Analysis and development of finite volume methods for the new generation of cubed sphere dynamical cores for the atmosphere

Luan da Fonseca Santos

Report presented to the
Institute of Mathematics and Statistics
of the University of São Paulo
for the Doctor of Science
qualifying examination

Program: Applied Mathematics
Advisor: Prof. Pedro da Silva Peixoto

During the development of this work the author was supported by CAPES and FAPESP (grant number 20/10280-4)

São Paulo
November, 2022

Analysis and development of finite volume methods for the new generation of cubed sphere dynamical cores for the atmosphere

Luan da Fonseca Santos

This is the original version of the
qualifying text prepared by candidate
Luan da Fonseca Santos, as submitted
to the Examining Committee.

Resumo

Luan da Fonseca Santos. **Análise e desenvolvimento de métodos de volumes finitos para modelos da nova geração da dinâmica atmosférica baseados na esfera cubada**. Exame de Qualificação (Doutorado). Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2022.

O modelo atmosférico global FV3 do GFDL-NOAA-USA, inicialmente desenvolvido para malhas do tipo latitude-longitude, foi adaptado para a esfera cubada visando atingir melhor escalabilidade em super-computadores massivamente paralelos. Entretanto, neste tipo de malhas estamos mais sujeitos à problemas como o grid imprinting. Além disso, o modelo carece de algumas propriedades miméticas, que são altamente desejáveis. Este projeto de doutorado propõe-se a analisar as propriedades das discretizações de volumes finitos utilizadas no modelo FV3 na esfera cubada. Iremos investigar como propriedades das células da esfera cubada interferem na precisão dos esquemas numéricos. O estudo irá começar com a implementação de um código para gerar a esfera cubada e calcular os operados discretos do FV3. Então, iremos analisar como a malha interfere nos modelos de advecção e de águas rasas na esfera.

Palavras-chave: Núcleo dinâmico da atmosfera, esfera cubada, volumes finitos.

Abstract

Luan da Fonseca Santos. **Analysis and development of finite volume methods for the new generation of cubed sphere dynamical cores for the atmosphere**. Qualifying Exam (Doctorate). Institute of Mathematics and Statistics, University of São Paulo, São Paulo, 2022.

The global atmospheric model FV3 from GFDL-NOAA-USA, which was originally designed for latitude-longitude grids, was adapted to the cubed sphere aiming to improve its scalability in massively parallel supercomputers. However, in this kind of grid, we are more likely to have grid imprinting problems. Besides that, the FV3 model lacks some highly desirable mimetic properties. This work aims to analyze the properties of the finite volume discretizations employed in the global atmospheric model FV3 on the cubed-sphere. We will investigate how the properties of the cells may impact on the accuracy of the numerical schemes. This study will firstly implement a cubed-sphere grid generator and the FV3 discrete operators on this grid. Then, we will and analyze how the cubed-sphere grid properties influence in the numerical schemes by assessing it using the advection and shallow-water equations on the sphere. We will study the numerical dispersion and conservations properties of the scheme aiming to propose modifications in the numerical schemes to develop a mimetic finite volume version of the model.

Keywords: Dynamical core, cubed-sphere, finite-volume.

Contents

Appendixes

Chapter 1

Introduction

1.1 Background

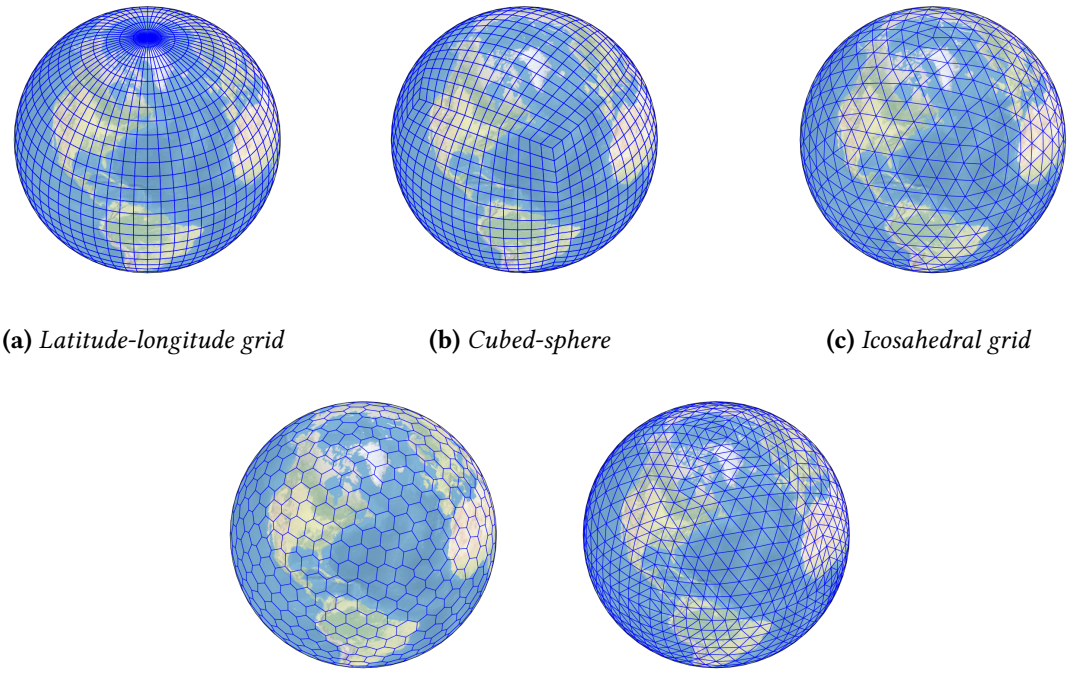
Weather and climate predictions are recognized as a good for mankind, due to the information they yield for diverse activities. For instance, short-range forecasts are useful for public use, while medium-range forecasts are helpful for industrial activities and agriculture. Seasonal forecasts (one up to three months) are important to energy planning and agriculture. At last, longer-range forecasts (one century, for instance) are useful for climate change projections that are important for government planning.

The first global Numerical Weather Prediction models emerged in the 1960s with applications to weather, seasonal and climate forecasts. All these applications are essentially based on the same set of Partial Differential Equations (PDEs) but with distinct time scales (D. L. Williamson, 2007). These PDEs are defined on the sphere and model the evolution of the atmospheric fluid given the initial conditions. One important component of global models is the dynamical core, which is responsible for solving the PDEs that governs the atmosphere dynamics on grid-scale. The development of numerical methods for dynamical cores has been an active research area since the 1960s.

Global models use the sphere as the computational domain and therefore they require a discretization of the sphere. The first global models used the latitude-longitude grid (Figure 1.1a), which is very suitable for finite-differences schemes due to its orthogonality. The major drawback of the latitude-longitude grid is the clustering of points at the poles, known as the "pole problem", which leads to extremely small time steps for explicit-in-time schemes due to the Courant-Friedrichs-Lewy (CFL) condition, making these schemes computationally very expensive.

The most successful method adopted in global atmospheric dynamical cores that overcomes the CFL restriction is the Semi-Implicit Semi-Lagrangian (SI-SL) scheme (Randall et al., 2018), which emerged in the 1980s and consists of the Lagrangian advection scheme applied at each time-step and the solution of fast gravity waves implicitly, allowing very large time steps despite the pole problem. The SI-SL approach combined with finite differences is still used nowadays, for instance in the UK Met Office global model ENDGame (Benacchio & Wood, 2016; Wood et al., 2014). The expensive part of the SI-SL approach is to

solve an elliptic equation at each time step, that comes from the semi-implicit discretization, which requires global data communication, being inefficient to run in massive parallel supercomputers. Besides that, Semi-Lagrangian schemes are inherently non-conservatives for mass, which is critical for climate forecasts (D. L. Williamson, 2007).



