INTERNATIONAL JOURNAL OF
NUMERICAL ANALYSIS AND MODELING
Volume 8, Number 3, Pages 518–541

ON THE INACCURACIES OF SOME FINITE VOLUME DISCRETIZATIONS OF THE LINEARIZED SHALLOW WATER PROBLEM

S. FAURE [#], M. PETCU [♭◇], R. TEMAM [♯], AND J. TRIBBIA [†]

Abstract. In this article we are interested in the study of the errors introduced by different finite volume discretizations in the study of the wave frequencies. This study is made in the context of hyperbolic systems arising in meteorology and oceanography. We show the existence of significant errors in the dispersion relation, only the long waves being computed accurately; this conclusion is similar to the finite differences case described in the article of Grotjahn and O'Brien quoted below. For the case of inertia-gravity waves, we study three often used schemes which are based on the upwind, centered or Lax-Wendroff fluxes. Moreover, the Total Variation Diminishing method (TVD) made from these fluxes (which usually provides an efficient way to eliminate spurious numerical oscillations) will give the same errors in the dispersion relation.

Key Words. finite volume discretizations, spurious caustics, numerical error, dispersion relation, hyperbolic systems

1. Introduction

The present work examines the errors introduced in low order finite volume discretizations velocity in second order schemes using methods which are typical of finite volume methods, e.g. the Lax-Wendroff method and upwinding. We observe that in addition to the well-known problems in phase propagation and group velocity in waves near the grid scale, the differential errors in the group velocity as a function of wave vector gives rise to the spurious formation of caustics in the inertia gravity part of the spectrum. Serious physical ramifications of spurious caustics in the divergent modes could thus result through the interaction of such local focusing and moist processes in models used for weather and climate prediction.

These results were to some extent foreshadowed in the in the context of finite difference discretizations: in Grotjahn and O'Brien (1976) the authors showed that the finite difference methods are causing significant angular and magnitude errors for the group velocity for different ocean models and only the long waves are represented with reasonable accuracy; Vichnevetsky (1987) studied the error introduced by the time and space discretizations, as well as by the boundaries; Blayo (2000) studied the error introduced by finite-difference schemes of higher orders, in the case of inertia-gravity waves; David, et. al. (2006) examined the linear dispersive mechanism for error focusing and the existence of spurious caustics for the case of discretized propagation-type equations. The motivation for considering the study of finite volume methods is that finite volume methods differ conceptually from finite

Received by the editors February 10, 2011.
2000 *Mathematics Subject Classification.* 35L60, 35Q35, 76B15, 76B65.

differences in their emphasis on estimating grid cell boundary fluxes as opposed to estimating derivatives for the finite difference techniques.

This study has been limited to low order spatial and temporal discretizations for finite volume methods. A more detailed analysis of finite volume discretizations using higher order numerics is the topic of ongoing research. The study of finite volume methods for hyperbolic problems is a subject of great interest and we mention in particular the books and articles by Levesque (2002), Toro (2006) and Toro (2001) (see also Toro and Clarke (1998)). We also cite here the work of Morton (2001/02) where first, second and third order accurate algorithms are developed for non-uniform one-dimensional grids, as well as for unstructured triangular meshes.

Of high interest in the recent years was the use of high-order finite volume methods for hyperbolic problems and we mention some of the recent works: Dumbser et.al. (2008), Martí-Müller (1996), Noelle et. al. (2007), Popov and Ustyugov (2007). It is already known that important group velocity errors exist in higher order schemes (see David and Sagaut (2008)) which have consequences for the prediction and simulation of coherent structures in the atmosphere related to solitons. The existence of spurious caustics in the shallow water equations for higher order FV discretizations, like those discussed here for second order schemes, has not yet been shown and is the subject of current research.

Of primary interest in this study will be the linearized Shallow Water problem with space periodical boundary conditions:

$$\frac{\partial u}{\partial t} + \bar{u}_0 \frac{\partial u}{\partial x} + \bar{v}_0 \frac{\partial u}{\partial y} + g \frac{\partial \Phi}{\partial x} - fv = 0, \tag{1.1}$$

$$\frac{\partial v}{\partial t} + \bar{u}_0 \frac{\partial v}{\partial x} + \bar{v}_0 \frac{\partial v}{\partial y} + g \frac{\partial \Phi}{\partial y} + fu = 0, \tag{1.2}$$

$$\frac{\partial \Phi}{\partial t} + \bar{u}_0 \frac{\partial \Phi}{\partial x} + \bar{v}_0 \frac{\partial \Phi}{\partial y} + \bar{\Phi}_0 \left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \right) = 0, \tag{1.3}$$

which can be rewritten as a system:

$$\frac{\partial q}{\partial t} + \mathbf{A} \frac{\partial q}{\partial x} + \mathbf{B} \frac{\partial q}{\partial y} + \mathbf{C} q = 0, \tag{1.4}$$

with $q(x, y, t) = (u(x, y, t), v(x, y, t), \Phi(x, y, t))$, g the gravity and the matrices $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ defined by:

$$\mathbf{A} = \begin{pmatrix} \bar{u}_0 & 0 & g \\ 0 & \bar{u}_0 & 0 \\ \bar{\Phi}_0 & 0 & \bar{u}_0 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} \bar{v}_0 & 0 & 0 \\ 0 & \bar{v}_0 & g \\ 0 & \bar{\Phi}_0 & \bar{v}_0 \end{pmatrix}, \quad \mathbf{C} = \begin{pmatrix} 0 & -f & 0 \\ f & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Of special interest below will be the case where $\bar{u}_0 = \bar{v}_0 = 0$ and f constant describing the inertia gravity waves. The properties of inertia-gravity waves are important, for example, in the process of geostrophic adjustment in which the atmosphere responds to changes in surface forcing with a continuous spectrum of high-frequency waves. In the numerical primitive equation models, these waves are generated during initialization and after convective events.

The continuous case. To find the dispersion relation in the continuous case, we replace q in equation (1.4) by the wave $q(x, y, t) = (\bar{u}, \bar{v}, \bar{\Phi}) e^{\mathbf{i}(kx + ly - \omega t)}$, and we

find:

$$(-\mathbf{i}\omega\mathbf{I} + \mathbf{i}k\mathbf{A} + \mathbf{i}\ell\mathbf{B} + \mathbf{C})\, q = 0, \tag{1.5}$$

with $\mathbf{i} = \sqrt{-1}$ and $\mathbf{I}$ the identity matrix. Solving the dispersion equation for (1.5) we obtain three wave frequencies:

$$\omega_0 = k\bar{u}_0 + l\ell\bar{v}_0, \quad \omega_\pm = \omega_0 \pm f\sqrt{R_D^2\kappa^2 + 1}, \tag{1.6a}$$

where $\kappa = \sqrt{k^2 + \ell^2}$ is the wave-number magnitude and R_D is the Rossby radius of deformation, $R_D = \sqrt{\bar{\Phi}_0 g}/f$.

The group velocities are $v_g = (\bar{u}_0, \bar{v}_0)$ for ω_0 and for $\omega_\pm$,

$$v_g = \begin{pmatrix} \dfrac{\partial\omega}{\partial k} \\ \dfrac{\partial\omega}{\partial l} \end{pmatrix} = \begin{pmatrix} \bar{u}_0 \pm \dfrac{R_D^2 fk}{\sqrt{R_D^2\kappa^2 + 1}} \\ \bar{v}_0 \pm \dfrac{R_D^2 fl}{\sqrt{R_D^2\kappa^2 + 1}} \end{pmatrix}, \tag{1.7}$$

Inertia-gravity waves

In the case of inertia gravity waves, the governing equations are (1.1)-(1.4) with $\bar{u}_0 = \bar{v}_0 = 0$ and f constant. The dispersion relation provides the wave frequencies in terms of the wave-number magnitude κ and of the constants of the problem, that are the same as in (1.6a) with $\bar{u}_0 = \bar{v}_0 = 0$. The first root, $\omega_0 = 0$ corresponds to the steady state and the remaining two roots, $\omega_\pm = \pm f\sqrt{R_D^2\kappa^2 + 1}$, correspond to bona fide traveling waves, called Poincaré waves.

The behavior of these waves is a mixed behavior between gravity waves (in the limit of no rotation $f = 0$) and inertial oscillations, so they are also called inertia-gravity waves. This is easy to be seen: for κ large (which corresponds to small wavelengths) we find $\kappa \gg 1/R_D$, which actually translates the fact that for κ too large, the wavelength is too small to feel the rotation. Since in this case we do not have the rotation, the gravity dominates. For κ small (which corresponds to large wavelengths) we find $\kappa \ll 1/R_D$, which means that the wavelength is much greater than the Rossby radius of deformation and in the Shallow Water equations the inertial terms dominate.

The frequencies ω_-, ω_0 and ω_+ are plotted in Figure 1.1 as a function of the ratio $R_D/\Delta x$. The parameters used are: $f = 1.5 \cdot 10^{-4} s^{-1}, g = 9.8 ms^{-2}, L = 6000km, \bar{\Phi}_0 = 10km$ and $\Delta x (= \Delta y)$ such that $R_D/\Delta x = 20, 10, 5$ (i.e. $\Delta x = L/60, L/30$ and $L/15$). The three exact inertia-gravity waves represented in Figure 1.1 will be used for comparison with the wave frequencies provided by the numerical schemes. The group velocities are represented in Figure 1.2.

2. Dispersion relations in the discrete case

In this section, we first recall the finite volume context. In order to discretize the doubly periodic domain $\Omega = (0, L_x) \times (0, L_y)$, we use rectangular finite volumes. We assume that all the volumes have the same dimensions $\Delta x \Delta y$ with $M\Delta x = L_x$ and $N\Delta y = L_y$ where M, N are given integers. Hence, we have MN control volumes defined by:

$$K_{ij} = (x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}) \times (y_{j-\frac{1}{2}}, y_{j+\frac{1}{2}}),$$

with

$$x_{i+\frac{1}{2}} = i\Delta x, \text{ for } i = \overline{0, M} \quad \text{and} \quad y_{j+\frac{1}{2}} = j\Delta y, \text{ for } j = \overline{0, N}.$$

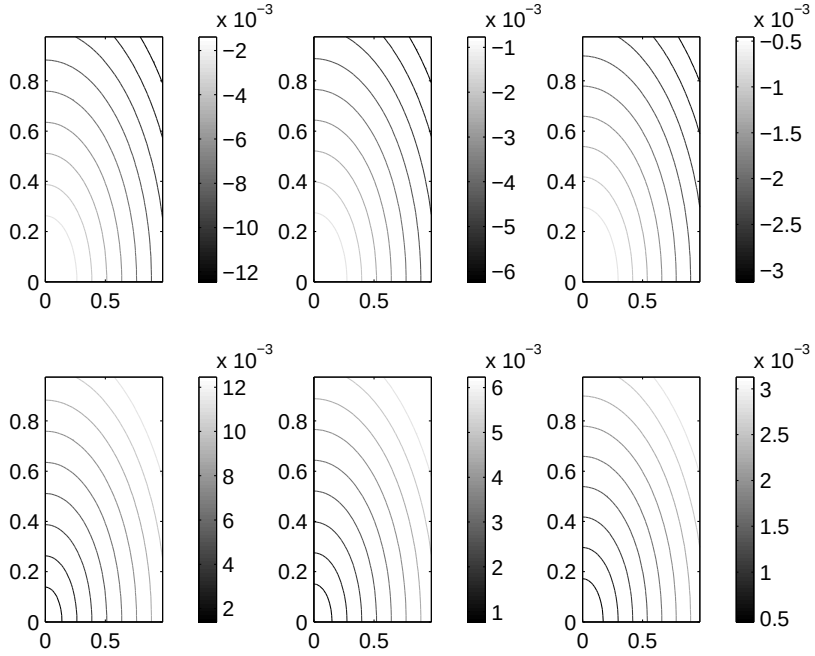

FIGURE 1.1. Inertia-gravity waves: ω_- (top) and ω_+ (bottom) as functions of $k\Delta x/\pi$ and $l\Delta y/\pi$ (with $\Delta x = \Delta y$) for three values of $R_D/\Delta x$: 20, 10, 5. Due to the symmetry of the problem, only the upper right quadrant is shown.

