# Analysis of modified Godunov type schemes for the two-dimensional linear wave equation with Coriolis source term on cartesian meshes

Emmanuel Audusse, Do Minh Hieu, Pascal Omnes, Yohan Penel

# Analysis of modified Godunov type schemes for the two-dimensional linear wave equation with Coriolis source term on cartesian meshes

Emmanuel Audusse, Do Minh Hieu, Pascal Omnes, Yohan Penel

October 18, 2017

**Abstract**

The study deals with colocated Godunov type finite volume schemes applied to the two-dimensional linear wave equation with Coriolis source term. The purpose is to explain the wrong behaviour of the classical scheme and to modify it in order to avoid accuracy issues around the geostrophic equilibrium and in geostrophic adjustment processes. To do so, a Hodge-like decomposition is introduced. Then three different well-balanced strategies are introduced. Some properties of the associated modified equation are proven and then extended to the semi-discrete case. Stability of fully discrete schemes under a classical CFL condition is established thanks to a Von Neumann analysis. Some numerical results reinforce the purpose.

## Contents

# 1 Introduction

The primitive equations of the ocean are widely used to model oceanographic flows at global or regional scales, see [19] and [5, 15] for a mathematical study. The presence of specific source terms, in particular those accounting for Coriolis force and non trivial bathymetry, plays an important role and is not obvious to deal with in numerical simulations. A good model to study the impact of the discretisation of these source terms on the quality of numerical solutions is the shallow water system (1) presented below. It is simpler than the primitive equations, in particular due to the reduction of dimension from 3D to 2D, but still contains most of the issues raised by the presence of source terms. In this context, the discretisation of the topographic source term in a collocated finite volume framework has been dealt with in many works over the last two decades, see the reference book [6] or [2] for a more recent review. In the present work, we focus on the collocated finite volume discretisation of the Coriolis source term.

At large scales, many oceanographic flows are perturbations of the so-called geostrophic equilibrium (3) that corresponds to a balance between the pressure gradient and the Coriolis force. It follows that the accuracy of numerical methods is strongly related in many situations to their ability to maintain this balance. In the collocated finite volume framework, very few works were devoted to this question. In [7], see also [6,24], the authors propose to extend a technique originally developed for the topographic term in [1]. The resulting scheme is named the *Apparent Topography* method. It turns out to yield good results for one dimensional experiments. In [8] the technique is extended to higher order schemes and assessed on two-dimensional problems. One of the main results in the present work is to prove that the Apparent Topography method alone is not accurate around the two-dimensional geostrophic equilibrium and has to be supplemented by other developments. In [18] the authors propose an alternative method, namely the Finite Volume Evolution Galerkin (FVEG) method, but they also mainly consider the one-dimensional geostrophic equilibrium, *i.e.* when the two-dimensional velocity is function of only one space variable. Very recently [23], an RS-IMEX scheme (for Reference Solution IMplicit EXplicit scheme) was designed for shallow water equations with Coriolis force and proven to be asymptotically consistent with the Quasi-Geostrophic Equations. Here we only consider time discretisations that lead to explicit computations, *i.e.* with no linear systems to solve. As previously mentioned, the present work is also restricted to the collocated finite volume framework, we refer to [17, 20] for other approaches.

The velocity field associated with the geostrophic equilibrium (3) is obviously divergence free. This implies that our study will share important properties with works devoted to the study of low Mach number (for Euler equations) or low Froude number (for shallow water equations) regimes. The reader is referred to [9–11, 13, 14, 16, 21] where some accurate numerical schemes are proposed. In particular, we shall often refer to the framework introduced in [9].

In the present work we investigate the preservation of the geostrophic equilibrium in the context of the two-dimensional wave equation with Coriolis force (2), that is the linearised version of the shallow water equations. It generalises a study initiated in [3, 4] in the one dimensional context. In Section 2 we recall the main characteristics of the wave equation with Coriolis force. In Section 3 we show that the classical collocated finite volume Godunov scheme is not accurate in the vicinity of the geostrophic equilibrium. This inaccuracy is mainly related to the numerical diffusion terms that make the stationary states of the scheme not consistent with those of the continuous model. In Section 4, we show that the Apparent Topography (AT) method proposed in [7] and a Divergence Penalisation (DP) method mentioned in [9] can be used to cure the problem, provided they are combined together or to other Low Froude strategies inspired from [9–11]. For that, we analyse the modified equations related to the aforementioned corrections. In Section 5 we turn to the related semi-discrete (in space) analysis. In particular we construct discrete operators that possess mimetic properties that are proven to be necessary for accuracy, see also [20]. We also propose two consistent discretisations of the continuous steady states and we design the corresponding numerical schemes. Finally we exhibit the dispersion relations associated to the different numerical schemes and we prove that one of the proposed scheme is energy dissipative. In Section 6 we introduce the time discretisation and we prove the main result of this work which is that some of the proposed schemes are accurate and linearly stable under some CFL conditions. In Section 7 we illustrate the previous properties through some two dimensional numerical results.

## 2 Properties of the linear wave equation with Coriolis source term in 2D

In order to study the dimensionless shallow water equation in the rotating frame

$$\begin{cases} \mathsf{St}\,\partial_t h + \nabla\cdot(h\bar{\mathbf{u}}) = 0, \\ \mathsf{St}\,\partial_t(h\bar{\mathbf{u}}) + \nabla\cdot(h\bar{\mathbf{u}}\otimes\bar{\mathbf{u}}) + \dfrac{1}{\mathsf{Fr}^2}\nabla\left(\dfrac{h^2}{2}\right) = -\dfrac{1}{\mathsf{Fr}^2}h\nabla b - \dfrac{1}{\mathsf{Ro}}h\bar{\mathbf{u}}^\perp, \end{cases} \tag{1}$$

we focus on the linear wave equation with Coriolis source term

$$\begin{cases} \partial_t r + a_\star \nabla\cdot\boldsymbol{u} = 0 \\ \partial_t \boldsymbol{u} + a_\star \nabla r = -\omega \boldsymbol{u}^\perp \end{cases} \quad \Longleftrightarrow \quad \partial_t q + L_w q = 0 \tag{2}$$

with $\boldsymbol{u} = (u,v)^T$, $\boldsymbol{u}^\perp = (-v,u)^T$, $q = (r,u,v)$ and

$$L_\omega q = \begin{pmatrix} a_\star \nabla\cdot\boldsymbol{u} \\ a_\star\nabla r + \omega\boldsymbol{u}^\perp \end{pmatrix}.$$

In the sequel, we assume that $a_* > 0$ is a constant.

System (2) is obtained from the shallow water equation (1) when the Froude number ($\mathsf{Fr}$) and the Rossby number ($\mathsf{Ro}$) are of order $\mathcal{O}(M)$ and the Strouhal number ($\mathsf{St}$) is of order $\mathcal{O}(M^{-1})$, *i.e.* for short times, for a small parameter $M \ll 1$, and $b \equiv cte$.

To begin with, let us introduce the Hilbert space

$$\left(L^2(\mathbb{T}^2)\right)^3 = \left\{ q=(r,u,v) \;\middle|\; \int_{\mathbb{T}^2} r^2 \;\mathrm{d}\boldsymbol{x} + \int_{\mathbb{T}^2} \left(u^2+v^2\right) \;\mathrm{d}\boldsymbol{x} < \infty \right\}$$

equipped with the scalar product

$$\langle q_1, q_2\rangle = \int_{\mathbb{T}^2} r_1 r_2 \;\mathrm{d}\boldsymbol{x} + \int_{\mathbb{T}^2} (u_1u_2 + v_1v_2) \;\mathrm{d}\boldsymbol{x}.$$

## 2.1 Structure of the kernel of the original model

Since the preservation of steady-states of (2) is crucial in the design of accurate numerical schemes, especially in the limit $M \to 0$, we recall some well-known results about those steady-states. Let us define the kernel of the linear operator $L_\omega$ as

$$\mathcal{E}_{\omega\neq 0} := \ker L_{\omega\neq 0} = \left\{ (r, \boldsymbol{u}) \in H^1(\mathbb{T}^2) \times \left(L^2(\mathbb{T}^2)\right)^2 \;\middle|\; a_\star \nabla r = -\omega \boldsymbol{u}^\perp \right\}. \tag{3}$$

Since $\omega\, \boldsymbol{u} = a_\star (\nabla r)^\perp$ implies that $\nabla \cdot \boldsymbol{u} = 0$, being a steady-state of (2) is equivalent to belonging to $\mathcal{E}_{\omega\neq 0}$. Let us mention that $\mathcal{E}_{\omega=0}$ is named the incompressible subspace (see [9] for more details). We shall keep the same terminology in the present work. As we shall see later on, the orthogonal space of the kernel plays an important role in the analysis of the behaviour of numerical schemes. Hence the following statement:

**Proposition 1.** *The orthogonal space of $\mathcal{E}_{\omega\neq 0}$ is given by*

$$\mathcal{E}_{\omega\neq 0}^\perp = \left\{ (p, \boldsymbol{v}) \in L^2(\mathbb{T}^2) \times \boldsymbol{H}(\mathrm{curl}, \mathbb{T}^2) \;\middle|\; \omega p = a_\star \nabla \times \boldsymbol{v} \right\}, \tag{4}$$

*where $\nabla \times \boldsymbol{u} := \partial_x u_y - \partial_y u_x$ and $\boldsymbol{H}(\mathrm{curl}, \mathbb{T}^2) := \left\{ \boldsymbol{u} \in \left(L^2(\mathbb{T}^2)\right)^2 \;\middle|\; \nabla \times \boldsymbol{u} \in L^2(\mathbb{T}^2) \right\}$.*

*Moreover, we have $\mathcal{E}_{\omega\neq 0} \oplus \mathcal{E}_{\omega\neq 0}^\perp = \left(L^2(\mathbb{T}^2)\right)^3$. In other words, any $q \in \left(L^2(\mathbb{T}^2)\right)^3$ can be decomposed into*

$$q = \hat{q} + \tilde{q}$$

*where $\hat{q} \in \mathcal{E}_{\omega\neq 0}$ and $\tilde{q} \in \mathcal{E}_{\omega\neq 0}^\perp$ and this decomposition is unique.*

The proof of this proposition can be found in Appendix A.

**Remark 1.** *The periodic boundary condition implies that for all elements $\tilde{q} = (p, \boldsymbol{v}) \in \mathcal{E}_{\omega\neq 0}^\perp$, we have*

$$p \in L_0^2(\mathbb{T}^2) := \left\{ f \in L^2(\mathbb{T}^2) \;\middle|\; \int_{\mathbb{T}^2} f \,\mathrm{d}\boldsymbol{x} = 0 \right\}.$$

## 2.2 Energy conservation

Let us define the energy as $E = \langle q, q \rangle$.

**Proposition 2.** *Let $q$ be a solution of System (2) on $\mathbb{T}^2$. Then, the energy is preserved*

$$E(t > 0) = E(t = 0).$$

*Proof.* To compute the energy estimate associated to System (2), we directly obtain

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\langle q, q \rangle = -\langle L_\omega q, q \rangle = 0$$

since $L_\omega$ is antisymmetric. □

**Remark 2.** *Energy conservation and linearity imply uniqueness of the solution of System (2).*

## 2.3 Behaviour of solutions

**Proposition 3.** *Let $q$ be the solution of System (2) with initial condition $q^0(\boldsymbol{x})$. Then:*

*i.* $\forall\, q^0(\boldsymbol{x}) \in \mathcal{E}_{\omega\neq 0}$, *we have* $q(t > 0, \boldsymbol{x}) = q^0(\boldsymbol{x}) \in \mathcal{E}_{\omega\neq 0}$.

*ii.* $\forall\, q^0(\boldsymbol{x}) \in \mathcal{E}_{\omega\neq 0}^\perp$, *we have* $q(t > 0, \boldsymbol{x}) \in \mathcal{E}_{\omega\neq 0}^\perp$.

*Proof.* Any initial condition $q^0 = (r^0, u^0, v^0) \in \mathcal{E}_{\omega\neq 0}$ is obviously a solution of (2); by the uniqueness property mentioned above, Point *i.* is proven.

As far as Point *ii.* is concerned, we consider $q^0 \in \mathcal{E}^{\perp}_{\omega\neq 0}$. For all $\hat{q} = (\hat{r}, \hat{\boldsymbol{u}})$ belonging to the kernel $\mathcal{E}_{\omega\neq 0}$, due to periodic boundary conditions, we obtain

$$\begin{aligned}\left\langle \frac{\mathrm{d}}{\mathrm{d}t} q, \hat{q} \right\rangle &= -a_\star \int_{\mathbb{T}^2} \hat{r} \nabla \cdot \boldsymbol{u} \, \mathrm{d}\boldsymbol{x} - a_\star \int_{\mathbb{T}^2} \nabla r \cdot \hat{\boldsymbol{u}} \, \mathrm{d}\boldsymbol{x} - \omega \int_{\mathbb{T}^2} \boldsymbol{u}^\perp \cdot \hat{\boldsymbol{u}} \, \mathrm{d}\boldsymbol{x} \\ &= \int_{\mathbb{T}^2} \boldsymbol{u} \cdot \left( a_\star \nabla \hat{r} + \omega \hat{\boldsymbol{u}}^\perp \right) \, \mathrm{d}\boldsymbol{x} + a_\star \int_{\mathbb{T}^2} r \nabla \cdot \hat{\boldsymbol{u}} \, \mathrm{d}\boldsymbol{x} = 0\end{aligned}$$

which implies that

$$\forall\ \hat{q} \in \mathcal{E}_{\omega\neq 0},\ \frac{\mathrm{d}}{\mathrm{d}t} \langle q, \hat{q} \rangle = 0 \implies \langle q(t,\cdot), \hat{q} \rangle = \langle q(t=0,\cdot), \hat{q} \rangle = 0$$

that is to say

$$q(t,\cdot) \in \mathcal{E}^{\perp}_{\omega\neq 0}.$$

This proves Point *ii*. □

**Corollary 1.** *Let $q$ be the solution of System (2) with initial condition $q^0$. Let $\mathbb{P}q^0$ be the orthogonal projection of $q^0$ onto the incompressible subspace $\mathcal{E}_{\omega\neq 0}$. Then, $q$ can be decomposed into*

$$q(t,\cdot) = \mathbb{P}q^0 + \tilde{q}(t,\cdot) \in \mathcal{E}_{\omega\neq 0} \oplus \mathcal{E}^{\perp}_{\omega\neq 0}$$

*where $\tilde{q}$ is the solution of System (2) with initial condition $(q^0 - \mathbb{P}q^0)$.*

*Moreover, the conservation of the energy for $\tilde{q}$ implies that for all times $t > 0$, $\|q(t,\cdot) - \mathbb{P}q^0\| = \|q^0 - \mathbb{P}q^0\|$ which allows to say that*

$$\|q^0 - \mathbb{P}q^0\| = \mathcal{O}(M) \implies \forall t > 0, \|q - \mathbb{P}q^0\|(t) = \mathcal{O}(M). \tag{5}$$

*In other words, when the initial condition $q^0$ is close to the incompressible subspace $\mathcal{E}_{\omega\neq 0}$, the solution of the linear wave equation (2) is still close to the projection of the initial condition onto $\mathcal{E}_{\omega\neq 0}$.*

One of the problems encountered with the Godunov scheme applied to (2) is that it does not reproduce this closeness to the projection of the initial condition on $\mathcal{E}_{\omega\neq 0}$ for values of $M \ll 1$. This inaccurate behaviour is explained in the next section. A numerical scheme for which the solution satisfies relation (5) will be said *accurate at low Froude number at any time*, as defined in [3].

# 3 Inaccuracy of the classical Godunov scheme

## 3.1 Numerical highlighting

We consider a cartesian mesh with mesh sizes $\Delta x$ (*resp.* $\Delta y$) in the $x$ (*resp.* $y$) direction. The semi-discrete Godunov scheme applied to the linear wave equation (2) can be written

$$\begin{cases} \frac{\mathrm{d}}{\mathrm{d}t} r_{i,j} + a_\star \left( \frac{u_{i+1,j} - u_{i-1,j}}{2\Delta x} + \frac{v_{i,j+1} - v_{i,j-1}}{2\Delta y} \right) - \frac{\kappa_r a_\star}{2} \left( \frac{r_{i+1,j} - 2r_{i,j} + r_{i-1,j}}{\Delta x} + \frac{r_{i,j+1} - 2r_{i,j} + r_{i,j-1}}{\Delta y} \right) = 0, \\ \frac{\mathrm{d}}{\mathrm{d}t} u_{i,j} + a_\star \frac{r_{i+1,j} - r_{i-1,j}}{2\Delta x} - \frac{\kappa_u a_\star}{2} \frac{u_{i+1,j} - 2u_{i,j} + u_{i-1,j}}{\Delta x} = \omega v_{i,j}, \\ \frac{\mathrm{d}}{\mathrm{d}t} v_{i,j} + a_\star \frac{r_{i,j+1} - r_{i,j-1}}{2\Delta y} - \frac{\kappa_v a_\star}{2} \frac{v_{i,j+1} - 2v_{i,j} + v_{i,j-1}}{\Delta y} = -\omega u_{i,j}, \end{cases} \tag{6}$$

where parameters $\kappa_r$, $\kappa_u$ and $\kappa_v$ lie in $[0,1]$ and represent the standard numerical diffusion of the Godunov type schemes. The classical Godunov scheme corresponds to $\kappa_r = \kappa_u = \kappa_v = 1$. The following facts are now well-known:

- In the 1D case, the classical Godunov scheme applied to the homogeneous system (*i.e.* with no Coriolis source term) is accurate for low $M$. It is no more the case when the Coriolis force is involved, see [3, 4]. Indeed, in that case the diffusion on the pressure equation is shown to be responsible for the inaccuracy. A simple correction consists in setting $\kappa_r = 0$ and is proven to be a stable strategy in [3]. This scheme is referred to in the sequel as the LF-C strategy, since we adapt the diffusion in the pressure equation to the Low-Froude (LF) case and we keep the classical (C) diffusion on the velocity equation.
- In the 2D case, the classical Godunov scheme applied to the homogeneous system (*i.e.* with no Coriolis source term) is inaccurate for low $M$ on cartesian meshes. This is due to the numerical viscosity on the velocity equation, see [9, 11] for more details. It is corrected in [9] by setting $\kappa_u = \kappa_v = 0$ to obtain a stable and accurate scheme. This scheme is referred to in the sequel as the C-LF strategy, since we keep the classical (C) diffusion on the pressure equation and we adapt the diffusion in the velocity equation to the Low-Froude (LF) case.

The purpose here is to show that none of these corrections (neither the LF-C nor the C-LF strategies) cures the inaccuracy of the Godunov scheme applied to the 2D wave equation with a Coriolis source term for low values of $M$. To do so, we consider the classical Godunov scheme and the modified versions proposed in [3, 9] when the initial condition is at the geostrophic equilibrium (see Fig. 1)

$$\begin{cases} r(t=0,x,y) = 1 - \exp\left[-\left(\frac{3x}{0.5}\right)^2 - \left(\frac{3y}{0.5}\right)^2\right] \\ u(t=0,x,y) = -\frac{6y}{0.5}\exp\left[-\left(\frac{3x}{0.5}\right)^2 - \left(\frac{3y}{0.5}\right)^2\right] \\ v(t=0,x,y) = \frac{6x}{0.5}\exp\left[-\left(\frac{3x}{0.5}\right)^2 - \left(\frac{3y}{0.5}\right)^2\right]. \end{cases} \tag{7}$$

in the periodic domain $\mathbb{T}^2 = [-0.5, 0.5] \times [-0.5, 0.5]$. This initial condition is obviously a steady state of System (2).