(a) *Latitude-longitude grid* **(b)** *Cubed-sphere* **(c)** *Icosahedral grid*

(d) *Pentagonal/Hexagonal grid* **(e)** *Octahedral grid*

Figure 1.1: *Examples of spherical grids: latitude-longitude grid (a) and grids based on Platonic solids (b)-(d).*

The emergence of the Fast Fourier Transform (FFT) in the 1960s with the work from Cooley and Tukey (1965) allowed the computation of discrete Fourier transforms with $N \log(N)$ complexity. The viability of the usage of FFTs for solving atmospheric flows was shown by Orszag (1970), using the barotropic vorticity equation on the sphere, and by Eliasen et al. (1970), using the primitive equations. The spectral transform method expresses latitude-longitude grid values, that represent some scalar field, using truncated spherical harmonics expansions, which consists of Fourier expansions in latitude circles and Legendre functions expansions in longitude circles. The coefficients in the spectral expansions are known as spectral coefficients and are usually thought to live in the so-called spectral space. Given the grid values, the spectral coefficients are obtained by performing a FFT followed by a Legendre Transform (LT). Conversely, given the spectral coefficients, the grid values are obtained by performing an inverse LT followed by an inverse FFT. The main idea of the spectral method is to apply the spectral transform, in order to go the spectral space, and evaluate spatial derivatives in the spectral space, which consists of multiplying the spectral coefficients by constants. Then, the method performs the inverse spectral transform in order to get back to grid space, and the nonlinear terms are treated on the grid space (Krishnamurti et al., 2006).

The spectral transform makes the use of SI-SL methods computationally cheap, since the solution to elliptic problems becomes easy, once the spherical harmonics are eigenfunctions of the Laplacian operator on the sphere. Therefore, the spectral transform method gets faster when combined with the SI-SL approach due to the larger times-steps allowed in this case. Due to these enhancements, the spectral transform dominated global atmospheric modeling (Randall et al., 2018) since the 1980s. Indeed, the spectral method is still used in many current operational Weather Forecasting models such as the Integrated Forecast System (IFS) from European Centre for Medium-Range Weather Forecasts (ECMWF), Global Forecast System (GFS) from National Centers for Environmental Prediction (NCEP) and the Brazilian Global Atmospheric Model (BAM) (Figueroa et al., 2016) from Center for Weather Forecasting and Climate Research [Centro de Previsão de Tempo e Estudos Climáticos (CPTEC)].

With the beginning of the multicore era in the 1990s, the global atmospheric models started to move towards parallel efficiency aiming to run at very high resolutions. Even though the spectral transform expansions have a global data dependency, some parallelization is feasible among all the computations of FFTs, LTs and their inverses (Barros et al., 1995). However, the parallelization of the spectral method requires data transpositions in order to compute FFTs and LTs in parallel. These transpositions demand a lot of global communication using, for instance, the Message Passing Interface (MPI) (Zheng & Marguinaud, 2018). Indeed, the spectral transform becomes the most expensive component of global spectral models when the resolution is increased due to the amount of MPI communications (Müller et al., 2019).

The adiabatic and frictionless continuous equations that govern the atmospheric flow have conserved quantities. Among them, some of the most important are mass, total energy, angular momentum and potential vorticity (Thuburn, 2011). Numerical schemes that are known for having discrete analogous of these conservative properties are known as mimetic schemes. As we pointed out, Semi-Lagrangian schemes lack mass conservation. Nevertheless, these schemes have been employed in dynamical cores for better computational performance. However, dynamical cores should have discrete analogous of the continuous conserved quantities, especially concerning for longer simulation runs.

Aiming for better performance in massively parallel computers and conservation properties, new dynamical cores have been developed since the beginning of the 2000s. Novel spherical grids have been proposed, in order to avoid the pole problem. A popular choice are grids based on Platonic solids (Staniforth & Thuburn, 2012). The construction of these grids relies on a Platonic circumscribed on the sphere and the projection of its faces onto the sphere, which leads to quasi-uniform and more isotropic spherical grids. Some examples of spherical grids based on Platonic solids employed in the new generation of dynamical cores are the cubed-sphere (Figure 1.1b), icosahedral grid (Figure 1.1c), the pentagonal/hexagonal or Voronoi grid (Figure 1.1d) and octahedral grid (Figure 1.1e), which are based on the cube, icosahedron, dodecahedron and octahedron, respectively (Ullrich et al., 2017).

1.2 Motivations

The cubed-sphere became a popular quasi-uniform grid for the new generation of dynamical cores. It was originally proposed by Sadourny (1972) and it was reviseted by Ronchi et al. (1996). Some of the cubed-sphere advantages are: uniformity; quadrilateral structure, making the grid indexing trivial; no overlappings; it is cheap to generate. However, the major drawbacks of the cubed-sphere are: non-orthogonal coordinate system, which leads to metric terms on the differetial operator; discontinuity of the coordinate system at the cube edges, which may generate numerical noise and demands special treatment of discrete operators at the cube edges.

Despite of its drawbacks, the cubed-sphere has been adpoted in some of the new generation dynamical cores. For instance, the cubed-sphere is used in the Community Atmosphere Model (CAM-SE) from the NCAR using spectral elements (Dennis et al., 2012) and in the Nonhydrostatic Unified Model of the Atmosphere (NUMA) from the US Navy using Discontinuous Galerkin methods (Giraldo et al., 2013). The cubed-sphere was also chosen to be used in the next UK Met Office global model using mixed finite elements (Kent et al., 2022). At last, the Finite Volume Cubed-Sphere dynamical core (FV3) from the Geophysical Fluid Dynamics Laboratory (GFDL) and the National Oceanic and Atmospheric Administration (NOAA) (L. M. Harris & Lin, 2013; Putman & Lin, 2007) is another example of new generation dynamical core based on the cubed-sphere.

The FV3 model is an extension of the Finite-Volume dynamical core (FVcore) from latitude-longitude grids to the cubed-sphere. The numerical methods from FVcore started to be developed with the transport scheme from the work Lin et al. (1994), which is based on the piecewise linear scheme from Van Leer (1977). This scheme was later improved, using the Piecewise Parabolic Method (PPM) (Carpenter et al., 1990; Colella & Woodward, 1984) using dimension splitting techniques that guarantee monotonicity and mass conservation, for the transport equation (Lin & Rood, 1996) and the shallow-water equations (Lin & Rood, 1997). An important feature is that the FVcore combines the Arakawa C- and D-grids (Arakawa & Lamb, 1977), where the C-grid values are computed in and intermediate time step. The full global model was then presented by Lin (2004).

The FVcore was adapted to the cubed-sphere grid (Putman, 2007; Putman & Lin, 2007), to reach better performance in parallel computers, leading to the FV3 model. Later, the FV3 also was improved to allow locally refinement grids through grid-nesting or grid-stretching (L. M. Harris & Lin, 2013). Currently, the FV3 model is capable of perfoming hydrostatic and non-hydrostatic atmospheric simulations and it was chosen as the new US global weather prediction model, indeed, it replaced the spectral transform Global Forecast System (GFS) in June, 2019 (Samenow, 2019).