The finite volume discretizations considered in this article can be formulated in the following general form:

$$\begin{aligned} \frac{1}{2\Delta t}(\beta q_{ij}^{n+1} - \beta(\beta-1)q_{ij}^{n} - (2-\beta)q_{ij}^{n-1}) \quad &+ \quad \frac{1}{\Delta x}(F^n_{i+\frac{1}{2}j} - F^n_{i-\frac{1}{2}j}) \\ &+ \quad \frac{1}{\Delta y}(G^n_{ij+\frac{1}{2}} - G^n_{ij-\frac{1}{2}}) + \mathbf{C}q_{ij}^n = 0, \end{aligned} \tag{2.1}$$

where

$$q_{ij}^n = \frac{1}{\Delta x \Delta y}\int_{K_{ij}} q(x,y,n\Delta t)dxdy, \quad i = 1,..,M, \quad j = 1,..,N, \quad n = 1,..,N_t,$$

and

$$F^n_{i+\frac{1}{2}j} = \begin{pmatrix} F^{u,n}_{i+\frac{1}{2}j} \\ F^{v,n}_{i+\frac{1}{2}j} \\ F^{\Phi,n}_{i+\frac{1}{2}j} \end{pmatrix} \text{ and } G^n_{ij+\frac{1}{2}} = \begin{pmatrix} G^{u,n}_{ij+\frac{1}{2}} \\ G^{v,n}_{ij+\frac{1}{2}} \\ G^{\Phi,n}_{ij+\frac{1}{2}} \end{pmatrix}$$

are the numerical fluxes.
The time discretization schemes depend on the value of β; for $\beta = 2$ we have the Euler scheme and for $\beta = 1$ the time discretization is the Leap-Frog scheme.

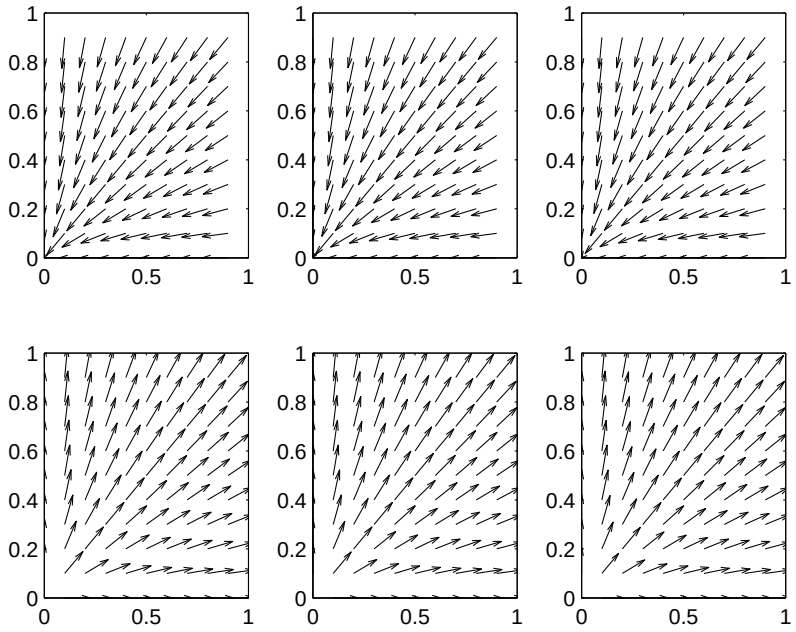

FIGURE 1.2. Inertia-gravity waves: $v_{g,-}$ (top) and $v_{g,+}$ (bottom) as functions of $k\Delta x/\pi$ and $l\Delta y/\pi$ (with $\Delta x = \Delta y$) for three values of $R_D/\Delta x$: 20, 10, 5. Due to the symmetry of the problem, only the upper right quadrant is shown.

2.1. Leap-Frog time discretization and centered fluxes. First, we recall the discrete dispersion relations when the numerical fluxes are obtained by a centered approximation, i.e.:

$$F^n_{i+\frac{1}{2}j} = \mathbf{A}\left(\frac{q^n_{i+1j} + q^n_{ij}}{2}\right) \text{ and } G^n_{ij+\frac{1}{2}} = \mathbf{B}\left(\frac{q^n_{ij+1} + q^n_{ij}}{2}\right). \tag{2.2}$$

Near the boundary, the periodical boundary conditions give for example $F_{M+\frac{1}{2}j} = F_{\frac{1}{2}j}$, that means, using one layer of fictitious control volumes $q_{M+1j} = q_{1j}$ and $q_{0j} = q_{Mj}$ for all $j = 1, .., N$.
Then, to obtain the dispersion relation, we replace q^n_{ij} by the discrete wave $(\bar{u}, \bar{v}, \bar{\Phi})e^{\mathbf{i}(ki\Delta x + lj\Delta y - \omega n\Delta t)}$ in the numerical scheme (2.1).
From the relations:

$$q^n_{i+1j} - q^n_{i-1j} = q^n_{ij} 2\mathbf{i}\sin(k\Delta x), \tag{2.3}$$

$$q^n_{ij+1} - q^n_{ij-1} = q^n_{ij} 2\mathbf{i}\sin(l\Delta y), \tag{2.4}$$

$$q^{n+1}_{ij} - q^{n-1}_{ij} = -q^n_{ij} 2\mathbf{i}\sin(\omega\Delta t), \tag{2.5}$$

we find that the dispersion relation is given by

$$det\left(-\mathbf{I}\tilde{\omega} + \mathbf{A}\tilde{k} + \mathbf{B}\tilde{l} - \mathbf{iC}\right) = 0, \tag{2.6}$$

where $\mathbf{i} = \sqrt{-1}$ and where we used, for an easier readability, the notations:

$$\tilde{\omega} = \frac{1}{\Delta t}\sin(\omega \Delta t), \quad \tilde{k} = \frac{1}{\Delta x}\sin(k\Delta x), \quad \tilde{l} = \frac{1}{\Delta y}\sin(l\Delta y). \tag{2.7}$$

Solving the dispersion relation, we find the wave frequencies ω_0, ω_- and ω_+ as:

$$\omega_0 = \frac{1}{\Delta t}\arcsin(\Delta t \tilde{\omega}_0), \quad \text{with} \quad \tilde{\omega}_0 = \tilde{k}\bar{u}_0 + \tilde{l}\bar{v}_0, \tag{2.8a}$$

$$\omega_\pm = \frac{1}{\Delta t}\arcsin\left(\Delta t\left(\tilde{\omega}_0 \pm f\sqrt{R_D^2\tilde{\kappa}^2 + 1}\right)\right), \text{ with } \tilde{\kappa}^2 = \tilde{k}^2 + \tilde{l}^2. \tag{2.8b}$$

Hence, the discrete group velocities $v_{d,g} = (\partial\omega_\pm/\partial k, \partial\omega_\pm/\partial l)$ are given by:

$$v_{d,g} = \begin{pmatrix} \dfrac{1}{\sigma_\pm}\{\cos(k\Delta x)\bar{u}_0 \pm \dfrac{R_D^2 f \sin(2k\Delta x)}{2\Delta x\sqrt{R_D^2\tilde{\kappa}^2+1}}\} \\ \dfrac{1}{\sigma_\pm}\{\cos(l\Delta y)\bar{v}_0 \pm \dfrac{R_D^2 f \sin(2l\Delta y)}{2\Delta y\sqrt{R_D^2\tilde{\kappa}^2+1}}\} \end{pmatrix}, \tag{2.9}$$

with $\sigma_\pm = \sqrt{1 - \Delta t^2\left(\tilde{\omega}_0 \pm f\sqrt{R_D^2\tilde{\kappa}^2+1}\right)^2}$.

We are also interested in studying the numerical error introduced by the dispersive properties of the numerical scheme. The dispersive mechanism results in local error focusing, so in a sudden local error growth. This phenomenon is referred to as a *spurious caustic phenomenon*. We will see that while the exact solution of the continuous model does not exhibit caustics, the discrete solution associated with the numerical scheme will admit spurious caustics.

More exactly, a caustic is defined as a focusing curve of different rays in one location which in mathematical terms is translated by the fact that the velocity group exhibits an extremum. For more details about this phenomenon, we refer the interested reader to Lighthill (2001) and Whitham (1999).

Hence we need to search for a wave number k_c such that:

$$\frac{\partial v_g}{\partial k}(k_c) = 0. \tag{2.10}$$

We can easily check that there are no caustics for the exact inertia-gravity waves. In fact, we found that the continuous velocity group is of the form:

$$v_g = \pm\frac{R_D^2 f k}{\sqrt{R_D^2\kappa^2+1}}.$$

By a simple computation we can find:

$$\frac{\partial v_g}{\partial k} = \mp R_D^2 f\frac{R_D^2 l^2 + 1}{(\sqrt{R_D^2\kappa^2+1})^3},$$

so there is no k_c such that $(\partial v_g/\partial k)(k_c) = 0$.

For the discrete case, we can easily see the spurious caustics appearing in the wave frequency corresponding to the steady state:

$$\omega_0 = \frac{1}{\Delta t}\arcsin(\Delta t\tilde{\omega}_0).$$

So, the first component of the velocity group is:

$$v_{g,d} = \frac{\bar{u}_0\cos(k\Delta x)}{\sqrt{1 - \Delta t^2\tilde{\omega}_0^2}},$$

and differentiating in terms of k we find:

$$\frac{\partial v_{g,d}}{\partial k} = \frac{\bar{u}_0 \Delta x}{\sqrt{1-\Delta t^2 \tilde{\omega}_0^2}} \Big\{ -\sin(k\Delta x) + \frac{\Delta t^2 \bar{u}_0 \tilde{k}' \tilde{\omega}_0}{1-\Delta t^2 \tilde{\omega}_0^2} \Big\}.$$

An obvious solution is obtained for $\tilde{k} = 0$ and $\tilde{l} = 0$ (which means that periodic peaks of energy appear), and a similar result also holds for $\partial v_{g,d}/\partial k$. This result is also illustrated in Figure 2.1.

Illustration: Inertia-gravity waves

Figure 2.2 shows the wave frequency values computed with a linear interpolation for the fluxes. We notice that the scheme is efficient for long waves, while for medium and small length waves we can notice that directional errors appear. In Figure 2.2 we can see that only the wave frequencies appearing in the left corner of the graphs, which correspond to long waves, are similar to the analytic wave frequencies represented in Figure 1.1. The same result is emphasized comparing Figure 1.2 where the exact group velocities are represented, to Figure 2.3 where the approximate group velocities are shown; we notice that important angular and directional errors appear for the small length waves.

Another property that we want to emphasize is that the scheme is more precise for a larger $R_D/\Delta x$. This can be seen from Figure 2.4 where we can see a zoom on the relative frequency error for the first modes.

We also show numerically the existence of the spurious caustics. From theoretical computations we saw that the velocity in the continuous case does not exhibit a local extremum. Figure 2.1 shows that the velocity presents periodic local extremum, which is in accordance to our computations where we showed that periodic peaks of energy appear.

