which consists of both linear and non-linear constraints. The linear constraints are formulated with the maximum degree of freedom to minimize resolution errors of the boundary closure schemes. The non-linear constraints are included in order to ensure the numerical stability of the implicit pentadiagonal schemes by mapping the eigenvalues of progressively larger systems onto a single non-linear constraint. Due to the non-linear and non-continuous nature of the constraints, the Genetic Algorithm is suggested as the only feasible optimization technique for this problem. The optimized schemes in this paper are tested in several benchmark problems including one-dimensional linear wave convection and two-dimensional vortex convection in an inviscid flow. Significantly increased accuracy and robustness of the present schemes are demonstrated.

Keywords: Compact finite differences, Boundary closure schemes,

Email address: J.W.Kim@soton.ac.uk (Jae Wook Kim)

1. Introduction

Compact finite difference schemes have been widely used in computational fluid dynamics in past two decades particularly after the seminal publication of Lele [1] where the properties of different families of high-order compact schemes were extensively studied. Some of the application areas relating to the flow simulation problems include CAA (computational aeroacoustics) [2, 3], LES (large eddy simulation) [4–6] and DNS (direct numerical simulation) [7–9]. Implicit compact schemes require an inversion of a banded matrix for the calculation of the derivatives and hence are usually more expensive than their explicit counterparts. However, implicit compact schemes can provide significantly higher accuracy than the explicit schemes for a given stencil with the same size. Especial attention should be paid to the boundary stencils that close the banded matrix system, in order to maintain the superiority of compact schemes throughout the domain. This is usually done by writing one-sided differences for N_b boundary points where N_b depends on the number of non-zero diagonals. It is, however, a formidable task to provide these boundary points the same level of spectral resolution and the same order of accuracy as those of the interior points in addition to ensuring the numerical stability of the entire system.

Kim [2] presented an optimized fourth-order accurate pentadiagonal compact scheme for CAA applications. The optimization was performed by using an integral error function similar to Kim and Lee [10], which provided a maximum resolution error of 0.1% for the wavenumbers in the range $[0, r]$ with $r = 0.839\pi$. He then proposed a set of boundary closures for this pentadiagonal system ($N_b = 3$) preserving the fourth-order accuracy consistently on all the boundary points. In this process, he first extrapolated the internal field (for both f and f') beyond the boundary by combining a fourth-order polynomial and a trigonometric series. The extrapolation used near the boundaries converted the original central differences to a set of three one-sided differences. A free parameter was introduced in the trigonometric series to optimize the final one-sided differences after some algebraic manipulations. Following Carpenter et al. [11], Kim used an eigenvalue analysis to assess the linear stability of the entire pentadiagonal system. Using grid refinement he showed that the system satisfied linear stability in an asymptotic manner.

Based on the same optimization strategy used in [2], Liu et al. [12] iteratively changed the upper bound of resolution range r and used sequential quadratic programming (SQP) technique to find the optimum values for the parameters. They also used the same iterative technique for the optimization of the boundary schemes. Liu et al. [12] showed that the optimization procedure can cause serious stability issues and hence reduced one order of accuracy for the first and third boundary points to provide a stable scheme. Carpenter et al. [11] previously noted that the stability of a numerical scheme and its optimality for spectral resolution do not necessarily go hand in hand and some families of stable schemes may not have a high level of spectral resolution.

Jordan [13] proposed to use composite templates to assess the real dispersive and dissipative properties of compact schemes. A composite template is constructed by applying Fourier transform to the entire matrix system of interior and boundary compact schemes for a specific number of grid points, which provides pseudo-wavenumber curves varying with the position in the grid. This differs from the conventional Fourier analysis that treats individual (boundary and interior) point separately. He used a least-square strategy to optimize the composite templates (minimizing the overall resolution error across the domain) and applied this strategy to tridiagonal systems. Later he applied a similar method to pentadiagonal systems [14] and suggested a set of three boundary schemes ($N_b = 3$) for pentadiagonal systems to be used in conjunction with the interior scheme of Kim [2]. Although the idea seems appealing, the authors have found that there are a few question marks associated with this technique. Firstly, it is not a trivial task at all, to define the most effective objective function (the error measure to be minimized) for these templates. Secondly, the templates are a function of the number of grid points used for their calculation, which means the optimization should be carried out on a case-by-case basis. In addition, the optimization of the composite templates was carried out without a constraint for numerical stability. The authors however, acknowledge that the composite template approach could be increasingly effective if some of these issues are resolved later on.

This paper aims to present a comprehensive optimization strategy that systematically incorporates constraints for both numerical stability and accuracy. In all the previous studies, as mentioned above, the numerical stability of a developed scheme is only examined after performing the optimization for accuracy. This inevitably requires a trial-and-error routine until a stable system is found, which is genuinely not an optimization process. In the

meantime, imposing a certain order of accuracy, p, automatically puts n constraints (generally $p \neq n$) on the m available coefficients, which leaves $m-n$ free parameters. The choice of these free parameters is arbitrary and therefore requires a trial-and-error as well. In this paper, the authors propose a new optimization strategy that requires no such trial-and-error routines outside the optimization process. This is achieved by adopting an advanced evolutionary technique, Genetic Alrogithm (GA), and including the stability criterion as a non-linear constraint in the optimization process. This guarantees that the outcome of the optimization is unconditionally stable and no further tests are required. In addition, all the equations to satisfy p-order accuracy are formulated as general linear constraints and all the m parameters are made available for the optimization, which provides maximum flexibility without the need of manual (trial-and-error) choices of the free parameters. The complex set of linear and non-linear constraints are effectively handled by Genetic Algorithm. This strategy is then applied to the optimization of the boundary compact schemes based on the pentadiagonal platform and the results are tested through a number of benchmark cases. Significant improvements are observed in terms of both accuracy and stability compared to previously optimized schemes.

2. Compact finite difference schemes

A general implicit pentadiagonal compact finite difference scheme can be presented by [2]

$$\beta\bar{f}'_{i-2}+\alpha\bar{f}'_{i-1}+\bar{f}'_i+\alpha\bar{f}'_{i+1}+\beta\bar{f}'_{i+2} = \frac{1}{h}\sum_{m=1}^{3} a_m \left(f_{i+m} - f_{i-m}\right) \quad \text{for} \quad 3 \leq i \leq N-3, \tag{1}$$

where $\bar{f}'$ is used to represent the numerical approximation to the analytical derivative f' and h is the grid spacing. This scheme on a grid with N nodes require $N_b = 3$ additional boundary schemes to close the system. Assuming a 7-point stencil for the right hand side similar to Eq.(1), the general form of the required boundary closures can be written by

$$\bar{f}'_i + \gamma_{01}\bar{f}'_{i\pm1} + \gamma_{02}\bar{f}'_{i\pm2} = \pm\frac{1}{h}\sum_{m=1}^{6} b_{0m}\left(f_{i\pm m} - f_i\right) \text{ for } i = \begin{cases} 0, \\ N, \end{cases} \tag{2}$$

$$\gamma_{10}\bar{f}'_{i\mp1} + \bar{f}'_i + \gamma_{12}\bar{f}'_{i\pm1} + \gamma_{13}\bar{f}'_{i\pm2} = \pm\frac{1}{h}\sum_{\substack{m=0\\m\neq1}}^{6} b_{1m}\left(f_{i\pm m\mp1} - f_i\right) \text{ for } i = \begin{cases}1,\\ N-1,\end{cases} \tag{3}$$

$$\gamma_{20}\bar{f}'_{i\mp2} + \gamma_{21}\bar{f}'_{i\mp1} + \bar{f}'_i + \gamma_{23}\bar{f}'_{i\pm1} + \gamma_{24}\bar{f}'_{i\pm2} = \pm\frac{1}{h}\sum_{\substack{m=0\\m\neq2}}^{6} b_{2m}\left(f_{i\pm m\mp2} - f_i\right) \text{ for } i = \begin{cases}2,\\ N-2.\end{cases} \tag{4}$$

A classical method to analyse the dissipative and dispersive errors of finite difference methods is through the use of Fourier analysis which is extensively described in [15] and later used by others to characterize the errors of their finite difference approximations. Fourier series on a periodic interval $[0, L]$ can be used to decompose the dependant function values into a set of simple oscillatory functions. Assuming that the domain is discretized into $N+1$ points, by truncating the infinite sum to the available $N+1$ points $(0, \cdots, N)$ the Fourier expansion can be written by

$$f(x) = \sum_{k=-N/2}^{N/2} \hat{f}(k)\exp\left(\frac{\jmath 2\pi k x}{L}\right), \tag{5}$$

where $\hat{f}(k)$ are the Fourier coefficients and $\jmath = \sqrt{-1}$. Similar to [1], by defining $\omega = 2\pi kh/L$ and $x^* = x/h$ the Fourier modes, $\exp(\jmath 2\pi kx/L)$, conveniently reduce to $\exp(\jmath\omega x^*)$. Noting that $\bar{f}'_{i\pm m} \equiv \bar{f}'(x_i \pm mh)$ and $f_{i\pm m} \equiv f(x_i \pm mh)$, following equations can be derive by applying Eq.(5) to Equations (1)–(4):

$$\begin{aligned}
&\jmath\bar{\omega}\hat{f}(k)\left(1 + 2\alpha\cos(\omega) + 2\beta\cos(2\omega)\right) = \\
&2\jmath\hat{f}(k)\sum_{m=1}^{3} a_m\sin(m\omega) \quad \text{for} \quad 3 \le i \le N-3
\end{aligned} \tag{6}$$

$$\begin{aligned}
&\jmath\bar{\omega}_i\hat{f}(k)\left[A_i(\omega) + \jmath B_i(\omega)\right] = \\
&\hat{f}(k)\left[C_i(\omega) + \jmath D_i(\omega)\right] \quad \text{for} \quad 0 \le i \le 2, \quad \text{or} \quad N-2 \le i \le N,
\end{aligned} \tag{7}$$

where

$$i=0, i=N: \begin{cases} A_i(\omega) = 1 + \gamma_{01}\cos(\omega) + \gamma_{02}\cos(2\omega) \\ B_i(\omega) = \gamma_{01}\sin(\omega) + \gamma_{02}\sin(2\omega) \\ C_i(\omega) = \sum\limits_{m=1}^{6} b_{0m}\left(\cos[m\omega] - 1\right) \\ D_i(\omega) = \sum\limits_{m=1}^{6} b_{0m}\sin(m\omega), \end{cases} \tag{8}$$

$$i=1, i=N-1: \begin{cases} A_i(\omega) = 1 + (\gamma_{10}+\gamma_{12})\cos(\omega) + \gamma_{13}\cos(2\omega) \\ B_i(\omega) = (\gamma_{12}-\gamma_{10})\sin(\omega) + \gamma_{13}\sin(2\omega) \\ C_i(\omega) = \sum\limits_{\substack{m=0 \\ m\neq 1}}^{6} b_{1m}\left(\cos\left[(m-1)\omega\right] - 1\right) \\ D_i(\omega) = \sum\limits_{\substack{m=0 \\ m\neq 1}}^{6} b_{1m}\sin\left[(m-1)\omega\right], \end{cases} \tag{9}$$

$$i=2, i=N-2: \begin{cases} A_i(\omega) = 1 + (\gamma_{21}+\gamma_{23})\cos(\omega) + (\gamma_{20}+\gamma_{24})\cos(2\omega) \\ B_i(\omega) = (\gamma_{23}-\gamma_{21})\sin(\omega) + (\gamma_{24}-\gamma_{20})\sin(2\omega) \\ C_i(\omega) = \sum\limits_{\substack{m=0 \\ m\neq 2}}^{6} b_{2m}\left(\cos\left[(m-2)\omega\right] - 1\right) \\ D_i(\omega) = \sum\limits_{\substack{m=0 \\ m\neq 2}}^{6} b_{2m}\sin\left[(m-2)\omega\right]. \end{cases} \tag{10}$$

In Eq.(6) and (7), $\bar{\omega}$ is a modified wave number which is different from ω due to the numerical approximation by the pentadiagonal FD scheme, i.e. Eq.(1), and the corresponding boundary schemes, Equations (2)–(4). An explicit equation for the modified wavenumbers of the interior points only has a real part and can simply be derived from Eq.(6) which is given by:

$$\bar{\omega}^I = \frac{2\sum_{m=1}^{3} a_m \sin(m\omega)}{1 + 2\alpha\cos(\omega) + 2\beta\cos(2\omega)}. \tag{11}$$

Additionally following explicit expression can be derived for the modified wavenumbers of the boundary points by multiplying both sides of Eq.(7) by $A(\omega) - \jmath B(\omega)$ to yield:

$$\bar{\omega}_i^{BC} = \frac{A_i(\omega)D_i(\omega) - B_i(\omega)C_i(\omega)}{A_i^2(\omega) + B_i^2(\omega)} - \jmath\frac{A_i(\omega)C_i(\omega) + B_i(\omega)D_i(\omega)}{A_i^2(\omega) + B_i^2(\omega)}. \tag{12}$$

3. Formulation of the optimization problem

In this section the optimization of the boundary schemes for pentadiagonal systems is discussed. The GA is succinctly discussed in Section 3.1 which is mainly provided to introduce the terminology and a detailed description of the method is not intended. A fitness function that satisfies several criteria for a high quality definition of the error measure is defined in Section 3.2. General linear constraints that enforce the 4^{th}-order accuracy of the method are discussed in Section 3.3 and the stability criteria which is implemented as a non-linear constraint is discussed in Section 3.4.

3.1. Optimization technique

The main optimization technique used in this paper is the genetic algorithm (GA) which can be classified as an evolutionary global optimization. Although the method is time consuming compared to the traditional gradient based techniques, to our experience, is the only feasible strategy for the optimizations performed in this paper due to the fact that some of the constraints are non-linear and non-continuous. The Matlab implementation of the genetic algorithm is used which conveniently handles both linear and non-linear constraints. We do not intent to delve into the details of the algorithm in this paper however some terms that are frequently used in this paper are explained next.