However, a well-known problem that occurs on cubed-sphere models that use low-order numerical methods is the grid imprinting visible due to the coordinate system discontinuity, especially at larger scales, leading to the emergence of a wavenumber 4 pattern. This was reported in the paper of Rančić et al. (2017), where the authors employ a finite-difference numerical scheme on the Uniform Jacobian cubed-sphere using a Arakawa B-grid. The unpublished report from Whitaker (2015) shows grid imprinting in other models, including the FV3. Generally speaking, grid imprinting is the presence of artificial behaviors on

the numerical solution that is associated with the grid employed. It is important to stress out that other quasi-uniform grids may also suffer from grid imprinting. For instance, a popular mimetic method, known as TRiSK, was proposed in the literature by Thuburn et al. (2009) and Ringler et al. (2010) using finite difference and finite volume schemes. This scheme is designed for general orthogonal grids, such as the Voronoi and icosahedral grids, and ensures mass and total energy conservation. This method has been employed in the dynamical core of the Model for Prediction Across Scales (MPAS) from National Center for Atmospheric Research (NCAR) (Skamarock et al., 2012), which intended to work on general Voronoi grids, including locally refined Voronoi grids. However, the TRiSK scheme is a low-order scheme and also suffers from grid imprinting, *i.e.*, geometric properties of the grid, such as cell alignment, interfere with the method accuracy (Peixoto, 2016; Peixoto & Barros, 2013; Weller, 2012). Furthermore, in locally refined Voronoi grids, the scheme may become unstable due to ill-aligned cells and numerical dissipation is needed (Santos & Peixoto, 2021), breaking the total energy conservation of the method.

Despite being chosen as the new US global weather prediction model, there is a lack of numerical studies of the FV3 discretizations in the literature, especially regarding the grid imprinting problem and its mimetic properties. Numerical results for the advection equation on the cubed-sphere using the FV3 dynamical core was presented in Putman and Lin (2007) and some shallow-water simulations were presented in L. M. Harris and Lin (2013), considering cubed-spheres with local refinement through grid nesting. From the work L. M. Harris and Lin (2013) we can notice that the FV3 dynamical lack convergence on the maximum norm for the shallow-water model considering the classical balanced geostrophic flow test case from D. Williamson et al. (1992). The authors attribute these errors to the abrupt change in the grid resolution near the nested grid, but no quantitative results are shown considering the quasi-uniform grid. Many other papers available in the literature use the complete FV3 model in three-dimensional frameworks which make it harder to perform a numerical analysis study due not only to its computational cost but also due to the complexity of three-dimensional atmospheric models. There are no detailed works published in intermediate two-dimensional frameworks, using, for instance, the shallow-water equations on the sphere. Even though the advection equation on the sphere plays a key role in the dynamical core development, since it models the transport of scalar fields on the sphere, important features captured by the shallow-water equations on the sphere, such as the Coriolis effect, inertia-gravity waves, geostrophic adjustment, Rossby waves, among others, are not captured by a simple advection model. Hence, shallow-water equations provide an excellent benchmark to assess dynamical cores in general, since it is only two-dimensional but is a complex enough geophysical model for atmosphere dynamics.

1.3 Goals

The aim of this work is to fill the gap in the literature regarding numerical studies of the FV3 discrete operators that we pointed out before. More explicitly, the goals of this work are:

- Investigate the occurrence of grid imprinting on the cubed-sphere using the advection equations and the shallow-water equations on the sphere;

- Propose improvements on the FV3 discrete operators and modifications on the cubed-sphere that alleviate grid imprinting;
- Investigate how we can add more mimetic properties to the FV3 discretizations.

1.4 Outline

This report is outlined as follows. Chapter 2 is dedicated to review the Piecewise Parabolic Method (PPM) for the one-dimensional advection equation. Chapter 3 reviews the dimension splitting method, which allow us to use one-dimensional methods, such as the PPM, to solve the two-dimensional advection equation. Chapter 4 introduces the cubed-sphere grid and shows some of its geometric properties. Chapter 5 extends the ideas of Chapter 3 to the cubed-sphere grid. The dimension-splitting method on each cubed-sphere panel works as in the plane, with the addtion of metric terms, due to non-orthogonality of the grid, and interpolation between panels to obtain ghost cells values needed for stencil computations.

Chapter 2

One-dimensional finite-volume methods

The aim of this chapter is to give a detailed description of the celebrated Piecewise-Parabolic Method (PPM) proposed by Colella and Woodward (1984). As we shall see, the PPM is a one-dimensional finite-volume method for hyperbolic conservation laws that at each time step requires two tasks. The first task may be stated as: given the estimates of average values of the conservation laws solution, find a Piecewise-Parabolic function that approximates the function and preserves its local integral value (also referred as local mass). The second task is the following: given the Piecewise-Parabolic approximation (also known as reconstruction), solve the conservation law using the parabolas to obtain the solution at the next time-step. For instance, if the conservation law is the advection equation, the second step consists of advecting the parabolas. In the first step, we may also require some monotonization constraints on the parabolas, to ensure that no new extreme value is created in the Piecewise-Parabolic reconstruction, ensuring that the scheme is free of numerical oscillations. The steps required for PPM make it an REA (reconstruct, evolve, and average) algorithm, or also referred to as a Godonuv-type method, which was originally proposed by Godunov (1959).

The PPM approach has become popular in the literature for gas dynamics simulations, astrophysical phenomena modeling (Woodward, 1986) and later on atmospheric simulations (Carpenter et al., 1990). Indeed, the PPM has been implemented in the FV3 dynamical core on its latitude-longitude grid (Lin, 2004) and cubed-sphere (Putman & Lin, 2007) versions. We point out that the reconstruction function may be built using other basis functions rather than parabolas. In fact, PPM may be thought of as an extension of the Piecewise-Linear method from Van Leer (1977), which, on the other hand, was inspired by the Piecewise-Constant method attributed to Godunov (1959). Besides that, other schemes inspired by PPM were proposed in the literature using higher-order polynomials, such as quartic polynomials (White & Adcroft, 2008). For a review of general piecewise-polynomial reconstruction we refer to the technical report from Engwirda and Kelley (2016), Lauritzen et al. (2011) and the references therein. Even though many other shapes for the basis functions are available in the literature, as well higher order schemes, L. Harris et al. (2021) points out that the PPM scheme suits well the FV3 needs in the sense of being a

flexible method that can be modified to ensure low diffusivity or shape-preservation, for example. Besides that, a finite-volume numerical method usually requires monotonicity constraints, which by Godunov's theorem, limits the order of convergence to at most 1. So, a higher-order scheme needs to be well-balanced on the trade-off between computational cost increasing and potential benefits.

This chapter starts with a basic review of one-dimensional conservation laws in the integral form in Section 2.1, and in Section 2.2 we set the framework of general one-dimensional finite-volumes schemes, where we also introduce concepts such as consistency, convergence and stability. Section 2.3 describes the PPM method and its convergence order analysis of its reconstruction in given in Subsection 2.3.1. Subsection 2.3.2 is dedicated to introducing possible ways to monotonizate the parabolas. Subsection 2.3.3 is dedicated to the description and investigation of the PPM flux computation considering the one-dimensional advection equation as the conservation law. Section 2.4 shows some numerical results using the PPM scheme for the advection equation. At last, Section 2.5 presents some conclusions. The usage of PPM to solve two-dimensional problems will be addressed in Chapter 3.

2.1 One-dimensional system of conservation laws in integral form

In this section, we are going to present the derivation of one-dimensional system of conservation laws in the integral form. The derivation presented here follows LeVeque (1990) and LeVeque (2002) closely and will be useful to fix some notation. Let us assume that x and t represent the spatial and time coordinates, respectively. Given $[x_1, x_2] \subset \mathbb{R}$, $x_1 \leq x_2$, and a time interval $[t_1, t_2] \subset]0, +\infty[$, $t_1 \leq t_2$, we aim to describe how m state variable densities given by functions $q_1, \cdots, q_m : \mathbb{R} \times [0, +\infty[\to \mathbb{R}$ evolve within time in the considered time interval, assuming that we have neither sinks nor sources for the mass of each state variable and also assuming that the mass flow rate is known for all the state variables.