2.2. Euler time discretization and Lax-Wendroff fluxes. With $\mathbf{C} = 0$, the usual Lax-Wendroff method for a linear system like (1.4) can be interpreted as a method of the form (2.1) where the numerical fluxes are given by:

$$\begin{aligned} F^n_{i+\frac{1}{2}j} =& \mathbf{A}\left(\frac{q^n_{i+1j}+q^n_{ij}}{2}\right) - \frac{\Delta t}{2\Delta x}\mathbf{A}^2\left(q^n_{i+1j}-q^n_{ij}\right) \\ &- \frac{\Delta t}{8\Delta y}\mathbf{AB}\left((q^n_{i+1j+1}-q^n_{i+1j})+(q^n_{ij+1}-q^n_{ij})\right. \\ &\left.+(q^n_{i+1j}-q^n_{i+1j-1})+(q^n_{ij}-q^n_{ij-1})\right), \end{aligned} \tag{2.11}$$

$$\begin{aligned} G^n_{ij+\frac{1}{2}} =& \mathbf{B}\left(\frac{q^n_{ij+1}+q^n_{ij}}{2}\right) - \frac{\Delta t}{2\Delta y}\mathbf{B}^2\left(q^n_{ij+1}-q^n_{ij}\right) \\ &- \frac{\Delta t}{8\Delta x}\mathbf{BA}\left((q^n_{i+1j+1}-q^n_{ij+1})+(q^n_{i+1j}-q^n_{ij})\right. \\ &\left.+(q^n_{ij+1}-q^n_{i-1j+1})+(q^n_{ij}-q^n_{i-1j})\right). \end{aligned} \tag{2.12}$$

In order to take into account the term $\mathbf{C}q$ in(1.4), we have to modify the above scheme. The time derivative is approximated by

$$\left(\frac{\partial q}{\partial t}\right)^n \simeq \frac{q^{n+1}-q^n}{\Delta t} - \frac{\Delta t}{2}\left(\frac{\partial^2 q}{\partial t^2}\right)^n \tag{2.13}$$

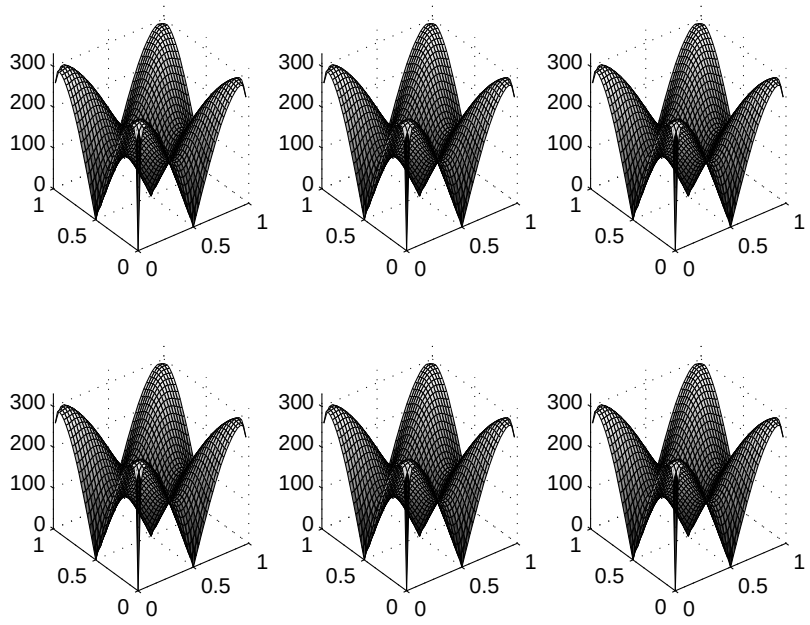

FIGURE 2.1. Inertia-gravity waves using Leap-Frog time discretization and linear interpolation for the fluxes: $||v_{g,-}||_2$ (top) and $||v_{g,+}||_2$ (bottom) as functions of $k\Delta x/\pi$ and $l\Delta y/\pi$ (with $\Delta x = \Delta y$) for three values of $R_D/\Delta x$: 20, 10, 5. Due to the symmetry of the problem, only the upper right quadrant is shown.

where the second time derivative is obtained from

$$\begin{aligned}\frac{\partial^2 q}{\partial t^2} =& \mathbf{A}^2\frac{\partial^2 q}{\partial x^2} + \mathbf{B}^2\frac{\partial^2 q}{\partial y^2} + (\mathbf{AB}+\mathbf{BA})\frac{\partial^2 q}{\partial xy}\\ &+ (\mathbf{AC}+\ \mathbf{CA})\frac{\partial q}{\partial x} + (\mathbf{BC}+\mathbf{CB})\frac{\partial q}{\partial y} + \mathbf{C}^2 q.\end{aligned} \tag{2.14}$$

Consequently, we obtain the following Lax-Wendroff fluxes:

$$\begin{aligned}F^n_{i+\frac{1}{2}j} =& \left(\mathbf{A} - \frac{\Delta t}{2}(\mathbf{AC}+\mathbf{CA})\right)\left(\frac{q^n_{i+1j}+q^n_{ij}}{2}\right) - \frac{\Delta t}{2\Delta x}\mathbf{A}^2\left(q^n_{i+1j}-q^n_{ij}\right)\\ &- \frac{\Delta t}{8\Delta y}\mathbf{AB}\left((q^n_{i+1j+1}-q^n_{i+1j})+(q^n_{ij+1}-q^n_{ij})\right.\\ &\left.+(q^n_{i+1j}-q^n_{i+1j-1})+(q^n_{ij}-q^n_{ij-1})\right),\end{aligned} \tag{2.15}$$

$$\begin{aligned}G^n_{ij+\frac{1}{2}} =& \left(\mathbf{B} - \frac{\Delta t}{2}(\mathbf{BC}+\mathbf{CB})\right)\left(\frac{q^n_{ij+1}+q^n_{ij}}{2}\right) - \frac{\Delta t}{2\Delta y}\mathbf{B}^2\left(q^n_{ij+1}-q^n_{ij}\right)\\ &- \frac{\Delta t}{8\Delta x}\mathbf{BA}\left((q^n_{i+1j+1}-q^n_{ij+1})+(q^n_{i+1j}-q^n_{ij})\right.\\ &\left.+(q^n_{ij+1}-q^n_{i-1j+1})+(q^n_{ij}-q^n_{i-1j})\right).\end{aligned} \tag{2.16}$$

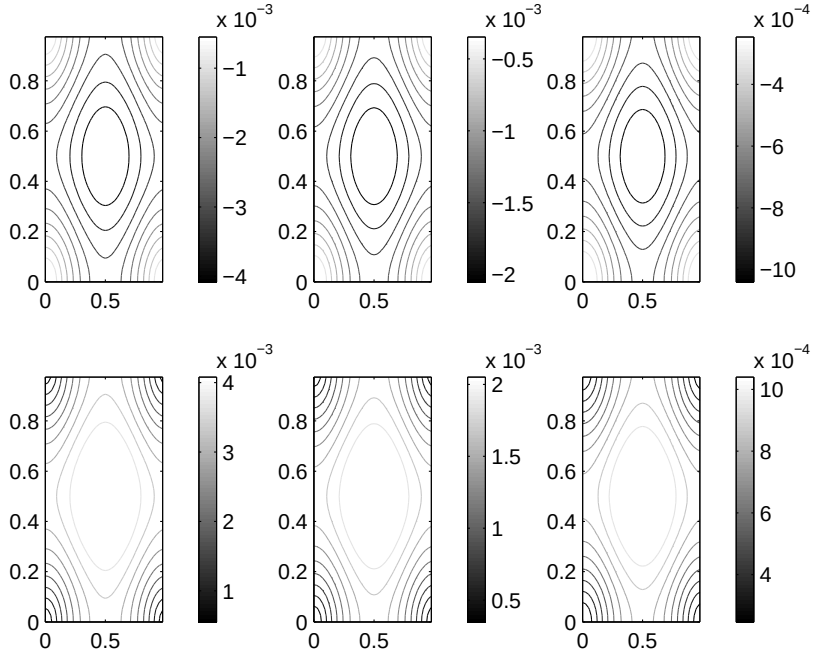

FIGURE 2.2. Inertia-gravity waves using Leap-Frog time discretization and linear interpolation for the fluxes: ω_- (top) and ω_+ (bottom) as functions of $k\Delta x/\pi$ and $l\Delta y/\pi$ (with $\Delta x = \Delta y$) for three values of $R_D/\Delta x$: 20, 10, 5. Due to the symmetry of the problem, only the upper right quadrant is shown. Compare to Figure 1.1.

Near the boundary, we use the same convention as for the centered scheme. From the above numerical flux definitions, we obtain the scheme

$$\begin{aligned}
&\frac{1}{\Delta t}\left(q_{ij}^{n+1}-q_{ij}^{n}\right)+\frac{1}{2\Delta x}(\mathbf{A}-\Delta t\frac{(\mathbf{AC}+\mathbf{CA})}{2})\left(q_{i+1j}^{n}-q_{i-1j}^{n}\right)\\
+\quad&\frac{1}{2\Delta y}(\mathbf{B}-\Delta t\frac{(\mathbf{BC}+\mathbf{CB})}{2})\left(q_{ij+1}^{n}-q_{ij-1}^{n}\right)+(\mathbf{C}-\Delta t\frac{\mathbf{C}^2}{2})q_{ij}^{n}\\
-\quad&\frac{\Delta t}{2\Delta x^2}\mathbf{A}^2\left(q_{i+1j}^{n}-2q_{ij}^{n}+q_{i-1j}^{n}\right)-\frac{\Delta t}{2\Delta y^2}\mathbf{B}^2\left(q_{ij+1}^{n}-2q_{ij}^{n}+q_{ij-1}^{n}\right)\\
-\quad&\frac{\Delta t}{8\Delta x\Delta y}(\mathbf{AB}+\mathbf{BA})\left(q_{i+1j+1}^{n}-q_{i+1j-1}^{n}-q_{i-1j+1}^{n}+q_{i-1j-1}^{n}\right)=0,
\end{aligned}\tag{2.17}$$

where the matrices are:

$$\mathbf{AC}+\mathbf{CA}=\begin{pmatrix}0 & -2f\bar{u}_0 & 0\\ 2f\bar{u}_0 & 0 & fg\\ 0 & -f\bar{\Phi}_0 & 0\end{pmatrix},\ \mathbf{BC}+\mathbf{CB}=\begin{pmatrix}0 & -2f\bar{v}_0 & -fg\\ 2f\bar{v}_0 & 0 & 0\\ f\bar{\Phi}_0 & 0 & 0\end{pmatrix},$$

$$\mathbf{A}^2=\begin{pmatrix}\bar{u}_0^2+g\bar{\Phi}_0 & 0 & 2g\bar{u}_0\\ 0 & \bar{u}_0^2 & 0\\ 2\bar{u}_0\bar{\Phi}_0 & 0 & \bar{u}_0^2+g\bar{\Phi}_0\end{pmatrix},\ \mathbf{B}^2=\begin{pmatrix}\bar{v}_0^2 & 0 & 0\\ 0 & \bar{v}_0^2+g\bar{\Phi}_0 & 2g\bar{v}_0\\ 0 & 2\bar{v}_0\bar{\Phi}_0 & \bar{v}_0^2+g\bar{\Phi}_0\end{pmatrix},$$

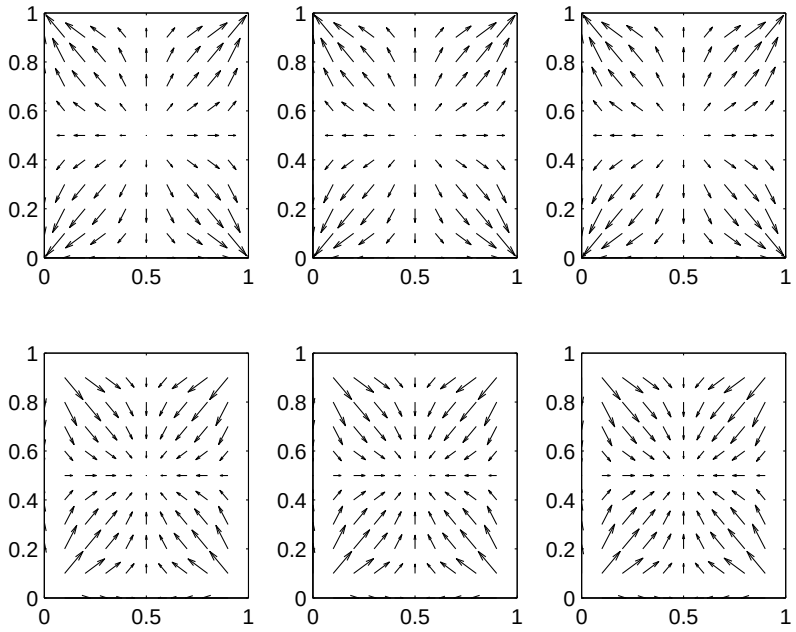

FIGURE 2.3. Inertia-gravity waves using Leap-Frog time discretization and linear interpolation for the fluxes: $v_{g,-}$ (top) and $v_{g,+}$ (bottom) as functions of $k\Delta x/\pi$ and $l\Delta y/\pi$ (with $\Delta x = \Delta y$) for three values of $R_D/\Delta x$: 20, 10, 5. Due to the symmetry of the problem, only the upper right quadrant is shown. Compare to Figure 2.1.

$$\mathbf{AB}+\mathbf{BA} = \begin{pmatrix} 2\bar{u}_0\bar{v}_0 & g\bar{\Phi}_0 & 2g\bar{v}_0 \\ g\bar{\Phi}_0 & 2\bar{u}_0\bar{v}_0 & 2g\bar{u}_0 \\ 2\bar{\Phi}_0\bar{v}_0 & 2\bar{\Phi}_0\bar{u}_0 & 2\bar{u}_0\bar{v}_0 \end{pmatrix}, \quad \mathbf{C}^2 = \begin{pmatrix} -f^2 & 0 & 0 \\ 0 & -f^2 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

The last three terms are diffusive and they stabilize the scheme.