The term GA was first used by Holland [16] and refers to a class of search techniques that are based on the principles of genetics and natural selection. Later Goldberg [17] properly formulates and applies the method to the optimization problems. The GA allows an initial population of N_p chromosomes to evolve to a population that maximizes the fitness. In the GA terminology the fitness function is simply the the objective function (dependant variable) that we are trying to minimize (maximize) and a chromosome translate to a vector of N_{var}, real valued independent variables at which the objective function should be evaluated [18]. These chromosomes are then coded and ranked based on their fitness (value of the objective function at each point). One can crudely think of the coding as creating a vector of binary presentations of each of the independent variables. For example if the objective function is a function of three independent variables x_1, x_2 and x_3 which are presented by 01, 10 and 11 respectively a chromosome can be the vector $C = (0, 1, 1, 0, 1, 1)$ and a population is the vector $P = (C_1, \cdots, C_{N_p})$, however discussing the actual encoding process and the implementation issues

are outside the scope of this paper, c.f. [18].

After randomly choosing an initial population, chromosome are selected two by two based on a probability calculated from their rank, to produce offspring (new points). This process which is called crossover then continues until a set of new points is created. Mutation is the other important component of the algorithm where one or more bits of a number of generated chromosomes are randomly changed (mutated) which prevents the algorithm from getting stuck in local minima. Points with the highest ranks in the current population, i.e the points corresponding to the current best values of the objective function, are usually called elites. It is also guaranteed that during the course of the algorithm, a number of elites, usually 2, are always survive for the next iteration. This process continues until some convergence criteria such as a tolerance for the minimum change in the objective function values or the constraints is met.

3.2. Formulation of the fitness function

In this section the fitness (objective) function used for the optimization of the resolution error is discussed in detail. The objective of an optimization strategy is to force the real and imaginary parts of the modified wavenumber to follow the true wavenumbers, i.e.

$$\mathrm{Re}(\bar{\omega}) \to \omega \tag{13}$$

$$\mathrm{Im}(\bar{\omega}) \to 0. \tag{14}$$

Considering Eq.(12) it is easy to show the equivalence of the requirements given by Equations (13) and (14) and requiring that $\mathscr{E}_i^L \to 0$, where $\mathscr{E}_i^L$ is an integral error measure defined by

$$\mathscr{E}_i^L|_{r_l}^{r_u} = \left\{ \int_{r_l}^{r_u} \left([\omega A_i(\omega) - D_i(\omega)]^2 + [\omega B_i(\omega) + C_i(\omega)]^2 \right) d\omega \right\}^{\frac{1}{2}} \text{ for } i \in \{0, 1, 2\}. \tag{15}$$

The parameters r_l and r_u are ideally set to 0 and π and $A_i \cdots D_i$ are given by Equations (8)–(10). Also note that the superscript L is used to refer to an error that is defined between the real and imaginary parts of the modified wavenumber curves and the straight lines $y = x, \forall x \in [0, 1]$ and $y = 0, \forall x \in [0, 1]$ respectively. Another possible objective function that can be defined for the boundary schemes is to require them to follow the modified wavenumber

of the interior points as closely as possible, i.e. $\mathscr{E}_i^{eI} \to 0$, where $\mathscr{E}_i^{eI}$ is defined by

$$\mathscr{E}_i^{eI}|_{r_l}^{r_u} = \left\{ \int_{r_l}^{r_u} \left(\left[\bar{\omega}^I - \mathrm{Re}(\bar{\omega}_i^{BC}) \right]^2 + \left[\mathrm{Im}(\bar{\omega}_i^{BC}) \right]^2 \right) d\omega \right\}^{\frac{1}{2}}, \tag{16}$$

and $\bar{\omega}^I$ and $\bar{\omega}^{BC}$ are given by Eq.(11) and Eq.(12) respectively. Also note that the superscript I is used to indicate that the error is defined between the interior and the boundary curves.

It is important to ensure that Eq.(13) and Eq.(14) are satisfied with a very high precision ($0.1\% - 0.5\%$ error [2, 10]) up to at least a critical wavenumber $\omega^c < \pi$. By setting $r_u = \omega^c$, the integral error meassure defined by Eq.(15) ensures the best possible performance up to $r_u = \omega^c$. It is obviously desirable to choose maximum value for ω^c, however setting $\omega^c = \pi$ and applying the other required constraints, which are discussed later in this section, the GA produces very large overshoots near ω^c. It is possible to remove these high frequencies by filtering operations [19], however it seems these large overshoots require strong filters with long cut-off frequencies which deteriorate the quality of the final solution. Note that Eq.(16) does not suffer from this problem however it also does not gaurantee the best performance up to the critical wavenumebr ω^c. A remedy would be to define the following blended error meassure

$$\mathscr{E}_{\omega_i^c} = w_i \mathscr{E}_i^{eL}|_0^{\omega_i^c} + (1 - w_i)\mathscr{E}_i^{eI}|_{\omega_i^c}^{\pi}, \tag{17}$$

where $w_i > 0.95$ is a weighting factor. Note that this parameter is not an optimization parameter and if added to the optimization procedure will always assume the smallest value since this relaxes the strict condition that requires the modified wavenumber curves to precisely follow the true wavenumber curves up to some ω^c. In this paper w_0 is set to 0.97 and $w_{1,2} = 0.99$. Similarly the value of ω^c is not an optimization parameter since again it will always assume the smallest possible value enforced by the lower bound constraint or zero if no constraint imposed. However after fixing a value for w_i the best value for ω^c can quickly be determined by a bisection type strategy by choosing two values for ω^c and calculating the mid-point then by assessing the errors best of these three points are used for the next iteration. In this paper ω_0^c is set to 0.75π and $\omega_{1,2}^c = 0.8$. Following single error measure is finally defined by adding the the errors from the three boundary schemes which is used as the main fitness function for the optimization procedure

$$\mathscr{E} = \frac{1}{3}\sum_{i=0}^{2} \mathscr{E}_{\omega_i^c} \tag{18}$$

3.3. Formulation of the constraints: Linear constraints

Certain relations should hold between the variables of the optimization procedure to retain the order of accuracy of the final optimized scheme. Since number of optimization variables (m) is larger than the number of constraints (n) it is customary to arbitrarily set $(m-n)$ variables [1, 2, 14] and use the remaining ones for the optimization. However it is not obvious which of these parameters should be fixed and hence at least a few iteration is required and still it is not guaranteed the final solution is actually the optimum solution. Note that in the current optimization procedure $m = 27$ and there are $n = 12$ constraints to be satisfied to preserve the 4^{th}-order accuracy of the method near the three boundary points. It is obviously not possible to test all the $\binom{m}{m-n}$ combinations and hence general linear constraints are derived to maximize the flexibility of the optimization procedure. These constraints can be written in the following simple form

$$\mathbf{A}\mathbf{x} = \mathbf{b}, \tag{19}$$

where $\mathbf{x}$ is the vector of optimization variable given by

$$\begin{aligned}
\mathbf{x}(1:8) &= [\gamma_{01}\ \gamma_{02}\ b_{01}\ \cdots\ b_{06}]^T \\
\mathbf{x}(9:17) &= [\gamma_{10}\ \gamma_{12}\ \gamma_{13}\ b_{10}\ b_{12}\ \cdots\ b_{16}]^T \\
\mathbf{x}(18:27) &= [\gamma_{20}\ \gamma_{21}\ \gamma_{23}\ \gamma_{24}\ b_{20}\ b_{21}\ b_{23}\ \cdots\ b_{26}]^T,
\end{aligned} \tag{20}$$

and superscript T means a transpose. Components of matrix $\mathbf{A}$ are derived by matching the corresponding terms in the Taylor series expansion of the boundary schemes given by Equations (2)–(4) to satisfy the 4^{th}-order accuracy. Sections of A with $A_{ij} \neq 0$ can be written as

$$\mathbf{A}(1:4,1:8) = \begin{pmatrix}
1 & 1 & -1 & -2 & -3 & -4 & -5 & -6 \\
2 & 2\cdot 2 & -1 & -2^2 & -3^2 & -4^2 & -5^2 & -6^2 \\
3 & 3\cdot 2^2 & -1 & -2^3 & -3^3 & -4^3 & -5^3 & -6^3 \\
4 & 4\cdot 2^3 & -1 & -2^4 & -3^4 & -4^4 & -5^4 & -6^4
\end{pmatrix}, \tag{21}$$

$$\mathbf{A}(5:8,9:17) = \begin{pmatrix} 1 & 1 & 1 & 1 & -1 & -2 & -3 & -4 & -5 \\ -2 & 2 & 2\cdot 2 & -1 & -1 & -2^2 & -3^2 & -4^2 & -5^2 \\ 3 & 3 & 3\cdot 2^2 & 1 & -1 & -2^3 & -3^3 & -4^3 & -5^3 \\ -4 & 4 & 4\cdot 2^3 & -1 & -1 & -2^4 & -3^4 & -4^4 & -5^4 \end{pmatrix}, \tag{22}$$

$$\mathbf{A}(9:12,18:27) = \begin{pmatrix} 1 & 1 & 1 & 1 & 2 & 1 & -1 & -2 & -3 & -4 \\ -2\cdot 2 & -2 & 2 & 2\cdot 2 & -2^2 & -1 & -1 & -2^2 & -3^2 & -4^2 \\ 3\cdot 2^2 & 3 & 3 & 3\cdot 2^2 & 2^3 & 1 & -1 & -2^3 & -3^3 & -4^3 \\ -4\cdot 2^3 & -4 & 4 & 4\cdot 2^3 & -2^4 & -1 & -1 & -2^4 & -3^4 & -4^4 \end{pmatrix}. \tag{23}$$

Vector $\mathbf{b}$ is zero except for $b(1) = -1$, $b(5) = -1$ and $b(9) = -1$. This concludes the set of 12 general linear constraints that enforce the 4^{th}-order accuracy of the final scheme.

3.4. Formulation of the constraints: non-Linear constraints

One of the main issues in optimization of the finite difference schemes is that the optimized scheme may accurately resolve a large section of the spectrum, however the final scheme is not necessarily stable. In all previous optimization techniques stability analysis were performed after optimizing the FD scheme which reduces the optimization process to a trial and error experiment. Stability of a finite difference scheme is usually evaluated by considering the one dimensional linear wave equation

$$\frac{\partial f}{\partial t} + c\frac{\partial f}{\partial x} = 0, \tag{24}$$

over a domain $x \in [0,1]$ with the boundary conditions $f(x=0,t) = g(t)$. However it was shown in Carpenter et al. [11] that for the purpose of stability analysis $g(t)$ can be set to zero without loss of generality. Same assumption is also used by others to analyse the stability of their schemes [2, 12, 13]. Applying the iterior and the boundary schemes, Eq.(1)–(4), with the boundary condition $f(x=0,t) = 0$, and assuming that the domain is discretized using $N+1$ nodes such that $h = 1/N$, results in the following system of linear equations

$$\mathbf{R}\bar{\mathbf{f}}' = -\frac{c}{h}\mathbf{S}\mathbf{f}. \tag{25}$$

Note that by applying the boundary condition $f(x=0,t) = 0$, first point can be eliminated from the system of equations. Therefore $\mathbf{R}$ and $\mathbf{S}$ are $N \times N$ banded matrices, $\mathbf{f}$ is a vector of function values at N nodes and $\bar{\mathbf{f}}'$ is the

vector of numerical approximations to the derivatives. The expanded form of matrices $\mathbf{R}$ and $\mathbf{S}$ can be found in Kim [2] and is not repeated here. The solution to a generic system of N linear differential equations

$$\mathbf{x}' = \mathbf{M}\mathbf{x}, \tag{26}$$

is given by [20]

$$\mathbf{x} = \sum_{m=1}^{N} c_m \tilde{\mathbf{x}}_m \exp(\lambda_m t), \tag{27}$$

where λ_m are the eigenvalues and $\tilde{\mathbf{x}}_m$ are the corresponding eigenvectors and c_m are constants determined by the initial conditions. By examining Eq.(27) it is obvious that that solutions are bounded only if $\mathrm{Re}(\lambda_m) < 0, \forall m \in \{1, \cdots, N\}$. Comparing Eq.(25) with Eq.(26) shows that the real parts of the scaled eigenvalues $\bar{\lambda}_m = -h\lambda_m/c$ of the matrix $\mathbf{M} = \mathbf{R}^{-1}\mathbf{S}$ should all be negative for the solutions of the system of equations (25) to remain bounded and hence stable in time.

In this paper we propose adding the stability condition as a non-linear constraint to the optimization process by requiring

$$\lambda_{\max} = \max\left(\mathrm{Re}(\bar{\lambda}_m), \forall m \in \{1, \cdots, N\}\right) < 0. \tag{28}$$

To save the computational time one might be tempted to choose the smallest possible system, which is a 7×7 system in this case, and directly use Eq.(28) as the functional form of the non-linear constraints. However there are two issues associated with this approach. Firstly satisfying the stability constraint on one grid level does not ensure the stability on larger grid sizes. Secondly a non-linear constraint for Eq.(28) is naturally implemented by defining

$$C(\lambda_{\max}) = H(\lambda_{\max}) - \frac{1}{2}, \tag{29}$$

where $H(x)$ is the Heaviside step function which can simply be implemented numerically by the following piecewise constant function [21]:

$$H(x) = \begin{cases} 1 & x > 0, \\ 0.5 & x = 0, \\ 0 & x < 0. \end{cases} \tag{30}$$

However this function is not continuous in the vicinity of zero and results in a slow convergence to poor solutions. To address the first problem q

progressively larger systems are used, with $N_1 < \cdots < N_q$ grid points, and we require that

$$\bar{\lambda}^q_{\max} = \max\left(\mathrm{Re}(\bar{\lambda}^i_m), \forall m \in \{1, \cdots, N_i\}, \forall i \in \{1, \cdots, q\}\right) < 0. \tag{31}$$

Additionally by noting that

$$H(x) = \frac{1}{2} \lim_{k\to\infty} \left[1 + \tanh(kx)\right], \tag{32}$$

a smooth approximation to the Heaviside function can be $\frac{1}{2}(1 + \tanh(kx))$, where k controls the steepness near zero and is set to $k = 1$ in this study. Then instead of using $C(\bar{\lambda}^q_{\max}) = H(\bar{\lambda}^q_{\max}) - \frac{1}{2}$ to implement this constraint, we propose using

$$C(\bar{\lambda}^q_{\max}) = \frac{1}{2}\tanh(\bar{\lambda}^q_{\max}). \tag{33}$$

This definition introduces a notion of continuity in the sense that it is a continuous function of its argument and to our experience, improves the convergence rate and quality of the final results. However it should be noted that although $\frac{1}{2}\tanh(x)$ is a continuous function of its argument, the argument to this function $\bar{\lambda}^q_{\max}$ is not a continuous function of the optimization variables $\mathbf{x}$, defined by Eq.(20), and hence the GA still seems to be the only feasible optimization strategy.