To set the problem in more mathematical terms, let us denote by $q : \mathbb{R} \times [0, +\infty[\to \mathbb{R}^m$, $q = q(x, t)$, the vector of state variables, *i.e.*, $q_k = q_k$ for $k = 1, \cdots, m$. The mass of q in $[x_1, x_2]$ at time t is defined by:

$$M_{[x_1,x_2]}(t) := \int_{x_1}^{x_2} q(x,t)\, dx \in \mathbb{R}^m. \tag{2.1}$$

Thus, the mass in $[x_1, x_2]$ of the k-th state variable q_k is equal to $(M_{[x_1,x_2]}(t))_k$, $\forall k = 1, \cdots, m$. We are going to assume the following physical constraints concerning the total mass of each state variable:

1. No mass is created;
2. No mass is destroyed.

Also, let us assume that the mass flow rate in a point x and at a time $t > 0$ is given by $f(q(x,t))$, where $f : \mathbb{R}^m \to \mathbb{R}^m$ is a continuously differentiable (C^1) function. This function

f is known as flux function. With the physical constraints that we imposed, the following equation must hold for the mass:

$$\frac{d}{dt}\left(\int_{x_1}^{x_2} q(x,t)\,dx \right) = f(q(x_1,t)) - f(q(x_2,t)). \tag{2.2}$$

Equation (2.2) is known as a conservation law written in integral form and tell us how the mass $M_{[x_1,x_2]}(t)$ varies with time. Another integral form of the conservation law may be obtained integrating Equation (2.2) with respect to time in $[t_1, t_2]$ leading to:

$$\int_{x_1}^{x_2} q(x,t_2)\,dx = \int_{x_1}^{x_2} q(x,t_1)\,dx + \int_{t_1}^{t_2} f(q(x_1,t))\,dt - \int_{t_1}^{t_2} f(q(x_2,t))\,dt. \tag{2.3}$$

Assuming that q is a C^1 function, we may write:

$$\int_{t_1}^{t_2} \frac{\partial q}{\partial t}(x,t)\,dt = q(x,t_2) - q(x,t_1), \tag{2.4}$$

and

$$\int_{x_1}^{x_2} \frac{\partial f}{\partial x}(q(x,t))\,dx = f(q(x_2,t)) - f(q(x_1,t)). \tag{2.5}$$

Replacing Equations (2.4) and (2.5) in (2.3) we get the differential form of the conservation law:

$$\int_{t_1}^{t_2} \int_{x_1}^{x_2} \left(\frac{\partial q}{\partial t}(x,t) + \frac{\partial f}{\partial x}(q(x,t)) \right) dx\,dt = 0. \tag{2.6}$$

Since Equation (2.6) must hold for all x_1, x_2, t_1 and t_2 such that $[x_1, x_2] \times [t_1, t_2] \subset \mathbb{R}\times]0, +\infty[$, we obtain the differential form of the conservation law:

$$\frac{\partial q}{\partial t}(x,t) + \frac{\partial f}{\partial x}(q(x,t)) = 0, \quad \forall (x,t) \in \mathbb{R}\times]0, +\infty[. \tag{2.7}$$

We shall assume that the eigenvalues of the Jacobian matrix of the flux function $Df(q)$ are all real and that $Df(q)$ is a diagonalizable matrix, $\forall q \in \mathbb{R}^m$, so that Equation (2.7) is a hyperbolic partial differential equation (LeVeque, 1990). As we will specify latter, some initial conditions will also be supposed to be known as well.

Many physically relevant equations may be written as Equation (2.7). Some examples are the Euler equations for gas dynamics, obtained when $m = 3$, and the one-dimensional shallow-water equations, obtained $m = 2$. Other relevant equation is the Burgers equation, which is obtained when $m = 1$ and $f(q) = q^2$. The Burgers equation is well known for developing shocks, even for smooth initial conditions, and is a simple prototype to study shock formation. At last, the linear advection equation is another interesting example, which is obtained when $m = 1$ and $f(q(x,t)) = u(x,t)q(x,t)$, where $u(x,t)$ is a given velocity. Strictly speaking, the linear advection is not in the form given by the Equation (2.7) since f depends on q but also on (x,t). But, one may check that Equation (2.7) is still hyperbolic in this case. The linear advection equation will play a key role in this work due

to its importance to the development of atmospheric dynamical cores.

We say that q is a strong or classical solution to the conservation law (2.7) if it is C^1 and satisfies the Equation (2.7). Applying the steps from Equation (2.3) to Equation (2.7) in reverse order, one may check that if q is a strong solution, then it satisfies the integral form (2.3) for all x_1, x_2, t_1 and t_2 such that $[x_1, x_2] \times [t_1, t_2] \subset \mathbb{R}\times]0, +\infty[$. Therefore, Equations (2.3) and (2.7) are equivalent when q is C^1. However, the problem (2.3) can be formulated to functions that are not C^1 and have discontinuities. More generally speaking, we say that $q \in L^\infty(D, \mathbb{R}^m)$ [1] if it satisfies the Equation (2.3) for all x_1, x_2, t_1 and t_2 such that $[x_1, x_2]\times[t_1, t_2] \subset \mathbb{R}\times]0, +\infty[$. It can be shown that this notion of weak solution is equivalent to requiring that (LeVeque, 1990):

$$\int_{-\infty}^{+\infty} \int_{0}^{+\infty} \left(\frac{\partial \phi}{\partial t}(x,t) q(x,t) + \frac{\partial \phi}{\partial x}(x,t) f(q(x,t)) \right) dt\, dx = \int_{-\infty}^{+\infty} \phi(x,0) q(x,0)\, dx, \tag{2.8}$$

$\forall \phi \in C_0^1(\mathbb{R}\times[0, +\infty[)$ where $C_0^1(\mathbb{R}\times[0, +\infty[)$ denotes the set of all continuously differentiable functions with compact support in $\mathbb{R} \times [0, +\infty[$. This formulation of weak solution is more commonly employed in the construction of Discontinuous Galerkin methods (Nair et al., 2011).

In order to develop finite-volume methods for a system of conservation laws, it is useful to define the vector of average values of the state variable vector q in the interval $[x_1, x_2]$ at a time t by:

$$Q(t) = \frac{1}{\Delta x} \int_{x_1}^{x_2} q(x,t)\, dx \in \mathbb{R}^m, \tag{2.9}$$

where $\Delta x = x_2 - x_1$. The Equation (2.2) may be rewritten in terms of Q as:

$$\frac{dQ}{dt}(t) = \frac{1}{\Delta x}(f(q(x_1,t)) - f(q(x_2,t))), \tag{2.10}$$

and so is Equation (2.3):

$$Q(t_2) = Q(t_1) + \frac{1}{\Delta x}\left(\int_{t_1}^{t_2} f(q(x_1,t))\, dt - \int_{t_1}^{t_2} f(q(x_2,t))\, dt \right). \tag{2.11}$$

To move towards finite volume schemes, we will restrict our attention to a conservation law in a bounded domain of the form $D = [a, b] \times [0, T]$, $a < b, T > 0$. However, we must impose some boundary conditions. One possible way that we will adopt in the text are the periodic boundary conditions:

$$q(a,t) = q(b,t), \quad \forall t \in [0,T]. \tag{2.12}$$

Also, we assume that an initial condition $q_0(x) = q(x, 0)$, $q_0 \in L^\infty([a, b], \mathbb{R}^m)$, is given. Thus, we have specified a Cauchy problem. We notice that Equations (2.10) and (2.11) hold for all x_1, x_2, t_1 and t_2 such that $[x_1, x_2] \times [t_1, t_2] \subset D$. So, let us discretize the domain D and write Equations (2.10) and (2.11) in terms of this discretization. Given a positive integer N_T,