Then, replacing, in (2.17), q_{ij}^n by the discrete wave $(\bar{u},\bar{v},\bar{\Phi})e^{\mathbf{i}(ki\Delta x+lj\Delta y-\omega n\Delta t)}$ and using the relations:

$$q_{ij}^{n+1} - q_{ij}^n = q_{ij}^n(e^{-\mathbf{i}\omega\Delta t} - 1), \tag{2.18}$$

$$q_{i+1j}^n - 2q_{ij}^n + q_{i-1j}^n = -4q_{ij}^n\tilde{\tilde{k}}\Delta x^2, \tag{2.19}$$

$$q_{ij+1}^n - 2q_{ij}^n + q_{ij-1}^n = -4q_{ij}^n\tilde{\tilde{l}}\Delta y^2, \tag{2.20}$$

$$q_{i+1j+1}^n - q_{i+1j-1}^n - q_{i-1j+1}^n + q_{i-1j-1}^n = q_{ij}^n(-4\sin(k\Delta x)\sin(l\Delta y)), \tag{2.21}$$

where we have used the previous notations $\tilde{k}$, $\tilde{l}$ and the new ones:

$$\tilde{\tilde{k}} = \frac{4}{\Delta x^2}\sin^2(k\Delta x/2), \quad \tilde{\tilde{l}} = \frac{4}{\Delta y^2}\sin^2(l\Delta y/2).$$

we obtain the dispersion relation as:

$$det\ N = 0, \tag{2.22}$$

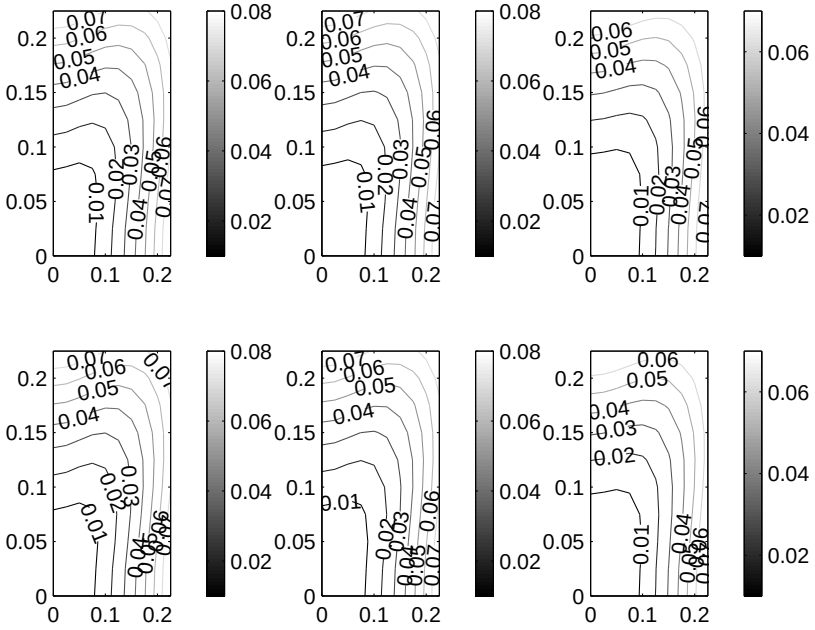

FIGURE 2.4. Inertia-gravity waves using Leap-Frog time discretization and linear interpolation for the fluxes: **zoom on the rms error** for the first modes of ω_- (top) and ω_+ (bottom) as functions of $k\Delta x/\pi$ and $l\Delta y/\pi$ (with $\Delta x = \Delta y$) for three values of $R_D/\Delta x$: 20, 10, 5. Due to the symmetry of the problem, only the upper right quadrant is shown.

where

$$N = \tfrac{1}{\Delta t}\mathbf{I}(e^{-\mathbf{i}\omega\Delta t} - 1) + \mathbf{i}\left(\mathbf{A} - \tfrac{\Delta t}{2}(\mathbf{AC} + \mathbf{CA})\right)\tilde{k} + \mathbf{i}\left(\mathbf{B} - \tfrac{\Delta t}{2}(\mathbf{BC} + \mathbf{CB})\right)\tilde{l}$$
$$+ \left(\mathbf{C} - \tfrac{\Delta t}{2}\mathbf{C}^2\right) + \tfrac{\Delta t}{2}\mathbf{A}^2\tilde{\tilde{k}} + \tfrac{\Delta t}{2}\mathbf{B}^2\tilde{\tilde{l}} + \tfrac{\Delta t}{2}\left(\mathbf{AB} + \mathbf{BA}\right)\tilde{k}\tilde{l}. \tag{2.23}$$

Observe that if we denote by $\tilde{k}(\Delta x)$, $\tilde{l}(\Delta y)$ the wave numbers $\tilde{k}$, $\tilde{l}$ in (2.7), then $\tilde{\tilde{k}} = \tilde{k}^2(\Delta x/2)$ and $\tilde{\tilde{l}} = \tilde{l}^2(\Delta y/2)$.

We now need to compute the determinant from (2.22). We can write $N = M + \alpha\mathbf{I}$ where the coefficients of the matrix M are:

$$
\begin{aligned}
m_{11} &= g\frac{\Delta t}{2}\tilde{\tilde{k}}\bar{\Phi}_0 + \frac{\Delta t}{2}f^2,\\
m_{12} &= -f + g\frac{\Delta t}{2}\tilde{k}\tilde{l}\bar{\Phi}_0 + \Delta t f\bar{u}_0\mathbf{i}\tilde{k} + \Delta t f\bar{v}_0\mathbf{i}\tilde{l},\\
m_{13} &= \mathbf{i}g\tilde{k} + g\Delta t\bar{u}_0\tilde{\tilde{k}} + \Delta t g\bar{v}_0\tilde{k}\tilde{l} + \frac{\Delta t}{2}gf\mathbf{i}\tilde{l},\\
m_{21} &= f + g\frac{\Delta t}{2}\tilde{k}\tilde{l}\bar{\Phi}_0 - \Delta t f\bar{u}_0\mathbf{i}\tilde{k} - \Delta t f\bar{v}_0\mathbf{i}\tilde{l},\\
m_{22} &= g\frac{\Delta t}{2}\tilde{\tilde{l}}\bar{\Phi}_0 + \frac{\Delta t}{2}f^2,\\
m_{23} &= \mathbf{i}g\tilde{l} + g\Delta t\bar{v}_0\tilde{\tilde{l}} + g\Delta t\bar{u}_0\tilde{k}\tilde{l} - \frac{\Delta t}{2}gf\mathbf{i}\tilde{k},\\
m_{31} &= \mathbf{i}\bar{\Phi}_0\tilde{k} + \bar{\Phi}_0\Delta t\bar{u}_0\tilde{\tilde{k}} + \Delta t\bar{\Phi}_0\bar{v}_0\tilde{k}\tilde{l} - \frac{\Delta t}{2}\bar{\Phi}_0 f\mathbf{i}\tilde{l},\\
m_{32} &= \mathbf{i}\bar{\Phi}_0\tilde{l} + \bar{\Phi}_0\Delta t\bar{v}_0\tilde{\tilde{l}} + \bar{\Phi}_0\Delta t\bar{u}_0\tilde{k}\tilde{l} + \frac{\Delta t}{2}\bar{\Phi}_0 f\mathbf{i}\tilde{k},\\
m_{33} &= g\frac{\Delta t}{2}\tilde{\tilde{k}}\bar{\Phi}_0 + g\frac{\Delta t}{2}\tilde{\tilde{l}}\bar{\Phi}_0,
\end{aligned}
\tag{2.24}
$$

and $\alpha = \dfrac{1}{\Delta t}(e^{-\mathbf{i}\omega\Delta t} - 1) + \mathbf{i}\tilde{k}\bar{u}_0 + \mathbf{i}\tilde{l}\bar{v}_0 + \dfrac{\Delta t}{2}\tilde{\tilde{k}}\bar{u}_0^2 + \dfrac{\Delta t}{2}\tilde{\tilde{l}}\bar{v}_0^2 + \Delta t\bar{u}_0\bar{v}_0\tilde{k}\tilde{l}.$

We then find that α satisfies the following equation:

$$\alpha^3 + a\alpha^2 + b\alpha + c = 0, \tag{2.25}$$

with

$$
\begin{aligned}
a &= m_{11} + m_{22} + m_{33},\\
b &= m_{11}m_{22} + m_{11}m_{33} + m_{22}m_{33} - m_{32}m_{23} - m_{21}m_{12} - m_{31}m_{13},\\
c &= m_{11}m_{22}m_{33} - m_{11}m_{32}m_{23} - m_{33}m_{21}m_{12} - m_{22}m_{31}m_{13}\\
&\quad + m_{21}m_{32}m_{13} + m_{31}m_{12}m_{23}.
\end{aligned}
\tag{2.26}
$$

We solve (2.25) using the classical transformation $\tilde{\tilde{\Omega}} = \alpha + a/3$. Equation (2.25) then becomes:

$$Y^3 + pY + q = 0,$$

where $p = b + a^2/3$ and $q = c - ab/3 + 2a^3/27$.

Following the Cardan method, the solutions are given by:

$$
\begin{aligned}
\tilde{\tilde{\Omega}}_- &= \frac{1}{3}\left(\sqrt[3]{\frac{-27q + \sqrt{27(27q^2 + 4p^3)}}{2}} + \sqrt[3]{\frac{-27q - \sqrt{27(27q^2 + 4p^3)}}{2}}\right),\\
\tilde{\tilde{\Omega}}_+ &= \frac{1}{3}\left(\mathbf{j}^2\sqrt[3]{\frac{-27q + \sqrt{27(27q^2 + 4p^3)}}{2}} + \mathbf{j}\sqrt[3]{\frac{-27q - \sqrt{27(27q^2 + 4p^3)}}{2}}\right),\\
\tilde{\tilde{\Omega}}_0 &= \frac{1}{3}\left(\mathbf{j}\sqrt[3]{\frac{-27q + \sqrt{27(27q^2 + 4p^3)}}{2}} + \mathbf{j}^2\sqrt[3]{\frac{-27q - \sqrt{27(27q^2 + 4p^3)}}{2}}\right),
\end{aligned}
$$

with $\mathbf{j} = -\frac{1}{2} + \mathbf{i}\frac{\sqrt{3}}{2}$.
Finally, recalling the definition of α, we find:

$$\begin{aligned}\frac{1}{\Delta t}\left(e^{-\mathbf{i}\omega\Delta t}-1\right) = \tilde{\tilde{\Omega}} - \mathbf{i}\,\tilde{k}\bar{u}_0 - \mathbf{i}\,\tilde{l}\bar{v}_0 - \frac{\Delta t}{\Delta x}\tilde{\tilde{k}}\bar{u}_0^2 - \frac{\Delta t}{\Delta y}\tilde{\tilde{l}}\bar{v}_0^2 - \Delta t\bar{u}_0\bar{v}_0\tilde{k}\tilde{l} \\ - \frac{2}{3}\left(\tilde{\tilde{k}}\frac{\Delta t}{\Delta x} + \tilde{\tilde{l}}\frac{\Delta t}{\Delta y}\right)g\bar{\Phi}_0 - \frac{\Delta t}{3}f^2,\end{aligned} \tag{2.27}$$

where ω is a complex number.