4. Results and discussions

In this section first the results of the optimization of the three boundary points that was discussed in Section 3 are presented. The resolution limits of the current scheme are presented and compared to the other currently proposed schemes. Then the stability of numerical scheme is presented and it is shown that the proposed strategy for adding the stability condition ensures the unconditional stability on different grid sizes. Convection of a 1D modulated wave equation is used to show the 4^{th}-order accuracy of the proposed scheme. The scheme is further applied to two benchmark problems namely 1D scalar wave convection and 2D vortex convection problem. The 2D vortex convection is further extended to the problem of two co-rotating vortices to test for unsteady effects.

4.1. Optimization results

Optimization is performed using the GA by implementing the fitness function defined by Eq.(18) and the set of linear constraints, Eq.(19). In addition non-linear constraint $C(\bar{\lambda}^q_{\max})$, defined by Eq.(33), is used with $q = 4$ grid levels, $N \in \{2^3, \cdots, 2^6\}$, which will be shown in this section to be adequate for the current optimization. Initial population of 900 chromosomes are used for the optimization which are chosen randomly by the solver to satisfy the constraints but are not necessarily optimum. Crossover is used to produce 80% of the offsprings and the rest are produced by mutation. In addition 2 elit points are gauranteed to migrate to the next iteration. Tolerance for the minimum relative changes in the solution vector and the function values are set to 10^{-15} and the tolerance for satisfying both linear and non-linear constraints is set to 10^{-12}. The final values of the coefficients are provided in Appendix A. In the next few paragraphs the dissipative and dispersive errors of the current scheme are discussed and compared to a few recently suggested papers [2, 12]. However the set of coefficient provided by Jordan [13] are not presented here since he used composite templates to optimize the scheme and hence comparison of the modified wavenumbers of the individual graphs is irrelevant.

Figure 1 shows the real and imaginary parts of the modified wavenumber curve for the first boundary point $i = 0$. The real part of the first boundary point has a resolution of 1% for $\omega^c_1 < 0.75\pi$ which is smaller than both previous studies [2, 12] that provide a resolution of 1% up to $\omega^c_1 \approx 0.88\pi$. The imaginary part provides a better resolution especially after $\omega = 0.56\pi$ than those provided by Kim [2] however they remain higher than those provided by Liu et al. [12] after $\omega = 0.2\pi$.

Figure 2 shows the error analysis for the second boundary point $i = 1$. The error in the real part of the modified wavenumber up to $\omega = 0.4\pi$ is larger than that proposed by Kim and smaller than the Liu's scheme. The resolution remain within the 1% limit up to $\omega = 0.79\pi$ which is slightly lower than those proposed by the both previous studies. However the error in the imaginary part of the second boundary point of the new scheme never exceeds 0.44% whereas the 1% limit for the Kim and Liu schemes is at 0.81π and 0.826π respectively.

Figure 3 shows the analysis for the third boundary point $i = 2$. The real part of the current scheme has a 1% resolution up to $\omega = 0.8\pi$ which is less than both the previous studies that provided a 1% resolution up to 0.926π. For the imaginary part the 1% resolution limit for the current scheme is at

0.796π whereas for the other two schemes is at 0.83π. However the errors for the other two schemes are increased with a higher slope after this point and the current scheme provides much smaller errors after 0.88π.

4.2. Accuracy and stability analysis

Figure 4 shows the $\mathrm{Im}(\bar{\lambda}_m)$ versus $-\mathrm{Re}(\bar{\lambda}_m)$, plotted for three grid levels $N \in \{64, 128, 256\}$. Which shows that the $-\mathrm{Re}(\bar{\lambda}_m)$, remains on the right half plane which ensures that the real part of the eigenvalues remain negative and the scheme is unconditionally stable. Additionally this shows that the strategy used to ensure stability on all grids by using four levels of progressively larger grids ($N \in \{2^3, \cdots, 2^6\}$) is adequate for this problem. However it is always possible to add more grid levels if the results for larger grids remain unstable at the cost of a longer computational time.

Fourth order formal accuracy of the method is guaranteed by accurately (with a tolerance of 10^{-12}) satisfying the linear constraints discussed in Section 3.3. However it is also important to monitor the next order truncation errors (in this case $O(h^5)$) since to our experience, a general optimization procedure can produce large constants in front of the higher order terms that can mask the order of accuracy of the method for small to medium size grids. By matching the Taylor series terms of the next order similar to Section 3.3, following equations can be written for the truncation errors in Eq.(2)–(4) assuming 4^{th}-order accuracy

$$C_5^0 h^5 f^{(5)}, \quad C_5^0 = 5(\gamma_{01} + 16\gamma_{02}) - \sum_{m=1}^{6} m^5 b_{0m}, \tag{34}$$

$$C_5^1 h^5 f^{(5)}, \quad C_5^1 = 5(\gamma_{10} + \gamma_{12} + 16\gamma_{13}) - \sum_{\substack{m=0 \\ m\neq 1}}^{6} (m-1)^5 b_{1m}, \tag{35}$$

$$C_5^2 h^5 f^{(5)}, \quad C_5^2 = 5(16\gamma_{20} + \gamma_{21} + \gamma_{23} + 16\gamma_{24}) - \sum_{\substack{m=0 \\ m\neq 2}}^{6} (m-2)^5 b_{2m}. \tag{36}$$

Table 1 summarises the values of the constants C_5^0, C_5^1 and C_5^2 and compares the results to previous studies [2, 14]. Note that the optimized coefficients suggested by Liu et al. [12] are not presented since they used 3^{rd}-order accuracy for $i = 0$ and $i = 2$. The value of the constants are all smaller compared

to those proposed by Jordan [14] (single precision) but compared to the coefficients suggested by Kim [2], C_5^1 and C_5^2 are respectively 17% and 70% smaller but C_5^0 is 46% larger.

The stability and accuracy of the current scheme is tested further by integrating the linear wave convection problem discussed in Section 3.4 using a test function. For this problem both the inlet and outlet boundaries are active and domain length is L, i.e. $x \in [-0.5L, 0.5L]$. The initial and boundary conditions are respectively set to

$$f(x, t=0) = f_\infty \left[1 + M \cos\left(\frac{K_1 x}{L}\right)\right] \sin\left(\frac{K_2 x}{L}\right), \tag{37}$$

$$f(x=0, t) = f_\infty \left[1 + M \cos\left(\frac{-cK_1 t}{L}\right)\right] \sin\left(\frac{-cK_2 t}{L}\right), \tag{38}$$

which is a carrier wave with a high frequency $K_2 = 25K_1$ and an amplitude f_∞, modulated by a low frequency wave with amplitude $M = 1.5$ and frequency $K_1 = 2\pi$. The exact solution to this problem is given by

$$f_{\text{exact}}(x, t) = f_\infty \left[1 + M \cos\left(\frac{K_1 \hat{x}}{L}\right)\right] \sin\left(\frac{K_2 \hat{x}}{L}\right), \quad \hat{x} = x - ct. \tag{39}$$

Figure 5 shows the wave convection at four different stages during one full period of motion with $N = 400$ grid points. Even near the boundaries no errors can be detected from this figure. To test both the stability and convergence rate of the current scheme the time integration is carried out for a relatively long period up of $t = 150L/c$ for $N = 200$ and $N = 400$. To save the computational time integration is limited to $t = 10L/c$ for $N = 800$ and $N = 1200$. In addition, to control the time integration errors and concentrate on the spatial errors, time step size is set to be much smaller than the CFL stability threshold by setting CFL = 0.2. An ℓ_2 and an ℓ_∞-norm are then defined by

$$\ell_2(t_n) = \left\{ \sum_{i=1}^{N} [f_i(t_n) - f_{\text{exact}}(t_n)]^2 / (N f_\infty^2) \right\}^{\frac{1}{2}} \tag{40}$$

$$\ell_\infty(t_n) = \max\left(|f_i(t_n) - f_{\text{exact}}(t_n)| / f_\infty, \forall i \in \{0, \cdots, N\} \right). \tag{41}$$

Figure 6 shows the average value of the both ℓ_∞ and ℓ_2 norms during the whole period of motion. Fourth-order convergence rate of the current scheme

is retained in both norms. Only the current scheme preserves the shape of the wave on the coarsest grids grid level $N = 200$. For a finer grid the coefficients suggested by Jordan [14] provide a smaller mean error at CFL = 0.2. For larger CFL numbers or coarser grids ($N = 200$ and $N = 400$), we were unable to generate stable calculations using the coefficients of Jordan [14]. Figure 7 shows the time history of the ℓ_2-norm on the grid with $N = 400$. All schemes except the one suggested by Jordan [14] remain stables with errors oscillating around a mean value. This explains larger than fourth-order convergence rate of this method in Figure 6. Note also that compared to the next unconditionally stable set of coefficients [12] a 3.36 fold improvement is observed on average on the four levels of grids used. Also note that in this set of coefficients (Liu et al. [12]), the local fourth order accuracy is not preserved.

4.3. Benchmark problem: Scalar wave convection

In this benchmark problem an initial wave pulse in the domain is convected through the computational boundary. The main difference between this problem and the 1D wave case used in Section 4.2 is that the wave completely leaves the computational domain and the numerical method should provide a clean zero solution. Therefore this problem is used to examine the properties of the schemes in reflecting the high frequency errors at the boundary. The problem was originally proposed in the Fourth Computational Aeroacoustics Workshop on Benchmark Problems [22]. The problem consists of an initial pulse

$$f(x, t = 0) = f_\infty \left(2 + \cos(\frac{a_1 x}{L})\right) \exp\left(-\frac{a_2 \ln(2) x^2}{L^2}\right), \tag{42}$$

defined over the interval $x \in [-0.5L, L]$ which is convected by Eq.(24) with the constants $a_2 = 100$ and $a_1 = 1.7a_2$. The exact solution to this problem is given by

$$f_{\text{exact}}(x, t) = f_\infty \left(2 + \cos(\frac{a_1 \hat{x}}{L})\right) \exp\left(-\frac{a_2 \ln(2) \hat{x}^2}{L^2}\right), \tag{43}$$

where $\hat{x} = x - ct$. In this problem the outlet boundary is active however since no wave is entering the domain the interior discretization is used on the inlet by setting $f(x < -0.5L, t) = f'(x < -0.5L, t) = 0$ for $t \geq 0$ similar to [2]. Figure 8 shows the motion of the single initial wave in the domain

at three different stages using $CFL = 0.5$ and $N = 1000$. The errors can not be detected in this figure and hence the ℓ_2-norms defined by Eq.(40) (for $i = 0 \cdots N$) are presented in Figure 9. Errors of the current scheme show no overshoot during the time the wave is leaving the domain whereas all previous studies show significant overshoot in the ℓ_2-norm of the error when the wave is leaving the boundary. Another important factor specially in aeroacoustic applications is that the remaining errors after a wave leaves the domain approach zero. Figure 9 shows that this error is less that 10% of errors predicted by the Jordan [14] results and less than 16% of both Kim [2] and Liu et al. [12] results. Maximum error which happens around $tc/L = 1$ is calculated for all four schmemes and are plotted in Figure 10 for five grid levels. Again the current optimized scheme provides better results than all previous schemes with a fourth order convergence rate.

4.4. Benchmark problem: 2D vorticity convection

In this section the problem of a vorticity wave convection in a supersonic stream is chosen as the second benchmark problem. The problem involves the solution of the 2D compressible Euler equations in full conservative form in general coordinates

$$\frac{\partial \hat{\mathbf{Q}}}{\partial t} + \frac{\partial \hat{\mathbf{E}}}{\partial \xi} + \frac{\partial \hat{\mathbf{F}}}{\partial \eta} = \mathbf{0}, \tag{44}$$

where the vector of conservative variables and fluxes in general coordinates are given by

$$\begin{aligned} \hat{\mathbf{Q}} &= J\mathbf{Q}, \\ \hat{\mathbf{E}} &= J\left(\xi_x \mathbf{E} + \xi_y \mathbf{F}\right), \\ \hat{\mathbf{F}} &= J\left(\eta_x \mathbf{E} + \eta_y \mathbf{F}\right). \end{aligned} \tag{45}$$

In Eq.(45), $J = (x_\xi y_\eta - x_\eta y_\xi)$ is the determinant of Jocobian of the transformation in 2 spatial dimensions and we have

$$\begin{pmatrix} \xi_x & \xi_y \\ \eta_x & \eta_y \end{pmatrix} = J^{-1} \begin{pmatrix} y_\eta & -x_\eta \\ -y_\xi & x_\xi \end{pmatrix}. \tag{46}$$

The conservative variables and fluxes in Cartesian coordinates are given by

$$\mathbf{Q} = \begin{pmatrix} \rho \\ \rho u \\ \rho v \\ \rho e_t \end{pmatrix}, \quad \mathbf{E} = \begin{pmatrix} \rho u \\ \rho u^2 + p \\ \rho uv \\ \rho(e_t + p)u \end{pmatrix} \quad \text{and} \quad \mathbf{F} = \begin{pmatrix} \rho v \\ \rho uv \\ \rho v^2 + p \\ \rho(e_t + p)v \end{pmatrix}. \tag{47}$$

In Eq.(47), $e_t = p/[(\gamma-1)\rho] + (u^2+v^2)/2$ is the total energy per unit mass and $\gamma = c_p/c_v$ is the ratio of the specific heat in constant pressure to that in constant volume which is set to its value for air, $\gamma = 1.4$, for this test problem. The problem was first used by Yee et al. [23] for the validation of their high-order shock capturing scheme and filters. The field variables at $t = 0$ are initialized to

$$\left.\begin{aligned} \frac{\rho(x,y)}{\rho_\infty} &= \left(1 - \frac{\gamma-1}{2}\psi^2(x,y)\right)^{\frac{1}{\gamma-1}} \\ \frac{u(x,y)}{a_\infty} &= M_\infty + (K_1 y)\psi(x,y) \\ \frac{v(x,y)}{a_\infty} &= -(K_1 x)\psi(x,y) \\ \frac{p(x,y)}{p_\infty} &= \left(\frac{\rho}{\rho_\infty}\right)^\gamma \end{aligned}\right\}, \text{ for } \begin{cases} \forall x \in [-0.5L, L] \\ \forall y \in [-0.75L, 0.75L], \end{cases} \tag{48}$$

with

$$\psi(x,y) = \frac{K_2}{2\pi}\exp\left(\frac{1 - K_1^2(x^2+y^2)}{2}\right), \tag{49}$$

where $K_1 = 12.5$ and the value of K_2 is chosen to test for linear and non-linear cases. In addition the far field velocity is give by $u_\infty = M_\infty a_\infty$, where the Mach number $M_\infty = 2$ and a_∞ is the speed of sound in far field conditions given by $a_\infty = \sqrt{\gamma p_\infty/\rho_\infty}$. The exact solution to this problem is simply the same set of equations, Eq.(48), with x, replaced by $\hat{x} = x - u_\infty t$. One-sided differences are used in x- and y-directions and on both sides of the domain. Due to the supersonic nature of the flow no boundary conditions are implemented in x-direction, however non-reflective [24] boundary conditions are implemented on the both bottom and top boundaries in y-direction. In addition the grid is generated using the following equations

$$\left.\begin{aligned} x_{i,j} &= -\frac{L}{2} + \frac{1.5L}{N}[i + \epsilon\sin(4\pi j/N)] \\ y_{i,j} &= -\frac{3L}{4} + \frac{1.5L}{N}[j + \epsilon\sin(4\pi i/N)] \end{aligned}\right\}, \text{ for } \quad 0 \le i \le N, \quad 0 \le j \le N \tag{50}$$

where ϵ controls the grid deformation such that $\epsilon = 0$ generates a uniform Cartesian grid. Note that ϵ is multiplied by a factor of $1/N$ and hence it should be adjusted when the grid is refined to generate a similar deformation amplitude on different grid levels.