In Figures 2 and 3, we present the numerical results for a $50 \times 50$ grid at time $t = 10$. It indicates that the Godunov type schemes with standard diffusion (Fig. 2(b)), and both corrected LF-C (Fig. 2(c)) and C-LF (Fig. 2(d)) schemes are unable to capture the steady state. At the same time, it is not possible to use a LF-LF strategy and completely delete both diffusion terms (*i.e.* using $\kappa_r = \kappa_u = \kappa_v = 0$), because the resulting fully discrete explicit scheme would then obviously be unstable. Note that for this test, the modification on the diffusion in the velocity equation provides more substantial improvements than the modification on the diffusion in the pressure equation: in the first case, the 2D structure of the solution is more or less preserved, see Figure 2, and the solution remains not so far from the exact one, see Figure 3, whereas in the second case, the solution is quite close to the one of the classical scheme. Nevertheless, this behaviour is related to this particular test case and we need more investigations to obtain numerical schemes which are able to exactly preserve steady states and then be accurate in any situation.

Before doing that, let us analyze the discrete kernel of the semi-discrete Godunov scheme in order to point out the main reason of the inaccuracy problem.

### 3.2 Analysis of the discrete kernel

Let us denote by $L_{\omega,\kappa,h}$ the spatial operator in the semi-discrete scheme (6), so that (6) reads $q'_{i,j} + L_{\omega,\kappa,h} q_{i,j} = 0$.

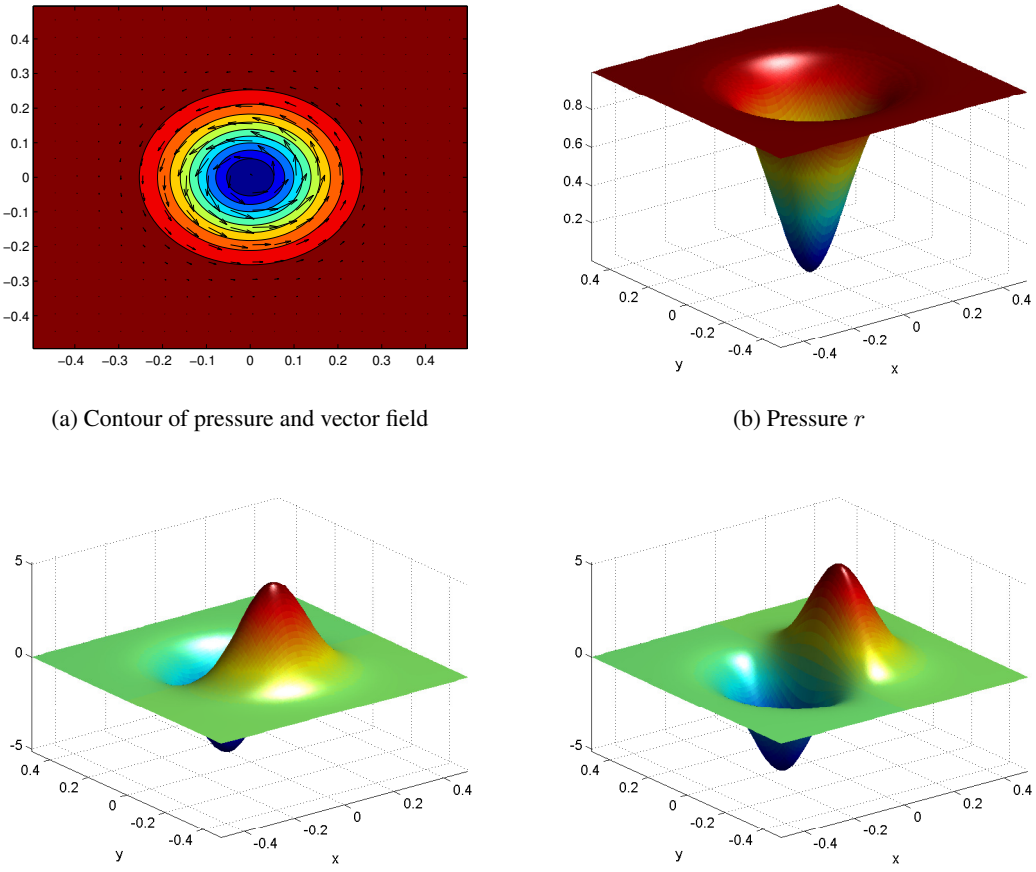
**Lemma 1.** *Let us define the discrete energy of System (6) by*

$$E_h(t) = \Delta x \Delta y \left[\sum_{i,j} r_{i,j}(t)^2 + \sum_{i,j} u_{i,j}(t)^2 + \sum_{i,j} v_{i,j}(t)^2\right].$$

*Then for any $\kappa_{r,u,v} \in [0,1]$*

$$\frac{\mathrm{d}}{\mathrm{d}t} E_h(t) \leq 0,$$

*which means that the energy associated to Godunov type schemes is dissipated.*



(a) Contour of pressure and vector field

(b) Pressure $r$

(c) Horizontal velocity $u$

(d) Vertical velocity $v$

Figure 1: Initial condition: stationary vortex.

*Proof.* Let us multiply the semi-discrete scheme (6) by $q_{i,j}\Delta x\Delta y$ and sum over all cells $(i,j)$. Due to periodic boundary conditions, we obtain by standard calculations

$$\left\langle \frac{\mathrm{d}q}{\mathrm{d}t}, q \right\rangle = -\frac{\kappa_r a_\star \Delta y}{2}\sum_{i,j}(r_{i+1,j}-r_{i,j})^2 - \frac{\kappa_r a_\star \Delta x}{2}\sum_{i,j}(r_{i,j+1}-r_{i,j})^2 \\ - \frac{\kappa_u a_\star \Delta y}{2}\sum_{i,j}(u_{i+1,j}-u_{i,j})^2 - \frac{\kappa_v a_\star \Delta x}{2}\sum_{i,j}(v_{i,j+1}-v_{i,j})^2 \leq 0. \quad (8)$$

□

Although the semi-discrete scheme is energy-dissipative, the fact still remains that it is not consistent with the incompressible space.

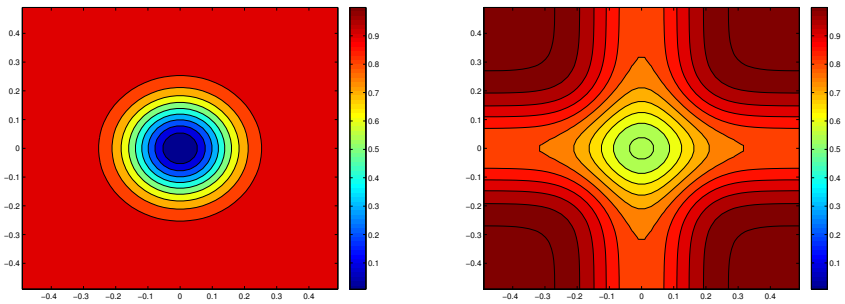
**Lemma 2.**

- *For $\kappa_r \neq 0$, $\kappa_u \neq 0$ and $\kappa_v \neq 0$, i.e. for the classical Godunov scheme, we have*

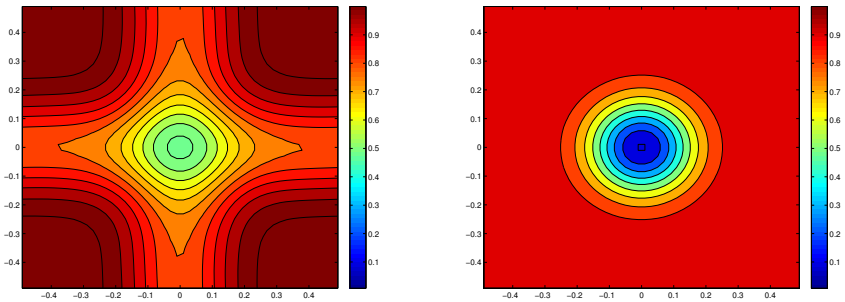
$$\ker L_{\omega,\kappa,h} = \left\{ q=(r,u,v)\in\mathbb{R}^{3N} \;\middle|\; \exists\, c\in\mathbb{R},\ r=c, u=0, v=0 \right\} \quad (9a)$$

- *For $\kappa_r \neq 0$, $\kappa_u = \kappa_v = 0$, i.e. for the C-LF strategy, we have*

$$\ker L_{\omega,\kappa,h} = \left\{ q=(r,u,v)\in\mathbb{R}^{3N} \;\middle|\; \exists\, c\in\mathbb{R},\ r=c, u=0, v=0 \right\} \quad (9b)$$



(a) Initial condition (b) Classical Godunov scheme



(c) LF-C scheme (d) C-LF scheme

Figure 2: Contours of the pressure $r$.

- *For $\kappa_r = 0$, $\kappa_u \neq 0$ and $\kappa_v \neq 0$, i.e. for the LF-C strategy, we have*

$$\ker L_{\omega,\kappa,h} = \left\{ q = (r,u,v) \in \mathbb{R}^{3N} \;\middle|\; \exists (u_j, v_i) \in \mathbb{R}^N \times \mathbb{R}^N \,,\, \forall\, (i,j),\; u_{i,j} = u_j,\; v_{i,j} = v_i, \right.$$
$$\left. a_\star \begin{pmatrix} \frac{r_{i+1,j} - r_{i-1,j}}{2\Delta x} \\ \frac{r_{i,j+1} - r_{i,j-1}}{2\Delta y} \end{pmatrix} = -\omega \begin{pmatrix} -v_i \\ u_j \end{pmatrix} \right\} \tag{9c}$$

**Remark 3.** *Although all kernels above correspond to discrete versions of the exact relation $a_* \nabla r = -\omega \boldsymbol{u}^\perp$, constraints upon the velocity field are too strong, so that those kernels do not match with the exact one $\mathcal{E}_{\omega \neq 0}$ defined in (3).*

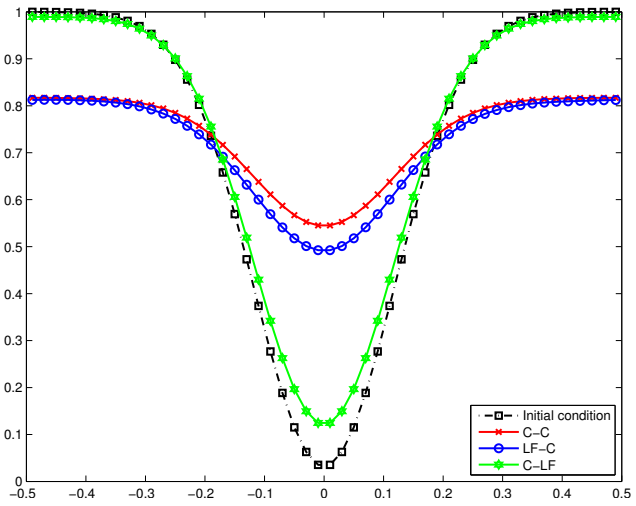



Figure 3: Cross-section ($y = 0$) of the pressure $r$.

*Proof.* Since any steady state of (6) satisfy $\frac{\mathrm{d}q_{i,j}}{\mathrm{d}t} = 0$, Equation (8) implies

$$\sum_{i,j} \left[\kappa_r \left(\Delta y (r_{i+1,j} - r_{i,j})^2 + \Delta x (r_{i,j+1} - r_{i,j})^2\right) + \kappa_u \Delta y (u_{i+1,j} - u_{i,j})^2 + \kappa_v \Delta x (v_{i,j+1} - v_{i,j})^2\right] = 0. \quad (10)$$

- When $\kappa_r \neq 0$, we easily get from (10)

$$\forall\ (i,j) \in [1, N_x] \times [1, N_y],\ r_{i+1,j} = r_{i,j} \text{ and } r_{i,j+1} = r_{i,j} \implies r = \text{const}. \quad (11)$$

  When $\kappa_u \neq 0$ and $\kappa_v \neq 0$, we also have $u_{i+1,j} = u_{i,j}$ and $v_{i,j+1} = v_{i,j}$ for all $(i,j)$, which implies that there exist $(u_j, v_i) \in \mathbb{R}^N \times \mathbb{R}^N$ such that

$$u_{i,j} = u_j \text{ and } v_{i,j} = v_i \ \ \forall (i,j).$$

  Going back to (6), $r = const$ implies that $u = 0$ and $v = 0$. Therefore, this leads to (9a).

- Likewise, when $\kappa_r \neq 0$ but $\kappa_u = \kappa_v = 0$, (11) still holds. Then we deduce from (6) that $u_{i,j} = v_{i,j} = 0$ and consequently (9b).

- Now, we consider the case $\kappa_r = 0$ (and $\kappa_u \neq 0$, $\kappa_v \neq 0$). We deduce from (10) that $u_{i,j} = u_j$ and $v_{i,j} = v_i$. Hence, the steady version of (6) now reads

$$\frac{a_\star}{2\Delta x}(r_{i+1,j} - r_{i-1,j}) = \omega v_i \quad \text{and} \quad \frac{a_\star}{2\Delta y}(r_{i,j+1} - r_{i,j-1}) = -\omega u_j$$

  which is nothing but (9c).

□

# 4 Properties of the first order modified equation with correction terms

In the previous section, we have shown that the classical Godunov scheme is inaccurate near the geostrophic equilibrium and that simple corrections consisting in deleting diffusion terms (LF-C or C-LF strategies) are not enough to ensure the accuracy. As it is not possible to delete all diffusion terms at the same time for stability reasons, it is essential to introduce some correction terms for the standard diffusion.

We aim at deriving a numerical scheme which is able to preserve steady states or to be accurate around steady states. It is worth pointing out that we not only have to deal with the balance between the pressure gradient and the Coriolis force but also to take into account the divergence free condition.

## 4.1 Definition of the schemes

We mention below two strategies to deal with diffusion terms. Each strategy leads to a different numerical scheme which will be referred to as X-Y scheme where X is related to the diffusion on the pressure equation and Y to the diffusion on the velocity equation.

i. The *Apparent Topography* scheme was introduced in [7], see also [8], to deal with the geostrophic equilibrium in the 1D nonlinear shallow water system. This strategy was proven to be stable in [4] for the 1D linear wave equation. Here we extend the procedure to the 2D linear wave equation. We notice that the steady state defined by $a_\star \nabla r = -\omega \boldsymbol{u}^\perp$ also satisfies

$$\nabla \cdot \left(\nabla r + \frac{\omega}{a_\star} \boldsymbol{u}^\perp\right) = 0.$$

It suggests that the numerical diffusion on the pressure equation can be modified into $\nabla \cdot (\nabla r + \frac{\omega}{a_\star} \boldsymbol{u}^\perp)$ instead of the classical operator $\Delta r$ – see (6). As for the velocity equations, either we keep the classical diffusion to obtain the *Apparent Topography-Classical scheme* (AT-C) or we delete diffusion terms which leads to the *Apparent Topography-Low Froude scheme* (AT-LF). In 1D, they are shown to be both stable and accurate [4].

| Schemes | $\kappa_r$ | $\kappa_u$ | $\kappa_v$ | $\eta_r$ | $\eta_u$ | $\eta_v$ |
|---|---|---|---|---|---|---|
| *AT-LF* | $\mathcal{O}(1)$ | 0 | 0 | $\kappa_r$ | 0 | 0 |
| *AT-C* | $\mathcal{O}(1)$ | $\mathcal{O}(1)$ | $\mathcal{O}(1)$ | $\kappa_r$ | 0 | 0 |
| *LF-DP* | 0 | $\mathcal{O}(1)$ | $\mathcal{O}(1)$ | 0 | $\kappa_u$ | $\kappa_v$ |
| *C-DP* | $\mathcal{O}(1)$ | $\mathcal{O}(1)$ | $\mathcal{O}(1)$ | 0 | $\kappa_u$ | $\kappa_v$ |
| *AT-DP* | $\mathcal{O}(1)$ | $\mathcal{O}(1)$ | $\mathcal{O}(1)$ | $\kappa_r$ | $\kappa_u$ | $\kappa_v$ |

Table 1: Parameters of Godunov type schemes with corrections.

ii. The *Divergence Penalisation* method consists in a modification on the diffusion on the velocity equation that is based on the operator $\nabla(\nabla\cdot\boldsymbol{u})$ instead of the classical diffusion in (6) since the equilibrium states satisfy $\nabla\cdot\boldsymbol{u}=0$. This idea was mentioned in [9, § 5.6] to be applied to the homogeneous linear wave equation, but not studied. We propose to extend it to the present case and to analyze its properties. As for the pressure equation, we can choose the classical diffusion to obtain the *Classical-Divergence Penalisation* scheme (C-DP) or delete this diffusion term to get the *Low Froude-Divergence Penalisation* scheme (LF-DP).

iii. Finally, we can combine both strategies to get the *Apparent Topography-Divergence Penalisation* method (AT-DP). It comes down to considering $\nabla\cdot(\nabla r+\frac{\omega}{a_\star}\boldsymbol{u}^\perp)$ for the diffusion on the pressure equation and $\nabla(\nabla\cdot\boldsymbol{u})$ for the velocity equation.

We shall prove below that the AT-LF, LF-DP and AT-DP approaches are accurate and stable. The AT-C and C-DP strategies, like the LF-C and C-LF ones previously mentioned in paragraph3.1, will not be able to cure the problem since only one issue is taken into account.

To carry out the accuracy analysis, we shall analyze the first order modified equation which is the common tool to study stability and accuracy of finite difference schemes. We refer to [22] for more details about this method. The first order modified equation corresponding to the aforementioned strategies is given by

$$\begin{cases}\partial_t r+a_\star(\partial_x u+\partial_y v)-\dfrac{\kappa_r^x a_\star\Delta x}{2}\dfrac{\partial^2 r}{\partial x^2}-\dfrac{\kappa_r^y a_\star\Delta y}{2}\dfrac{\partial^2 r}{\partial y^2}+\dfrac{\eta_r^x a_\star\Delta x}{2}\dfrac{\omega}{a_\star}\dfrac{\partial v}{\partial x}-\dfrac{\eta_r^y a_\star\Delta y}{2}\dfrac{\omega}{a_\star}\dfrac{\partial u}{\partial y}=0,\\ \partial_t u+a_\star\partial_x r-\dfrac{\kappa_u a_\star\Delta x}{2}\dfrac{\partial^2 u}{\partial x^2}-\dfrac{\eta_u a_\star\Delta x}{2}\dfrac{\partial^2 v}{\partial x\partial y}=\omega v,\\ \partial_t v+a_\star\partial_y r-\dfrac{\kappa_v a_\star\Delta y}{2}\dfrac{\partial^2 v}{\partial y^2}-\dfrac{\eta_v a_\star\Delta y}{2}\dfrac{\partial^2 u}{\partial y\partial x}=-\omega u,\end{cases}\tag{12}$$

where parameters $\eta_r^x\geq 0$, $\eta_r^y\geq 0$, $\eta_u\geq 0$ and $\eta_v\geq 0$ stand for the correction terms. We recall that $\kappa_{r,u,v}^{x,y}\in[0,1]$. The modified equation reads in a compact form

$$\begin{cases}\partial_t q+\mathcal{L}q=0,&\text{(13a)}\\ q(t=0,\boldsymbol{x})=q^0(\boldsymbol{x}).&\text{(13b)}\end{cases}$$

The spatial operator is defined by $\mathcal{L}=L_\omega-\mathcal{B}_{\kappa,\eta}$, with $L_\omega$ as in (2) and

$$\mathcal{B}_{\kappa,\eta}q=\begin{pmatrix}\frac{\kappa_r^x a_\star\Delta x}{2}\frac{\partial^2 r}{\partial x^2}+\frac{\kappa_r^y a_\star\Delta y}{2}\frac{\partial^2 r}{\partial y^2}\\ \frac{\kappa_u a_\star\Delta x}{2}\frac{\partial^2 u}{\partial x^2}\\ \frac{\kappa_v a_\star\Delta y}{2}\frac{\partial^2 v}{\partial y^2}\end{pmatrix}+\begin{pmatrix}-\frac{\eta_r^x a_\star\Delta x}{2}\frac{\omega}{a_\star}\frac{\partial v}{\partial x}+\frac{\eta_r^y a_\star\Delta y}{2}\frac{\omega}{a_\star}\frac{\partial u}{\partial y}\\ \frac{\eta_u a_\star\Delta x}{2}\frac{\partial^2 v}{\partial x\partial y}\\ \frac{\eta_v a_\star\Delta y}{2}\frac{\partial^2 u}{\partial y\partial x}\end{pmatrix}.$$

The choices of parameters in (12-13) corresponding to the aforementioned strategies are summarised in Table 1.