The real part of ω, which is the wave frequency that we wanted to find, is computed as:

$$Re(\omega) = \frac{1}{\Delta t}\arctan\frac{\tilde{k}\bar{u}_0 + \tilde{l}\bar{v}_0 - \ Im(\tilde{\tilde{\Omega}})}{Re(\tilde{\tilde{\Omega}}) + \frac{1}{\Delta t} - \tilde{\tilde{k}}\bar{u}_0^2\frac{\Delta t}{\Delta x} - \tilde{\tilde{l}}\bar{v}_0^2\frac{\Delta t}{\Delta y} - \Delta t\bar{u}_0\bar{v}_0\tilde{k}\tilde{l} - \frac{2}{3}\left(\tilde{\tilde{k}}\frac{\Delta t}{\Delta x} + \tilde{\tilde{l}}\frac{\Delta t}{\Delta y}\right)g\bar{\Phi}_0}.$$

Illustration: Inertia-gravity waves

In the case of inertia-gravity waves, the parameters $\bar{u}_0$ and $\bar{v}_0$ are equal to zero; hence $N = \alpha\mathbf{I} + M$ and the matrix M is simpler:

$$\begin{aligned}
&m_{11} = \tilde{\tilde{k}}g\bar{\Phi}_0\frac{\Delta t}{2} + \frac{\Delta t}{2}f^2, \quad m_{12} = -f + \tilde{k}\tilde{l}g\bar{\Phi}_0\frac{\Delta t}{2}, \quad m_{13} = \mathbf{i}g\tilde{k} + \frac{\Delta t}{2}gf\mathbf{i}\tilde{l},\\
&m_{21} = f + \tilde{k}\tilde{l}g\bar{\Phi}_0\frac{\Delta t}{2}, \quad m_{22} = \tilde{\tilde{l}}g\bar{\Phi}_0\frac{\Delta t}{2} + \frac{\Delta t}{2}f^2, \quad m_{23} = \mathbf{i}g\tilde{l} - \frac{\Delta t}{2}gf\mathbf{i}\tilde{k},\\
&m_{31} = \mathbf{i}\bar{\Phi}_0\tilde{k} - \frac{\Delta t}{2}\bar{\Phi}_0 f\mathbf{i}\tilde{l}, \quad m_{32} = \mathbf{i}\bar{\Phi}_0\tilde{l} + \frac{\Delta t}{2}\bar{\Phi}_0 f\mathbf{i}\tilde{k}, \quad m_{33} = \tilde{\tilde{k}}g\bar{\Phi}_0\frac{\Delta t}{2} + \tilde{\tilde{l}}g\bar{\Phi}_0\frac{\Delta t}{2},
\end{aligned}$$

with $\alpha = (\exp(-\mathbf{i}\omega\Delta t) - 1)/\Delta t$.

Then the real part of the wave frequency is then found as:

$$Re(\omega) = \frac{1}{\Delta t}\arctan\frac{-Im(\tilde{\tilde{\Omega}})}{Re(\tilde{\tilde{\Omega}}) + \frac{1}{\Delta t} - \frac{\Delta t}{3}\left(\tilde{\tilde{k}} + \tilde{\tilde{l}}\right)g\bar{\Phi}_0 - \frac{\Delta t}{3}f^2}, \tag{2.28}$$

where $\tilde{\tilde{\Omega}} = \alpha + a/3$ and $a = m_{11} + m_{22} + m_{33}$.

Figure 2.5 shows the wave frequency values computed using Lax-Wendroff fluxes. We notice the same thing as for the Leap-Frog discretization with centered fluxes, that the scheme is most accurate for very long waves, while for medium and small length waves we notice that directional errors appear. This conclusion also arises when comparing Figure 1.2 to Figure 2.6; we can easily see that for medium and small length waves important angular and directional errors appear in the computations of the velocity group.

This scheme is also more precise for a larger value of $R_D/\Delta x$. This conclusion arises from Figure 2.7 where we can see a zoom on the relative frequency error for the first modes.

Because of the difficulty of the computations, we do not study theoretically the existence of the spurious caustics for this scheme. However, we present a numerical result about the existence of this phenomenon, by representing the velocity obtained using the Euler-time discretization with Lax-Wendroff fluxes as a function of $k\Delta x/\pi$ (see Figure 2.8). We can notice that while the exact velocity was not presenting local extrema, the approximate velocity obtained with this scheme does exhibit periodic local extrema.

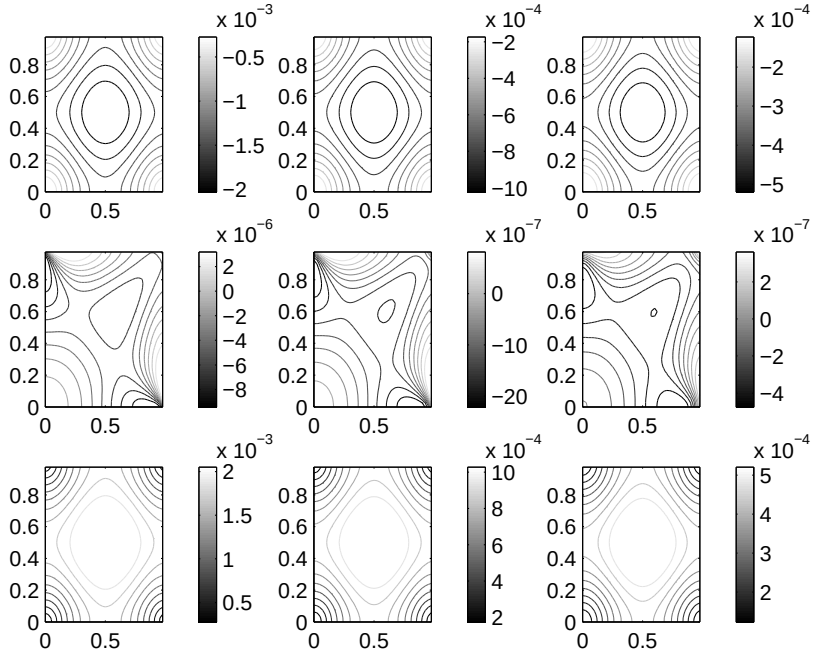

FIGURE 2.5. Inertia-gravity waves using Euler time discretization and Lax-Wendroff fluxes: ω_- (top), ω_0 (middle) and ω_+ (bottom) as functions of $k\Delta x/\pi$ and $l\Delta y/\pi$ (with $\Delta x = \Delta y$) for three values of $R_D/\Delta x$: 20, 10, 5. Due to the symmetry of the problem, only the upper right quadrant is shown. Compare to Figure 1.1.

2.3. Leap-Frog time discretization and upwind fluxes. The "upwind" method for the linear system (1.4) requires the knowledge of the eigenvalues and the eigenvectors of the matrices $\mathbf{A}$ and $\mathbf{B}$. We write:

$$\mathbf{A} = \mathbf{P_A}\mathbf{\Lambda_A}{\mathbf{P_A}}^{-1} \text{ and } \mathbf{B} = \mathbf{P_B}\mathbf{\Lambda_B}{\mathbf{P_B}}^{-1}$$

where $\mathbf{\Lambda_A}$ (respectively $\mathbf{\Lambda_B}$) is the diagonal matrix of the eigenvalues of $\mathbf{A}$ (respectively $\mathbf{B}$):

$$\mathbf{\Lambda_A} = Diag(\bar{u}_0 - \sqrt{\bar{\Phi}_0 g}, \bar{u}_0, \bar{u}_0 + \sqrt{\bar{\Phi}_0 g}), \mathbf{\Lambda_B} = Diag(\bar{v}_0 - \sqrt{\bar{\Phi}_0 g}, \bar{v}_0, \bar{v}_0 + \sqrt{\bar{\Phi}_0 g}),$$

and $\mathbf{P_A}$ (respectively $\mathbf{P_B}$) is the corresponding eigenvectors matrix:

$$\mathbf{P_A} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ -\sqrt{\bar{\Phi}_0/g} & 0 & \sqrt{\bar{\Phi}_0/g} \end{pmatrix}, \mathbf{P_B} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & -\sqrt{\bar{\Phi}_0/g} & \sqrt{\bar{\Phi}_0 g} \end{pmatrix}.$$

Moreover, we define ${\mathbf{\Lambda_A}}^+ = max(\mathbf{\Lambda_A}, 0)$ and ${\mathbf{\Lambda_A}}^- = max(-\mathbf{\Lambda_A}, 0)$ as the diagonal matrices associated with the positive and negative eigenvalues, so $\mathbf{\Lambda_A} = {\mathbf{\Lambda_A}}^+ - {\mathbf{\Lambda_A}}^-$. We do the same thing for $\mathbf{\Lambda_B}$, using ${\mathbf{\Lambda_B}}^+$ and ${\mathbf{\Lambda_B}}^-$, and we then introduce the following matrices:

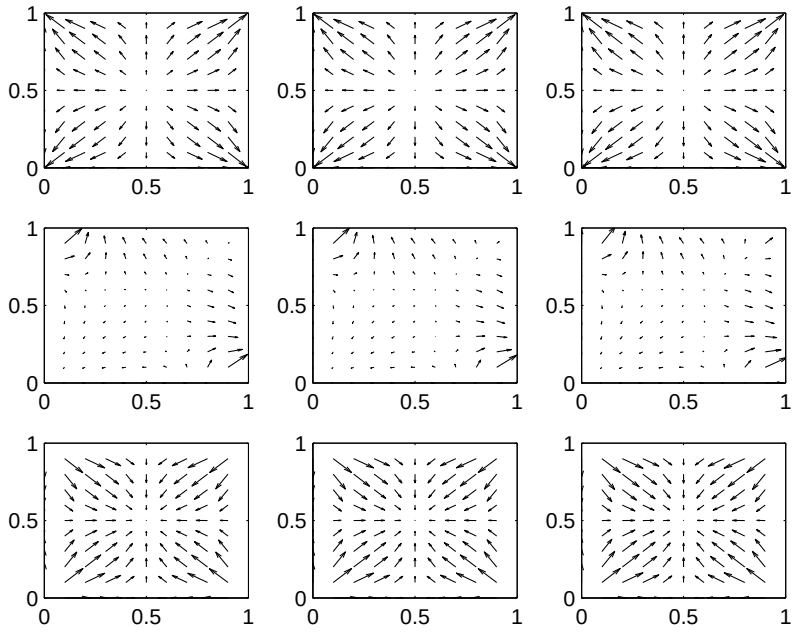

FIGURE 2.6. Inertia-gravity waves using Euler time discretization and Lax-Wendroff fluxes: $v_{g,-}$ (top), $v_{g,0}$ (middle) and $v_{g,+}$ (bottom) as functions of $k\Delta x/\pi$ and $l\Delta y/\pi$ (with $\Delta x = \Delta y$) for three values of $R_D/\Delta x$: 20, 10, 5. Due to the symmetry of the problem, only the upper right quadrant is shown. Compare to Figure 2.1.

$$
\begin{aligned}
&\mathbf{A}^+ = \mathbf{P_A}\mathbf{\Lambda_A}^+\mathbf{P_A}^{-1} \text{ and } \mathbf{B}^+ = \mathbf{P_B}\mathbf{\Lambda_B}^+\mathbf{P_B}^{-1},\\
&\mathbf{A}^- = \mathbf{P_A}\mathbf{\Lambda_A}^-\mathbf{P_A}^{-1} \text{ and } \mathbf{B}^+ = \mathbf{P_B}\mathbf{\Lambda_B}^-\mathbf{P_B}^{-1},\\
&\mathbf{A} = \mathbf{A}^+ - \mathbf{A}^- \text{ and } \mathbf{B} = \mathbf{B}^+ - \mathbf{B}^-,\\
&|\mathbf{A}| = \mathbf{A}^+ + \mathbf{A}^- \text{ and } |\mathbf{B}| = \mathbf{B}^+ + \mathbf{B}^-.
\end{aligned}
$$