To analyse the performance of the current FD scheme the z-vorticity is calculated in general coordinates using the conservative form

$$\omega_z = J^{-1}\left(\frac{\partial}{\partial\xi}[J(\xi_x v - \xi_y u)] - \frac{\partial}{\partial\eta}[J(\eta_y u - \eta_x v)]\right), \tag{51}$$

and is compared to the analytical value which is calculated by direct differentiation of velocity components $u(\hat{x}, y)$ and $v(\hat{x}, y)$ given in Eq.(48). Consequently an ℓ_2-norm and an ℓ_∞-norm can be defined by

$$\ell_2(t_n) = \frac{1}{N+1}\left(\sum_{i=0}^{N}\sum_{j=0}^{N}(\omega_z(x_i, y_j, t_n) - \omega_{z,\text{exact}}(x_i, y_j, t_n))\right)^{1/2}, \tag{52}$$

$$\ell_\infty(t_n) = \max\left(|\omega_z(x_i, y_j, t_n) - \omega_{z,\text{exact}}(x_i, y_j, t_n)|, \forall i, j \in \{0, \cdots, N\}\right), \tag{53}$$

where t_n is the current time step.

It is well established that the stability of this type of calculations, in addition to the generic stability of the scheme as was discussed in Section 3.4, depends on effective removal of unresolved wavenumbers. Among other method such as up-winding [25] and damping models [26], high-order filters [1, 27, 28] are usually used to stabilize the numerical solution. Recently Kim [19] proposed a set of filters with variable cut-off frequency which are used in conjunction with the set of the currently optimised boundary coefficients in this paper. These filter are given by [19]

$$\tilde{\Delta}f_i + \gamma_{01}^F\tilde{\Delta}f_{i\pm1} + \gamma_{02}^F\tilde{\Delta}f_{i\pm2} = 0, \text{ for } i = \begin{cases}0,\\ N,\end{cases} \tag{54}$$

$$\gamma_{10}^F\tilde{\Delta}f_{i\mp1} + \tilde{\Delta}f_i + \gamma_{12}^F\tilde{\Delta}f_{i\pm1} + \gamma_{13}^F\tilde{\Delta}f_{i\pm2} = 0, \text{ for } i = \begin{cases}1,\\ N-1,\end{cases} \tag{55}$$

$$\gamma_{20}^F\tilde{\Delta}f_{i\mp2} + \gamma_{21}^F\tilde{\Delta}f_{i\mp1} + \tilde{\Delta}f_i + \gamma_{23}^F\tilde{\Delta}f_{i\pm1} + \gamma_{24}^F\tilde{\Delta}f_{i\pm2} = \sum_{\substack{m=0\\m\neq2}}^{5} b_{2m}^F\left(f_{i\pm m\mp2} - f_i\right) \text{ for } i = \begin{cases}2,\\ N-2.\end{cases} \tag{56}$$

$$\beta^F \tilde{\Delta} f_{i-2} + \alpha^F \tilde{\Delta} f_{i-1} + \tilde{\Delta} f_i + \alpha^F \tilde{\Delta} f'_{i+1} + \beta^F \tilde{\Delta} f'_{i+2} = \sum_{m=1}^{3} a_m^F \left(f_{i+m} - 2 f_i + f_{i-m} \right) \quad \text{for} \quad 3 \leq i \leq N-3, \tag{57}$$

where $\tilde{f}_i = f_i + \tilde{\Delta} f_i$ is the filtered value calculated at the end of each time step. Choosing a single cut-off wave number, Ω^c and a boundary weighting factor w, such that

$$\Omega_i^c = \begin{cases} \Omega^c & \text{for} \quad 3 \leq i \leq N-3, \\ (1-w_2)\Omega^c & \text{for} \quad i = 2, N-2, \\ (1-w_1)\Omega^c & \text{for} \quad i = 1, N-1, \\ (1-w_0)\Omega^c & \text{for} \quad i = 0, N, \end{cases} \tag{58}$$

the values of the parameters γ_{im}^F, b_{2m}^F, a_m^F, α^F and β^F, in Eq.(54)–(57), can be calculated directly using explicit equations provided in [19]. Kim [19] performed a parametric study to choose the best values for w_0, w_1, w_2 and Ω^c by requiring a stable solution in time using the current test problem and minimum error. After the parametric study, he suggested $w_0/3 = w_1/2 = w_2 = 0.085$ and $\Omega^c = 0.88\pi$ for his set of coefficients. In this paper, instead of a parametric study, we allow for the boundary weighting factors to assume arbitrary values and perform an optimization by defining the independent variable vector $x = [\Omega^c, w_0, w_1, w_2]$ and the following objective function

$$\mathscr{E}^F = \begin{cases} C & \text{if} \quad \mathit{Unstable} \\ \ell_2(t_{n_f}) \max\left(\ell_2(t_n), n \in \{0, 1, \cdots, n_f\}\right) & \text{if} \quad \mathit{Stable}, \end{cases} \tag{59}$$

where C is a large arbitrary constants set to 10^3 in this study to provide a numeric value for degenerate x values where the calculation does not remain stable. The values of the Ω^c in addition to the boundary weighting factors are provided in Appendix A alongside the values of the FD boundary coefficients.

Figure 11 shows the contours of the normalized vorticity $\omega_z^* = \omega_z L/U_\infty$, convected through the domain for the Cartesian grid and the test case with maximum deformation, $\epsilon = 7$. Figure 12 compares the normalized $\ell_2(t)$ and $\ell_\infty(t)$-norms for $t \in [0, 2]$ calculated using the current coefficients and filters and those coefficient and filters suggested in [2] and [19]. Both values of the ℓ_2 and ℓ_∞ remain about one third the errors produced by the previously suggested coefficients and filters which translates into more than 150% increase in the efficiency of the current optimized FD and filters.

Values of the both ℓ_2 and ℓ_∞-norms are plotted in Figure 13 for different values of $K_2 \in \{0.1, 0.3, 1, 3, 5\}$ corresponding to both linear (for $K_2 < 1$) and non-linear ($K_2 > 1$) convection problems. The maximum value of the errors is calculated for an integration time up to $tU_\infty/L = 40$, long after the wave leaves the domain to ensure the stability of the suggested coefficient and filters. Error values are also compared to those calculated using the previous study [2, 19] in Figure 13 which shows more improvement for larger values of K_2. For different values of K_2 improvements in the range of 53%–177% (for ℓ_2) and in the range 35%–178% (for ℓ_∞) are observed where percentages are calculated by $100(\ell_{\text{old}} - \ell_{\text{current}})/\ell_{\text{current}}$.

Figure 14 shows the grid convergence study of the current method for the vortex convection problem, using four different grid levels of 150–600 and $K_2 = 5$. The 4^{th}-order accuracy is preserved both in ℓ_∞ and ℓ_2-norms when the suggested boundary coefficients are used in conjunction with the high order filters. Figure 15 shows the effects of the grid deformation on the quality of the results. To save the computational time the integration is performed in this case up to $tU_\infty/L = 3$. Both norms increase by increasing the deformation parameter ϵ however the results always remain less than those calculated using the coefficients and filters suggested in [2, 19]. The largest increased is observed in the maximum value of the ℓ_∞ norm for the previously suggested coefficients, where 71% increase is observed by changin from $\epsilon = 0$ to $\epsilon = 7$.

4.5. Benchmark problem: Co-rotating vortex convection

For the final test case the problem of two co-rotating vortices is considered. The initial conditions is set by superimposing two vortices given by Eq.(48), with y replaced by $y_1 = y + \delta$ and $y_2 = y - \delta$ respectively for each vortex, where δ controls the distance between the two vortex cores. This case is substantially different from the previous case in that the motion of the vortices is unsteady and is coupled to the pressure fluctuations. There is no analytical solution for this test case and hence all the errors are calculated with respect to a corresponding reference simulation. The reference problem for each test case consists of an identical test case performed on a domain two times larger in ξ direction. The errors are then calculated by comparing the results of a simulation with half the reference size in ξ-direction with boundary schemes implemented on the cut boundary, to the full size simulation at specific times. For the full simulation 400×200 grid points are used

in ξ and η directions respectively whereas for the half plane simulations 200 grid points are used in ξ direction.

For this case we set the distance between the two vortex cores to $\delta = 0.15L$ and two different grids are generated using Eq.(50), with $\epsilon = 0$ and $\epsilon = 5$. It should also be noted that for the full grid reference solution, $\sin(8\pi i/N)$ is used for $y_{i,j}$ to generate identical grid to that of the half plane case. The vortex parameters are set to $K_2 = 5$ and $K_1 = 8.33$ for both the top and bottom vortices. Figure 16 shows the vorticity contours at three different times during the course of motion at $tU_\infty/L = 0$, 0.48 and 1.0. The contours are compared to the reference solution and no differences can be observed in this figure, however the the unsteady motion of the vortices is clear. Figure 17 shows the non-dimensional pressure contours P/P_∞, at $tU_\infty/L = 1$ and the pressure propagation toward the boundary which interacts with the vortices. However in Figure 17 no artificial reflection or wiggles is observed near the boundary compared to the reference numerical solution. Figure 18 shows the ℓ_2-norms calculated using the double size domain as the reference solution for two grids with $\epsilon = 0$ and $\epsilon = 5$. In this test case the pressure field characterizes the solution and is used to calculate the errors. Significant improvement is observed up to $tU_\infty/L \approx 1$ when the top vortex leaves the domain. After that both set of coefficients provide similar errors up to $tU_\infty/L = 1.6$, i.e while the second core leaves the domain and after that the current scheme again produces less error. Also note that more improvement is achieved for the deformed grid with $\epsilon = 5$.

5. Conclusion

The Genetic Algorithm is successfully used to construct a comprehensive optimization platform for pentadiagonal compact finite difference schemes. It is demonstrated that the new optimization strategy efficiently deals with both linear and non-linear constraints at the same time leading to highly accurate stable boundary schemes without the need of additional ad-hoc trial-and-error routines that were commonly required in the old optimization approaches. The enhanced accuracy and guaranteed stability of the schemes obtained in this paper are confirmed through benchmark test cases where the formerly optimized schemes failed to maintain numerical stability or to perform better than the present ones. In addition, a new set of optimized filter coefficients are obtained also by using the Genetic Algorithm in order to enhance the numerical stability of the proposed schemes for generic non-

linear problems on multi-dimensional curvilinear grids. It is envisaged that the proposed optimization strategy will contribute later on to improving further other types of compact schemes as well. The authors also considers that an extension of the work to accommodate the composite templates will have a significant potential.

Acknowledgement

The authors gratefully acknowledge that the present work has been supported by EPSRC (Engineering and Physical Sciences Research Council) under EP/J007633/1.

Appendix A. Parameters of the new compact FD scheme and filters

Table A.1 summarizes the new 4^{th}-order boundary coefficients for the pentadiagonal systems to be used in conjunction with the interior coefficients provided in [2]. In Table A.2 we have summarized the filter parameters as discussed in Section 4.4.

Table A.1: Summary of the boundary coefficients appearing in Eq. (2)–(4).

Coefficient	$i = 0$	$i = 1$	$i = 2$
γ_{i0}	-	9.486703622867607e-2	4.127253978047144e-2
γ_{i1}	5.590226531590711	-	4.708395755079016e-1
γ_{i2}	3.911115464821060	1.852980118858077	-
γ_{i3}	-	7.841681122699989e-1	5.713690208719099e-1
γ_{i4}	-	-	6.287995158522702e-2
b_{i0}	-	-3.469447847494813e-1	-1.534532664885535e-1
b_{i1}	-3.320861355280472	-	-6.866311200147498e-1
b_{i2}	5.452259004221430	1.652135357932134e-1	-
b_{i3}	1.150275611660523	1.379421330446014	7.176431952789228e-1
b_{i4}	-1.839611359673221e-1	1.789691155384915e-1	2.186728528907302e-1
b_{i5}	5.771607628595115e-2	-2.142195128295235e-2	-1.419994100359792e-3
b_{i6}	-1.431288821544212e-2	1.958948887672967e-3	5.236289985873258e-4

Table A.2: Summary of the filter parameters.

Filter Parameter	Value
Ω^c	0.85009765625
w_0	0.05
w_1	0.05
w_2	0.07

References

[1] S. K. Lele, Compact finite difference schemes with spectral-like resolution, Journal of Computational Physics 103 (1992) 16–42.

[2] J. W. Kim, Optimised boundary compact finite difference schemes for computational aeroacoustics, Journal of Computational Physics 225 (2007) 995 – 1019.

[3] A. Sescu, R. Hixon, A. A. Afjeh, Multidimensional optimization of finite difference schemes for computational aeroacoustics, Journal of Computational Physics 227 (2008) 4563 –88.

[4] E. Johnsen, J. Larsson, A. V. Bhagatwala, W. H. Cabot, P. Moin, B. J. Olson, P. S. Rawat, S. K. Shankar, B. Sjgreen, H. Yee, X. Zhong, S. K. Lele, Assessment of high-resolution methods for numerical simulations of compressible turbulence with shock waves, Journal of Computational Physics 229 (2010) 1213 –37.