**Remark 4.** *For the numerical diffusion on the velocity equation to be consistent with $\nabla(\nabla\cdot\boldsymbol{u})$, we have to take $\kappa_u\Delta x=\kappa_v\Delta y$ and $\eta_u=\kappa_u$, $\eta_v=\kappa_v$. Likewise, to recover the operator $\nabla\cdot(\nabla r+\frac{\omega}{a_\star}\boldsymbol{u}^\perp)$, we must consider the case $\kappa_r^x\Delta x=\kappa_r^y\Delta y$, $\eta_r^x=\kappa_r^x$ and $\eta_r^y=\kappa_r^y$.*

From now on, we shall denote and assume that

$$\begin{aligned}\nu_r &= \frac{\kappa_r^x a_\star \Delta x}{2} = \frac{\kappa_r^y a_\star \Delta y}{2}, \quad \nu_u = \frac{\kappa_u a_\star \Delta x}{2} = \frac{\kappa_v a_\star \Delta y}{2}, \\ \gamma_r &= \frac{\eta_r^x a_\star \Delta x}{2} = \frac{\eta_r^y a_\star \Delta y}{2}, \quad \gamma_u = \frac{\eta_u a_\star \Delta x}{2} = \frac{\eta_v a_\star \Delta y}{2}\end{aligned} \tag{14}$$

in the correction terms. With such assumptions, the action of the diffusion operator may be rewritten as follows

$$B_{\kappa,\eta} q = B_{\nu,\gamma} q = \begin{pmatrix} \nabla \cdot (\nu_r \nabla r + \gamma_r \frac{\omega}{a_\star} \boldsymbol{u}^\perp) \\ \frac{\partial}{\partial x}(\nu_u \frac{\partial u}{\partial x} + \gamma_u \frac{\partial v}{\partial y}) \\ \frac{\partial}{\partial y}(\gamma_u \frac{\partial u}{\partial x} + \nu_u \frac{\partial v}{\partial y}) \end{pmatrix}.$$

Then we can study the behaviour of the schemes for an initial solution in the incompressible space $\mathcal{E}_{\omega\neq 0}$ or in its orthogonal $\mathcal{E}_{\omega\neq 0}^\perp$, see Prop. 3.

**Proposition 4.**

*i. A solution in the incompressible space $\mathcal{E}_{\omega\neq 0}$ is preserved for all time by the LF-DP, AT-LF and AT-DP schemes.*

*ii. The orthogonal subspace $\mathcal{E}_{\omega\neq 0}^\perp$ is preserved by the LF-DP scheme.*

*Proof.* All schemes such that $\gamma_r = \nu_r$ and $\gamma_u = \nu_u$ (namely LF-DP, AT-LF and AT-DP) satisfy $q \in \mathcal{E}_{\omega\neq 0} \implies \mathcal{B}_{\nu,\gamma} q = 0$ (Point *i*).
The proof for Point *ii* is very similar to the one in Prop. 3 up to the term

$$\nu_u \int_{\mathbb{T}^2} \hat{\boldsymbol{u}} \cdot \nabla(\nabla \cdot \boldsymbol{u}) \, \mathrm{d}\boldsymbol{x} = -\nu_u \int_{\mathbb{T}^2} (\nabla \cdot \hat{\boldsymbol{u}})(\nabla \cdot \boldsymbol{u}) \, \mathrm{d}\boldsymbol{x} = 0$$

since $\hat{\boldsymbol{u}}$ is in the incompressible subspace $\mathcal{E}_{\omega\neq 0}$. □

For the LF-DP strategy, we can also study the evolution of the energy, see Prop 2.

**Proposition 5.** *The LF-DP and C-DP schemes are energy-dissipative.*

*Proof.* Due to the fact that $\langle L_\omega q, q\rangle = 0$ as $L_\omega$ is antisymmetric, when $\gamma_r = 0$ and $\nu_u = \gamma_u$, we have

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|q\|^2 = \langle \mathcal{B}_{\nu,\gamma} q, q\rangle = -\nu_r \|\nabla r\|^2 - \nu_u \|\nabla \cdot \boldsymbol{u}\|^2 \leq 0.$$

This means that the modified equation associated to the LF-DP and C-DP schemes is dissipative. □

Hence the LF-DP strategy enables to mimic Corollary 1.

**Corollary 2.** *The solution $q_{\nu,\gamma}$ of the modified equation for the LF-DP parameters satisfies the inequality*

$$\forall\, t \geq 0,\ \|q_{\nu,\gamma} - \mathbb{P}q^0\|(t) \leq \|q^0 - \mathbb{P}q^0\|,$$

*which means the solution is accurate at low Froude number at any time, as defined at the end of Section 2.*

*Proof.* Let us first notice that $\mathbb{P}q^0$ is the solution of Eq. (13a) for the LF-DP parameters ($\nu_r = \gamma_r = 0$ and $\nu_u = \gamma_u$) with initial condition $\mathbb{P}q^0$ due to Prop. 4.*i*. We deduce by linearity that any solution of (13) reads $q_{\nu,\gamma}(t,\boldsymbol{x}) = \mathbb{P}q^0(\boldsymbol{x}) + \tilde{q}(t,\boldsymbol{x})$ where $\tilde{q}$ satisfies

$$\begin{cases} \partial_t \tilde{q} + \mathcal{L}\tilde{q} = 0, \\ \tilde{q}(t=0,\boldsymbol{x}) = q^0(\boldsymbol{x}) - \mathbb{P}q^0(\boldsymbol{x}). \end{cases}$$

As the energy is decreasing (Prop. 4.*ii*), we have

$$\|q_{\nu,\gamma} - \mathbb{P}q^0\|(t) = \|\tilde{q}\|(t) \leq \|\tilde{q}^0\| = \|q^0 - \mathbb{P}q^0\|.$$

□

### 4.2 Stability properties

For the AT strategy, we are not able to establish energy estimates as for the LF-DP strategy. So we investigate the stability of this approach by studying the behaviour of the Fourier modes of the solution.

**Lemma 3.** *Fourier modes associated to the LF-DP, C-DP, AT-LF and AT-DP schemes are damped.*

*Proof.* We now look for plane wave solutions of the modified equation (12) under the form

$$q(t,\boldsymbol{x}) = \exp\left[\imath(\tau t + \boldsymbol{k}\cdot\boldsymbol{x})\right]\hat{q} \tag{15}$$

where $\boldsymbol{k} = (k_x, k_y)$ is the wave number and $\tau$ is the wave frequency. To ensure that these waves are captured by the scheme, we assume

$$|\boldsymbol{k}| \leq \frac{\pi}{\Delta x}. \tag{16}$$

Such functions are generally solutions of the modified equation under a dispersion relation, *i.e.* a relation between $\tau$ and $\boldsymbol{k}$ commonly written as $\tau = \tau(\boldsymbol{k})$. In the present case, the Fourier modes (15) are some solutions provided

$$\mathcal{A}\hat{q} = -\imath\tau\hat{q} \tag{17}$$

and the matrix $\mathcal{A}$ is given by

$$\mathcal{A} = \begin{pmatrix} \nu_r k_x^2 + \nu_r k_y^2 & a_\star \imath k_x - \gamma_r \frac{\omega}{a_\star}\imath k_y & a_\star \imath k_y + \gamma_r \frac{\omega}{a_\star}\imath k_x \\ a_\star \imath k_x & \nu_u k_x^2 & \gamma_u k_x k_y - \omega \\ a_\star \imath k_y & \gamma_u k_x k_y + \omega & \nu_u k_y^2. \end{pmatrix}$$

The statement of the lemma is equivalent to saying that the real part of all eigenvalues are positive. Indeed, $-\imath\tau$ is an eigenvalue due to (17). The decrease for long times in (15) requires a negative coefficient for $t$.

The characteristic polynomial $\mathcal{P}(\lambda)$ reads

$$\begin{aligned}\mathcal{P}(\lambda) = \lambda^3 - (\nu_r+\nu_u)\,|\boldsymbol{k}|^2\lambda^2 + \left[a_\star^2|\boldsymbol{k}|^2 + \omega^2 + \nu_r\nu_u|\boldsymbol{k}|^4 + (\nu_u^2-\gamma_u^2)k_x^2k_y^2\right]\lambda \\ - (\nu_r-\gamma_r)\omega^2|\boldsymbol{k}|^2 - (\nu_u^2-\gamma_u^2)\nu_r k_x^2k_y^2|\boldsymbol{k}|^2 - (\nu_u-\gamma_u)2a_\star^2k_x^2k_y^2 - \gamma_r(\nu_u-\gamma_u)k_xk_y\omega(k_x^2-k_y^2).\end{aligned}$$

Let us mention that Prop. 4.*i* corresponds to the fact that $\lambda = 0$ is a root of $\mathcal{P}$ for the LF-DP, AT-LF and AT-DP schemes.

- For the *LF-DP* scheme ($\nu_r = \gamma_r = 0$ and $\nu_u = \gamma_u$), $\lambda_0 = 0$ is an eigenvalue while the other two are given by

$$\lambda_c = \frac{\nu_u|\boldsymbol{k}|^2}{2} \pm \imath\sqrt{\omega^2 + a_\star^2|\boldsymbol{k}|^2 - \left(\frac{\nu_u}{2}\right)^2|\boldsymbol{k}|^4}.$$

  Hypothesis (16) and $\kappa_u \in [0,1]$ ensure that the term in the square root is positive. Hence $\Re(\lambda_c) > 0$.

- For the *C-DP* scheme ($\gamma_r = 0$ and $\nu_u = \gamma_u$), the linear system $\mathcal{A}q = \lambda q$ reads

$$\nu_r|\boldsymbol{k}|^2\, r + \imath a_\star k_x u + \imath a_\star k_y v = \lambda\, r, \tag{18a}$$
$$\imath a_\star k_x r + \nu_u k_x^2\, u + (\nu_u k_x k_y - \omega)\, v = \lambda\, u, \tag{18b}$$
$$\imath a_\star k_y r + (\nu_u k_x k_y + \omega)\, u + \nu_u k_y^2\, v = \lambda\, v. \tag{18c}$$

  We now multiply (18a) by $\bar{r}$, (18b) by $\bar{u}$ and (18c) by $\bar{v}$ in order to obtain

$$\begin{aligned}\lambda\left(|r|^2+|u|^2+|v|^2\right) = \nu_r|\boldsymbol{k}|^2|r|^2 + \nu_u(k_x^2|u|^2 + k_y^2|v|^2) + 2\nu_u k_x k_y \Re(u\bar{v}) \\ + 2\imath\left[a_\star\Re\left(r(k_x\bar{u}+k_y\bar{v})\right) + \omega\Im\left(u\bar{v}\right)\right]\end{aligned}$$

  which implies that $\Re(\lambda) > 0$ by using the fact that $k_x^2|u|^2 + k_y^2|v|^2 \geq 2|k_xk_y||uv|$ and $\Re(u\bar{v}) \geq -|uv|$.

- For the *AT-LF* scheme ($\nu_r = \gamma_r$ and $\nu_u = \gamma_u = 0$), $\lambda_0 = 0$ is an eigenvalue. The other two are given by

$$\lambda_c = \frac{\nu_r|\boldsymbol{k}|^2}{2} \pm \imath\sqrt{\omega^2 + a_\star^2|\boldsymbol{k}|^2 - \left(\frac{\nu_r}{2}\right)^2|\boldsymbol{k}|^4} \implies \Re(\lambda_c) \geq 0.$$

- Finally, we consider the *AT-DP* scheme ($\nu_r = \gamma_r$ and $\nu_u = \gamma_u$). It is obvious that $\lambda = 0$ is a solution. The other solutions satisfy the following equation

$$\lambda^2 - (\nu_r + \nu_u)|\boldsymbol{k}|^2\lambda + \nu_r\nu_u|\boldsymbol{k}|^4 + \left[\omega^2 + a_\star^2|\boldsymbol{k}|^2\right] = 0.$$

  The solution of the above equation is given by

$$\lambda = \frac{\nu_r + \nu_u}{2}|\boldsymbol{k}|^2 \pm \imath\sqrt{\omega^2 + a_\star^2|\boldsymbol{k}|^2 - \left(\frac{\nu_r - \nu_u}{2}\right)^2|\boldsymbol{k}|^4}.$$

  which means that the Fourier modes are damped with speed $\frac{\nu_r+\nu_u}{2}|\boldsymbol{k}|^2$.

□

**Remark 5.** *The exact Fourier modes of the linear wave equation (2) are such that*

$$\lambda_{wave} = \pm\imath\sqrt{\omega^2 + a_*^2|\boldsymbol{k}|^2}. \tag{19}$$

*Consequently, we notice that the AT-DP scheme is the only one that can recover the exact imaginary part by taking $\nu_r = \nu_u$. This choice will be done in the sequel.*

**Remark 6.** *For the* AT-C *scheme ($\nu_r = \gamma_r$ and $\gamma_u = 0$), we are not able to prove the Fourier modes are damped. Nevertheless the Fourier analysis provides some information on the behaviour of the solution when the diffusion on the velocity equation is small. In that case, the characteristic polynomial becomes*

$$\chi(\lambda, \nu_u) = \lambda^3 - (\nu_r + \nu_u)|\boldsymbol{k}|^2\lambda^2 + \left[\omega^2 + a_\star^2|\boldsymbol{k}|^2 + \nu_r\nu_u|\boldsymbol{k}|^4 + \nu_u^2k_x^2k_y^2\right]\lambda \\ - \nu_r|\boldsymbol{k}|^2\nu_u^2k_x^2k_y^2 - 2\nu_u a_\star^2k_x^2k_y^2 - \nu_r\nu_u\omega k_xk_y(k_x^2 - k_y^2).$$

*We note that under (16) and due to $\kappa_{r,u} \in [0,1]$, $\partial_\lambda\chi(\lambda, \nu_u)$ does not vanish which means there is a single real root. Therefore, by the implicit function theorem, we can define for $\nu_u$ small enough a function $\nu_u \longmapsto \lambda_0(\nu_u)$ corresponding to the unique root of the polynomial. We have*

$$\lambda_0(\nu_u) \underset{\nu_u\to 0}{\sim} \lambda_0'(\nu_u = 0)\nu_u = -\frac{\partial_{\nu_u}\chi(0,0)}{\partial_\lambda\chi(0,0)}\nu_u = \frac{2a_\star^2k_x^2k_y^2 + \nu_r\omega k_xk_y(k_x^2 - k_y^2)}{\left(k_x^2 + k_y^2\right)a_\star^2 + \omega^2}\nu_u.$$

*As a result, we deduce that when $\kappa_u = \mathcal{O}(M)$, then $\lambda_0(\nu_u) = \mathcal{O}(M)$. Note that the sign of the eigenvalue remains undetermined.*

# 5 Analysis of the semi-discrete Godunov type schemes

In this section, we investigate some ways to construct "well-balanced" schemes, *i.e.* that preserve a discrete version of the incompressible subspace. The study of the modified equation leads us to focus on the numerical viscosity on both pressure and velocity equations. As a result, it is essential to consider suitable diffusion terms for the Godunov scheme. More specifically, we proposed to introduce the diffusion operators $\nabla\cdot(\nabla r + \omega\boldsymbol{u}^\perp)$ and $\nabla(\nabla\cdot\boldsymbol{u})$ for pressure and velocity equations respectively.

We now turn to the space discretisation of the aforementioned strategies. We consider a collocated finite volume framework. To ensure that the resulting schemes satisfy properties similar to those proved at the continuous level, one first has to construct some discrete differential operators and corresponding discrete kernels consistent with $\mathcal{E}_{\omega\neq 0}$. This requires to choose the location where the differential relation defining the kernels holds: either at the cell centers or at the vertices. One finally has to verify that the resulting schemes still satisfy stability properties.

## 5.1 Cell-centered scheme

A first possibility consists in locating the kernels at the same place as the unknowns, namely at the cell centers. Let us denote $r_h = (r_{i,j})$, $u_h = (u_{i,j})$ and $v_h = (v_{i,j})$ be in $\mathbb{R}^N$ where $N = N_x \times N_y$.

### 5.1.1 Discrete operators

We define the gradient $\nabla^c_{2h}$, the divergence $\nabla^c_{2h}\cdot$ and the curl $\nabla^c_{2h}\times$ as

$$
\begin{aligned}
[\nabla^c_{2h} r_h]_{i,j} &= \begin{pmatrix} \frac{r_{i+1,j}-r_{i-1,j}}{2\Delta x} \\ \frac{r_{i,j+1}-r_{i,j-1}}{2\Delta y} \end{pmatrix} \\
[\nabla^c_{2h} \cdot \boldsymbol{u}_h]_{i,j} &= \frac{u_{i+1,j} - u_{i-1,j}}{2\Delta x} + \frac{v_{i,j+1} - v_{i,j-1}}{2\Delta y}, \\
[\nabla^c_{2h} \times \boldsymbol{u}_h]_{i,j} &= -\nabla^c_{2h} \cdot \boldsymbol{u}_h^\perp = \frac{v_{i+1,j} - v_{i-1,j}}{2\Delta x} - \frac{u_{i,j+1} - u_{i,j-1}}{2\Delta y}.
\end{aligned}
$$

**Lemma 4.** *These operators satisfy the following mimetic properties:*

*i.* $\langle \nabla^c_{2h} r_h, \boldsymbol{u}_h \rangle = -\langle r_h, \nabla^c_{2h} \cdot \boldsymbol{u}_h \rangle$ *which implies that* $\langle r_h, \nabla^c_{2h} \times \boldsymbol{u}_h \rangle = -\left\langle (\nabla^c_{2h} r_h)^\perp, \boldsymbol{u}_h \right\rangle$*;*

*ii.* $\nabla^c_{2h} \times \nabla^c_{2h} r_h = 0$.