With these notations, the "upwind" numerical fluxes read:

$$
F^n_{i+\frac{1}{2}j} = \mathbf{A}\left(\frac{q^n_{i+1j} + q^n_{ij}}{2}\right) - \frac{|\mathbf{A}|}{2}\left(q^n_{i+1j} - q^n_{ij}\right), \tag{2.29}
$$

$$
G^n_{ij+\frac{1}{2}} = \mathbf{B}\left(\frac{q^n_{ij+1} + q^n_{ij}}{2}\right) - \frac{|\mathbf{B}|}{2}\left(q^n_{ij+1} - q^n_{ij}\right). \tag{2.30}
$$

Near the boundary, we still use the same convention than for the previous schemes. From the above flux definition, we obtain the scheme

$$
\begin{aligned}
&\tfrac{1}{2\Delta t}\left(q^{n+1}_{ij} - q^{n-1}_{ij}\right) + \tfrac{1}{2\Delta x}\mathbf{A}\left(q^n_{i+1j} - q^n_{i-1j}\right) + \tfrac{1}{2\Delta y}\mathbf{B}\left(q^n_{ij+1} - q^n_{ij-1}\right) + \mathbf{C}q^n_{ij}\\
&-\tfrac{1}{2\Delta x}|\mathbf{A}|\left(q^n_{i+1j} - 2q^n_{ij} + q^n_{i-1j}\right) - \tfrac{1}{2\Delta y}|\mathbf{B}|\left(q^n_{ij+1} - 2q^n_{ij} + q^n_{ij-1}\right) = 0.
\end{aligned} \tag{2.31}
$$

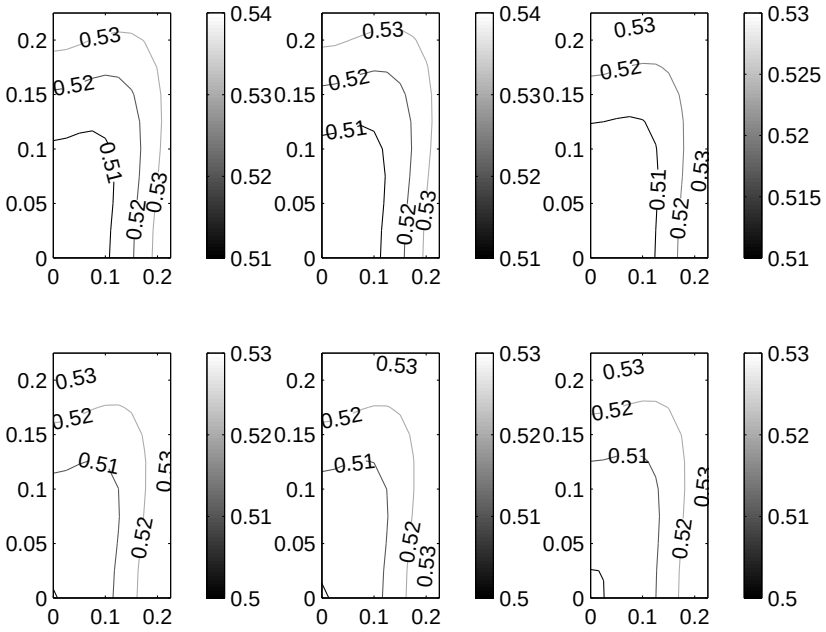

FIGURE 2.7. Inertia-gravity waves using Euler time discretization and Lax-Wendroff fluxes: **zoom on the rms error** for the first modes of ω_- (top) and ω_+ (bottom) as functions of $k\Delta x/\pi$ and $l\Delta y/\pi$ (with $\Delta x = \Delta y$) for three values of $R_D/\Delta x$: 20, 10, 5. Due to the symmetry of the problem, only the upper right quadrant is shown.

Then, replacing q_{ij}^n by the discrete wave $(\bar{u}, \bar{v}, \bar{\Phi})e^{\mathbf{i}(ki\Delta x + lj\Delta y - \omega n\Delta t)}$, we obtain the dispersion relation by solving:

$$det\left(-\mathbf{i}\tilde{\omega}\mathbf{I} + \mathbf{i}\tilde{k}\mathbf{A} + \mathbf{i}\tilde{l}\mathbf{B} + \mathbf{C} + |\mathbf{A}|\tilde{\tilde{k}}\frac{\Delta x}{2} + |\mathbf{B}|\tilde{\tilde{l}}\frac{\Delta y}{2}\right) = 0. \tag{2.32}$$

In the particular case of a subcritical flow, the constants $\bar{u}_0, \bar{v}_0$ and $\bar{\Phi}_0$ are such that

$$\bar{u}_0 > 0, \quad \bar{v}_0 > 0, \quad \bar{u}_0^2 + \bar{v}_0^2 < \bar{\Phi}_0 g. \tag{2.33}$$

Hence, the matrices $|\mathbf{A}|$ and $|\mathbf{B}|$ are

$$|\mathbf{A}| = \begin{pmatrix} \sqrt{\bar{\Phi}_0 g} & 0 & \dfrac{\bar{u}_0}{\sqrt{\bar{\Phi}_0/g}} \\ 0 & \bar{u}_0 & 0 \\ \bar{u}_0\sqrt{\bar{\Phi}_0/g} & 0 & \sqrt{\bar{\Phi}_0 g} \end{pmatrix}, \; |\mathbf{B}| = \begin{pmatrix} \bar{v}_0 & 0 & 0 \\ 0 & \sqrt{\bar{\Phi}_0 g} & \dfrac{\bar{v}_0}{\sqrt{\bar{\Phi}_0/g}} \\ 0 & \bar{v}_0\sqrt{\bar{\Phi}_0/g} & \sqrt{\bar{\Phi}_0 g} \end{pmatrix},$$

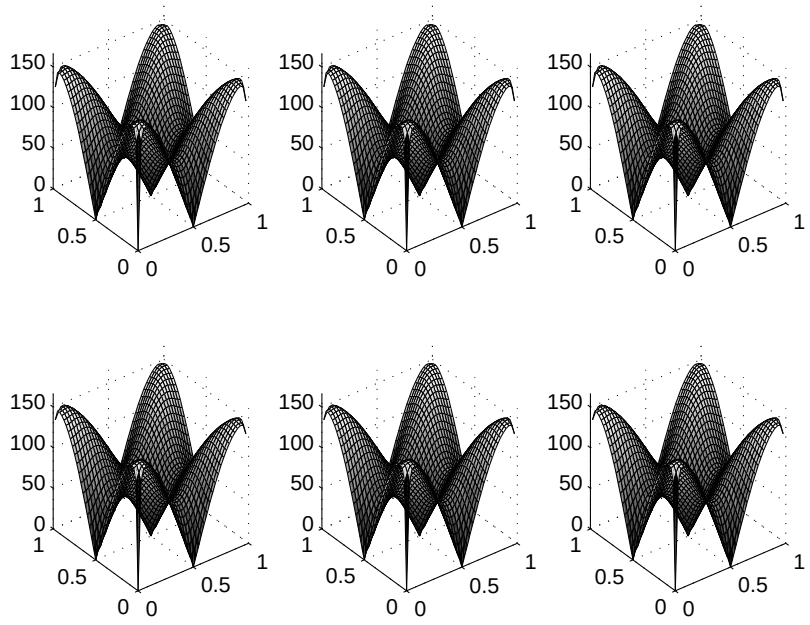

FIGURE 2.8. Inertia-gravity waves using Euler time discretization and Lax-Wendroff fluxes: $||v_{g,-}||_2$ (top) and $||v_{g,+}||_2$ (bottom) as functions of $k\Delta x/\pi$ and $l\Delta y/\pi$ (with $\Delta x = \Delta y$) for three values of $R_D/\Delta x$: 20, 10, 5. Due to the symmetry of the problem, only the upper right quadrant is shown.

and the dispersion relation (2.32) becomes:

$$\begin{vmatrix} \tilde{\Omega} - \mathbf{i}(\bar{v}_0 - \lambda)\tilde{\bar{l}} & \mathbf{i}f & g(\tilde{k} - \mathbf{i}\frac{\bar{u}_0}{\lambda}\tilde{\bar{k}}) \\ -\mathbf{i}f & \tilde{\Omega} - \mathbf{i}(\bar{u}_0 - \lambda)\tilde{\bar{k}} & g(\tilde{l} - \mathbf{i}\frac{\bar{v}_0}{\lambda}\tilde{\bar{l}}) \\ \bar{\Phi}_0(\tilde{k} - \mathbf{i}\frac{\bar{u}_0}{\lambda}\tilde{\bar{k}}) & \bar{\Phi}_0(\tilde{l} - \mathbf{i}\frac{\bar{v}_0}{\lambda}\tilde{\bar{l}}) & \tilde{\Omega} \end{vmatrix} = 0,$$

where $\lambda = \sqrt{\bar{\Phi}_0 g}$ and $\tilde{\Omega} = -\tilde{\omega} + \tilde{k}\bar{u}_0 + \tilde{l}\bar{v}_0 - \mathbf{i}\lambda(\tilde{\bar{k}} + \tilde{\bar{l}})$.
This is equivalent to:

$$\tilde{\Omega}^3 - \mathbf{i}\tilde{\Omega}^2\left[(\bar{u}_0 - \lambda)\tilde{\bar{k}} + (\bar{v}_0 - \lambda)\tilde{\bar{l}}\right] - \tilde{\Omega}\left[f^2 + (\bar{u}_0 - \lambda)(\bar{v}_0 - \lambda)\tilde{\bar{k}}\tilde{\bar{l}} + (\lambda\tilde{k} - \mathbf{i}\bar{u}_0\tilde{\bar{k}})^2 + (\lambda\tilde{l} - \mathbf{i}\bar{v}_0\tilde{\bar{l}})^2\right] + \mathbf{i}(\bar{u}_0 - \lambda)\tilde{\bar{k}}(\lambda\tilde{k} - \mathbf{i}\bar{u}_0\tilde{\bar{k}})^2 + \mathbf{i}(\bar{v}_0 - \lambda)\tilde{\bar{l}}(\lambda\tilde{l} - \mathbf{i}\bar{v}_0\tilde{\bar{l}})^2 = 0.$$

Then, using the transformation $\tilde{\bar{\Omega}} = \tilde{\Omega} - \frac{\mathbf{i}}{3}((\bar{u}_0 - \lambda)\tilde{\bar{k}} + (\bar{v}_0 - \lambda)\tilde{\bar{l}})$, the second order term disappears and $\tilde{\bar{\Omega}}$ satisfies:

$$\tilde{\bar{\Omega}}^3 + p\tilde{\bar{\Omega}} + q = 0,$$

where

$$p = \frac{1}{3}\left((\bar{u}_0-\lambda)^2\tilde{\tilde{k}}^2 + (\bar{v}_0-\lambda)^2\tilde{\tilde{l}}^2 - (\bar{u}_0-\lambda)(\bar{v}_0-\lambda)\tilde{\tilde{k}}\tilde{\tilde{l}}\right) - (\lambda\tilde{k}-\mathbf{i}\bar{u}_0\tilde{\tilde{k}})^2 - (\lambda\tilde{l}-\mathbf{i}\bar{v}_0\tilde{\tilde{l}})^2 - f^2,$$

and

$$\begin{aligned} q &= \frac{\mathbf{i}}{27}\tilde{\tilde{k}}(\bar{u}_0-\lambda)\left[18(\lambda\tilde{k}-\mathbf{i}\bar{u}_0\tilde{\tilde{k}})^2 - 9(\lambda\tilde{l}-\mathbf{i}\bar{v}_0\tilde{\tilde{l}})^2 - 9f^2 - 3(\bar{v}_0-\lambda)^2\tilde{\tilde{l}}^2 + 2(\bar{u}_0-\lambda)^2\tilde{\tilde{k}}^2\right], \\ &+ \frac{\mathbf{i}}{27}\tilde{\tilde{l}}(\bar{v}_0-\lambda)\left[18(\lambda\tilde{l}-\mathbf{i}\bar{v}_0\tilde{\tilde{l}})^2 - 9(\lambda\tilde{k}-\mathbf{i}\bar{u}_0\tilde{\tilde{k}})^2 - 9f^2 - 3(\bar{u}_0-\lambda)^2\tilde{\tilde{k}}^2 + 2(\bar{v}_0-\lambda)^2\tilde{\tilde{l}}^2\right]. \end{aligned}$$

Following the Cardan method, the solutions are given by:

$$\begin{aligned} \tilde{\tilde{\Omega}}_0 &= \frac{1}{3}\left(\sqrt[3]{\frac{-27q+\sqrt{27(27q^2+4p^3)}}{2}} + \sqrt[3]{\frac{-27q-\sqrt{27(27q^2+4p^3)}}{2}}\right), \\ \tilde{\tilde{\Omega}}_- &= \frac{1}{3}\left(\mathbf{j}^2\sqrt[3]{\frac{-27q+\sqrt{27(27q^2+4p^3)}}{2}} + \mathbf{j}\sqrt[3]{\frac{-27q-\sqrt{27(27q^2+4p^3)}}{2}}\right), \\ \tilde{\tilde{\Omega}}_+ &= \frac{1}{3}\left(\mathbf{j}\sqrt[3]{\frac{-27q+\sqrt{27(27q^2+4p^3)}}{2}} + \mathbf{j}^2\sqrt[3]{\frac{-27q-\sqrt{27(27q^2+4p^3)}}{2}}\right), \end{aligned}$$

with $\mathbf{j} = -\frac{1}{2} + \mathbf{i}\frac{\sqrt{3}}{2}$.