[5] S. Kawai, K. S. Shankar, S. K. Lele, Assessment of localized artificial diffusivity scheme for large-eddy simulation of compressible turbulent flows, Journal of Computational Physics 229 (2010) 1739–62.

[6] S. Nagarajan, S. K. Lele, J. H. Ferziger, A robust high-order compact method for large eddy simulation, Journal of Computational Physics 191 (2003) 392–419.

[7] E. Lamballais, V. Fortun, S. Laizet, Straightforward high-order numerical dissipation via the viscous term for direct and large eddy simulation, Journal of Computational Physics 230 (2011) 3270 –5.

[8] Z.-S. Sun, Y.-X. Ren, C. Larricq, S.-Y. Zhang, Y.-C. Yang, A class of finite difference schemes with low dispersion and controllable dissipation for dns of compressible turbulence, Journal of Computational Physics 230 (2011) 4616 –35.

[9] A. W. Cook, J. J. Riley, Direct numerical simulation of a turbulent reactive plume on a parallel computer, Journal of Computational Physics 129 (1996) 263 –83.

[10] J. W. Kim, D. J. Lee, Optimized compact finite difference schemes with maximum resolution, AIAA Journal 34 (1996) 887–93.

[11] M. H. Carpenter, D. Gottlieb, S. Abarbanel, Stable and accurate boundary treatments for compact, high-order finite-difference schemes, Applied Numerical Mathematics 12 (1993) 55 – 87.

[12] Z. Liu, Q. Huang, Z. Zhao, J. Yuan, Optimized compact finite difference schemes with high accuracy and maximum resolution, International journal of aeroacoustics 7 (2008) 123–46.

[13] S. A. Jordan, The spatial resolution properties of composite compact finite differencing, Journal of Computational Physics 221 (2007) 558 –76.

[14] S. A. Jordan, Optimization, resolution and application of composite compact finite difference templates, Applied Numerical Mathematics 61 (2011) 108 –30.

[15] R. Vichnevetsky, J. B. Bowles, Fourier Analysis of Numerical Approximations of Hyperbolic Equations, SIAM, 1982.

[16] J. Holland, Adaptation in Natural and Artificial Systems, University of Michigan Press, 1975.

[17] D. E. Goldberg, Genetic Algorithms in Search, Optimization, and Machine Learning, Addison-Wesley, 1989.

[18] R. L. Haupt, S. H. Haupt, Practical Genetic Algorithms, second edition ed., Wiley, 2004.

[19] J. W. Kim, High-order compact filters with variable cut-off wavenumber and stable boundary treatment, Computers & Fluids 39 (2010) 1168 –82.

[20] W. E. Boyce, R. C. DiPrima, Elementary Differential Equations and Boundary Value Problems, ninth edition ed., Wiley, 2009.

[21] M. Abramowitz, I. Stegun, Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, Dover Publications, 1972.

[22] C. K. W. Tam, Aliasing problem: category 1 problem 1 analytic solutions, in: Fourth computational aeroacoustics workshop on benchmark problems, NASA Glenn Research Centre, Cleveland. USA, 2003, pp. 31–2.

[23] H. C. Yee, N. D. Sandham, M. J. Djomehri, Low-dissipative high-order shock-capturing methods using characteristic-based filters, Journal of Computational Physics 150 (1999) 199 – 238.

[24] J. W. Kim, D. J. Lee, Generalized characteristic boundary conditions for computational aeroacoustics, AIAA Journal 38 (2000) 2040–9.

[25] S. Yamamoto, H. Daiguji, Higher-order-accurate upwind schemes for solving the compressible euler and navier-stokes equations, Computers & Fluids 22 (1993) 259 –70.

[26] C. K. W. Tam, J. C. Webb, Z. Dong, A study of the short wave components in computational acoustics, in: J. Hardin, M. Hussaini (Eds.), Computational Aeroacoustics, ICASE/NASA LaRC Series, Springer New York, 1993, pp. 116–30.

[27] D. V. Gaitonde, M. R. Visbal, Pade-type higher-order boundary filters for the navier-stokes equations, Aiaa Journal 38 (2000) 2103–12.

[28] L. Zhanxin, H. Qibai, H. Li, Y. Jixuan, Optimized compact filtering schemes for computational aeroacoustics, International journal for numerical methods in fluids 60 (2009) 827–45.

Table 1: Truncation error constants for the current scheme compared to the previous studies. The absolute value of the constants remain small and in the same order of magnitude as the previous studies.

Study	C_5^0	C_5^1	C_5^2
Current	9.482453030544406	0.1439116261302793	0.0397946275581891
Kim [2]	6.572807481949297	-0.1776706065593920	0.107241450832102
Jordan [14]	30.5649277	-14.7385031	2.9068336

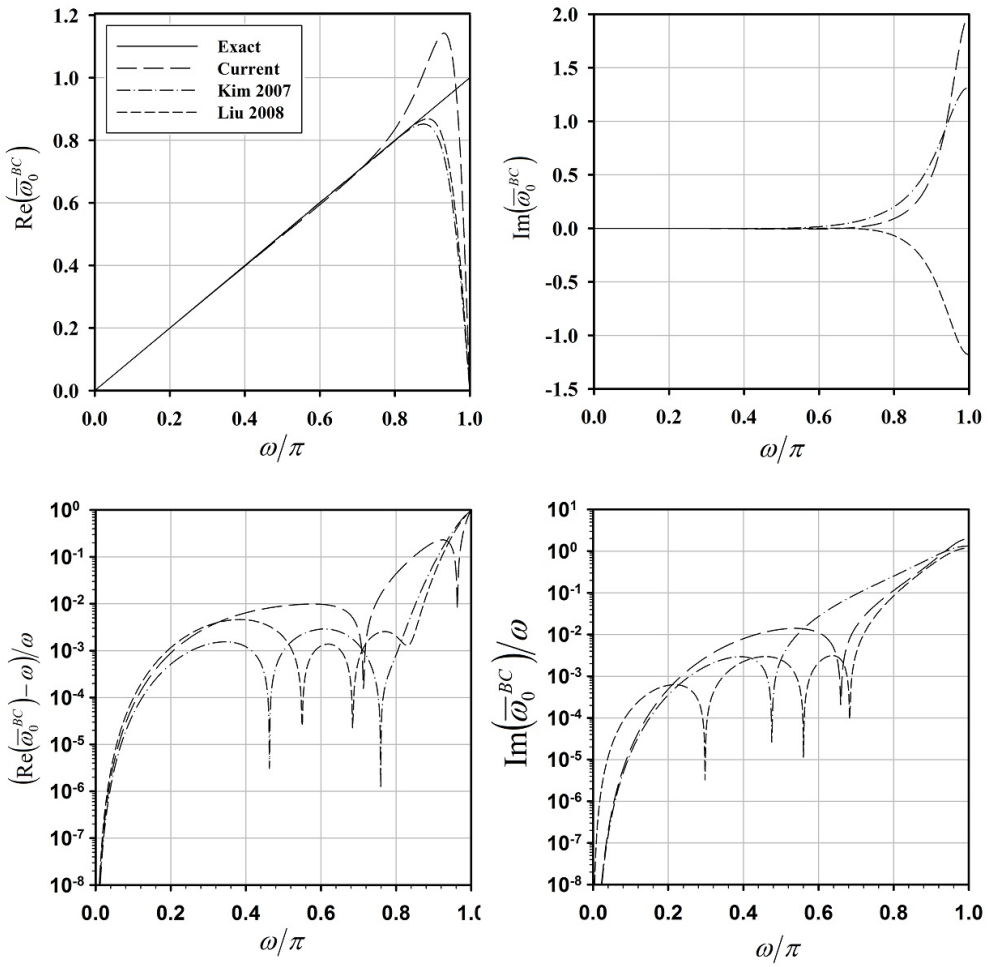

Fig. 1: Modified wavenumber plots for the first boundary point $i = 0$, dissipative and dispersive errors are compared to previous studies Kim [2] and Liu et al. [12].

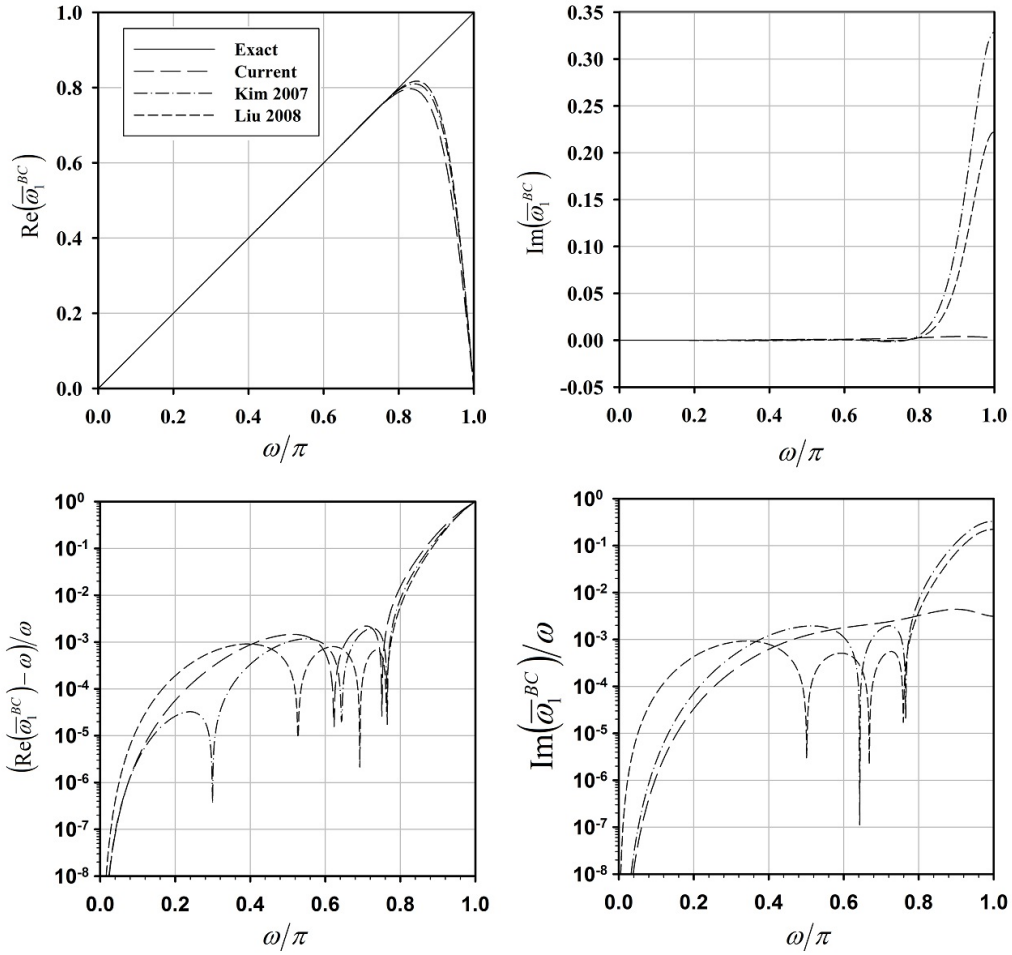

Fig. 2: Modified wavenumber plots for the second boundary point $i = 1$, See the legends of Figure 1.

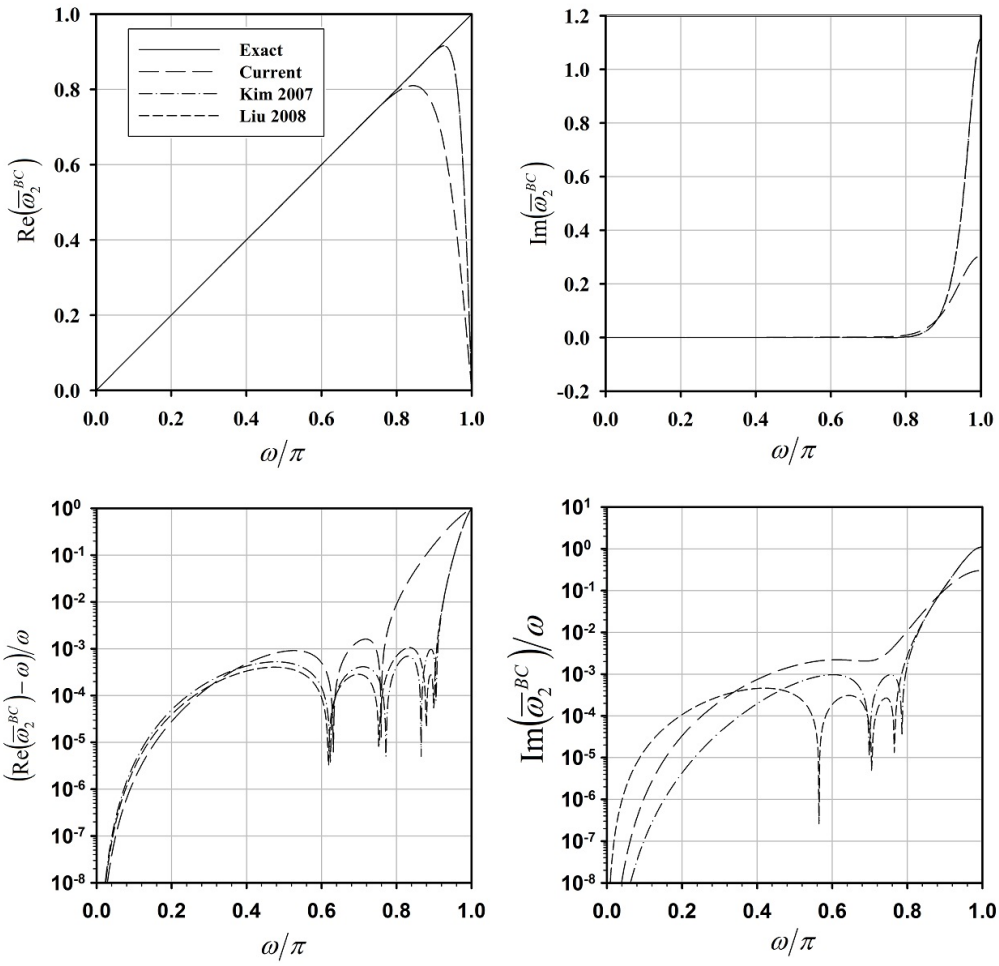

Fig. 3: Modified wavenumber plots for the third boundary point $i = 2$, See the legends of Figure 1.