[1] $L^\infty(D, \mathbb{R}^m) = \{q : D \to \mathbb{R}^m$ such that q is bounded.$\}$

we define the time step $\Delta t = \frac{T}{N_T}$, $t^n = n\Delta t$, for $n = 0, 1, \cdots, N_T$. For the spatial discretization, we consider a uniformly spaced partition of $[a, b]$ given by:

$$[a, b] = \bigcup_{i=1}^{N} X_i, \text{ where } X_i = [x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}] \text{ and } a = x_{\frac{1}{2}} < x_{\frac{3}{2}} < \cdots < x_{N-\frac{1}{2}} < x_{N+\frac{1}{2}} = b. \quad (2.13)$$

Each interval X_i is referred to as the control volume. We shall use the notations $\Delta x = x_{i+\frac{1}{2}} - x_{i-\frac{1}{2}}$ and $x_i = \frac{1}{2}(x_{i+\frac{1}{2}} + x_{i-\frac{1}{2}})$, $\forall i = 1, \cdots, N$, to define the control volume length and centroid, respectively. We also denote by $Q_i(t) \in \mathbb{R}^m$ as the vector of average values of state variable vector at time t in the control volume X_i, $\forall i = 1, \cdots, N$. Replacing t_1, t_2, x_1 and x_2 by $t^n, t^{n+1}, x_{i-\frac{1}{2}}$ and $x_{i+\frac{1}{2}}$, respectively, in Equation (2.10), we get:

$$\frac{dQ_i}{dt}(t) = \frac{1}{\Delta x}(f(q(x_{i-\frac{1}{2}}, t)) - f(q(x_{i+\frac{1}{2}}, t))), \quad \forall i = 1, \cdots, N. \quad (2.14)$$

Similarly, Equation (2.11) becomes:

$$Q_i(t^{n+1}) = Q_i(t^n) + \frac{1}{\Delta x}\left(\int_{t^n}^{t^{n+1}} f(q(x_{i-\frac{1}{2}}, t))\, dt - \int_{t^n}^{t^{n+1}} f(q(x_{i+\frac{1}{2}}, t))\, dt \right), \quad \forall i = 1, \cdots, N, \quad \forall n = 0, \cdots, N_T - 1. \quad (2.15)$$

In order to use a more compact notation, it is helpful to use the following centered difference notation:

$$\delta_x g(x_i, t) = g(x_{i+\frac{1}{2}}, t) - g(x_{i-\frac{1}{2}}, t), \quad (2.16)$$

for an arbitrary vector valued function g. Using this notation, Equations (2.14) and (2.15) lead to:

$$\frac{dQ_i}{dt}(t) = -\frac{1}{\Delta x}\delta_x f(q(x_i, t)) \quad \forall i = 1, \cdots, N, \quad (2.17)$$

and

$$Q_i(t^{n+1}) = Q_i(t^n) - \frac{\Delta t}{\Delta x}\delta_x \left(\frac{1}{\Delta t} \int_{t^n}^{t^{n+1}} f(q(x_i, t))\, dt \right), \quad \forall i = 1, \cdots, N, \quad \forall n = 0, \cdots, N_T - 1, \quad (2.18)$$

respectively. It is worth pointing out that we have made no approximation in Equations (2.17) and (2.18). Indeed, if q satisfies Equation (2.2), $\forall [x_1, x_2] \subset [a, b]$ and $\forall t \in [0, T]$, then Equation (2.17) is just Equation (2.2) evaluated in the control volumes and written in terms of the average values Q. Similarly, if q satisfies Equation (2.3), $\forall [x_1, x_2] \times [t_1, t_2] \subset D$, then Equation (2.18) is just Equation (2.3) evaluated in the control volumes, at the time instants t^n, and written in terms of the average values Q.

Notice that in Equation (2.18) we divided and multiplied by Δt, so that we can interpret $\frac{1}{\Delta t}\int_{t^n}^{t^{n+1}} f(q(x_i, t))\, dt$ as a mean-time average flux. This interpretation is very handy for the derivation of finite-volume schemes.

The formulations given by Equations (2.17) and (2.18) are the cornerstone of the development of finite volume methods for conservation laws. On the right-hand side of

Equation (2.17), the flux function f may be discretized leading to an ordinary differential equation (ODE) that might be solved using classical ODE integrators. These methods are known as semi-discrete methods (LeVeque, 2002), since only the spatial coordinate is discretized. In this work, we shall restrict our attention to methods based on Equation (2.18), even though the PPM approach is applicable for semi-discrete methods (*e.g.* Suresh and Huynh (1997)).

2.2 The finite-volume approach

We summarize the problem of the system of conservation laws in the integral form discussed in Section 2.1 in Problem 2.1. For simplicity, hereafter we shall constrain our attention to the one-dimensional advection equation, that is, we are going to assume $m = 1$ and that the flux function has the form $f(q(x,t)) = u(x,t)q(x,t)$, where $u(x,t)$ is the velocity which is assumed to be given.

2.2.1 Discretization of the problem

Problem 2.1. *Given $D = [a,b] \times [0,T]$, a C^1 velocity function $u : D \to \mathbb{R}$, we would like to find a weak solution $q \in L^\infty(D, \mathbb{R})$ of the advection equation in the integral form:*

$$\int_{x_1}^{x_2} q(x,t_2)\,dx = \int_{x_1}^{x_2} q(x,t_1)\,dx + \int_{t_1}^{t_2} f(q(x_1,t))\,dt - \int_{t_1}^{t_2} f(q(x_2,t))\,dt,$$

$\forall [x_1,x_2] \times [t_1,t_2] \subset D$, given the initial condition $q(x,0) = q_0(x)$, $\forall x \in [a,b]$, assuming periodic boundary conditions, i.e., $q(a,t) = q(b,t)$, $\forall t \in [0,T]$, and $f(q(x,t)) = u(x,t)q(x,t)$.

We point out that, for Problem 2.1, the total mass in $[a,b]$ satisfies:

$$M_{[a,b]}(t) = M_{[a,b]}(0), \quad \forall t \in [0,T]. \tag{2.19}$$

This is the conservation of total mass propriety and is highly desirable for any numerical scheme that intends to give a robust approximation of a system of conservation laws solution. In Section 2.1 we introduced a version of Problem 2.1 considering a discretization of the domain D. This idea is summarized in Problem 2.2.

Problem 2.2. *Assume the framework of Problem 2.1. We consider positive integers N and N_T, a spatial discretization of [a,b] given by $X_i = [x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}], \forall i = 1, \cdots, N$, $a = x_{\frac{1}{2}} < x_{\frac{3}{2}} < \cdots < x_{N-\frac{1}{2}} < x_{N+\frac{1}{2}} = b$, $\Delta x = x_{i+\frac{1}{2}} - x_{i-\frac{1}{2}}$, a time discretization $t^n = n\Delta t$, $\Delta t = \frac{T}{N_T}$, $\forall n = 0, \cdots, N_T$. Let us also assume that the values of N and N_T are always chosen in such a way that $N_t = c \cdot N$, for some c fixed and therefore $\frac{\Delta t}{\Delta x} = \sigma$ for an also fixed $\sigma \in \mathbb{R}$. Since we are in the framework of Problem 2.1, it follows that:*

$$Q_i(t^{n+1}) = Q_i(t^n) - \sigma \delta_x \left(\frac{1}{\Delta t} \int_{t^n}^{t^{n+1}} f(q(x_i,t))\,dt \right), \quad \forall i = 1, \cdots, N, \quad \forall n = 1, \cdots, N_T - 1, \tag{2.20}$$

where $Q_i(t) = \frac{1}{\Delta x} \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} q(x,t)\,dx$.

Our problem now consists of finding the values $Q_i(t^n)$, $\forall i = 1, \cdots, N$, $\forall n = 0, \cdots, N_T - 1$,

given the initial values $Q_i(0)$, $\forall i = 1, \cdots N$. In other words, we would like to find the average values of q in each control volume X_i at the considered time instants.