Such properties turn out to be crucial for stability purposes as claimed in [12, 20].

### 5.1.2 Discrete kernel

We now define the discrete kernel at the cell centers as the natural equivalent to $\mathcal{E}_{\omega\neq 0}$ defined in (3)

$$
\mathcal{E}^c_{\omega\neq 0,h} = \left\{ \hat{q}_h = (\hat{r}_h, \hat{u}_h, \hat{v}_h) \in \mathbb{R}^{3N} \;\middle|\; a_\star \nabla^c_{2h} \hat{r}_h = -\omega \hat{\boldsymbol{u}}_h^\perp \right\}. \tag{20}
$$

In particular, we prove the following lemma which is the semi-discrete counterpart to Proposition 1.

**Lemma 5.** *The orthogonal space of* $\mathcal{E}^c_{\omega\neq 0,h}$ *is*

$$
\mathcal{E}^{c,\perp}_{\omega\neq 0,h} = \left\{ \tilde{q}_h = (\tilde{r}_h, \tilde{u}_h, \tilde{v}_h) \in \mathbb{R}^{3N} \;\middle|\; a_\star \nabla^c_{2h} \times \tilde{\boldsymbol{u}}_h = \omega \tilde{r}_h \right\}. \tag{21}
$$

*This implies the following discrete Hodge decomposition*

$$
\mathbb{R}^{3N} = \mathcal{E}^c_{\omega\neq 0,h} \oplus \mathcal{E}^{c,\perp}_{\omega\neq 0,h}.
$$

*Proof.* By definition, an element $\tilde{q}_h = (\tilde{r}_h, \tilde{\boldsymbol{u}}_h)$ of the orthogonal of $\mathcal{E}^c_{\omega\neq 0,h}$ verifies, for all $\hat{q}_h = (\hat{r}_h, \hat{\boldsymbol{u}}_h)$ in $\mathcal{E}^c_{\omega\neq 0,h}$:

$$
\langle \tilde{r}_h, \hat{r}_h \rangle + \langle \tilde{\boldsymbol{u}}_h, \hat{\boldsymbol{u}}_h \rangle = \langle \tilde{r}_h, \hat{r}_h \rangle + \langle \tilde{\boldsymbol{u}}_h^\perp, \hat{\boldsymbol{u}}_h^\perp \rangle = 0.
$$

Using the definition of $\mathcal{E}^c_{\omega\neq 0,h}$ and Lemma 4 this implies

$$
\langle \tilde{r}_h, \hat{r}_h \rangle - \frac{a_\star}{\omega} \langle \tilde{\boldsymbol{u}}_h^\perp, \nabla^c_{2h} \hat{r}_h \rangle = \langle \tilde{r}_h, \hat{r}_h \rangle + \frac{a_\star}{\omega} \langle \nabla^c_{2h} \cdot \tilde{\boldsymbol{u}}_h^\perp, \hat{r}_h \rangle = \langle \tilde{r}_h - \frac{a_\star}{\omega} \nabla^c_{2h} \times \tilde{\boldsymbol{u}}_h, \hat{r}_h \rangle = 0.
$$

Since $\hat{r}_h$ can be arbitrary in $\mathbb{R}^N$, this is exactly equivalent to $\omega \tilde{r}_h - a_\star \nabla^c_{2h} \times \tilde{\boldsymbol{u}}_h$. □

**Remark 7.** *For any $q_h \in \mathbb{R}^{3N}$, the unique decomposition*

$$q_h = \hat{q}_h + \tilde{q}_h \quad \textit{with} \quad \hat{q}_h = (\hat{r}_h, \hat{u}_h, \hat{v}_h) \in \mathcal{E}^c_{\omega\neq 0,h} \quad \textit{and} \quad \tilde{q}_h = (\tilde{r}_h, \tilde{u}_h, \tilde{v}_h) \in \mathcal{E}^{c,\perp}_{\omega\neq 0,h}$$

*may be found by the following process: Let $\hat{r}_h$ satisfy the equation*

$$\hat{r}_h - \frac{a_*^2}{\omega^2}\nabla^c_{2h}\cdot(\nabla^c_{2h}\hat{r}_h) = r_h - \frac{a_\star}{\omega}\nabla^c_{2h}\times \boldsymbol{u}_h. \tag{22}$$

*It can be shown that (22) has a unique solution since it amounts to solving a linear system involving an $M$-matrix. Then, let us define $\hat{\boldsymbol{u}}_h$ by*

$$\hat{\boldsymbol{u}}_h = \frac{a_\star}{\omega}(\nabla^c_{2h}\hat{r}_h)^\perp \tag{23}$$

*so that $\hat{q}_h = (\hat{r}_h, \hat{\boldsymbol{u}}_h) \in \mathcal{E}^c_{\omega\neq 0,h}$. Finally, we set $\tilde{q}_h = q_h - \hat{q}_h$ and it remains to prove that $\tilde{q}_h \in \mathcal{E}^{c,\perp}_{\omega\neq 0,h}$. It suffices to notice that*

$$\begin{aligned} a_\star\nabla^c_{2h}\times\tilde{\boldsymbol{u}}_h = a_\star\left(\nabla^c_{2h}\times\boldsymbol{u}_h + \nabla^c_{2h}\cdot\hat{\boldsymbol{u}}_h^\perp\right) &\overset{(23)}{=} a_\star\left(\nabla^c_{2h}\times\boldsymbol{u}_h - \frac{a_\star}{\omega}\nabla^c_{2h}\cdot(\nabla^c_{2h}\hat{r}_h)\right) \\ &\overset{(22)}{=} a_\star\nabla^c_{2h}\times\boldsymbol{u}_h - \omega\left(\hat{r}_h - r_h + \frac{a_\star}{\omega}\nabla^c_{2h}\times\boldsymbol{u}_h\right) = \omega\tilde{r}_h. \end{aligned}$$

### 5.1.3 Semi-discrete scheme

The cell-centered semi-discrete scheme reads

$$\begin{cases} \dfrac{\mathrm{d}}{\mathrm{d}t}r_{i,j}(t) + a_\star[\nabla^c_{2h}\cdot\boldsymbol{u}_h]_{i,j} - \nu_r\left[\nabla^c_{2h}\cdot\left(\nabla^c_{2h}r_h + \dfrac{\omega}{a_\star}\boldsymbol{u}_h^\perp\right)\right]_{i,j} = 0, & \text{(24a)} \\ \dfrac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{u}_{i,j}(t) + a_\star[\nabla^c_{2h}r_h]_{i,j} - \nu_u\left[\nabla^c_{2h}(\nabla^c_{2h}\cdot\boldsymbol{u}_h)\right]_{i,j} = -\omega\boldsymbol{u}^\perp_{i,j}. & \text{(24b)} \end{cases}$$

The modified equation associated to the scheme (24) is (12) for coefficients chosen as in Remark 4. The stencil associated to the scheme (24) is a 13-point stencil: it involves the 8 points around the considered one (*i.e.* at a distance $\Delta x$ or $\Delta y$) and 4 points to a distance $2\Delta x$ (or $2\Delta y$) in the definition of both diffusion terms. Moreover the definition of the diffusion terms induces no relation between odd and even cells. This may be the reason for checkerboard type oscillations. The interface scheme (26) we propose in the sequel will not be affected by this drawback.

**Proposition 6.**

*i. Steady states of the semi-discrete scheme (24) are the discrete geostrophic equilibria from (20).*

*ii. The pressure gradient and Coriolis forces are energy conservative.*

*iii. The discrete energy of the $LF - DP$ scheme ($\nu_r = 0$) is decreasing.*

*Proof.* On the one hand, by construction, discrete geostrophic equilibria (20) are steady states of (24). On the other hand, let us consider steady states of (24). Applying the operator $\nabla^c_{2h}\times$ to (24b), we obtain $\nabla^c_{2h}\cdot\boldsymbol{u}_h = 0$ due to Lemma 4.*ii*. This proves Point *i*.

Point *ii* is a straightforward consequence of Lemma 4 *i*. Moreover, when $\nu_r = 0$, the scalar product with $q_h$ leads to

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}E_h(t) = -a_\star\langle\nabla^c_{2h}\cdot\boldsymbol{u}_h, r_h\rangle - a_\star\langle\nabla^c_{2h}r_h, \boldsymbol{u}_h\rangle + \nu_u\langle\nabla^c_{2h}[\nabla^c_{2h}\cdot\boldsymbol{u}_h], \boldsymbol{u}_h\rangle = -\nu_u\|\nabla^c_{2h}\cdot\boldsymbol{u}_h\|^2 \leq 0$$

thanks to Lemma 4.*i*. This proves Point *iii*. □

## 5.2 Vertex-based scheme

The original *Apparent Topography* scheme [7] was designed in 1D so that equilibrium states are located at the interfaces while the unknowns are still at the cell centers. That is why we are interested in this part in investigating another version of the scheme.

### 5.2.1 Discrete kernel

Let us define the discrete kernel by imposing the geostrophic equilibrium at the interfaces of each cell

$$\mathcal{E}^v_{\omega\neq 0,h} = \left\{ \hat{q}_h = (\hat{r}_h, \hat{u}_h, \hat{v}_h) \in \mathbb{R}^{3N} \;\middle|\; a_\star \begin{pmatrix} \frac{\hat{r}_{i+1,j}-\hat{r}_{i,j}}{\Delta x} \\ \\ \frac{\hat{r}_{i,j+1}-\hat{r}_{i,j}}{\Delta y} \end{pmatrix} = -\omega \begin{pmatrix} -\frac{\hat{v}_{i+1,j}+\hat{v}_{i,j}}{2} \\ \\ \frac{\hat{u}_{i,j+1}+\hat{u}_{i,j}}{2} \end{pmatrix} \right\}. \tag{25}$$

### 5.2.2 Discrete operators

To design the numerical scheme, we first define the discrete operators at the vertices of each cell $(i,j)$ – see Figure 4

$$\begin{aligned}
[\nabla^{\mathrm{v}}_h r_h]_{i+1/2,j+1/2} &= \begin{pmatrix} \frac{(r_{i+1,j+1}+r_{i+1,j})-(r_{i,j+1}+r_{i,j})}{2\Delta x} \\ \\ \frac{(r_{i+1,j+1}+r_{i,j+1})-(r_{i+1,j}+r_{i,j})}{2\Delta y} \end{pmatrix} \\
[\nabla^{\mathrm{v}}_h \cdot \boldsymbol{u}_h]_{i+1/2,j+1/2} &= \frac{(u_{i+1,j+1}+u_{i+1,j})-(u_{i,j+1}+u_{i,j})}{2\Delta x} + \frac{(v_{i+1,j+1}+v_{i,j+1})-(v_{i+1,j}+v_{i,j})}{2\Delta y} \\
[\nabla^{\mathrm{v}}_h \times \boldsymbol{u}_h]_{i+1/2,j+1/2} &= -\nabla^{\mathrm{v}}_h \cdot \boldsymbol{u}^\perp_h, \\
[f^{\mathrm{v}}_h(u_h)]_{i+1/2,j+1/2} &= \frac{u_{i+1,j+1}+u_{i,j+1}+u_{i+1,j}+u_{i,j}}{4}.
\end{aligned}$$

We shall also need dual operators that enable to switch from the vertex grid to the center grid. For $\boldsymbol{\varphi}_h = (\varphi_h, \psi_h)$ defined at the vertices, we define the following operators

$$\begin{aligned}
[\nabla^{\mathrm{c}}_h \varphi_h]_{i,j} &= \frac{1}{2}\begin{pmatrix} \frac{\varphi_{i+1/2,j+1/2}-\varphi_{i-1/2,j+1/2}}{\Delta x} \\ \\ \frac{\varphi_{i+1/2,j+1/2}-\varphi_{i+1/2,j-1/2}}{\Delta y} \end{pmatrix} + \frac{1}{2}\begin{pmatrix} \frac{\varphi_{i+1/2,j-1/2}-\varphi_{i-1/2,j-1/2}}{\Delta x} \\ \\ \frac{\varphi_{i-1/2,j+1/2}-\varphi_{i-1/2,j-1/2}}{\Delta y} \end{pmatrix} \\
[\nabla^{\mathrm{c}}_h \cdot \boldsymbol{\varphi}_h]_{i,j} &= \frac{(\varphi_{i+1/2,j+1/2}+\varphi_{i+1/2,j-1/2})-(\varphi_{i-1/2,j+1/2}+\varphi_{i-1/2,j-1/2})}{2\Delta x} \\
&\quad + \frac{(\psi_{i+1/2,j+1/2}+\psi_{i-1/2,j+1/2})-(\psi_{i+1/2,j-1/2}+\psi_{i-1/2,j-1/2})}{2\Delta y} \\
[f^{\mathrm{c}}_h(\varphi_h)]_{i,j} &= \frac{\varphi_{i+1/2,j+1/2}+\varphi_{i-1/2,j+1/2}+\varphi_{i+1/2,j-1/2}+\varphi_{i-1/2,j-1/2}}{4}.
\end{aligned}$$

With such operators, we have the following compatibility property:

**Lemma 6.** *Any $\hat{q}_h \in \mathcal{E}^v_{\omega\neq 0,h}$ satisfies the geostrophic equilibrium and the divergence free condition at the vertices:*

$$\begin{cases} [\nabla^v_h \hat{r}_h]_{i+1/2,j+1/2} = -\omega [f^v_h(\hat{\boldsymbol{u}}^\perp_h)]_{i+1/2,j+1/2}, \\ \\ [\nabla^v_h \cdot \hat{\boldsymbol{u}}_h]_{i+1/2,j+1/2} = 0. \end{cases}$$

Moreover, they satisfy mimetic properties:

**Lemma 7.**

*i.* $\nabla^v_h \times \nabla^c_h \left[f^v_h(r_h)\right] = \nabla^v_h \times f^c_h \left[\nabla^v_h r_h\right] = 0$;

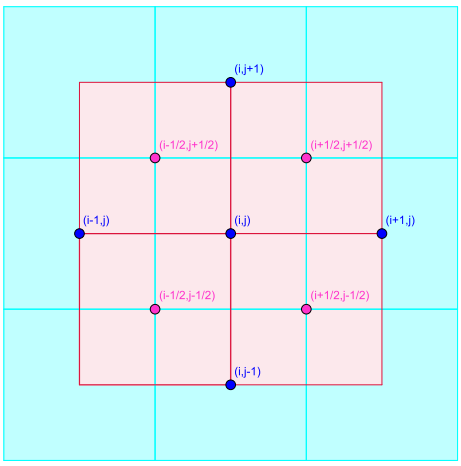



Figure 4: Cell centers $(i, j)$ and vertices $(i + 1/2, j + 1/2)$.

*ii.* $\left\langle f_h^c\left[\nabla_h^v r_h\right], \boldsymbol{u}_h\right\rangle = -\left\langle r_h, f_h^c\left[\nabla_h^v \cdot \boldsymbol{u}_h\right]\right\rangle$ *and* $\left\langle f_h^c\left[f_h^v(u_h)\right], v_h\right\rangle = \left\langle u_h, f_h^c\left[f_h^v(v_h)\right]\right\rangle$.

*Proof.* Each property results from direct computations. For instance:

$$
\begin{aligned}
&\left\langle f_h^{\mathrm{c}}\left[\nabla_h^{\mathrm{v}} r_h\right], \boldsymbol{u}_h\right\rangle \\
&\quad= \sum_{i,j}\left[\frac{1}{4}\left(\frac{r_{i+1,j+1}-r_{i-1,j+1}}{2\Delta x}\right)+\frac{1}{2}\left(\frac{r_{i+1,j}-r_{i-1,j}}{2\Delta x}\right)+\frac{1}{4}\left(\frac{r_{i+1,j-1}-r_{i-1,j-1}}{2\Delta x}\right)\right] u_{i,j} \\
&\qquad+\sum_{i,j}\left[\frac{1}{4}\left(\frac{r_{i+1,j+1}-r_{i+1,j-1}}{2\Delta y}\right)+\frac{1}{2}\left(\frac{r_{i,j+1}-r_{i,j-1}}{2\Delta y}\right)+\frac{1}{4}\left(\frac{r_{i-1,j+1}-r_{i-1,j-1}}{2\Delta y}\right)\right] v_{i,j} \\
&\quad= \sum_{i,j}\left(\frac{r_{i+1,j}-r_{i-1,j}}{2\Delta x}\right)\left(\frac{u_{i,j+1}+2u_{i,j}+u_{i,j-1}}{4}\right)+\sum_{i,j}\left(\frac{r_{i,j+1}-r_{i,j-1}}{2\Delta y}\right)\left(\frac{v_{i+1,j}+2v_{i,j}+v_{i-1,j}}{4}\right) \\
&\quad= -\sum_{i,j} r_{i,j}\left[\frac{1}{4}\left(\frac{u_{i+1,j+1}-u_{i-1,j+1}}{2\Delta x}\right)+\frac{1}{2}\left(\frac{u_{i+1,j}-u_{i-1,j}}{2\Delta x}\right)+\frac{1}{4}\left(\frac{u_{i+1,j-1}-u_{i-1,j-1}}{2\Delta x}\right)\right] \\
&\qquad-\sum_{i,j} r_{i,j}\left[\frac{1}{4}\left(\frac{v_{i+1,j+1}-v_{i+1,j-1}}{2\Delta y}\right)+\frac{1}{2}\left(\frac{v_{i,j+1}-v_{i,j-1}}{2\Delta y}\right)+\frac{1}{4}\left(\frac{v_{i-1,j+1}-v_{i-1,j-1}}{2\Delta y}\right)\right] \\
&\quad= -\left\langle f_h^{\mathrm{c}}\left[\nabla_h^{\mathrm{v}} \cdot \boldsymbol{u}_h\right], r_h\right\rangle .
\end{aligned}
$$

□

### 5.2.3 Semi-discrete scheme

The semi-discrete scheme with the kernel at the interface is given by

$$
\begin{cases}
\dfrac{\mathrm{d}}{\mathrm{d}t} r_{i,j}(t) + a_\star f_h^{\mathrm{c}}\left[\nabla_h^{\mathrm{v}} \cdot \boldsymbol{u}_h\right]_{i,j} - \nu_r \nabla_h^{\mathrm{c}} \cdot \left[\nabla_h^{\mathrm{v}} r_h + \omega f_h^{\mathrm{v}}(\boldsymbol{u}_h^{\perp})\right]_{i,j} = 0, \\
\dfrac{\mathrm{d}}{\mathrm{d}t} \boldsymbol{u}_{i,j}(t) + a_\star f_h^{\mathrm{c}}\left[\nabla_h^{\mathrm{v}} r_h\right]_{i,j} - \nu_u \nabla_h^{\mathrm{c}}\left[\nabla_h^{\mathrm{v}} \cdot \boldsymbol{u}_h\right]_{i,j} = -\omega f_h^{\mathrm{c}}\left[f_h^{\mathrm{v}}(\boldsymbol{u}_h^{\perp})\right]_{i,j} .
\end{cases}
\tag{26}
$$

The modified equation associated to the scheme (26) is still (12) for coefficients chosen as in Remark 4. The stencil associated to this second scheme (26) is a classical 9-point stencil since it only involves the 8 points around the considered one. It is then more compact than the one of the cell-centered scheme (24).