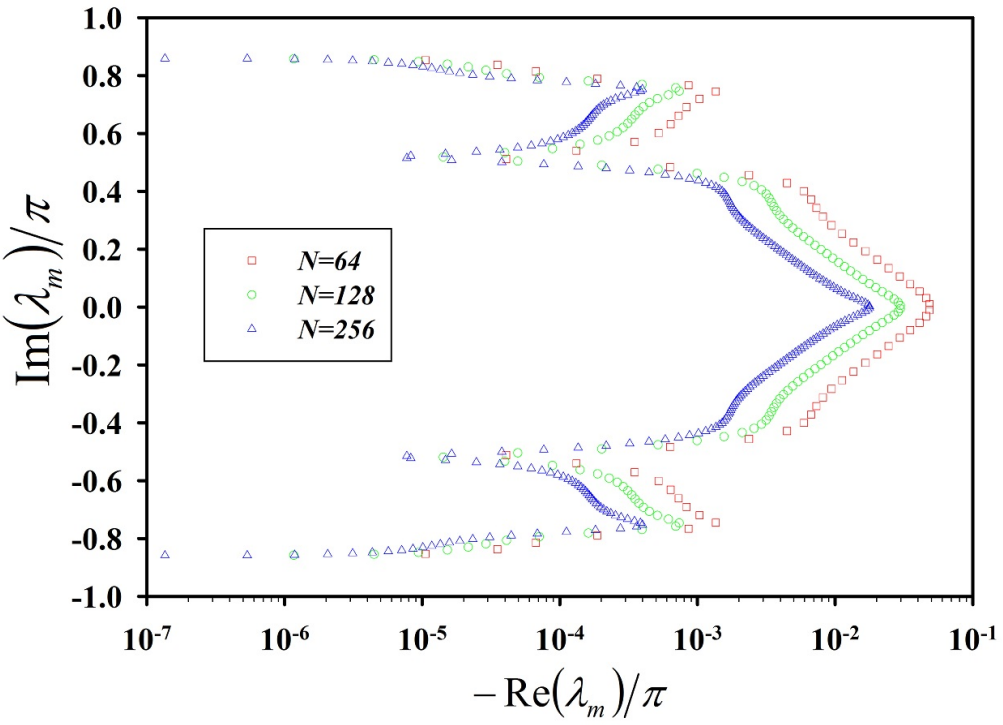

Fig. 4: Eigenvalues of the final optimized scheme presented for three grid levels $N = 64, 128, 256$. Enforcing the stability criteria on four smaller grid sizes also ensures unconditional stability on larger grids.

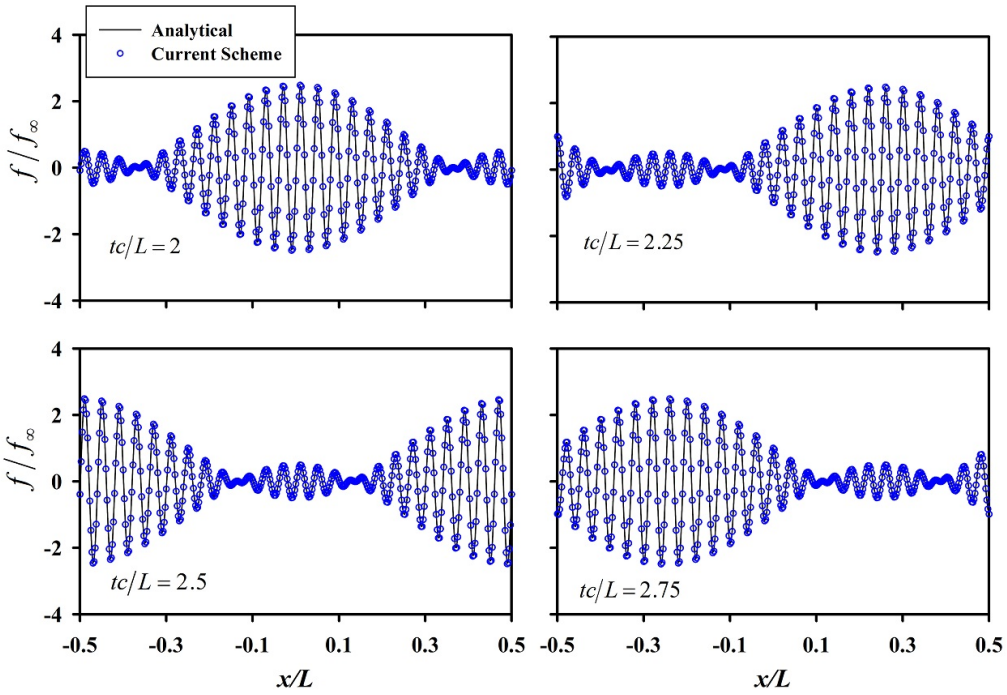

Fig. 5: One period of motion of the modulated wave convection problem, presented at four different stages. The solution is virtually error free even near the boundaries.

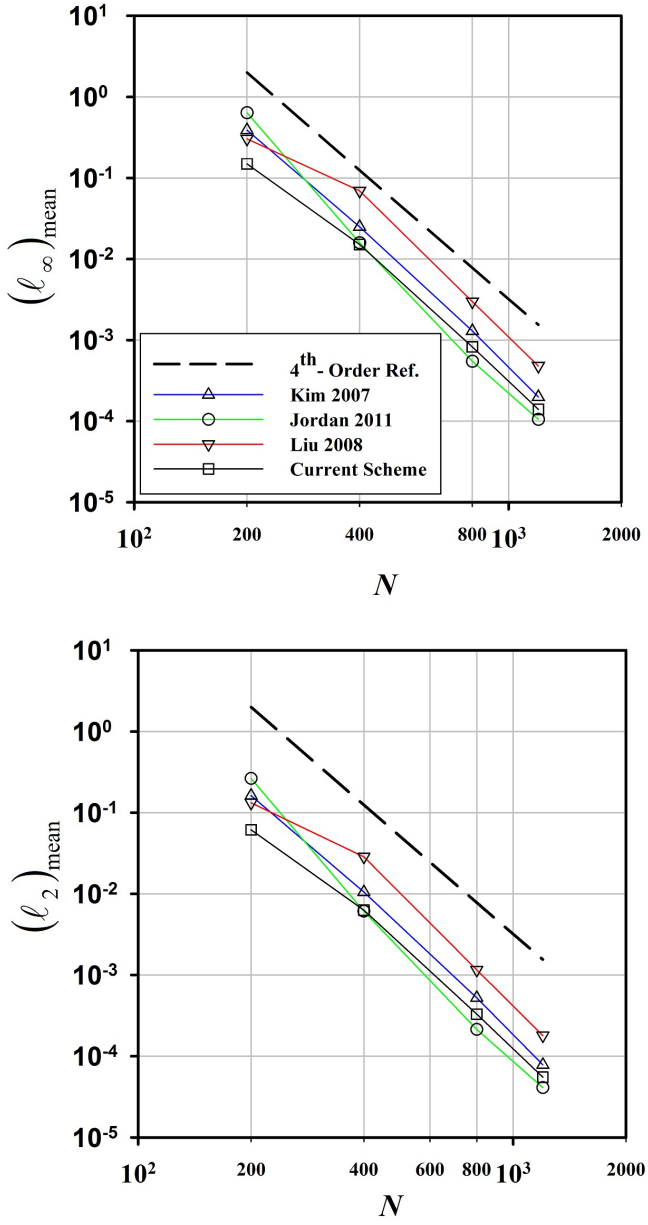

Fig. 6: Comparison of the convergence rates for the modulated wave convection. Average values for both ℓ_∞ and ℓ_2 are presented on different grid levels. Note that the coefficients of Jordan [14] provide better results on fine grids ($N = 800, 1200$) with small time-step sizes (CFL $= 0.2$), otherwise they fail to provide stable solutions, see Figure 7.

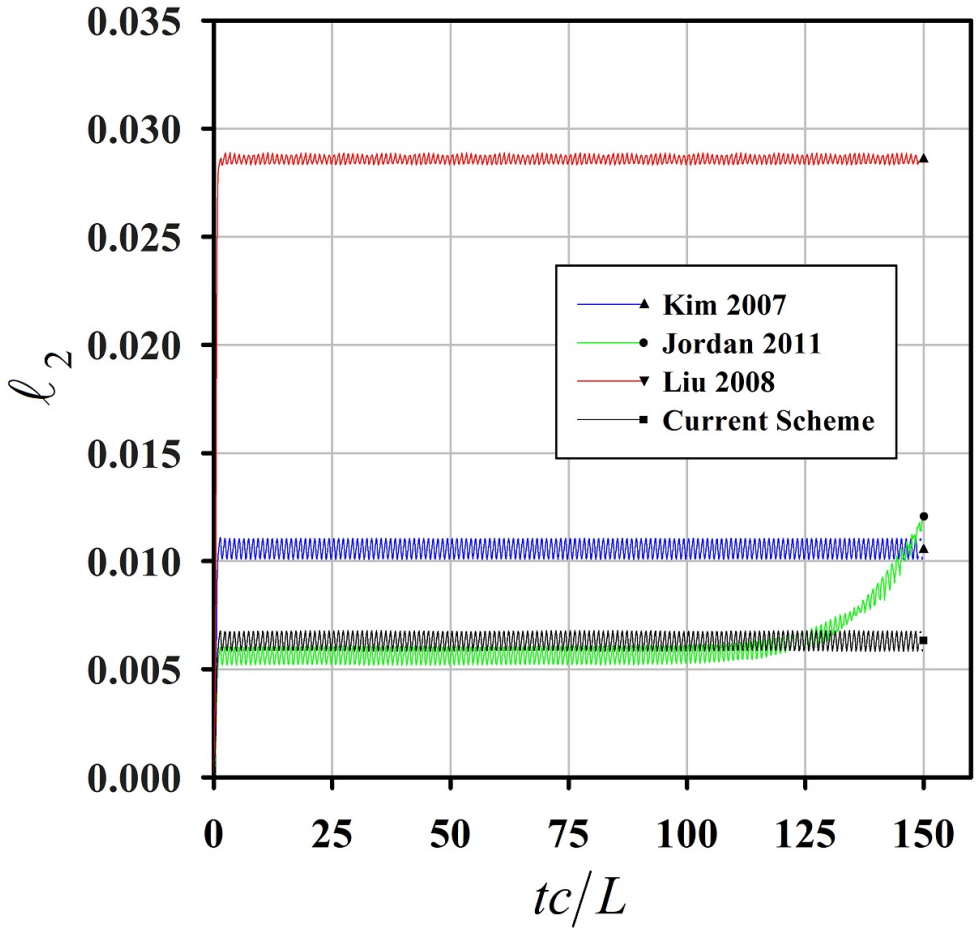

Fig. 7: Time history of the ℓ_2-norm on the grid with $N = 400$ points.

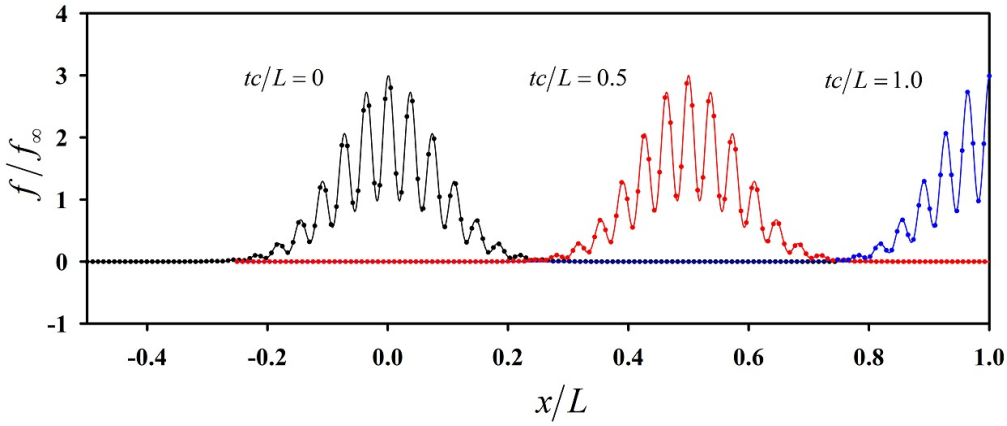

Fig. 8: Motion of the scalar wave convection problem presented at three different stages for $N = 1000$. Errors can not be detected even when the wave is leaving the domain. •, FD approximation. —, Analytical solution.

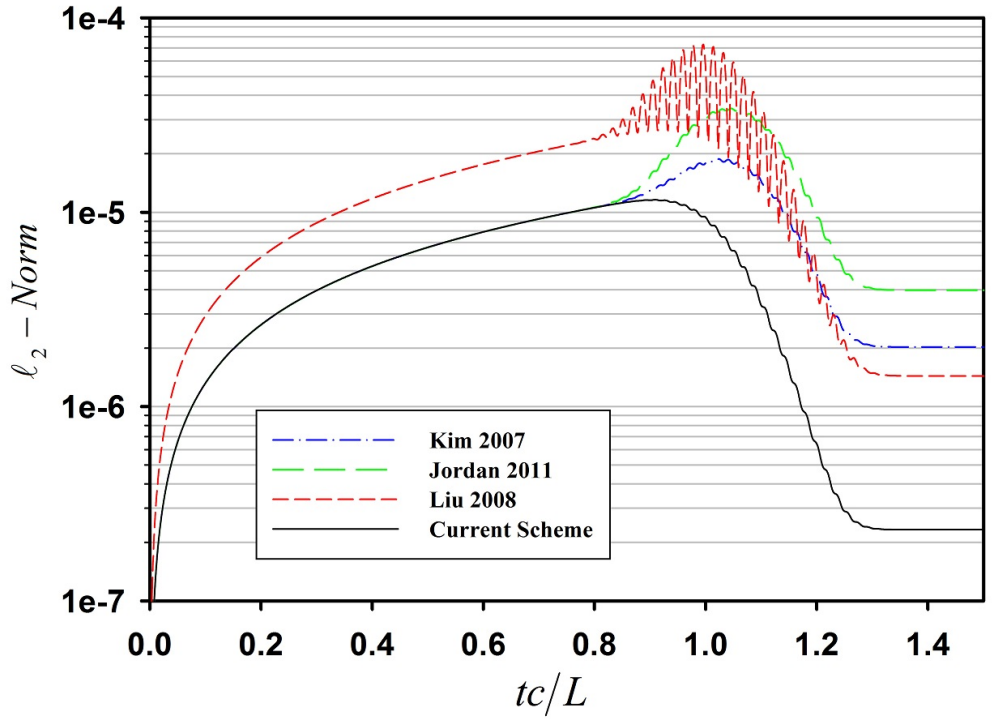

Fig. 9: Time history of the ℓ_2-norms of the errors for $N = 1000$. No overshoot is observed in the error history of the current scheme and the remaining error is the domain after the wave leaves the domain shows significant improvement.