Finally, $\tilde{\omega} = \sin(\omega\Delta t)/\Delta t$ is given by:

$$\begin{aligned} \tilde{\omega}_+ &= \tilde{k}\bar{u}_0 + \tilde{l}\bar{v}_0 - \frac{\mathbf{i}}{3}\left((\bar{u}_0+2\lambda)\tilde{\tilde{k}} + (\bar{v}_0+2\lambda)\tilde{\tilde{l}}\right) \\ &- \frac{1}{3}\left(\mathbf{j}\sqrt[3]{\frac{-27q+\sqrt{27(27q^2+4p^3)}}{2}} + \mathbf{j}^2\sqrt[3]{\frac{-27q-\sqrt{27(27q^2+4p^3)}}{2}}\right), \\ \tilde{\omega}_- &= \tilde{k}\bar{u}_0 + \tilde{l}\bar{v}_0 - \frac{\mathbf{i}}{3}\left((\bar{u}_0+2\lambda)\tilde{\tilde{k}} + (\bar{v}_0+2\lambda)\tilde{\tilde{l}}\right) \\ &- \frac{1}{3}\left(\mathbf{j}^2\sqrt[3]{\frac{-27q+\sqrt{27(27q^2+4p^3)}}{2}} + \mathbf{j}\sqrt[3]{\frac{-27q-\sqrt{27(27q^2+4p^3)}}{2}}\right), \\ \tilde{\omega}_0 &= \tilde{k}\bar{u}_0 + \tilde{l}\bar{v}_0 - \frac{\mathbf{i}}{3}\left((\bar{u}_0+2\lambda)\tilde{\tilde{k}} + (\bar{v}_0+2\lambda)\tilde{\tilde{l}}\right) \\ &- \frac{1}{3}\left(\sqrt[3]{\frac{-27q+\sqrt{27(27q^2+4p^3)}}{2}} + \sqrt[3]{\frac{-27q-\sqrt{27(27q^2+4p^3)}}{2}}\right). \end{aligned}$$

We obtain the wave frequencies ω_0, ω_- and ω_+ from:

$$\omega_0 = \frac{1}{\Delta t}\mathcal{R}e(\arcsin(\Delta t\tilde{\omega}_0)),\ \omega_\pm = \frac{1}{\Delta t}\mathcal{R}e(\arcsin(\Delta t\tilde{\omega}_\pm)).$$

where arcsin is the complex inverse trigonometric function.

Illustration: Inertia-gravity waves

In the case of inertia-gravity waves, the parameters $\bar{u}_0$ and $\bar{v}_0$ are equal to zero, hence the expressions of p and q become:

$$p = \frac{\lambda^2}{3}\left((\frac{\Delta x}{2}\tilde{\tilde{k}})^2 + (\frac{\Delta y}{2}\tilde{\tilde{l}})^2 + 5\frac{\Delta x}{2}\tilde{\tilde{k}}\frac{\Delta y}{2}\tilde{\tilde{l}}\right) - \lambda^2(\tilde{k}^2+\tilde{l}^2) + f^2,$$

and

$$q = -\frac{\lambda\Delta x\tilde{\tilde{k}}}{54}\left[18\lambda^2\tilde{k}^2 - 9\lambda^2\tilde{l}^2 - 9f^2 + 3\lambda^2(\frac{\Delta y}{2}\tilde{\tilde{l}})^2 + 2\lambda^2(\frac{\Delta x}{2}\tilde{\tilde{k}})^2\right],$$
$$-\frac{\lambda\Delta y\tilde{\tilde{l}}}{54}\left[18\lambda^2\tilde{l}^2 - 9\lambda^2\tilde{k}^2 - 9f^2 + 3\lambda^2(\frac{\Delta x}{2}\tilde{\tilde{k}})^2 + 2\lambda^2(\frac{\Delta y}{2}\tilde{\tilde{l}})^2\right].$$

The wave frequencies are given by:

$$\omega_0 = \frac{1}{\Delta t}\mathcal{R}e(\arcsin(\Delta t\tilde{\omega}_0)), \quad \omega_\pm = \frac{1}{\Delta t}\mathcal{R}e(\arcsin(\Delta t\tilde{\omega}_\pm)),$$

where

$$\tilde{\omega}_0 = -\frac{2\mathbf{i}\lambda}{3}\left(\tilde{\tilde{k}} + \tilde{\tilde{l}}\right) + \frac{\mathbf{i}}{3}\left(\sqrt[3]{\frac{-27q + \sqrt{27(27q^2+4p^3)}}{2}} + \sqrt[3]{\frac{-27q - \sqrt{27(27q^2+4p^3)}}{2}}\right),$$
$$\tilde{\omega}_+ = -\frac{2\mathbf{i}\lambda}{3}\left(\tilde{\tilde{k}} + \tilde{\tilde{l}}\right) + \frac{\mathbf{i}}{3}\left(\mathbf{j}\sqrt[3]{\frac{-27q + \sqrt{27(27q^2+4p^3)}}{2}} + \mathbf{j}^2\sqrt[3]{\frac{-27q - \sqrt{27(27q^2+4p^3)}}{2}}\right),$$
$$\tilde{\omega}_- = -\frac{2\mathbf{i}\lambda}{3}\left(\tilde{\tilde{k}} + \tilde{\tilde{l}}\right) + \frac{\mathbf{i}}{3}\left(\mathbf{j}^2\sqrt[3]{\frac{-27q + \sqrt{27(27q^2+4p^3)}}{2}} + \mathbf{j}\sqrt[3]{\frac{-27q - \sqrt{27(27q^2+4p^3)}}{2}}\right),$$

and where arcsin is again the complex inverse trigonometric function and $\mathbf{j} = -1/2 + \mathbf{i}\sqrt{3}/2$.

Figure 2.9 shows the discrete dispersion values computed with upwind fluxes. As the previous two schemes, this scheme is more precise for small wavenumbers (that is for long waves) while significant errors occur for medium and large wavenumbers. As before, in Figure 2.10 we represent the approximate velocity group and comparing to the exact velocity group represented in Figure 1.2 we see that significant angular and directional errors appear for the medium and small length waves.

As for the case of the scheme obtained by Euler time discretization with Lax-Wendroff fluxes, for the Leap-Frog time discretization with upwind fluxes the computations are too complicated in order to show the existence of spurious caustics. Nevertheless, the existence of this phenomena is seen numerically, by representing in Figure 2.11 the norm of the approximate group velocity as a function of $k\Delta x/\pi$. In this figure we can see that while the exact group velocity is not presenting local extrema, the approximate group velocity has very high local extrema when we work with small length waves, which is translated into sudden local error growths.

3. Conclusion

In this paper we have studied the errors introduced by some finite volume discretizations into the computation of the wave frequencies and of the group velocities. The linear models considered here are related to the Shallow Water problem, which is of much interest in geophysical fluid dynamics. It is well-known that the main source of the numerical errors are the dispersive and dissipative properties of the numerical schemes, both sources being underlined in our study for the following schemes: Leap-Frog time discretization with centered fluxes, Euler time discretization and Lax-Wendroff fluxes, Leap-Frog discretization and upwind fluxes.

In Section (2.1) we studied the behavior of the wave frequencies as a function of the wavenumbers for the Leap-Frog time discretization with centered fluxes scheme, as well as the dispersive phenomenon generated by the scheme. First, we noticed that, as for the schemes studied by Grotjahn and O'Brien, the scheme provides a

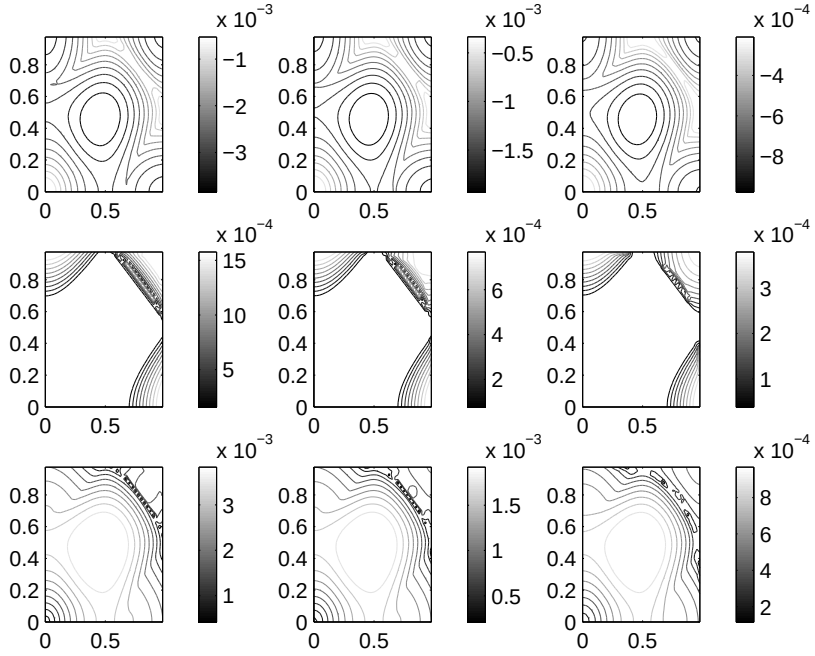

FIGURE 2.9. Inertia-gravity waves using Leap-Frog time discretization and upwind fluxes: ω_- (top), ω_0 (middle) and ω_+ (bottom) as functions of $k\Delta x/\pi$ and $l\Delta y/\pi$ (with $\Delta x = \Delta y$) for three values of $R_D/\Delta x$: 20, 10, 5. Due to the symmetry of the problem, only the upper right quadrant is shown. Compare to Figure 1.1.

good approximation for the small wavenumbers, while important errors appear for the large wavenumbers. We have also seen the dispersive property of the scheme, evidenced by the spurious caustics. This phenomenon may give rise to a sudden growth of the error, since we have spurious local energy maximum (the velocity group exhibits an extremum). The dispersive property evidenced here is a source of angular and magnitude errors in the computations of the group velocity for the numerical model. The existence of the spurious caustics is shown theoretically as well as numerically for the Leap-Frog discretization with centered fluxes (see Figure 2.1), while the Euler time discretization with Lax-Wendroff fluxes and for the Leap-Frog discretization with upwind fluxes the phenomenon was illustrated numerically (see Figures 2.8 and 2.11).

In Section (2.2) and (2.3), we studied the errors introduced by an Euler time-discretization with Lax-Wendroff fluxes scheme and the Leap-Frog time-discretization with upwind fluxes scheme. In both cases, we obtained complex discrete phase frequencies approximating real analytical frequencies. This is obviously a source of error, since the imaginary part of the phase frequency will generate a growth or a decay, and therefore large amplitude errors in the numerical schemes.