**Proposition 7.**

| *Godunov type scheme* | $\alpha$ | $\beta$ | $\eta$ |
|---|---|---|---|
| *Cell-centered* | $\sin(k_x\Delta x)$ | $\sin(k_y\Delta y)$ | $1$ |
| *Vertex-based* | $2\sin(\frac{k_x\Delta x}{2})\cos(\frac{k_y\Delta y}{2})$ | $2\sin(\frac{k_y\Delta y}{2})\cos(\frac{k_x\Delta x}{2})$ | $\cos(\frac{k_x\Delta x}{2})\cos(\frac{k_y\Delta y}{2})$ |

Table 2: Parameters $\alpha$, $\beta$ an $\eta$ in the Fourier analysis of the semi-discrete schemes.

*i. Steady states of the semi-discrete scheme (26) are the geostrophic equilibria from (25).*

*ii. The pressure gradient and Coriolis forces are energy conservative.*

*iii. The energy of the LF-DP scheme ($\nu_r = 0$) is decreasing.*

*Proof.* Point *i.* results from Lemma 6 and from Lemma 7.*i*. Moreover, according to Lemma 7.*ii*, we have

$$\langle f_h^{\rm c}\left[\nabla_h^{\rm v} r_h\right], \boldsymbol{u}_h\rangle + \langle f_h^{\rm c}\left[\nabla_h^{\rm v}\cdot \boldsymbol{u}_h\right], r_h\rangle = 0 \quad \text{and} \quad \left\langle f_h^{\rm c}\left[f_h^{\rm v}(\boldsymbol{u}_h^{\perp})\right], \boldsymbol{u}_h\right\rangle = 0$$

which proves Point *ii*.

After some computations, we have

$$\langle \nabla_h^{\rm c}\left[\nabla_h^{\rm v}\cdot \boldsymbol{u}_h\right], \boldsymbol{u}_h\rangle = -\sum_{i,j}\left[\frac{u_{i+1,j+1}-u_{i,j+1}}{2\Delta x}+\frac{u_{i+1,j}-u_{i,j}}{2\Delta x}+\frac{v_{i+1,j+1}-v_{i+1,j}}{2\Delta y}+\frac{v_{i,j+1}-v_{i,j}}{2\Delta y}\right]^2.$$

Therefore, when $\nu_r = 0$, we deduce that

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}E_h(t) = \nu_u\langle \nabla_h^{\rm c}\left[\nabla_h^{\rm v}\cdot \boldsymbol{u}_h\right], \boldsymbol{u}_h\rangle = -\nu_u\|\nabla_h^{\rm v}\cdot \boldsymbol{u}_h\|^2$$

which means that the semi-discrete LF-DP scheme is dissipative. This proves Point *iii*. □

### 5.3 Fourier analysis

Let us carry out a Fourier analysis of the semi-discrete schemes by considering the discrete Fourier modes

$$r_{i,j}(t) = \varphi_r(t)e^{\imath(k_x x_i + k_y y_j)}, \quad u_{i,j}(t) = \varphi_u(t)e^{\imath(k_x x_i + k_y y_j)} \quad \text{and} \quad v_{i,j}(t) = \varphi_v(t)e^{\imath(k_x x_i + k_y y_j)}.$$

that are substituted in the cell-centered scheme (24) and in the vertex-based scheme (26) to obtain the differential system

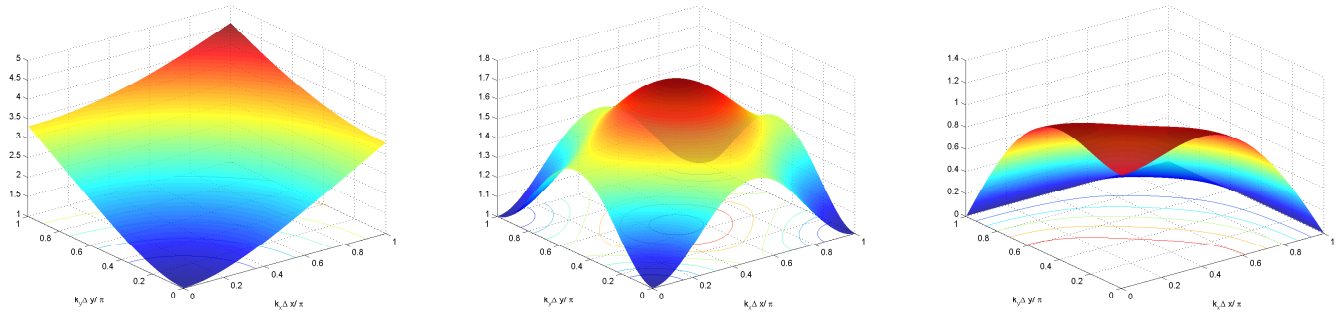
$$\begin{pmatrix}\varphi_r'(t)\\ \varphi_u'(t)\\ \varphi_v'(t)\end{pmatrix} = \begin{pmatrix}-\nu_r\left(\frac{\alpha^2}{\Delta x^2}+\frac{\beta^2}{\Delta y^2}\right) & -\imath\eta\left(a_\star\frac{\alpha}{\Delta x}-\nu_r\frac{\omega}{a_\star}\frac{\beta}{\Delta y}\right) & -\imath\eta\left(a_\star\frac{\beta}{\Delta y}+\nu_r\frac{\omega}{a_\star}\frac{\alpha}{\Delta x}\right)\\ -\imath a_\star\frac{\alpha}{\Delta x}\eta & -\nu_u\frac{\alpha^2}{\Delta x^2} & -\nu_u\frac{\alpha}{\Delta x}\frac{\beta}{\Delta y}+\omega\eta^2\\ -\imath a_\star\frac{\beta}{\Delta y}\eta & -\nu_u\frac{\alpha}{\Delta x}\frac{\beta}{\Delta y}-\omega\eta^2 & -\nu_u\frac{\beta^2}{\Delta y^2}\end{pmatrix}\begin{pmatrix}\varphi_r(t)\\ \varphi_u(t)\\ \varphi_v(t)\end{pmatrix} \tag{27}$$

where parameters $\alpha$, $\beta$ an $\eta$ are specified in Table 2 depending on the scheme under study. One eigenvalue of the amplification matrix in (27) is $\lambda_0 = 0$ which corresponds to the stationary state. The other eigenvalues are given by

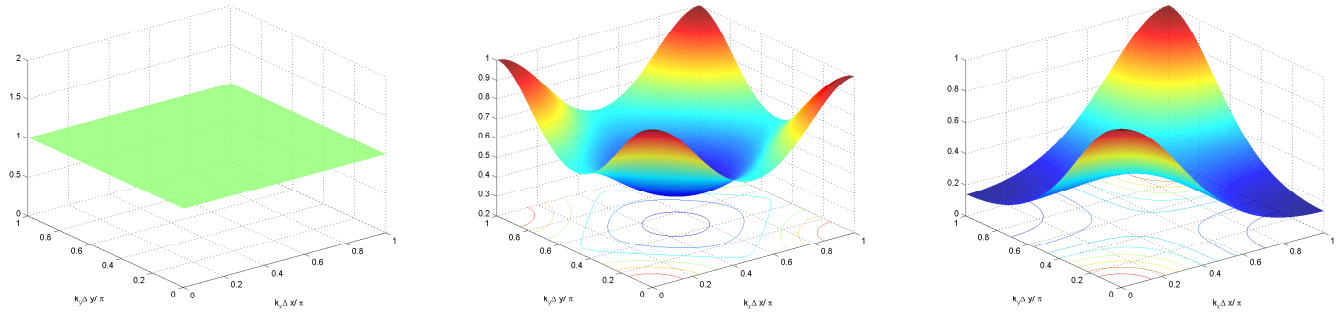
$$\lambda_c = \frac{\nu_r+\nu_u}{2}\left(\frac{\alpha^2}{\Delta x^2}+\frac{\beta^2}{\Delta y^2}\right) \pm \imath\sqrt{\omega^2\eta^4 + a_\star^2\eta^2\left(\frac{\alpha^2}{\Delta x^2}+\frac{\beta^2}{\Delta y^2}\right) - \left(\frac{\nu_r-\nu_u}{2}\right)^2\left(\frac{\alpha^2}{\Delta x^2}+\frac{\beta^2}{\Delta y^2}\right)^2}.$$

As mentioned above, it is essential with the $AT-DP$ scheme to take $\nu_r = \nu_u$ in order to be as close as possible to the exact dispersion relation (19), see Figure 5.

**Remark 8.** *We notice that the damping rate $\Re(\lambda)$ of the $AT-DP$ scheme is larger than those of the $AT-LF$ and $LF-DP$ schemes.*



(a) Exact dispersion law (b) Dispersion law (cell-centered) (c) Dispersion law (vertex-based)



(d) Exact damping (e) Damping error (cell-centered) (f) Damping error (vertex-based)

Figure 5: Dispersion relation and damping for the AT-DP scheme with $a_* = \omega \Delta x$.

# 6 Analysis of the fully discrete Godunov type schemes

We consider an explicit discretisation for the advection term. Nevertheless it is well known that a fully explicit discretisation of the Coriolis term leads in that case to unstable schemes, see [8]. Then let us set

$$\boldsymbol{u}^\theta = \begin{pmatrix} \theta_1 u^n + (1-\theta_1) u^{n+1} \\ \theta_2 v^n + (1-\theta_2) v^{n+1} \end{pmatrix}$$

for some $\theta_1, \theta_2 \in [0,1]$. In particular, for $\theta_1 = \theta_2 = \theta$, then $\boldsymbol{u}^\theta = \theta \boldsymbol{u}^n + (1-\theta)\boldsymbol{u}^{n+1}$.

## 6.1 Stability condition

For the sake of clarity, we shall assume in the sequel that

$$\Delta x = \Delta y = h.$$

As a consequence, due to (14), we have

$$\kappa_r := \kappa_r^x = \kappa_r^y = \eta_r^x = \eta_r^y, \qquad \kappa_u = \kappa_v = \eta_u = \eta_v \qquad \text{and} \qquad \nu_\# = \frac{\kappa_\# a_* \Delta x}{2}.$$

We propose the following time discretisation for the cell-centered scheme

$$\begin{cases} \dfrac{r_{i,j}^{n+1} - r_{i,j}^n}{\Delta t} + a_\star [\nabla_{2h}^c \cdot \boldsymbol{u}_h^n]_{i,j} - \nu_r \left[ \nabla_{2h}^c \cdot \left( \nabla_{2h}^c r_h^n + \dfrac{\omega}{a_*} \boldsymbol{u}_h^{n,\perp} \right) \right]_{i,j} = 0, \\ \dfrac{\boldsymbol{u}_{i,j}^{n+1} - \boldsymbol{u}_{i,j}^n}{\Delta t} + a_\star [\nabla_{2h}^c r_h^n]_{i,j} - \nu_u \left[ \nabla_{2h}^c (\nabla_{2h}^c \cdot \boldsymbol{u}_h^n) \right]_{i,j} = -\omega \boldsymbol{u}_{i,j}^{\theta,\perp}, \end{cases} \tag{28}$$

and for the vertex-based scheme

$$\begin{cases} \dfrac{r_{i,j}^{n+1}-r_{i,j}^{n}}{\Delta t}+a_\star f_h^{\mathrm{c}}\left[\nabla_h^{\mathrm{v}}\cdot\boldsymbol{u}_h^n\right]_{i,j}-\nu_r\nabla_h^{\mathrm{c}}\cdot\left[\nabla_h^{\mathrm{v}}r_h^n+\omega f_h^{\mathrm{v}}(\boldsymbol{u}_h^n)^\perp\right]_{i,j}=0,\\ \dfrac{\boldsymbol{u}_{i,j}^{n+1}-\boldsymbol{u}_{i,j}^{n}}{\Delta t}+a_\star f_h^{\mathrm{c}}\left[\nabla_h^{\mathrm{v}}r_h^n\right]_{i,j}-\nu_u\nabla_h^{\mathrm{c}}\left[\nabla_h^{\mathrm{v}}\cdot\boldsymbol{u}_h^n\right]_{i,j}=-\omega f_h^{\mathrm{c}}\big[f_h^{\mathrm{v}}(\boldsymbol{u}_h^\theta)\big]_{i,j}^\perp. \end{cases} \tag{29}$$

In order to avoid inverting a matrix with a large stencil in the computation of the scheme, the vertex-based scheme is restricted to the cases $(\theta_1=1,\theta_2=0)$ and $(\theta_1=0,\theta_2=1)$.

**Lemma 8.** *Any choice such that $\theta_1+\theta_2>1$ makes schemes (28) and (29) unstable. In particular, the explicit case $\theta_1=\theta_2=1$ is unstable, as mentioned before.*

The proof of this lemma is embedded in the proof of Theorem 1 below.

**Theorem 1.** *For a uniform mesh $\Delta x=\Delta y=h$, the LF-DP schemes* (i.e. (28) *and* (29) *with $\nu_r=0$) are stable under the following conditions*

$$\Delta t\le\min\{\Delta t_a,\Delta t_b\}\quad \textit{where}\quad \Delta t_a:=\frac{\kappa_u h}{2a_*}\ \textit{and}\ \Delta t_b:=\frac{2}{\omega|\theta_2-\theta_1|}.$$

**Remark 9.** *The restriction on the time step $\Delta t_a$ (resp. $\Delta t_b$) is the classical CFL condition for advection (resp. rotation) phenomena. Note that the choice $\theta_2=\theta_1$ makes the CFL condition independent from the Coriolis parameter.*

*Proof.* Let us denote

$$\varpi=\omega\Delta t,\quad \sigma=a_\star\frac{\Delta t}{h}.$$

We now perform the Fourier analysis for fully discrete Godunov type schemes by substituting the fully discrete Fourier mode

$$r_{i,j}^n=\varphi_r^n e^{\imath(k_x x_i+k_y y_j)},\quad u_{i,j}^n=\varphi_u^n e^{\imath(k_x x_i+k_y y_j)}\quad\text{and}\quad v_{i,j}^n=\varphi_v^n e^{\imath(k_x x_i+k_y y_j)}$$

into the fully discrete scheme to obtain

$$\mathcal{T}_\theta\varphi^{n+1}=\mathcal{M}_\theta\varphi^n$$

where

$$\mathcal{T}_\theta=\begin{pmatrix}1&0&0\\0&1&-(1-\theta_2)\varpi\eta^2\\0&(1-\theta_1)\varpi\eta^2&1\end{pmatrix}$$

and

$$\mathcal{M}_\theta=\begin{pmatrix}1-\frac{\kappa_r\sigma}{2}\left(\alpha^2+\beta^2\right)&-\imath\eta\left(\sigma\alpha-\frac{\kappa_r}{2}\varpi\beta\right)&-\imath\eta\left(\sigma\beta+\frac{\kappa_r}{2}\varpi\alpha\right)\\-\imath\sigma\alpha\eta&1-\frac{\kappa\sigma}{2}\alpha^2&-\frac{\kappa\sigma}{2}\alpha\beta+\theta_2\varpi\eta^2\\-\imath\sigma\beta\eta&-\frac{\kappa\sigma}{2}\alpha\beta-\theta_1\varpi\eta^2&1-\frac{\kappa\sigma}{2}\beta^2\end{pmatrix}.$$

Let us set $\Lambda(\theta_1,\theta_2)=1+\varpi^2\eta^4(1-\theta_1)(1-\theta_2)=\det\mathcal{T}_\theta$. The characteristic polynomial of this amplification matrix $\mathcal{T}_\theta^{-1}\mathcal{M}_\theta$ has one root $\lambda=1$ and the other roots are also roots of the second-order polynomial

$$P(\lambda):=\Lambda\lambda^2+\xi\lambda+\zeta \tag{30}$$

where

$$\xi=-2+\varpi^2\eta^4(\theta_1+\theta_2-2\theta_1\theta_2)+\frac{\kappa_u\sigma}{2}\left[\alpha^2+\beta^2-\varpi\eta^2\alpha\beta(\theta_2-\theta_1)\right]+\frac{\kappa_r\sigma}{2}\left(\alpha^2+\beta^2\right)\Lambda$$

and

$$\zeta = 1 + \varpi^2\eta^4\theta_1\theta_2 + \frac{\kappa_r\sigma}{2}\varpi^2\eta^4\left[\alpha^2\theta_1(1-\theta_2) + \beta^2\theta_2(1-\theta_1)\right] - \frac{\kappa_r\sigma}{2}\left(\alpha^2+\beta^2\right) \\ + \sigma\left(\sigma\eta^2 - \frac{\kappa_u}{2} + \sigma\frac{\kappa_r\kappa_u}{4}(\alpha^2+\beta^2)\right)\left[\alpha^2+\beta^2 - \varpi\eta^2\alpha\beta(\theta_2-\theta_1)\right].$$

Let us first prove Lemma 8 and consider for that the stationary state, $k_x = k_y = 0$, which implies $\alpha = \beta = 0$. The characteristic polynomial then reduces to

$$P(\lambda) = \Lambda\lambda^2 + \left[-2 + \varpi^2\eta^4(\theta_2+\theta_1-2\theta_2\theta_1)\right]\lambda + 1 + \varpi^2\eta^4\theta_2\theta_1.$$

For the scheme to be stable, all eigenvalues must satisfy $|\lambda| \leq 1$. In this simple case, a necessary condition is $|\lambda_1\lambda_2| \leq 1$, which is equivalent to

$$\frac{\zeta}{\Lambda} \leq 1 \iff \varpi^2(1-\theta_2-\theta_1) \geq 0.$$

This proves Lemma 8. Let us now turn to the proof of Theorem 1.

We now consider the fully discrete LF-DP cell-centered scheme:

$$\kappa_r = 0, \quad \eta = 1 \quad \text{and} \quad -1 \leq \alpha, \beta \leq 1.$$

Then parameters $\xi$ and $\zeta$ involved in (30) reduce to

$$\xi = -2 + \varpi^2(\theta_2+\theta_1-2\theta_2\theta_1) + \frac{\kappa_u\sigma}{2}\left[\alpha^2+\beta^2 - \varpi\alpha\beta(\theta_2-\theta_1)\right]$$

and

$$\zeta = 1 + \varpi^2\theta_2\theta_1 + \sigma\left(\sigma - \frac{\kappa_u}{2}\right)\left[\alpha^2+\beta^2 - \varpi\alpha\beta(\theta_2-\theta_1)\right].$$

Imposing $|\lambda| \leq 1$ is equivalent to

$$|\zeta| \leq \Lambda \quad \text{and} \quad |\xi| \leq \Lambda + \zeta.$$

- Firstly, the condition $\zeta \leq \Lambda$ can be written as

$$f_1(\alpha,\beta) = \varpi^2\left[1-(\theta_2+\theta_1)\right] + \sigma\left(\frac{\kappa_u}{2} - \sigma\right)\left[\alpha^2+\beta^2 - \varpi\alpha\beta(\theta_2-\theta_1)\right] \geq 0$$

  which in particular holds when

$$\sigma \leq \frac{\kappa_u}{2} \quad \text{and} \quad \varpi|\theta_2-\theta_1| \leq 2. \tag{31}$$

  Indeed, the latter constraint implies that $\alpha^2+\beta^2-\varpi\alpha\beta(\theta_2-\theta_1) \in [0,4]$ since $\alpha,\beta \in [-1,1]$.