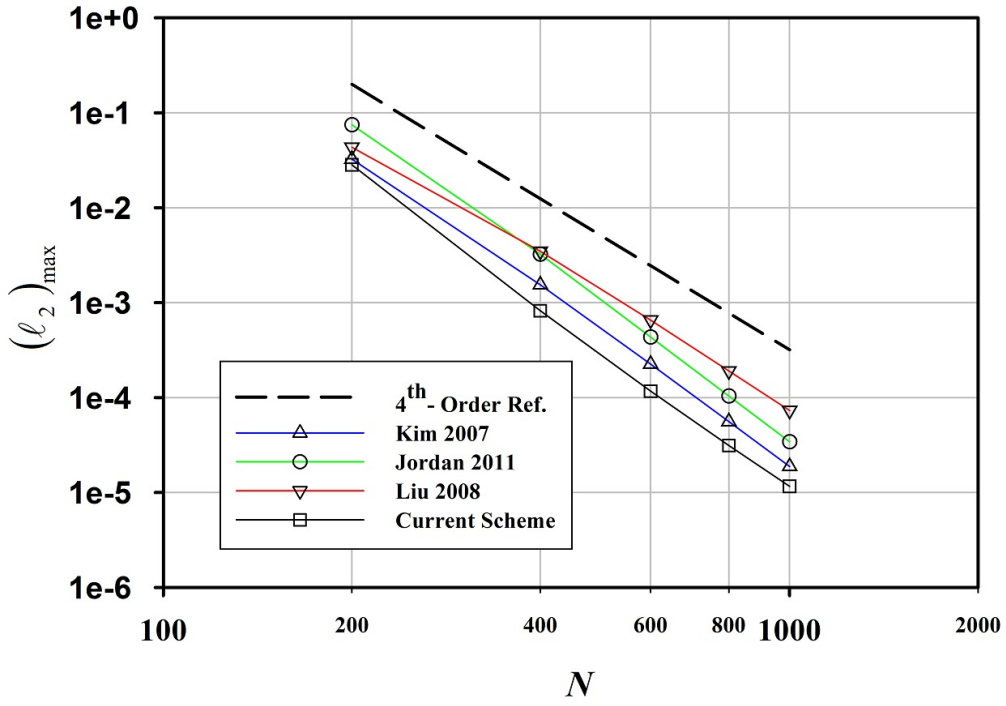

Fig. 10: Convergence results of the maximum ℓ_2-norm during the course of convection of the scalar wave.

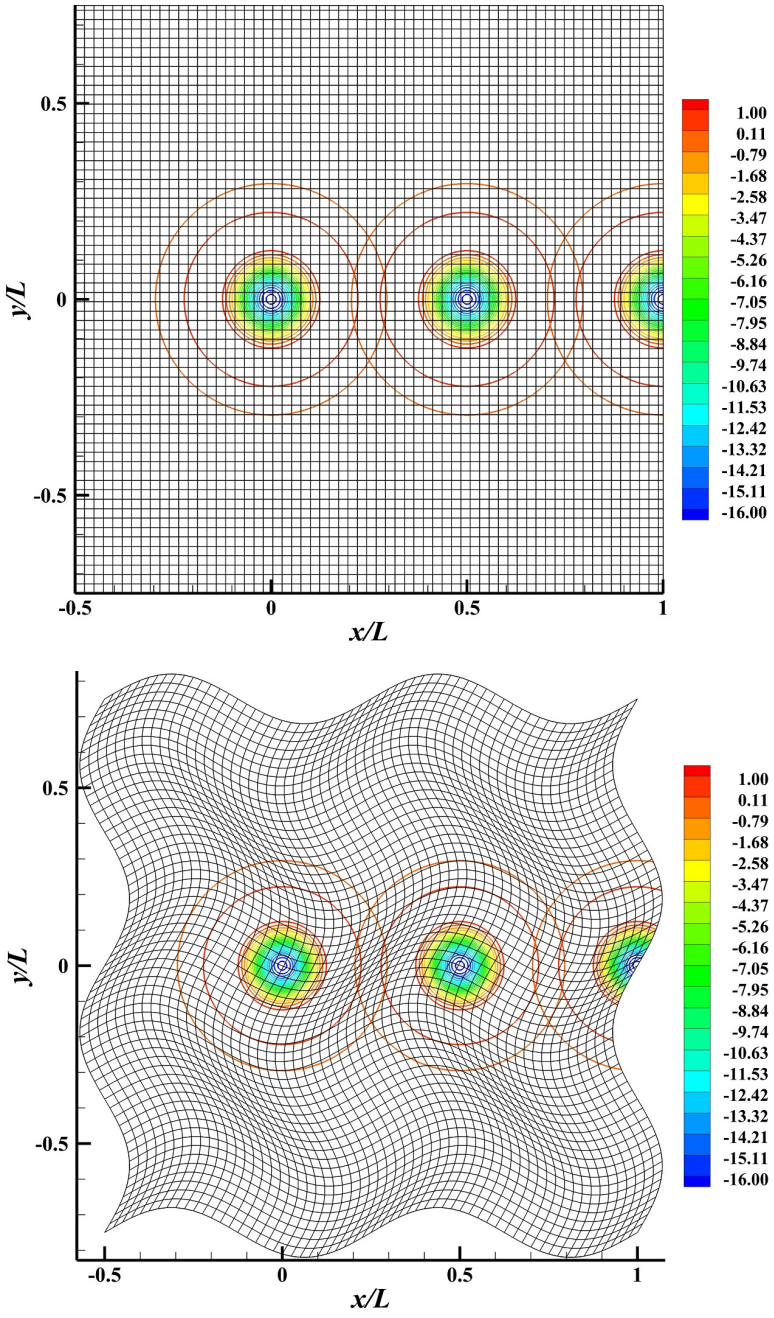

Fig. 11: Contours of the non-dimensional vortex wave, $\omega_z^* = \frac{\omega_z L}{U_\infty}$, plotted for $N = 600$ and $K_2 = 5$. Twenty levels of contours are presented between $\omega_z^* = 1$ and $\omega_z^* = -16$. Only 1/10 of the grid lines are presented for two cases with $\epsilon = 0$ and $\epsilon = 7$.

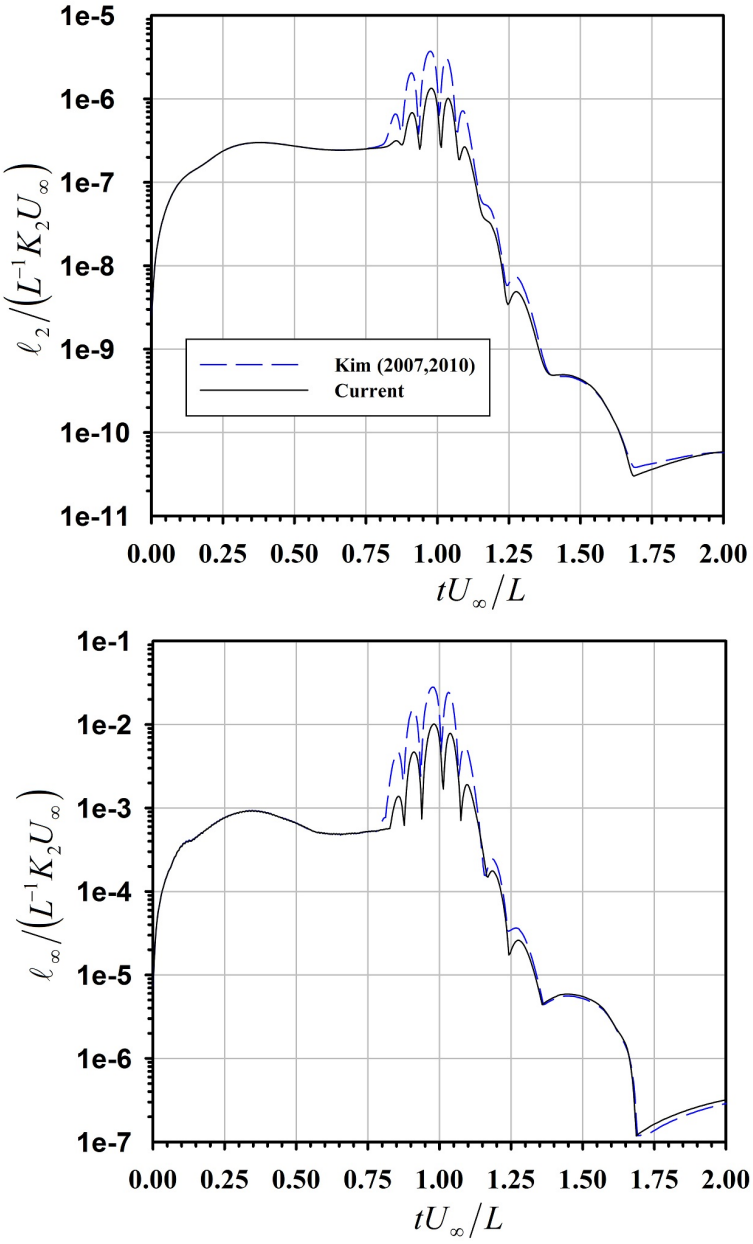

Fig. 12: Comparison of the ℓ_2 and ℓ_∞-norms of the 2D vortex wave for $N = 150$, $k_2 = 5$ and $\epsilon = 0$. Current error levels are compared with the errors produced by using the values of Ω^c and w in [19] suggested to be used for the boundary coefficients in [2].

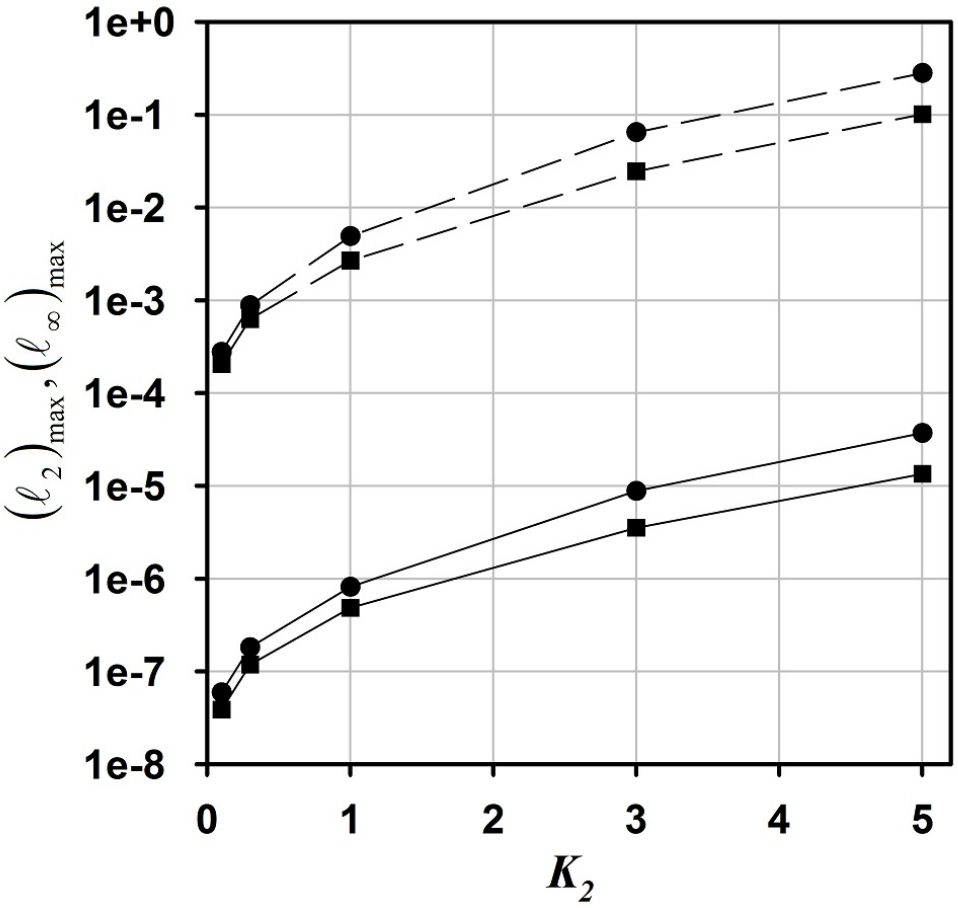

Fig. 13: Maximum value of the ℓ_2 (solid lines) and ℓ_∞-norms (dashed lines) of the 2D vortex wave for $N = 150$, $\epsilon = 0$ and different K_2 values calculated for a total integration time up to $t = 20$. ■, Current study; •, Previous study [2, 19].

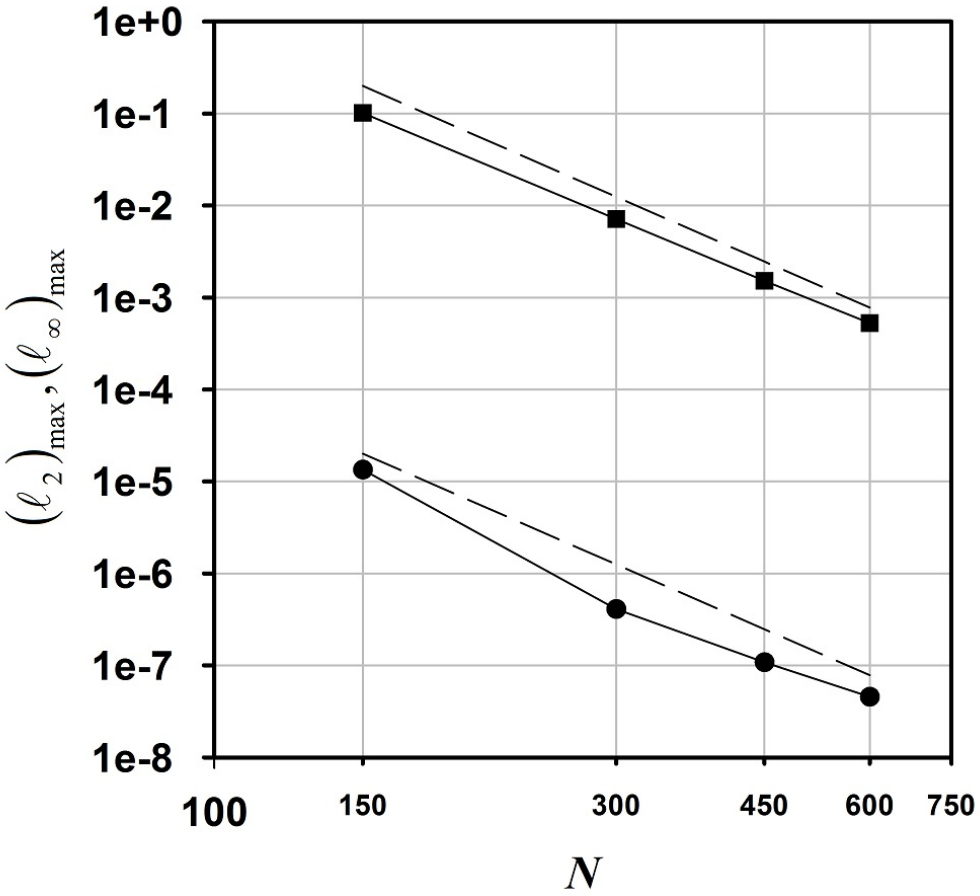

Fig. 14: Grid convergence study for the 2D vortex convection problem. Maximum value of the ℓ_2 (•) and ℓ_∞-norms (■) of the 2D vortex wave for $K_2 = 5$ and different grid levels calculated during the course of integration $t \in [0, 20]$.