For the Leap-Frog method, the errors are also caused by the fact that the method proposed here is highly diffusive. In the case of the Lax-Wendroff fluxes scheme, we increased the order of the method, but the accuracy of the method is still not

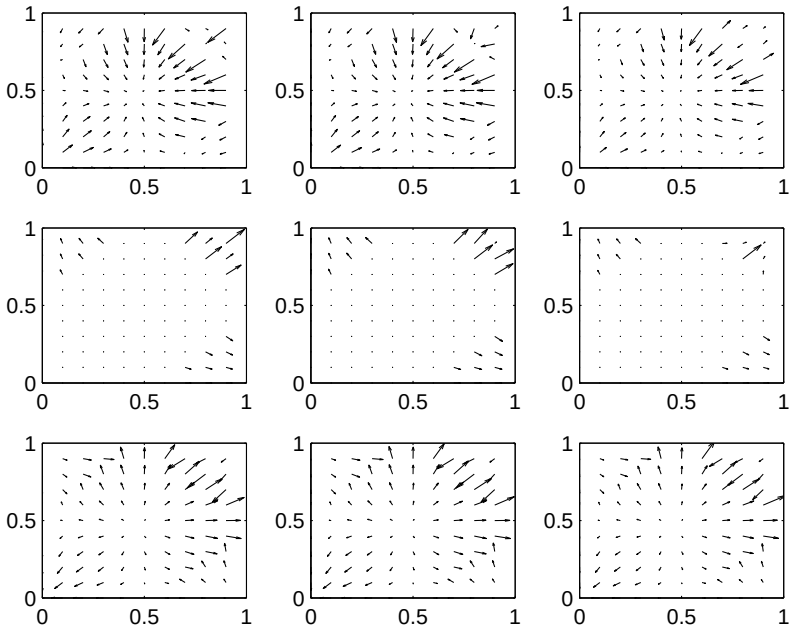

FIGURE 2.10. Inertia-gravity waves using Leap-Frog time discretization and upwind fluxes: $v_{g,-}$ (top), $v_{g,0}$ (middle) and $v_{g,+}$ (bottom) as functions of $k\Delta x/\pi$ and $l\Delta y/\pi$ (with $\Delta x = \Delta y$) for three values of $R_D/\Delta x$: 20, 10, 5. Due to the symmetry of the problem, only the upper right quadrant is shown. Compare to Figure 2.1.

increased; in fact comparing the results obtained for all the three proposed cases, this scheme is the least suitable for the study of our problem.

We conclude by noticing that for all the proposed schemes, for the large and medium wavenumbers (amplitude, angular and directional) important errors are introduced.

The relative errors introduced by all these schemes, for the first modes, are shown in Figures 2.4, 2.7, and 3.1. Taking a closer look at Figures 2.4, 2.7, and 3.1, where we represent the error exhibited by the schemes when we compute the first modes, we can see that the worst results are given by the Euler time discretization with Lax-Wendroff fluxes scheme. For this scheme, even for the first modes the error is considerably higher compared to the other schemes. The fact that the Lax-Wendroff fluxes scheme is the least adapted to the study of these problems is not surprising taking into account the diffusive character of this scheme. We can also see that all schemes are more precise for a larger $R_D/\Delta x$.

Finally we compare our work with the previous work of R. Grotjahn and J. O'Brien (1976), where the authors studied the error introduced by finite difference schemes for several linear hyperbolic systems used in meteorology and oceanography (starting from simple systems as the advection equation or the gravity waves, to more complex systems such as the shallow water model). The authors showed that the finite difference schemes produce dispersive waves even when the exact waves were

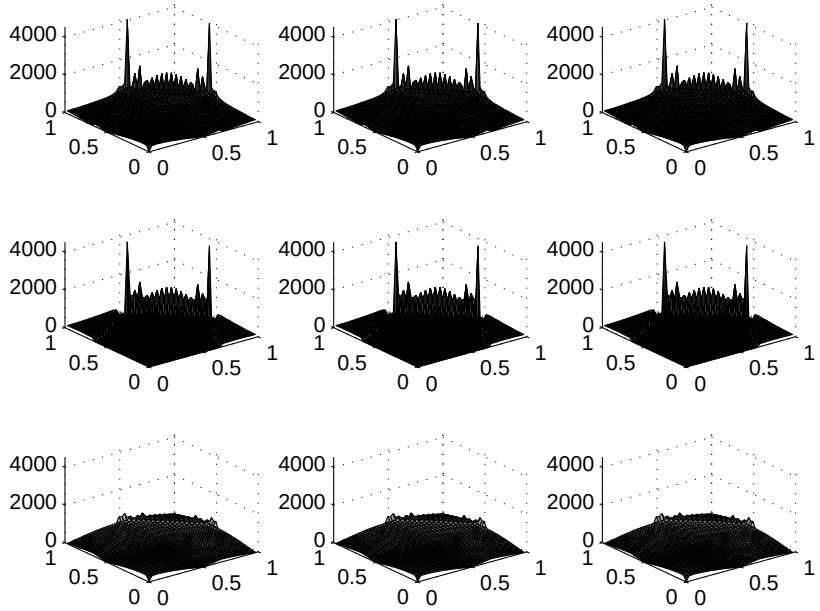

FIGURE 2.11. Inertia-gravity waves using Leap-Frog time discretization and upwind fluxes: $||v_{g,-}||_2$ (top), $||v_{g,0}||_2$ (middle) and $||v_{g,+}||_2$ (bottom) as functions of $k\Delta x/\pi$ and $l\Delta y/\pi$ (with $\Delta x = \Delta y$) for three values of $R_D/\Delta x$: 20, 10, 5. Due to the symmetry of the problem, only the upper right quadrant is shown.

not dispersive. The dispersive nature of the schemes introduces important angular and magnitude errors in the group velocity. It is also shown that the implicit schemes are underestimating the oscillations while the explicit schemes are overestimating them. The only waves for which the schemes provide a good approximation are the long waves. The similarity between our conclusions and the conclusions obtained in Grotjahn and O'Brien (1976) is not surprising since in both papers second order schemes are analyzed. Note that in Grotjahn and O'Brien (1976) the possible occurrence of caustics has not been discussed.

In our work, we examined some finite volume schemes typical for the study of hyperbolic problems, and we noticed that similar results as for the finite difference schemes are obtained. As mentioned before, the waves for which the schemes considered give a good approximation are the long waves. We conclude by underlining that we also evidenced in our article the appearance of spurious caustics for the finite volume methods considered here, and that the caustics form not only at grid limit but also for resolved waves. To the best of our knowledge, the possible occurrence of caustics has not been shown before. A natural continuation to this work, that we intend to address in the future, is the study of spurious caustics for high order schemes.

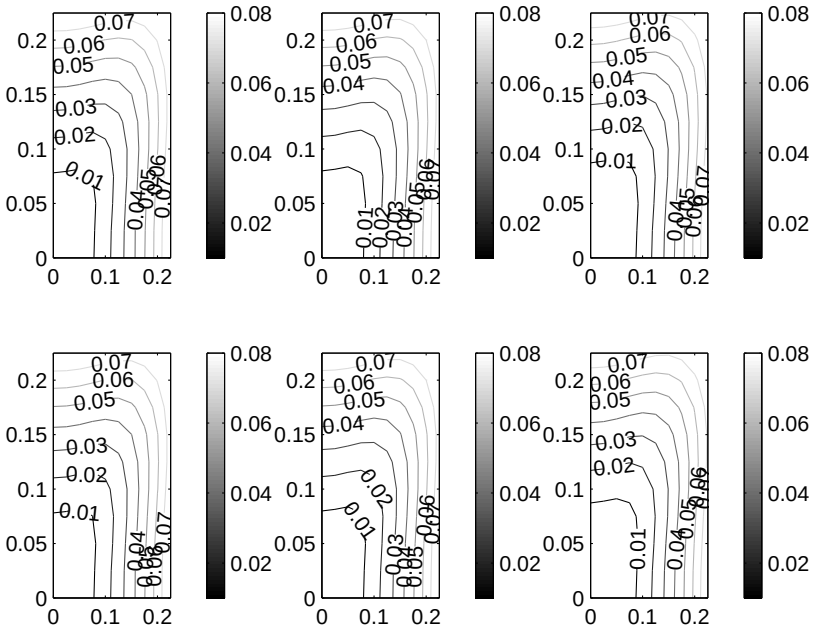

FIGURE 3.1. Inertia-gravity waves using Leap-Frog time discretization and upwind fluxes: **zoom on the rms error** for the first modes of ω_- (top) and ω_+ (bottom) as functions of $k\Delta x/\pi$ and $l\Delta y/\pi$ (with $\Delta x = \Delta y$) for three values of $R_D/\Delta x$: 20, 10, 5. Due to the symmetry of the problem, only the upper right quadrant is shown.

Acknowledgments

This work was partially supported by the National Science Foundation under the grant NSF-DMS-0906440, and by the Research Fund of Indiana University.

References

[Bla00] Eric Blayo. Compact finite difference schemes for ocean models. I. Ocean waves. *J. Comput. Phys.*, 164(2):241–257, 2000.

[DET08] Michael Dumbser, Cedric Enaux, and Eleuterio F. Toro. Finite volume schemes of very high order of accuracy for stiff hyperbolic balance laws. *J. Comput. Phys.*, 227(8):3971–4001, 2008.

[DS08] Claire David and Pierre Sagaut. Structural stability of the finite dispersion-relation preserving schemes. *Chaos, Solitons & Fractals*, 2008. in press.

[DSS06] Claire David, Pierre Sagaut, and Tapan Sengupta. A linear dispersive mechanism for numerical error growth: spurious caustics. arXiv:math.AP/0607407 v1, 2006.

[GO76] R. Grotjahn and J. O'Brien. Some inaccuracies in finite differencing hyperbolic equations. *Monthly weather review*, 104:180–194, 1976.

[LeV02] Randall J. LeVeque. *Finite volume methods for hyperbolic problems*. Cambridge Texts in Applied Mathematics. Cambridge University Press, Cambridge, 2002.

[Lig01] James Lighthill. *Waves in fluids*. Cambridge Mathematical Library. Cambridge University Press, Cambridge, 2001. Reprint of the 1978 original.

[MM96] José Ma. Martí and Ewald Müller. Extension of the piecewise parabolic method to one-dimensional relativistic hydrodynamics. *J. Comput. Phys.*, 123(1):1–14, 1996.

[Mor02] K. W. Morton. Discretization of unsteady hyperbolic conservation laws. *SIAM J. Numer. Anal.*, 39(5):1556–1597 (electronic), 2001/02.

[NXS07] Sebastian Noelle, Yulong Xing, and Chi-Wang Shu. High-order well-balanced finite volume WENO schemes for shallow water equation with moving water. *J. Comput. Phys.*, 226(1):29–58, 2007.

[PU07] M. V. Popov and S. D. Ustyugov. A piecewise-parabolic method on a local stencil for gas dynamic problems. *Zh. Vychisl. Mat. Mat. Fiz.*, 47(12):2055–2075, 2007.

[TC98] E. F. Toro and J. F. Clarke, editors. *Numerical methods for wave propagation*, volume 47 of *Fluid Mechanics and its Applications*. Kluwer Academic Publishers Group, Dordrecht, 1998. Selected papers from the International Workshop held in Manchester, May 1995.

[Tor01] E. F. Toro, editor. *Godunov methods*. Kluwer Academic/Plenum Publishers, New York, 2001. Theory and applications.

[Tor06] E. F. Toro. Riemann solvers with evolved initial conditions. *Internat. J. Numer. Methods Fluids*, 52(4):433–453, 2006.

[Vic87] Robert Vichnevetsky. Wave propagation analysis of difference schemes for hyperbolic equations: a review. *Internat. J. Numer. Methods Fluids*, 7(5):409–452, 1987.

[Whi99] G. B. Whitham. *Linear and nonlinear waves*. Pure and Applied Mathematics (New York). John Wiley & Sons Inc., New York, 1999. Reprint of the 1974 original, A Wiley-Interscience Publication.

[#] Laboratoire de Mathématiques, Université Paris-Sud, France

[♭] Laboratoire de Mathématiques et Applications, UMR 6086, Université de Poitiers, France

[♢] The Institute of Mathematics of the Romanian Academy, Bucharest, Romania

[♯]The Institute for Scientific Computing and Applied Mathematics, Indiana University, Bloomington, IN, USA

[†] National Center for Atmospheric Research, Boulder, USA