Finally, we define the one-dimensional (1D) finite-volume (FV) scheme problem as follows in Problem 2.3. We use the notation $q_i^n = q(x_i, t^n)$ to represent the values of q on the discretization of domain D and $u_{i+\frac{1}{2}}^n = u(x_{i+\frac{1}{2}}, t^n)$ to represent the velocity at the control volume edges.

Problem 2.3 (1D-FV scheme)**.** *Assume the framework defined in Problem 2.2. The finite-volume approach of Problem 2.2 consists of a finding a scheme of the form:*

$$Q_i^{n+1} = Q_i^n - \sigma \delta_i F_i^n, \quad \forall i = 1, \cdots, N, \quad \forall n = 0, \cdots, N_T - 1, \tag{2.21}$$

where $\delta_i F_i^n = F_{i+\frac{1}{2}}^n - F_{i-\frac{1}{2}}^n$ and $Q_i^n \in \mathbb{R}^m$ is intended to be an approximation of $Q_i(t^n)$ in some sense. We define by $Q_i^0 = Q_i(0)$ or $Q_i^0 = q_i^0$. The term $F_{i+\frac{1}{2}}^n = \mathcal{F}(Q^n, u^n; i)$, is known as numerical flux, where $\mathcal{F}$ is the numerical flux function, and it approximates $\frac{1}{\Delta t}\int_{t^n}^{t^{n+1}} f(q(x_{i+\frac{1}{2}}, t))\, dt$, $\forall i = 0, 1, \cdots, N$, or, in other words, it estimates the time-averaged fluxes at the control volume X_i boundaries.

Remark 2.1. *Notice that in the previous problem, we are using the notations $Q^n = (Q_1^n, \cdots, Q_N^n)$, $u^n = (u_{\frac{1}{2}}^n, \cdots, u_{N+\frac{1}{2}}^n)$.*

Remark 2.2. *A scheme of the form from Equation (2.21) is referred to as a 1D-FV scheme and it is also known as a conservative scheme.*

Remark 2.3. *When computing the numerical fluxes, we need values of Q_i that are out of the range $1, \cdots, N$. Since we are under the assumption of periodic boundary conditions, this problem is overcome by assuming periodicity on the data Q.*

For a 1D-FV the discrete total mass at the time-step n is given by

$$M^n = \Delta x \sum_{i=1}^{N} Q_i^n.$$

Therefore, the discrete total mass is constant for a 1D-FV scheme, which follows from a straightforward computation:

$$\begin{aligned} M^{n+1} &= \Delta x \sum_{i=1}^{N} Q_i^{n+1} \\ &= M^n - \Delta t \sum_{i=1}^{N} (F_{i+\frac{1}{2}}^n - F_{i-\frac{1}{2}}^n) \\ &= M^n - \Delta t (F_{N+\frac{1}{2}}^n - F_{\frac{1}{2}}^n) \\ &= M^n, \end{aligned}$$

where we are using that $F_{N+\frac{1}{2}}^n = F_{\frac{1}{2}}^n$, since we are assuming periodic boundary conditions.

2.2.2 Consistency and convergence

Before moving to the definition of convergence, we point out an important relation between the average values of q and its value at the cell centroids. We mentioned in Problem 2.3 that the initial condition may be considered as q_i^0 instead of $Q_i(0)$. Furthermore, when analyzing the convergence of a 1D-FV scheme, we may want to compare Q_i^n with q_i^n since $Q_i(t^n)$ requires the computation of an analytical integral, which may be too complicated to obtain in some cases. In the following Proposition 2.1, we give a simple proof of that q_i^n approximates Q_i^n with second order error when q is twice continuously differentiable.

Proposition 2.1. *If $q \in C^2$, then $Q_i(t^n) - q_i^n = C_1 \Delta x^2$, where C_1 is a constant that depends only on q.*

Proof. From Taylor's expansion, it follows that, for $x \in X_i$, we have:

$$q(x, t^n) = q(x_i, t^n) + \frac{\partial q}{\partial x}(x_i, t^n)(x - x_i) + \frac{\partial^2 q}{\partial x^2}(\xi, t^n)\frac{(x - x_i)^2}{2}, \tag{2.22}$$

for some ξ between x and x_i. Therefore:

$$\begin{aligned} Q_i(t^n) - q_i^n &= \frac{1}{\Delta x}\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} q(x, t^n)\, dx - q(x_i, t^n) \\ &= \frac{1}{\Delta x}\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} \left(\frac{\partial q}{\partial x}(x_i, t^n)(x - x_i) + \frac{\partial^2 q}{\partial x^2}(\xi, t^n)\frac{(x - x_i)^2}{2} \right) dx \\ &= \frac{1}{\Delta x}\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} \frac{\partial^2 q}{\partial x^2}(\xi, t^n)\frac{(x - x_i)^2}{2}\, dx \end{aligned}$$

Using the mean value theorem for integrals, we have:

$$Q_i(t^n) - q_i^n = \frac{\partial^2 q}{\partial x^2}(\eta, t^n)\frac{1}{\Delta x}\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} \frac{(x - x_i)^2}{2}\, dx = \frac{\partial^2 q}{\partial x^2}(\eta, t^n)\frac{\Delta x^2}{24}$$

for some $\eta \in X_i$, from which the proposition follows with.

$$C_1 = \frac{1}{24}\frac{\partial^2 q}{\partial x^2}(\eta, t^n). \tag{2.23}$$

□

To move towards the convergence of 1D-FV schemes, we introduce the local truncation error (LTE hereafter) τ_i^n following LeVeque (2002):

$$Q_i(t^{n+1}) = Q_i(t^n) - \frac{\Delta t}{\Delta x}\left(\mathcal{F}(Q(t^n), u^n, i) - \mathcal{F}(Q(t^n), u^n, i-1) \right) + \Delta t \tau_i^n. \tag{2.24}$$

Notice the LTE is obtained by replacing the exact solution in Equation (2.21). Since $Q_i(t^n)$ is the exact solution of Equation (2.20), the local truncation error may be rewritten

as

$$\tau_i^n = \frac{1}{\Delta x}\left[\left(\frac{1}{\Delta t}\int_{t^n}^{t^{n+1}} (uq)(x_{i+\frac{1}{2}}, t)\,dt - \mathcal{F}(Q(t^n), u^n; i)\right) + \left(\frac{1}{\Delta t}\int_{t^n}^{t^{n+1}} (uq)(x_{i-\frac{1}{2}}, t)\,dt - \mathcal{F}(Q(t^n), u^n; i-1)\right)\right]. \tag{2.25}$$

The LTE gives a measure of how well the 1D-FV scheme approximates the integral form of the considered conservation law. Another interpretation of the LTE is that the LTE gives the error obtained after applying the scheme for a single time-step using the exact solution. The 1D-FV scheme is said to be consistent if the LTE converges to zero.

Given $r = (r_1, \cdots, r_N) \in \mathbb{R}^N$, we define the p-norm by

$$\|r\|_{p,N} = \begin{cases} \left(\sum_{i=1}^N |r_i|^p\right)^{\frac{1}{p}} & \text{if } 1 \leq p < \infty, \\ \max_{i=1,\cdots,N} |r_i| & \text{otherwise .} \end{cases}$$

We the define $\tau^n = (\tau_1^n, \cdots, \tau_N^n)$, which represent the LTEs at the time-step n. Now we can define consistency.

Definition 2.1. *A 1D-FV scheme is said to be consistent in the p-norm if:*

$$\lim_{N_T, N\to\infty}\left[\max_{1\leq n\leq N_T} \|\tau^n\|_{p,N}\right] = 0,$$

and it is said to be consistent with order P in the $p-$norm if there exists a constant C that does not depend neither on Δt nor on Δx, such that

$$\max_{1\leq n\leq N_T} \|\tau^n\|_{p,N} = O(\Delta x^P).$$

From Equation (2.25), it follows that we basically need to ensure that the numerical flux function $\mathcal{F}$ converges to the time-averaged flux at edges when $\Delta x \to 0$ in order to guarantee consistency. In Section 2.3.3 we shall address how the numerical flux from PPM approximates the time-averaged flux at edges.