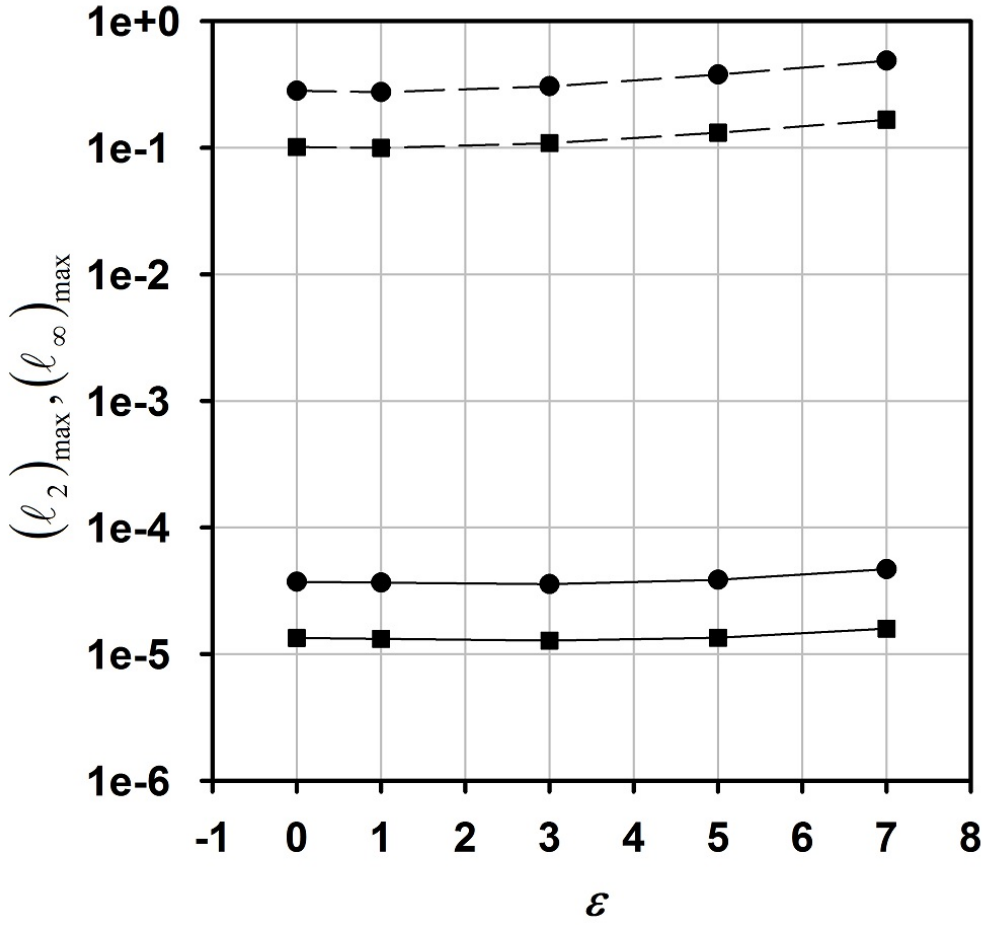

Fig. 15: Maximum value of the ℓ_2 (solid lines) and ℓ_∞-norms (dashed lines) of the 2D vortex wave for $N = 150$, $K_2 = 5$ and different ϵ values calculated for a total integration time up to $t = 1.5$. ■, Current study; •, Previous study [2, 19].

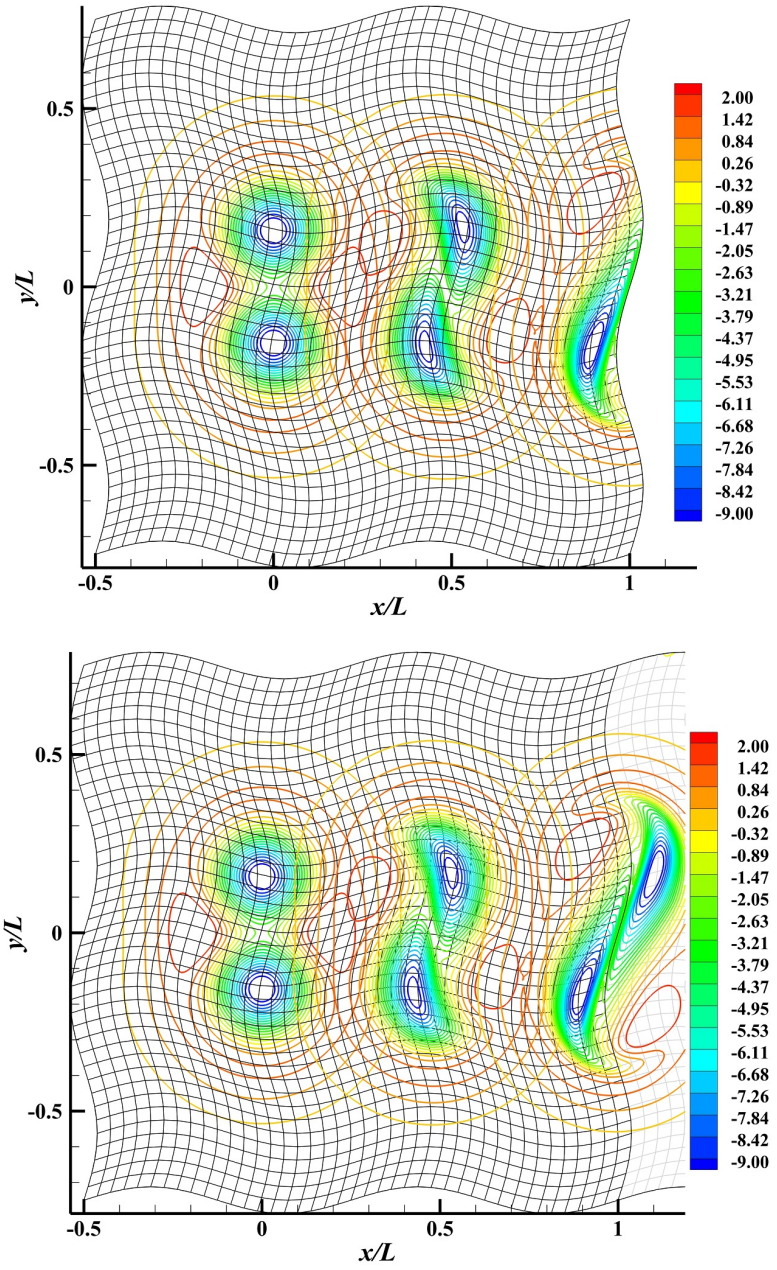

Fig. 16: Comparison of the normalized vorticity contours w_z^*, for the half plane simulation with the full plane simulation. 20 levels of contours and 1/5 of grid lines are presented at $tU_\infty/L = 0$, 0.48 and 1.0 for both cases.

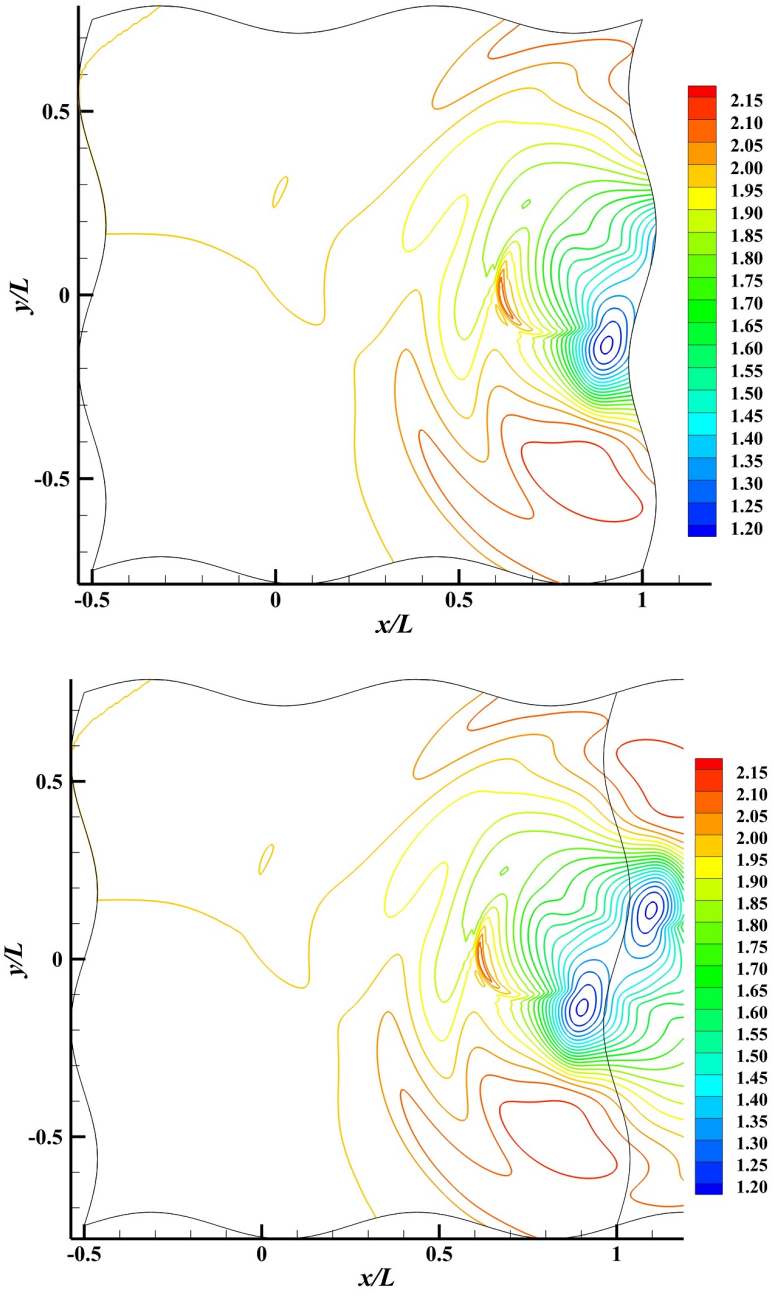

Fig. 17: Propagation of a pressure wave toward the boundaries resulting in unsteady behaviour of the test case. Non-dimensional pressure contours (P/P_∞) are compared at $tU_\infty/L = 1$.

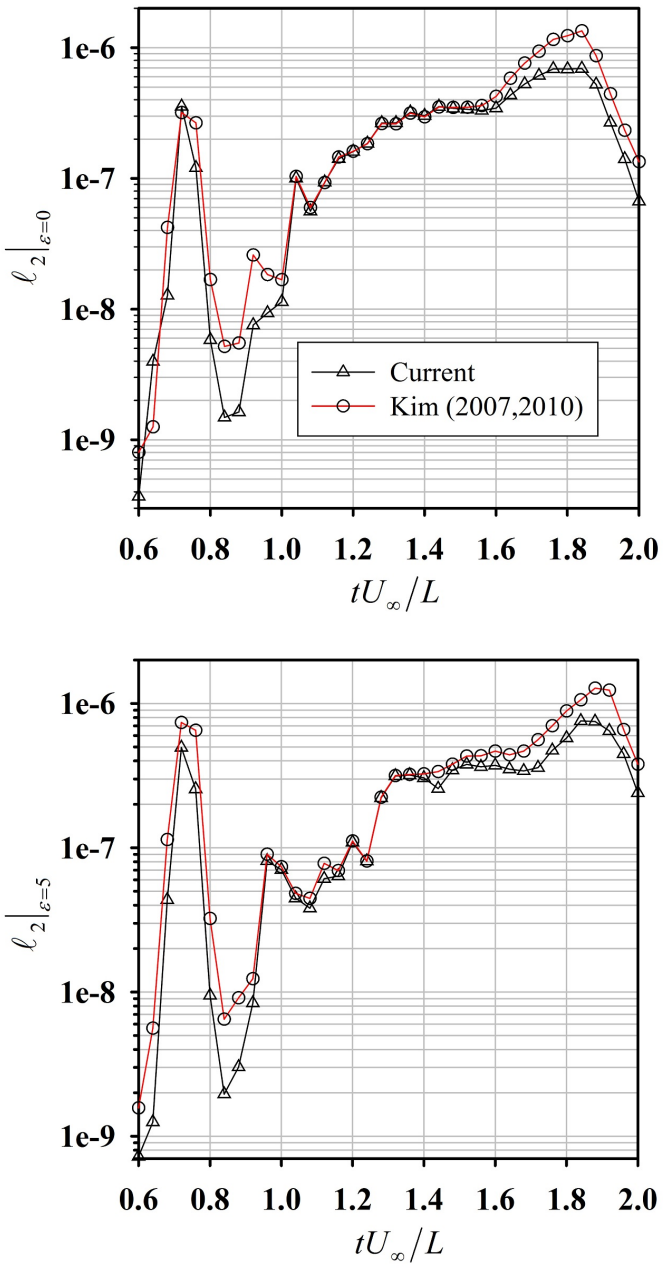

Fig. 18: Comparison of the ℓ_2-norms between the previous study [2, 19] and the current scheme at 41 points, using the double size domain as the reference solution.