- The condition $\zeta \geq -\Lambda$ is equivalent to

$$f_2(\alpha,\beta) = 2 + \varpi^2\left[1-(\theta_2+\theta_1)+2\theta_2\theta_1\right] - \sigma\left(\frac{\kappa_u}{2} - \sigma\right)\left[\alpha^2+\beta^2 - \varpi\alpha\beta(\theta_2-\theta_1)\right] \geq 0.$$

  Under (31) and due to the fact that $\kappa_u \in [0,1]$, we have

$$2 - \sigma\left(\frac{\kappa}{2} - \sigma\right)4 = 4\left(\sigma - \frac{\kappa_u}{2}\right)^2 + 2 - \frac{\kappa_u^2}{4} \geq 0$$

  which ensures that the requirement $f_2 \geq 0$ is always satisfied.

- The case $-\xi \leq \Lambda + \zeta$ reads

$$f_3(\alpha,\beta) = \varpi^2 + \sigma^2\left[\alpha^2+\beta^2 - \varpi\alpha\beta(\theta_2-\theta_1)\right] \geq 0$$

  which always holds under (31).

- Finally, the condition $\xi \leq \Lambda + \zeta$ reads

$$f_4(\alpha,\beta) = 4 + \varpi^2(1-2\theta_2)(1-2\theta_1) + \sigma(\sigma-\kappa_u)\left[\alpha^2+\beta^2-\varpi\alpha\beta(\theta_2-\theta_1)\right] \geq 0.$$

Let us notice that due to (31) and $\theta_2, \theta_1 \in [0,1]$, we have

$$f_4(\alpha,\beta) \geq p_4 := 4\left[\sigma^2 - \sigma\kappa_u - \frac{\varpi^2}{4} + 1\right].$$

Either $\omega^2h^2 < (4-\kappa_u)a_*^2$ and $p_4$ is a second-order polynomial with respect to $\Delta t$ that is always positive: there is no additional constraint upon the time step. Or $\omega^2h^2 \geq (4-\kappa_u)a_*^2$ and $\Delta t$ must be small enough to ensure that $p_4 \geq 0$, *i.e.*

$$\Delta t \leq \frac{h}{a_*\kappa_u} \times 2\frac{1-\sqrt{1-\frac{4a_*^2-\omega^2h^2}{a_*^2\kappa_u^2}}}{\frac{4a_*^2-\omega^2h^2}{a_*^2\kappa_u^2}}. \tag{32}$$

The convexity of the function $x \mapsto 1-\sqrt{1-x}$ shows that when $\omega^2h^2 \leq 4a_*^2$, the bound in (32) is greater than $\frac{h}{a_*\kappa_u} \geq \frac{h}{2a_*}$. Hence in that case (32) is less restrictive than (31). The study of the monotonicity of the bound with respect to $\kappa_u$ shows that it is also the case when $\omega^2h^2 > 4a_*^2$. Consequently, the only constraint upon the time step is (31) which ends the proof of Theorem 1.

In the vertex-based case, the only difference is that $\eta \in [0,1]$. It can be shown that $\eta = 1$ is always the most restrictive constraint and the same stability conditions hold. □

**Proposition 8.** *Let us set* $\varphi(x) = \frac{\sqrt{1+x^2}-1}{x^2}$. *The cell-centered/vertex-based AT-LF and AT-DP schemes are stable provided that the time step is smaller than*

| *Scheme* | $(\theta_1=0,\theta_2=0)$ | $(\theta_1=1,\theta_2=0)$ *or* $(\theta_1=0,\theta_2=1)$ | $(\theta_1=1/2,\theta_2=1/2)$ |
|---|---|---|---|
| *AT-LF* ($\kappa_u=0$) | $\frac{\kappa_r}{2}\frac{h}{a_*}$ | $\min\left\{\frac{2}{\omega}, \frac{\kappa_r h}{4a_*}\varphi\left(\frac{\kappa_r\omega h}{4a_*}\right), \frac{4h}{\kappa_r a_*}\varphi\left(\frac{2\omega h}{\kappa_r a_*}\right)\right\}$ | $\frac{\kappa_r h}{a_*}\varphi\left(\frac{\kappa_r\omega h}{2a_*}\right)$ |
| *AT-DP* ($\kappa_r=\kappa_u=\kappa$) | $\frac{2\kappa}{2+\kappa^2}\frac{h}{a_*}$ | $\min\left\{\frac{\kappa}{2+\kappa^2}\frac{h}{a_*}, \frac{1}{\omega}\right\}$ | $\min\left\{\frac{\kappa}{2+\kappa^2}\frac{h}{a_*}, \frac{2}{\omega}\right\}$ |

*Proof.* The proof relies on same kind of computations than Theorem 1. □

**Remark 10.** *Contrary to the result in Theorem 1, for the choice* $\theta_1=\theta_2=1/2$, *the CFL conditions in Prop. 8 still depend on the Coriolis parameter* $\omega$. *The only choice for which the CFL condition does not depend on the Coriolis parameter is a fully implicit discretisation of the Coriolis term, i.e.* $\theta_1=\theta_2=0$.

*We also notice that the stability conditions associated to the AT-LF scheme are more restrictive than the conditions for the LF-DP scheme.*

## 6.2 Orthogonality-preserving property

We now turn to another major aspect of the linear wave equation which is the preservation of the orthogonal subspace, see Prop. 3. It means that when the initial condition is in the orthogonal subspace, the numerical solution remains in this subspace at any time. If the numerical scheme satisfies such a property, we say that this scheme is an *orthogonality-preserving scheme*.

As we shall see below, the original schemes (28) and (29) are not orthogonality-preserving schemes. That is why we have to modify them. To do so, let us change the time discretisation of the velocity divergence on the pressure

equation in the cell-centered scheme as

$$\begin{cases} \dfrac{r_{i,j}^{n+1}-r_{i,j}^{n}}{\Delta t}+a_\star[\nabla_{2h}^{c}\cdot\boldsymbol{u}_h^{\tau}]_{i,j}-\nu_r\left[\nabla_{2h}^{c}\cdot\left(\nabla_{2h}^{c}r_h^n+\dfrac{\omega}{a_*}(\boldsymbol{u}_h^n)^{\perp}\right)\right]_{i,j}=0,\\ \dfrac{\boldsymbol{u}_{i,j}^{n+1}-\boldsymbol{u}_{i,j}^{n}}{\Delta t}+a_\star[\nabla_{2h}^{c}r_h^n]_{i,j}-\nu_u\left[\nabla_{2h}^{c}(\nabla_{2h}^{c}\cdot\boldsymbol{u}_h^n)\right]_{i,j}=-\omega\boldsymbol{u}_{i,j}^{\theta,\perp}, \end{cases} \tag{33}$$

and in the vertex-based scheme as

$$\begin{cases} \dfrac{r_{i,j}^{n+1}-r_{i,j}^{n}}{\Delta t}+a_\star f_h^{\mathrm{c}}\left[\nabla_h^{\mathrm{v}}\cdot\boldsymbol{u}_h^{\tau}\right]_{i,j}-\nu_r\nabla_h^{\mathrm{c}}\cdot\left[\nabla_h^{\mathrm{v}}r_h^n+\omega f_h^{\mathrm{v}}\left((\boldsymbol{u}_h^n)^{\perp}\right)\right]_{i,j}=0,\\ \dfrac{\boldsymbol{u}_{i,j}^{n+1}-\boldsymbol{u}_{i,j}^{n}}{\Delta t}+a_\star f_h^{\mathrm{c}}\left[\nabla_h^{\mathrm{v}}r_h^n\right]_{i,j}-\nu_u\nabla_h^{\mathrm{c}}\left[\nabla_h^{\mathrm{v}}\cdot\boldsymbol{u}_h^n\right]_{i,j}=-\omega f_h^{\mathrm{c}}\left[f_h^{\mathrm{v}}(\boldsymbol{u}_h^{\theta})\right]_{i,j}^{\perp}, \end{cases} \tag{34}$$

for $\boldsymbol{u}_h^{\tau}=(\tau_1u^n+(1-\tau_1)u^{n+1},\tau_2v^n+(1-\tau_2)v^{n+1})^T$ with $\tau_1,\,\tau_2\in[0,1]$. Note that these modified schemes are still explicit since the updated velocity can be computed first and then introduced in the pressure equation.

It is straightforward to prove that these modified schemes still preserve the corresponding discrete kernels (20) and (25). They also preserve the orthogonal subspace:

**Proposition 9.** *The fully discrete cell-centered* (33) *and vertex-based* (34) *schemes are orthogonality-preserving schemes provided that*

$$\kappa_r=0\quad\textit{and}\quad\tau_1=\theta_1,\quad\tau_2=\theta_2. \tag{35}$$

*Proof.* Let us assume that $q_h^n\in\mathcal{E}_{\omega\neq0,h}^{c,\perp}$ and show that $q_h^{n+1}\in\mathcal{E}_{\omega\neq0,h}^{c,\perp}$.

Taking the discrete scalar product of (33) with $\hat{q}_h\in\mathcal{E}_{\omega\neq0,h}^{c}$, we obtain

$$\begin{aligned}\langle q_h^{n+1},\hat{q}_h\rangle=\langle q_h^n,\hat{q}_h\rangle&-a_\star\Delta t\left(\langle\nabla_{2h}^{c}\cdot\boldsymbol{u}_h^{\tau},\hat{r}_h\rangle+\langle\nabla_{2h}^{c}r_h^n,\hat{\boldsymbol{u}}_h\rangle\right)\\&+\nu_u\Delta t\langle\nabla_{2h}^{c}[\nabla_{2h}^{c}\cdot\boldsymbol{u}_h^n],\hat{\boldsymbol{u}}_h\rangle+\nu_r\Delta t\left\langle\nabla_{2h}^{c}\cdot\left[\nabla_{2h}^{c}r_h^n+\frac{\omega}{a_\star}\boldsymbol{u}_h^{n,\perp}\right],\hat{r}_h\right\rangle-\omega\Delta t\langle\boldsymbol{u}_h^{\theta,\perp},\hat{\boldsymbol{u}}_h\rangle.\end{aligned}$$

Because of $\nabla_{2h}^{c}\cdot\hat{\boldsymbol{u}}_h=0$ and due to Lemma 4, we have

$$\begin{aligned}\langle\nabla_{2h}^{c}r_h^n,\hat{\boldsymbol{u}}_h\rangle&=-\langle r_h^n,\nabla_{2h}^{c}\cdot\hat{\boldsymbol{u}}_h\rangle=0,\\ \langle\nabla_{2h}^{c}[\nabla_{2h}^{c}\cdot\boldsymbol{u}_h],\hat{\boldsymbol{u}}_h\rangle&=-\langle\nabla_{2h}^{c}\cdot\boldsymbol{u}_h,\nabla_{2h}^{c}\cdot\hat{\boldsymbol{u}}_h\rangle=0.\end{aligned}$$

Moreover $\langle q_h^n,\hat{q}_h\rangle=0$ and

$$-a_\star\Delta t\,\langle\nabla_{2h}^{c}\cdot\boldsymbol{u}_h^{\tau},\hat{r}_h\rangle=a_\star\Delta t\,\langle\boldsymbol{u}_h^{\tau},\nabla_{2h}^{c}\hat{r}_h\rangle=-\omega\Delta t\langle\boldsymbol{u}_h^{\tau},\hat{\boldsymbol{u}}_h^{\perp}\rangle=\omega\Delta t\langle\boldsymbol{u}_h^{\tau,\perp},\hat{\boldsymbol{u}}_h\rangle.$$

As a result, we obtain

$$\langle q_h^{n+1},\hat{q}_h\rangle=\omega\Delta t\left\langle(\boldsymbol{u}_h^{\tau}-\boldsymbol{u}_h^{\theta})^{\perp},\hat{\boldsymbol{u}}_h\right\rangle+\nu_r\Delta t\left\langle\nabla_{2h}^{c}\cdot\left[\nabla_{2h}^{c}r_h^n+\frac{\omega}{a_\star}\boldsymbol{u}_h^{n,\perp}\right],\hat{r}_h\right\rangle.$$

Therefore, in order to ensure that $\forall\,\hat{q}_h\in\mathcal{E}_{\omega\neq0,h}^{c}$, $\langle q_h^{n+1},\hat{q}_h\rangle=0$, we need $\nu_r=0$ and $\tau_1=\theta_1$, $\tau_2=\theta_2$.

Similarly, for the vertex-based scheme (34), we have

$$\begin{aligned}\langle q_h^{n+1},\hat{q}_h\rangle=\langle q_h^n,\hat{q}_h\rangle-a_\star\Delta t\left(\left\langle f_h^{\mathrm{c}}\left[\nabla_h^{\mathrm{v}}r_h^n\right],\hat{\boldsymbol{u}}_h\right\rangle+\left\langle f_h^{\mathrm{c}}\left[\nabla_h^{\mathrm{v}}\cdot\boldsymbol{u}_h^{\tau}\right],\hat{r}_h\right\rangle\right)+\nu_u\Delta t\left\langle\nabla_h^{\mathrm{c}}\left[\nabla_h^{\mathrm{v}}\cdot\boldsymbol{u}_h^n\right],\hat{\boldsymbol{u}}_h\right\rangle\\+\nu_r\Delta t\left\langle\nabla_h^{\mathrm{c}}\cdot\left[\nabla_h^{\mathrm{v}}r_h^n+\omega f_h^{\mathrm{v}}(\boldsymbol{u}_h^{n,\perp})\right],\hat{r}_h\right\rangle-\omega\Delta t\left\langle f_h^{\mathrm{c}}\left[f_h^{\mathrm{v}}(\boldsymbol{u}_h^{\theta,\perp})\right],\hat{\boldsymbol{u}}\right\rangle\end{aligned}$$

and due to Lemma 7

$$\langle q_h^{n+1},\hat{q}_h\rangle=\nu_r\Delta t\left\langle\nabla_h^{\mathrm{c}}\cdot\left[\nabla_h^{\mathrm{v}}r_h^n+\omega f_h^{\mathrm{v}}(\boldsymbol{u}_h^{n,\perp})\right],\hat{r}_h\right\rangle+\omega\Delta t\left\langle f_h^{\mathrm{c}}\left[f_h^{\mathrm{v}}(\hat{\boldsymbol{u}}_h^{\perp})\right],\boldsymbol{u}_h^{\theta}-\boldsymbol{u}_h^{\tau}\right\rangle.$$

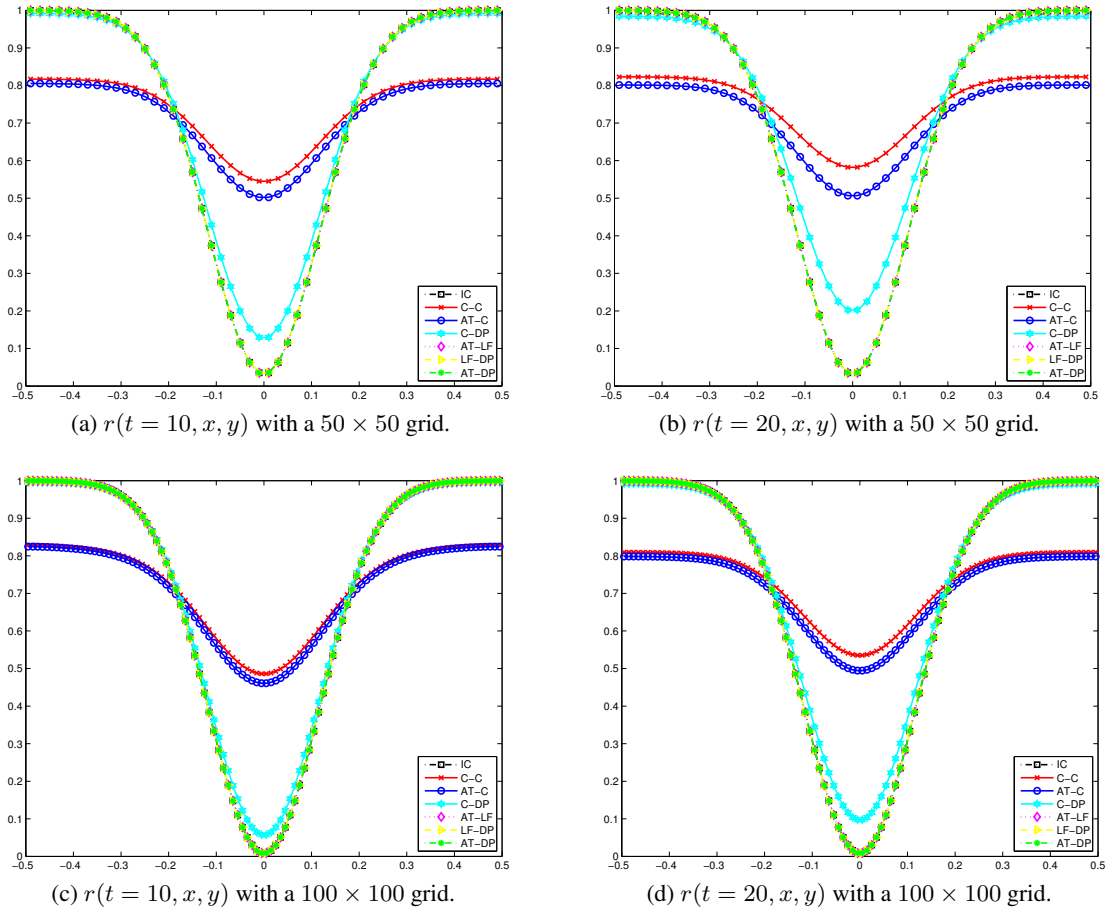
Therefore, under (35), we have $\langle q_h^{n+1},\hat{q}_h\rangle=0$ for any $\hat{q}_h\in\mathcal{E}_{\omega\neq0,h}^{v}$. □

# 7 Numerical results

## 7.1 Well-balanced test case with initial condition in the kernel

We first come back to the test case presented in Section 3 to explain the wrong behaviour of the classical scheme and of the naive corrections referred to as LF-C and C-LF strategies. In practice we define the initial discrete pressure $r$ by using relation (7) applied at the cell centers and the initial discrete velocity by using the definition of the discrete kernel (20). The initial state is then a discrete stationary solution when we use the scheme defined by (28) or (33). As expected, the AT-DP, AT-LF and LF-DP strategies exactly maintain the stationary state, whereas the results obtained with the AT-C and C-DP strategies are very similar to the ones obtained with the LF-C and C-LF strategies and are not able to preserve the stationary state, compare Fig. 6 and 3.

In Fig. 6 we present the results for two different grid sizes and two different final times. It is clear that the error decreases when the mesh is refined and increases with time, that is not surprising. As it has already been noticed, it clearly appears that, for this test case, the correction on the diffusion for the velocity equation, *i.e.* C-DP strategy, has a much larger impact than the correction on the diffusion for the pressure equation, *i.e.* AT-C strategy, but is not enough to preserve the stationary state. This behaviour will be investigated in more details in Section 7.3.




(a) $r(t = 10, x, y)$ with a $50 \times 50$ grid.

(b) $r(t = 20, x, y)$ with a $50 \times 50$ grid.

(c) $r(t = 10, x, y)$ with a $100 \times 100$ grid.

(d) $r(t = 20, x, y)$ with a $100 \times 100$ grid.

Figure 6: Cross-section of pressure.