At last, we define the pointwise error at time-step n by:

$$E_i^n = Q_i(t^n) - Q_i^n, \quad i = 1, \cdots, N,$$

and we define the vector of errors by $E^n = (E_1^n, \cdots, E_N^n)$.

Definition 2.2. *A 1D-FV scheme is said to be convergent in the p-norm if:*

$$\lim_{N_T, N\to\infty}\left[\max_{1\leq n\leq N_T} \|E^n\|_{p,N}\right] = 0,$$

and it is said to converge with order P in the $p-$norm if there exists a constant C that does

not depend neither on Δt nor on Δx, such that

$$\max_{1\leq n\leq N_T} \|E^n\|_{p,N} = O(\Delta x^P).$$

Subtracting Equation (2.21) from Equation (2.24) we get the following equation for the error:

$$E_i^{n+1} = E_i^n - \frac{\Delta t}{\Delta x}\left[\left(\mathcal{F}(Q(t^n), u^n; i) - \mathcal{F}(Q^n, u^n; i)\right) - \left(\mathcal{F}(Q(t^n), u^n; i-1) - \mathcal{F}(Q^n, u^n; i-1)\right)\right] + \tau_i^n \Delta t \tag{2.26}$$

2.2.3 Stability

In order to define the concept of stability, it is useful to introduce an operator representation of 1D-FV schemes. Recall that the framework of Problem 2.3, has the hypothesis that the values of N and N_T are always chosen in such a way that $N_t = c \cdot N$, for some c fixed and therefore $\frac{\Delta t}{\Delta x} = \sigma$ for an also fixed $\sigma \in \mathbb{R}$. In this context, we define the operators $A_{N,n} : \mathbb{R}^N \to \mathbb{R}^N$ whose i-th entry is given by:

$$[A_{N,n}(Q)]_i = Q_i - \sigma\left(\mathcal{F}(Q; u^n, i) - \mathcal{F}(Q; u^n, i-1)\right), \tag{2.27}$$

for $i = 1, \cdots, N$, $n = 0, \cdots, N_T - 1$. Notice that the dependence on n is due to the velocity that may be allowed to vary with time. As it is usual, we are assuming periodicity in the entries of Q when we apply the operator $A_{N,n}$. Thus, Equation (2.21) may be rewritten in a vector form by

$$Q^{n+1} = A_{N,n}(Q^n),$$

and Equation (2.24) in a vector form reads

$$Q(t^{n+1}) = A_{N,n}(Q(t^n)) + \Delta t \tau^n,$$

and the error equation (2.26) is given by

$$E^{n+1} = A_{N,n}(Q(t^n)) - A_{N,n}(Q^n) + \Delta t \tau^n. \tag{2.28}$$

The stability theory focus on uniformly bounding the norm of $A_{N,n}(Q(t^n)) - A_{N,n}(Q^n)$ (LeVeque, 2002). We define stability as follows.

Definition 2.3. *A 1D-FV scheme is stable in the $p-$norm if*

$$\|A_{N,n}(Q) - A_{N,n}(P)\|_{p,N} \leq (1 + \alpha \Delta t)\|Q - P\|_{p,N}, \tag{2.29}$$

for all $Q, P \in \mathbb{R}^N$ and α is a constant that does not depend neither on N, Δt and n.

Assuming that the scheme is stable in the $p-$norm, then it follows from Equation (2.28)

that:

$$\begin{aligned}\|E^{n+1}\|_{p,N} &\leq \|A_{N,n}(Q(t^n)) - A_{N,n}(Q^n)\|_{p,N} + \Delta t \max_{n=1,\cdots,N_T} \|\tau^n\|_{p,N} \\ &\leq (1+\alpha\Delta t)\|E^n\|_{p,N} + \Delta t \max_{n=1,\cdots,N_T} \|\tau^n\|_{p,N} \\ &\leq (1+\alpha\Delta t)^n\|E^0\|_{p,N} + \Delta t \max_{n=1,\cdots,N_T} \|\tau^n\|_{p,N} \sum_{k=0}^{n-1}(1+\alpha\Delta t)^k \\ &\leq e^{\alpha T}(\|E^0\|_{p,N} + T \max_{n=1,\cdots,N_T} \|\tau^n\|_{p,N})\end{aligned} \tag{2.30}$$

where we used $n\Delta t \leq T$, $T = N\Delta t$ and the inequality $e^t > 1 + t$. When computing the initial average values using the value at the cell centroid, the initial error E^0 converges to zero provided q is twice continuously differentiable by Proposition 2.1. Therefore, it follows that if the scheme is stable and consistent then it is convergent. Furthermore, if it is stable and consistent with order P, then the convergence order is at least equal to $\min\{P, 2\}$. In the case where both the conservation law and $A_{N,n}$ are linear, this result is a particular case of the Lax-Ritchmyer stability and the convergence is guaranteed by the Lax equivalence theorem (LeVeque, 2002). In this Chapter, we are interested only in the linear advection equation. However, as we shall see in Section 2.3.3, the operator $A_{N,n}$ may become non-linear when monotonicity constraints are activated.

Notice that, if $A_{N,n}$ is linear, then stability is equivalent to require that

$$\|A_{N,n}\|_{p,N} \leq 1 + \alpha\Delta t,$$

where

$$\|A_{N,n}\|_{p,N} = \sup_{Q\in\mathbb{R}^N} \frac{\|A_{N,n}(Q)\|_{p,N}}{\|Q\|_{p,N}}$$

is the operator p-norm.

For linear operators, we may use the discrete Fourier transform (Trefethen, 2000) to estimate the 2-norm of $A_{N,n}$. This approach is known as Von Neumann stability analysis. We define the nodes $\theta_i = i\frac{2\pi}{N}$, $i = 1, \cdots, N$, $\Delta\theta = \frac{2\pi}{N}$, $\theta = (\theta_1, \theta_2, \cdots, \theta_N)$. The imaginary unit is denoted by $\imath$. The Fourier modes are given by:

$$e^{\imath k\theta} = (e^{\imath k\theta_1}, e^{\imath k\theta_2}, \cdots, e^{\imath k\theta_N}) \in \mathbb{C}^N,$$

for $k = 1, \cdots, N$. Each k is referred to wavenumber and θ_k is called dimensionless wavenumber. The Fourier modes form an orthogonal basis of $\mathbb{C}^N$ with respect to the inner product

$$\langle Q, P\rangle = \frac{1}{N}\sum_{i=1}^{N} Q_i\overline{P_i}.$$

for $P, Q \in \mathbb{C}$ and $\overline{z}$ denotes the complex conjugate of z. Given $Q \in \mathbb{R}^N$, we may may express it in terms of the Fourier modes

$$Q = \sum_{k=1}^{N} a_k \exp(\imath k\theta),$$

where $a_k \in \mathbb{C}$. The 2-norm of Q is then given by:

$$\|Q\|_{2,N} = \sqrt{N \sum_{k=1}^{N} |a_k|^2}.$$

The idea of Von Neumann stability analysis is to apply the operator $A_{N,n}$ on each Fourier mode and analyze how it modifies its amplitude. For ease of analysis, we assume that the velocity is constant, which implies that the operator $A_{N,n}$ has constant coefficients and does not depend on n. For the general case, where the velocity is not constant, the stability can be ensured using the frozen coefficients method (Strikwerda, 2004, p. 59). This method boils down to performing multiple times the stability analysis with a constant velocity being equal to each one of the possible values of the velocity on the grid. If the scheme is stable for all the possible constant velocities, then stability is ensured. Since the operator is supposed to be linear with constant coefficients and we are assuming periodic boundaries conditions, we may write:

$$A_{N,n}(e^{\imath k\theta}) = \rho(k) e^{\imath k\theta},$$

where the term $\rho(k)$ is called amplification factor and it is an eigenvalue of $A_{N,n}$. The norm of $A_{N,n}(Q)$ is bounded by:

$$\|A_{N,n}(Q)\|_{2,N}^2 = N \sum_{k=1}^{N} |a_k|^2 |\rho(k)|^2 \leq \max_{k=1,\cdots,N} |\rho(k)|^2 \|Q\|_{2,N}^2.$$