### 7.2 Orthogonality-preserving test case with initial condition in the orthogonal subspace

In this test case, we consider periodic boundary conditions and an initial velocity field given by

$$\begin{cases} u(t=0,x,y) = \frac{1}{2}\exp\left[-\left(\frac{4x}{0.4}\right)^2 - \left(\frac{4y}{0.8}\right)^2\right] \\ v(t=0,x,y) = \frac{1}{2}\exp\left[-\left(\frac{4x}{0.8}\right)^2 - \left(\frac{4y}{0.4}\right)^2\right]. \end{cases}$$

in the domain $\mathbb{T}^2 = [-0.5, 0.5] \times [-0.5, 0.5]$. Then the initial pressure $r(t = 0, x, y)$ is constructed by using the definition of the discrete orthogonal subspace (21). Note that for this test case, we only present results for the cell-centered scheme (28) for which we can compute explicitly the orthogonal of the kernel. Note that for the vertex-based scheme (29), we can also prove that the LF-DP scheme with the appropriate $u^\tau$ velocity preserves the orthogonal, but we cannot provide an explicit expression for this subspace. The time discretisation parameter for Coriolis term is $\theta_1 = \theta_2 = 1/2$ for all the numerical results. A $50 \times 50$ grid is used.

As expected, Figure 7(a) indicates that, for the choice $\tau = 1$, no scheme is orthogonality-preserving (if ti were the case, the curves would remain exactly zero). Nevertheless it clearly appears that the projection $\hat{q}$ onto the kernel depends on the numerical strategy and is much larger for the C-C and the AT-C schemes than for the other ones. Figure 7(b) shows that the orthogonal part of the solution is less damped when the LF strategy is used, i.e. for AT-LF and LF-DP schemes, since the numerical diffusion is canceled for one equation. In Fig. 7(c) and Fig. 7(d), we present the same results, but focusing on the LF-DP scheme for different values of the parameter $\tau$ used for the time discretisation of the velocity in the pressure equation. It appears on Fig. 7(c) that the case with $\tau = \theta$ is the only one for which the projection of the solution in the kernel remains zero for all time, which means that the orthogonal subspace is stable for the scheme. In Figure 7(d), it appears that the damping increases when the time discretisation becomes more and more implicit, *i.e.* the parameter $\tau$ becomes smaller and smaller. Note that the choice $\tau = \theta = 1/2$ for which the orthogonal is a stable subspace corresponds to a mean damping: contrary to the previous test case, the solution evolves but remains in the orthogonal.

### 7.3 Behaviour of the solution with initial condition close to the kernel

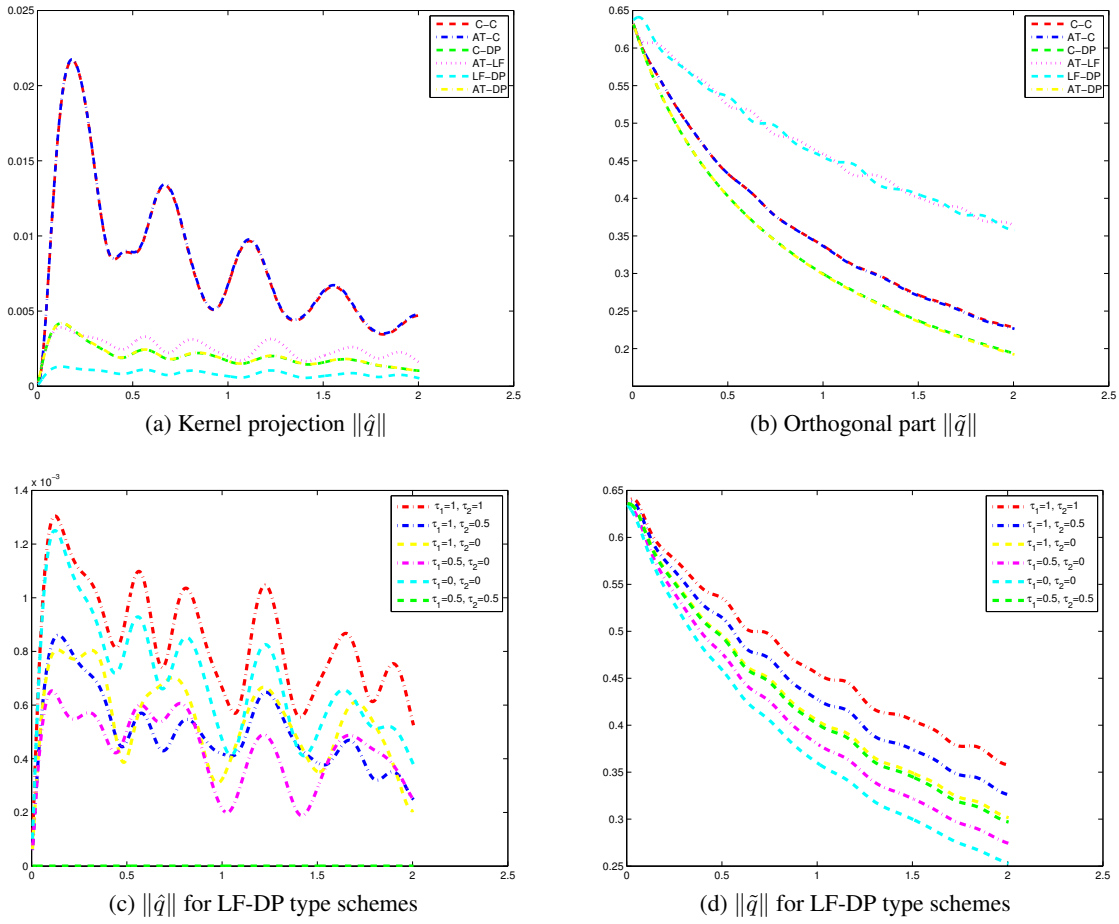
We now consider an initial condition close to the discrete kernel up to a perturbation of size $M \ll 1$

$$q_h^0 = \hat{q}_h^0 + M\frac{\tilde{q}_h^0}{\|\tilde{q}_h^0\|},$$

where $\hat{q}_h^0$ stands for the projection onto the kernel given in Section 7.1 and $\tilde{q}_h^0$ is the orthogonal part considered in Section 7.2. Here the Froude number $M$ is set equal to $10^{-3}$ and a $50 \times 50$ grid is used. In Figure 8(a) we present the evolution in time of the deviation from the initial projection $\hat{q}_h^0$. It appears that for the C-C, AT-C and C-DP schemes, that are not able to maintain steady states, the deviation increases regularly with time. Nevertheless it increases much faster for C-C and AT-C schemes than for C-DP schemes, which reinforces the conclusions of the first numerical example, see Section 7.1. For C-C and AT-C schemes, the deviation becomes almost constant when the discrete solution reaches a stationary state of the scheme, which is very different from the initial one since the kernels of those scheme are inaccurate approximations of the continuous ones, see Lemma 2. The same phenomenon should occur for the C-DP scheme but since the deviation increases slowly, one needs to wait for a long time.

In Fig. 8(b) we present the norm of the part of the solution that belongs to the orthogonal subspace. It appears that for each scheme, it is mostly decreasing in time, despite some oscillations, meaning that, for each scheme, the solution tends to a stationary state that belongs to the kernel of the considered scheme. Note that the solution of the AT-C scheme tends quite quickly to a stationary state in its kernel since the orthogonal part vanishes. For C-C and C-DP schemes, the decreasing of the orthogonal part is slower, which explains that the deviation is still increasing in Fig. 8(a), even if very slowly for large time for the C-C scheme.

In Fig. 9, we present for different values of $M$, the maximum value, over the time interval, of the deviation from the initial projection $\hat{q}_h^0$. It clearly exhibits that, for the well-balanced LF-DP, AT-LF and AT-DP strategies, the deviation




(a) Kernel projection $\|\hat{q}\|$

(b) Orthogonal part $\|\tilde{q}\|$

(c) $\|\hat{q}\|$ for LF-DP type schemes

(d) $\|\tilde{q}\|$ for LF-DP type schemes

Figure 7: Evolution of the kernel and orthogonal part for $\theta_1 = \theta_2 = \frac{1}{2}$.

is proportional to $M$ whereas it remains constant for the other strategies, even if the constant is smaller for the C-DP scheme than for the C-C and AT-C schemes. It emphasizes the importance of the well-balanced strategy to ensure the accuracy near the geostrophic equilibrium.

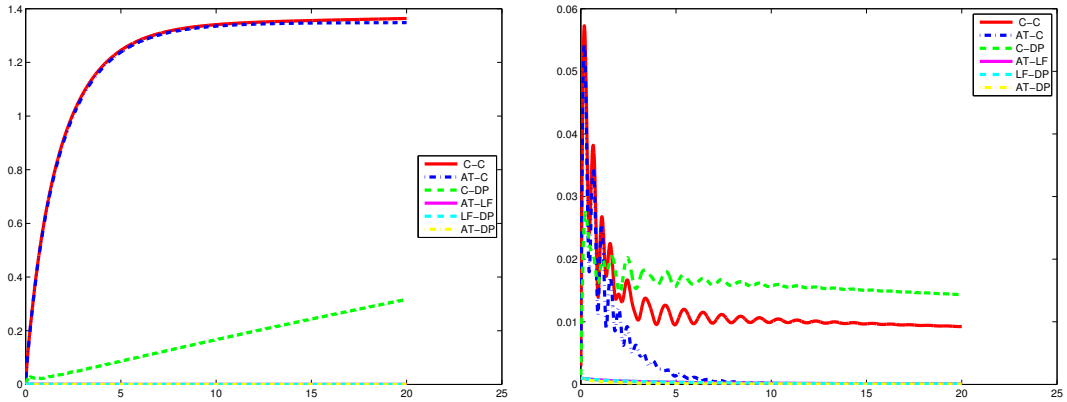
### 7.4 Water column test case and geostrophic adjustment

In this test case, we consider a discontinuous initial condition which is given by

$$\begin{cases} r(t=0,x,y) = \begin{cases} 2, & \text{if} \quad x^2+y^2 \leq 1 \\ 1, & \text{if} \quad x^2+y^2 > 1, \end{cases} \\ u(t=0,x,y) = 0, \\ v(t=0,x,y) = 0. \end{cases}$$

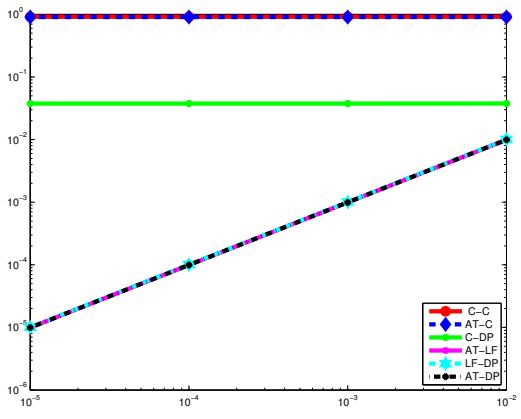
with periodic boundary conditions on the domain $[-5,5] \times [-5,5]$. This initial condition corresponds to a circular dam break and is very far from the geostrophic equilibrium (3). Hence the solution of the wave equation with Coriolis term (2) will contain a travelling wave that should go out of the domain (here due to periodic boundary conditions, the waves remain in the domain but will vanish for long time because of numerical diffusion) and the remaining stationary state will be the geostrophic equilibrium (3) corresponding to the initial data. Discrete solutions will exhibit the same behaviour but the remaining state will belong to the discrete kernel of each scheme.

In Fig. 10, we present the evolution in time of the pressure $r$ for different schemes. In Fig. 10(f), *i.e.* for long time, three groups can be exhibited: the one corresponding to the well-balanced schemes, *i.e.* the LF-DP, AT-LF and AT-DP schemes, the one corresponding to the schemes for which the kernel is given by (9a), *i.e.* the C-C and the




(a) $\|q - \mathbb{P}q^0\|(t)$. (b) $\|q - \mathbb{P}q\|(t)$.

Figure 8: Evolution in time of the deviation for an initial condition close to the discrete kernel.




(a)

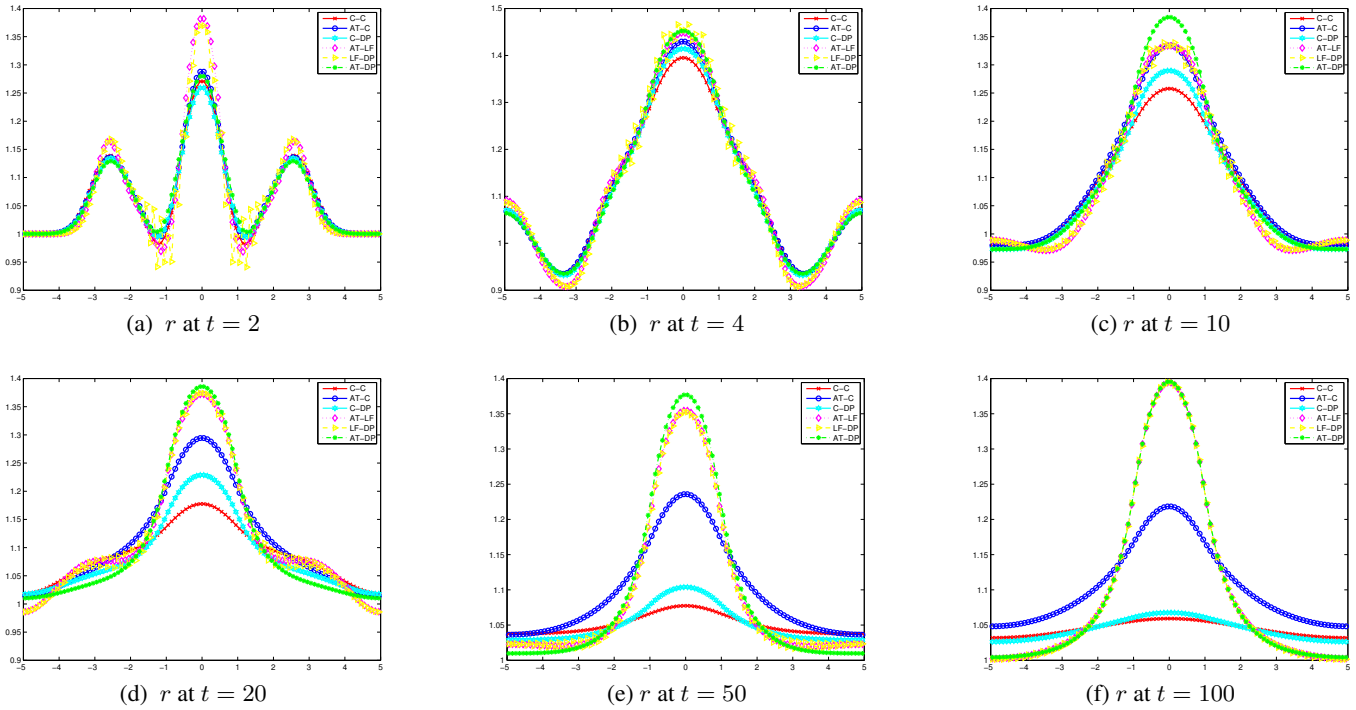
Figure 9: $\max_{t\in[0,2]} \|q - \mathbb{P}q^0\|(t)$ as a function of the Froude number (log-log scale)

C-DP schemes, and the AT-C scheme for which the kernel is given by (9c). In Fig. 12 we present at the final time the results for the quantities $r$, $u$ and $v$ for three schemes, corresponding to the three groups previously mentioned. Results appear to be very different (note the scale is not the same for the three figures).

On the left column, solutions of the C-C and C-DP schemes are close to a constant state (see the scale on the $z$-axis) that corresponds to the discrete kernel (9a). Note that the discrete kernels (9a) and (9b) are the same and correspond respectively to the C-C and C-DP schemes. On the center column, $u$-velocity (*resp.* $v$-velocity) corresponding to the solution of the AT-C scheme is almost constant in the $x$-direction (*resp.* in the $y$-direction). It is in agreement with the definition of the kernel (9c), that is neither a constant state, nor a good approximation of the continuous geostrophic equilibrium (3).

In the right column, solution for the AT-DP scheme is very similar to the geostrophic equilibrium plotted in Fig. 1, which may indicate that the solution is close to the discrete kernel (20), that has been proven to be a good approximation of the continuous geostrophic equilibrium (3). It is clearly exhibited in Fig. 11 where we show that, for long time, the gradient of the pressure along the $x$-axis balances exactly the $x$-component of the Coriolis force, which characterizes the geostrophic equilibrium (the result would be the same for any cross-section in any direction). Among the C-C, AT-C and C-DP schemes, that are not well-balanced, note that, whereas the C-DP scheme appeared to be preferable in the previous test cases since the deviation from the discrete geostrophic equilibrium remained relatively small, here, the solution of the C-DP scheme is very similar to the one of the classical C-C scheme and is totally inaccurate. It allows to conclude that the well-balanced property is absolutely necessary to obtain accurate solutions for a large range of test cases.

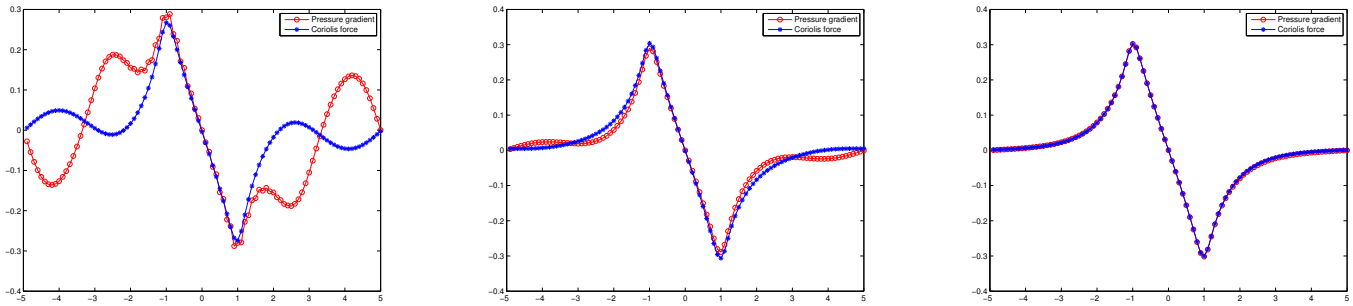
In Fig. 10(a) to 10(e) we present the transient part of this geostrophic adjustment. It appears that the time evolution of the solutions of the three well-balanced schemes, even if they converge to the same state, is not completely similar. In particular the solution for the AT-DP scheme is different from a group composed by the solutions corresponding to the AT-LF and LF-DP schemes. Note also that for short time, the solution of the LF-DP scheme presents some oscillations, that are due to the discontinuity of the initial solution. This difference is highlighted in Fig. 13 where we present the time evolution of the energy. Indeed, even if the final state is the same for the three well-balanced schemes, the time evolution is different for, on the one hand, the AT-DP scheme, and, on the other hand, the AT-LF and the LF-DP schemes, for which the energy decreases more slowly. Nevertheless note that, as expected from Th. 1, the energy is globally decreasing for all schemes, even if we consider a discontinuous initial condition.




(a) $r$ at $t = 2$ (b) $r$ at $t = 4$ (c) $r$ at $t = 10$

(d) $r$ at $t = 20$ (e) $r$ at $t = 50$ (f) $r$ at $t = 100$

Figure 10: Cross section of the pressure $r$ at $y = 0$ at different times.




(a) $t = 4$ (b) $t = 20$ (c) $t = 100$

Figure 11: Cross section of the pressure gradient and Coriolis force at $y = 0$ for AT-DP scheme.

# 8 Conclusion

In this work we propose new collocated finite volume Godunov type schemes to compute accurate approximate solutions of the wave equation with Coriolis term. The main ingredient of the method is to modify the numerical diffusion of the scheme to make the discrete kernel compatible with the so-called geostrophic equilibrium. It extends techniques proposed in [7] and [9]. We propose three different well-balanced schemes, namely the AT-LF (Apparent Topography & Low Froude) scheme, the LF-DP (Low Froude & Divergence Penalisation) scheme and the AT-DP (Apparent Topography & Divergence Penalization) scheme, and two different ways to discretise the geostrophic equilibrium, namely at the centers of the cell or at the interfaces.

The main result of the paper is the proof of stability, under classical CFL conditions, of all these modified schemes, see Th. 1. Moreover some numerical test cases allow us to investigate the behaviour of the schemes for different kinds of initial solutions, including discontinuous ones, and conclude that the well-balanced property is essential to ensure an accurate geostrophic adjustment. Future works will be dedicated to the extension of these results to the fully nonlinear two-dimensional shallow water equations with Coriolis term (1).

# A Proof of the Hodge decomposition in the continuous case (Prop. 1)

*Proof.* In order to prove (4), let us denote by $\mathbb{A}$ the space

$$\mathbb{A} := \left\{ (p, \boldsymbol{v}) \in \left(L^2(\mathbb{T}^2)\right)^3 \;\middle|\; \forall\, \varphi \in C_c^\infty(\mathbb{T}^2),\ \int_{\mathbb{T}^2} a_\star \boldsymbol{v}^\perp \cdot \nabla \varphi \ \mathrm{d}\boldsymbol{x} = \int_{\mathbb{T}^2} \omega p \varphi \ \mathrm{d}\boldsymbol{x} \right\}.$$

We first show that $\mathbb{A}$ is a subset of $\mathcal{E}^\perp_{\omega\neq 0}$. Let us take $\tilde{q} = (p, \boldsymbol{v}) \in \mathbb{A}$. Then for all $q = (r, \boldsymbol{u}) \in \mathcal{E}_{\omega\neq 0}$, we have

$$\begin{aligned}\langle \tilde{q}, q\rangle &= \int_{\mathbb{T}^2} rp \ \mathrm{d}\boldsymbol{x} + \int_{\mathbb{T}^2} \boldsymbol{u}\cdot\boldsymbol{v} \ \mathrm{d}\boldsymbol{x} = \int_{\mathbb{T}^2} rp \ \mathrm{d}\boldsymbol{x} + \frac{a_\star}{\omega}\int_{\mathbb{T}^2} \boldsymbol{v}\cdot\nabla^\perp r \ \mathrm{d}\boldsymbol{x} \\ &= \int_{\mathbb{T}^2} \frac{\omega}{a_\star} pr \ \mathrm{d}\boldsymbol{x} - \int_{\mathbb{T}^2} \boldsymbol{v}^\perp\cdot\nabla r \ \mathrm{d}\boldsymbol{x} = 0.\end{aligned}$$

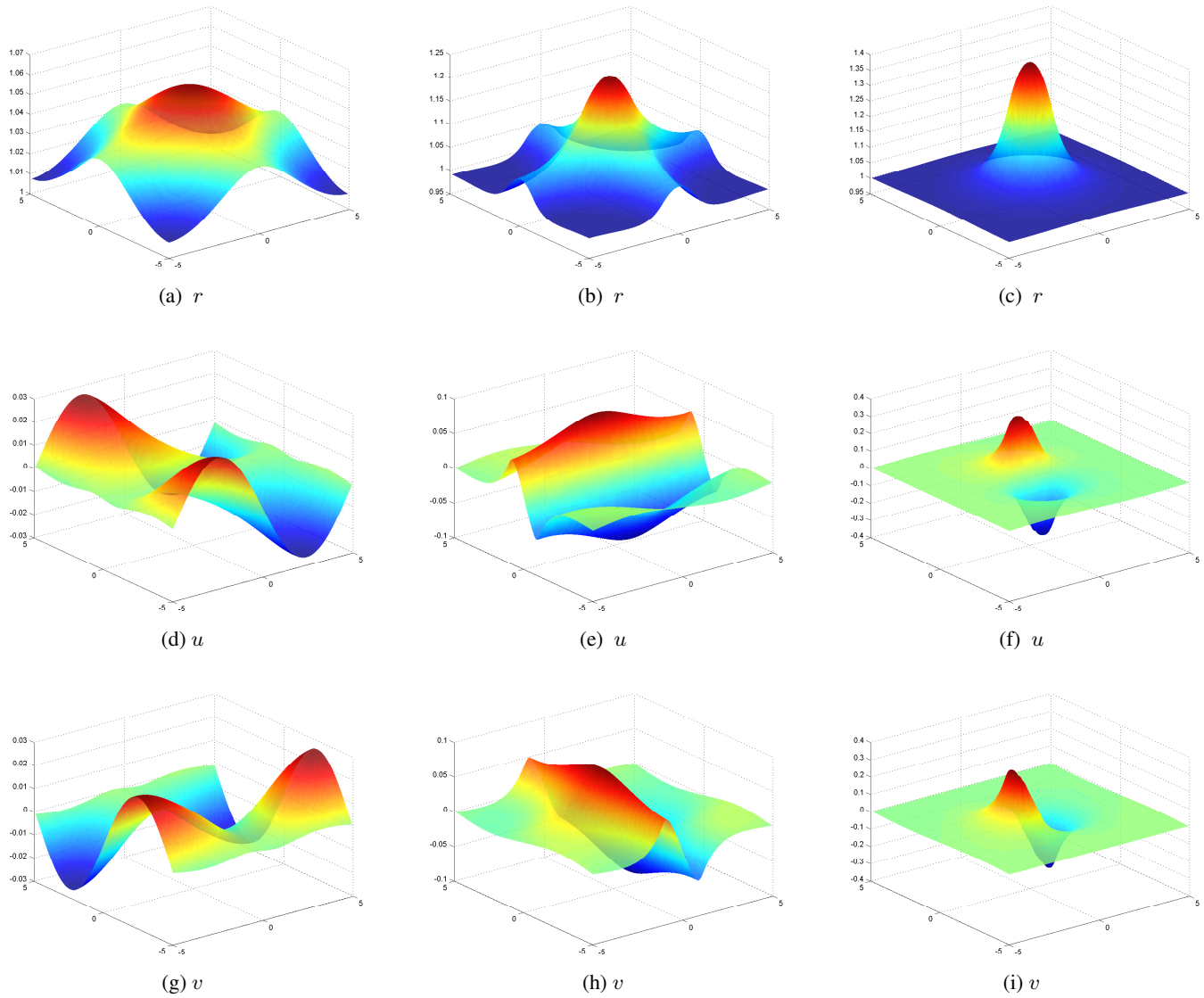
By density of $C_c^\infty(\mathbb{T}^2)$ in $H^1(\mathbb{T}^2)$, it follows that $\tilde{q} = (p, \boldsymbol{v}) \in \mathcal{E}^\perp_{\omega\neq 0}$. Therefore, we conclude that $\mathbb{A} \subset \mathcal{E}^\perp_{\omega\neq 0}$.

On the other hand, let us take $\tilde{q} = (p, \boldsymbol{v}) \in \mathcal{E}^\perp_{\omega\neq 0}$. For any $\phi \in H^1(\mathbb{T}^2)$ we have $\hat{q} := (\frac{\omega}{a_\star}\phi, \nabla^\perp\phi) \in \mathcal{E}_{\omega\neq 0}$ . This provides

$$\langle \tilde{q}, q\rangle = 0 \implies \int_{\mathbb{T}^2} \frac{\omega}{a_\star}\phi p \ \mathrm{d}\boldsymbol{x} - \int_{\mathbb{T}^2} \boldsymbol{v}^\perp\cdot\nabla\phi \ \mathrm{d}\boldsymbol{x} = 0.$$

As a result, we have

$$\forall\ \phi \in H^1(\mathbb{T}^2),\ \int_{\mathbb{T}^2} \frac{\omega}{a_\star}\phi p \ \mathrm{d}\boldsymbol{x} = \int_{\mathbb{T}^2} \boldsymbol{v}^\perp\cdot\nabla\phi \ \mathrm{d}\boldsymbol{x},$$



(a) $r$ (b) $r$ (c) $r$

(d) $u$ (e) $u$ (f) $u$

(g) $v$ (h) $v$ (i) $v$

Figure 12: Comparison between C-C (left), AT-C ( middle) and AT-DP (right) schemes at time $t = 100$.

which leads to

$$\forall\ \phi \in C_c^\infty(\mathbb{T}^2),\ \int_{\mathbb{T}^2} \frac{\omega}{a_\star}\phi p\ \mathrm{d}\boldsymbol{x} = \int_{\mathbb{T}^2} \boldsymbol{v}^\perp \cdot \nabla\phi\ \mathrm{d}\boldsymbol{x}.$$

It implies that $\tilde{q} \in \mathbb{A}$, that is to say $\mathcal{E}^\perp_{\omega\neq 0}$ is a subset of $\mathbb{A}$. In conclusion, we have

$$\mathcal{E}^\perp_{\omega\neq 0} = \mathbb{A} = \left\{ (p,\boldsymbol{v}) \in \left(L^2(\mathbb{T}^2)\right)^3 \,\middle|\, \forall\,\varphi \in C_c^\infty(\mathbb{T}^2),\ \int_{\mathbb{T}^2} a_\star \boldsymbol{v}^\perp \cdot \nabla\varphi\ \mathrm{d}\boldsymbol{x} = \int_{\mathbb{T}^2} \omega p\varphi\ \mathrm{d}\boldsymbol{x} \right\}.$$

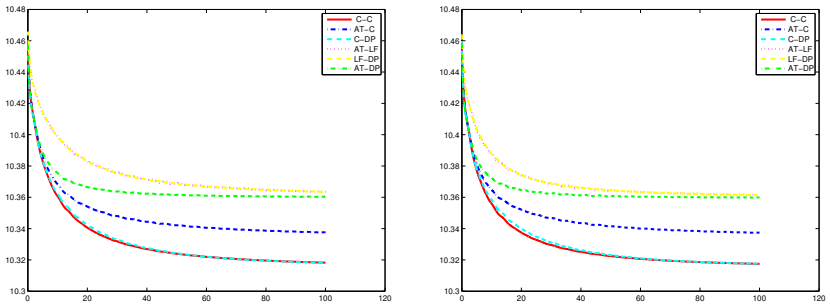
We eventually have to prove that

$$\mathcal{E}_{\omega\neq 0} \oplus \mathcal{E}^\perp_{\omega\neq 0} = \left(L^2(\mathbb{T}^2)\right)^3.$$

By the fact that $\mathcal{E}_{\omega\neq 0} \oplus \mathcal{E}^\perp_{\omega\neq 0} \subset \left(L^2(\mathbb{T}^2)\right)^3$ is trivial, we only have to check $\left(L^2(\mathbb{T}^2)\right)^3 \subset \mathcal{E}_{\omega\neq 0} \oplus \mathcal{E}^\perp_{\omega\neq 0}$ . We suppose $q \in (L^2(\mathbb{T}^2))^3$, we shall find $\hat{q} \in \mathcal{E}_{\omega\neq 0}$ and $\tilde{q} \in \mathcal{E}^\perp_{\omega\neq 0}$ such that $q = \hat{q} + \tilde{q}$. For $q = (r,u,v) \in \left(L^2(\mathbb{T}^2)\right)^3$, let us denote $\mu(r) = \frac{1}{|\mathbb{T}^2|}\int_{\mathbb{T}^2} r\ \ \mathrm{d}\boldsymbol{x}$ and consider the following variational form :

*Find* $h \in H^1(\mathbb{T}^2)$ *such that*: $\forall\ \ \varphi \in H^1(\mathbb{T}^2),\ a(h,\varphi) = F(\varphi)$, where

$$a(h,\varphi) := \int_{\mathbb{T}^2} \nabla h\cdot\nabla\varphi\ \mathrm{d}\boldsymbol{x} + \left(\frac{\omega}{a_\star}\right)^2 \int_{\mathbb{T}^2} h\varphi\ \mathrm{d}\boldsymbol{x},\ F(\varphi) := \frac{\omega}{a_\star}\int_{\mathbb{T}^2} \boldsymbol{u}^\perp\cdot\nabla\varphi\ \mathrm{d}\boldsymbol{x} - \left(\frac{\omega}{a_\star}\right)^2 \int_{\mathbb{T}^2} (r-\mu(r))\varphi\ \mathrm{d}\boldsymbol{x}. \quad (36)$$




(a) Energy of cell center scheme (b) Energy of interface scheme scheme

Figure 13: Evolution in time of the energy

The existence and uniqueness of $h \in H^1(\mathbb{T}^2)$ results from the Lax-Milgram theorem for $\omega \neq 0$. We consider the decomposition for $r$ given by

$$r = \hat{r} + \tilde{r} \quad \text{with} \quad \hat{r} = \mu(r) - h \quad \text{and} \quad \tilde{r} = r - \mu(r) + h.$$

For the decomposition of $\boldsymbol{u}$, we simply construct $\hat{\boldsymbol{u}}$ by setting

$$\hat{\boldsymbol{u}} = \frac{a_\star}{\omega} \nabla^\perp \hat{r} \quad \text{and} \quad \tilde{\boldsymbol{u}} = \boldsymbol{u} - \hat{\boldsymbol{u}},$$

which implies $(\hat{r}, \hat{\boldsymbol{u}}) \in \mathcal{E}_{\omega \neq 0}$ and

$$\hat{\boldsymbol{u}}^\perp = -\frac{a_\star}{\omega} \nabla \hat{r} = \frac{a_\star}{\omega} \nabla \hat{h}.$$

Therefore, (36) implies that for all $\varphi \in H^1(\mathbb{T}^2)$ we have

$$\frac{\omega}{a_\star} \int_{\mathbb{T}^2} (\hat{\boldsymbol{u}} - \boldsymbol{u})^\perp \cdot \nabla \varphi \, \mathrm{d}\boldsymbol{x} + \left( \frac{\omega}{a_\star} \right)^2 \int_{\mathbb{T}^2} \tilde{r} \varphi \, \mathrm{d}\boldsymbol{x} = 0$$

which implies that

$$\forall \, \varphi \in C_c^\infty(\mathbb{T}^2), \ \int_{\mathbb{T}^2} a_\star \tilde{\boldsymbol{u}}^\perp \cdot \nabla \varphi \, \mathrm{d}\boldsymbol{x} = \int_{\mathbb{T}^2} \omega \tilde{r} \varphi \, \mathrm{d}\boldsymbol{x}.$$

□

# References


[1] E. Audusse, F. Bouchut, M.-O. Bristeau, R. Klein, and B. Perthame. A fast and stable well-balanced scheme with hydrostatic reconstruction for shallow water flows. *SIAM J. Sci. Comput.*, 25(6):2050–2065, 2004.

[2] E. Audusse, C. Chalons, and P. Ung. A simple well-balanced and positive numerical scheme for the shallow-water system. *Commun. Math. Sci.*, 13(5):1317–1332, 2015.

[3] E. Audusse, S. Dellacherie, M. H. Do, P. Omnes, and Y. Penel. Godunov type scheme for the linear wave equation with Coriolis source term. In *ESAIM:ProcS.*, 2017.

[4] E. Audusse, M.H. Do, P. Omnes, and Y. Penel. Analysis of apparent topography scheme for the linear wave equation with Coriolis force. In *FVCA VIII. Hyperbolic, Elliptic and Parabolic Problems*, volume 200, pages 209–217, 2017.

[5] P. Azérad and F. Guillén. Mathematical justification of the hydrostatic approximation in the primitive equations of geophysical fluid dynamics. *SIAM J. Math. Anal.*, 33(4):847–859, 2001.

[6] F. Bouchut. *Nonlinear stability of finite volume methods for hyperbolic conservation laws and well-balanced schemes for sources*. Frontiers in Mathematics. Birkhäuser Verlag, Basel, 2004.

[7] F. Bouchut, J. Le Sommer, and V. Zeitlin. Frontal geostrophic adjustment and nonlinear wave phenomena in one-dimensional rotating shallow water. II. High-resolution numerical simulations. *J. Fluid Mech.*, 514:35–63, 2004.

[8] M.J. Castro, J.A. López, and C. Parés. Finite volume simulation of the geostrophic adjustment in a rotating shallow-water system. *SIAM J. Sci. Comput.*, 31(1):444–477, 2008.

[9] S. Dellacherie. Analysis of Godunov type schemes applied to the compressible Euler system at low Mach number. *J. Comput. Phys.*, 229(4):978–1016, 2010.

[10] S. Dellacherie, J. Jung, P. Omnes, and P.-A. Raviart. Construction of modified godunov type schemes accurate at any mach number for the compressible euler system. *Math. Models Methods Appl. Sci.*, 26(13):2525–2615, 2016.

[11] S. Dellacherie, P. Omnes, and F. Rieper. The influence of cell geometry on the Godunov scheme applied to the linear wave equation. *J. Comput. Phys.*, 229(14):5315–5338, 2010.

[12] Christopher Eldred. *Linear and nonlinear properties of numerical methods for rotating shallow water equations*. PhD thesis, Colorado State University, 2015.

[13] H. Guillard and A. Murrone. On the behavior of upwind schemes in the low Mach number limit: II. Godunov type schemes. *Comput. Fluids*, 33(4):655–675, 2004.

[14] H. Guillard and C. Viozat. On the behaviour of upwind schemes in the low Mach number limit. *Comput. Fluids*, 28(1):63–86, 1999.

[15] C. Hu, R. Temam, and M. Ziane. The primitive equations on the large scale ocean under the small depth hypothesis. *Discrete Contin. Dyn. Syst.*, 9(1):97–131, 2003.

[16] R. Klein. Semi-implicit extension of a Godunov-type scheme based on low Mach number asymptotics I: One-dimensional flow. *J. Comput. Phys.*, 121(2):213–237, 1995.

[17] D.Y. Le Roux. Spurious inertial oscillations in shallow-water models. *J. Comput. Phys.*, 231(24):7959–7987, 2012.

[18] M. Lukacova-Medvidova, S. Noelle, and M. Kraft. Well-balanced finite volume evolution Galerkin methods for the shallow water equations. *J. Comput. Phys.*, 221(1):122–147, 2007.

[19] D. Olbers, J. Willebrand, and C. Eden. *Ocean dynamics*. Springer Science & Business Media, 2012.

[20] J. Thuburn and C.J. Cotter. A framework for mimetic discretization of the rotating shallow-water equations on arbitrary polygonal grids. *SIAM J. Sci. Comput.*, 34(3):B203–B225, 2012.

[21] S. Vater and R. Klein. Stability of a cartesian grid projection method for zero Froude number shallow water flows. *Numer. Math.*, 113(1):123–161, 2009.

[22] R.F. Warming and B.J. Hyett. The modified equation approach to the stability and accuracy analysis of finite-difference methods. *J. Comput. Phys.*, 14(2):159–179, 1974.

[23] H. Zakerzadeh. The RS-IMEX scheme for the rotating shallow water equations with the Coriolis force. In *FVCA VIII. Hyperbolic, Elliptic and Parabolic Problems*, pages 199–207, 2017.

[24] V. Zeitlin. *Nonlinear dynamics of rotating shallow water: Methods and advances*, volume 2. Elsevier Science, 2007.