Therefore:

$$\|A_{N,n}\|_{2,N} \leq \max_{k=1,\cdots,N} |\rho(k)|.$$

If we show that $\max_{k=1,\cdots,N} |\rho(k)| \leq 1 + \alpha\Delta t$, with α independent of Δt, N and n, then we ensure the stability of $A_{N,n}$. Generally speaking, the numerical flux can be written as a stencil of the form

$$\mathcal{F}(Q; u^n, i) = \sum_{l=-q}^{p} \alpha_{l,i} Q_{i+l},$$

when no monotonicity constraint is imposed, where the coefficients $\alpha_{l,i}$ depend on $u^n, \Delta t$ and Δx. We can then express ρ in terms of $\alpha_{l,i}$. Indeed, when we apply the operator $A_{N,n}$ in a Fourier mode, we get:

$$\begin{aligned}
[A_{N,n}(e^{\imath k\theta})]_i &= e^{\imath k\theta_i} - \sigma\left(\sum_{l=-q}^{p} \alpha_{l,i} e^{\imath k\theta_{i+l}} - \sum_{l=-q}^{p} \alpha_{l,i-1} e^{\imath k\theta_{i-1+l}} \right) \\
&= e^{\imath k\theta_i} \left(1 - \sigma\left(\sum_{l=-q}^{p} \alpha_{l,i} e^{\imath k\theta_l} - \sum_{l=-q}^{p} \alpha_{l,i-1} e^{\imath k\theta_{l-1}} \right) \right)
\end{aligned}$$

Hence, the amplification factor has the form

$$\rho(k) = 1 - \sigma\left(\sum_{l=-q}^{p} \alpha_{l,i} e^{\imath k\theta_l} - \sum_{l=-q}^{p} \alpha_{l,i-1} e^{\imath k\theta_{l-1}} \right) \tag{2.31}$$

In Section 2.3.3 we shall analyse $|\rho(k)|$ in terms of the PPM coefficients.

2.3 The Piecewise-Parabolic Method

In this Section, we are going to review and analyze the Piecewise-Parabolic method (PPM). This method was proposed by Colella and Woodward (1984) for gas dynamic simulations and its viability for atmospheric simulations has been shown by Carpenter et al. (1990). This method is based on using parabolas to reconstruct the function from its average values, ensuring mass conservation and monotonicity. PPM is an extension of the Piecesiwe-Linear method from Van Leer (1977) and it is employed in the FV3 model using the dimension splitting method from Lin and Rood (1996). This section is organized as follows: in Subsection 2.3.1 we present and analyze the PPM reconstruction method and the monotonization and flux computation are presented and analyzed in Subsections 2.3.2 and 2.3.3, respectively.

2.3.1 Reconstruction

Let us consider a function $q \in L^\infty([a,b],\mathbb{R})$, a discretization of $[a,b]$ as in Problem 2.2 and assume that we are given the average values $Q_i = \frac{1}{\Delta x}\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} q(x)\,dx$ on each control volume X_i, $\forall i = 1,\cdots,N$. We make use of the indicator function of each control volume X_i defined by:

$$\chi_i(x) = \begin{cases} 1 & \text{if } x \in X_i \\ 0 & \text{otherwise} \end{cases}$$

Our task is to find a Piecewise-Parabolic (PP) function:

$$q_{PP}(x) = \sum_{i=1}^{N} \chi_i(x) q_i(x), \tag{2.32}$$

where $q_i \in \mathcal{P}_2$ [2] is such that:

1. $\frac{1}{\Delta x}\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} q_i(x)\,dx = Q_i$, that is, q_i preserves the mass on each control volume X_i;
2. No new extreme is generated

We shall assume that each q_i may be expressed as:

$$q_i(x) = q_{L,i} + z_i(x)(\Delta q_i + q_{6,i}(1 - z_i(x))), \quad \text{where } z_i(x) = \frac{x - x_{i-\frac{1}{2}}}{\Delta x}, \quad x \in X_i, \tag{2.33}$$

where the values $q_{L,i}$, Δq_i and $q_{6,i}$ will be specified latter. Note that each z_i is just a normalization function that maps X_i onto $[0,1]$. Under this assumption, it is easy to see that $\lim_{x\to x_{i-\frac{1}{2}}^+} q_i(x) = q_{L,i}$. If we define $q_{R,i} = \lim_{x\to x_{i+\frac{1}{2}}^-} q_i(x)$, then we have:

$$\Delta q_i = q_{R,i} - q_{L,i}. \tag{2.34}$$

[2] $\mathcal{P}_n$ stands for the space of real polynomials of degree $\leq$ n.

The average value of q_i is given by:

$$\frac{1}{\Delta x}\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} q_i(x)\,dx = \frac{(q_{L,i}+q_{R,i})}{2} + \frac{q_{6,i}}{6} \tag{2.35}$$

Under the hypothesis of mass conservation, we have:

$$q_{6,i} = 6\left(Q_i - \frac{(q_{L,i}+q_{R,i})}{2}\right). \tag{2.36}$$

Therefore, we have found the parameters Δq_i and $q_{6,i}$ as functions of the parameters $q_{L,i}$ and $q_{R,i}$, such that the polynomial p_i from (2.32) guarantees mass conservation. To completely determine the polynomial p_i, we need to set the values $q_{L,i}$ and $q_{R,i}$, which, as we have seen, represent the limits of q_i when x tends to the left and right boundaries of X_i, respectively. Hence, it is natural to seek for $q_{L,i}$ as an approximation of $q(x_{i-\frac{1}{2}})$ and $q_{R,i}$ as an approximation of $q(x_{i+\frac{1}{2}})$. So, let us describe a way to approximate $q(x_{i+\frac{1}{2}})$, and denote its estimation by $q_{i+\frac{1}{2}}$ $\forall i = 0, 1, \cdots, N$. We introduce the following function:

$$Q(x) = \int_a^x q(\xi)\,d\xi, \tag{2.37}$$

and we notice that:

$$Q(x_{i+\frac{1}{2}}) = \Delta x \sum_{k=1}^{i} Q_k \text{ and } Q'(x) = q(x). \tag{2.38}$$

Therefore $Q'(x_{i+\frac{1}{2}}) = q(x_{i+\frac{1}{2}})$, $\forall i = 0, 1, \cdots, N$. We introduce a quartic polynomial $Q_{i4} \in \mathcal{P}_4$ that interpolates the data $\left(x_{i+k+\frac{1}{2}}, Q(x_{i+k+\frac{1}{2}})\right)_{k=-2,-1,0,1,2}$. Then, we define $q_{i+\frac{1}{2}} = \frac{d}{dx}Q_{i4}(x_{i+k+\frac{1}{2}})$. An explicit expression for $q_{i+\frac{1}{2}}$ is given by (Colella & Woodward, 1984):

$$q_{i+\frac{1}{2}} = \frac{1}{2}\left(Q_{i+1} + Q_i\right) - \frac{1}{6}\left(\delta Q_{i+1} - \delta Q_i\right), \tag{2.39}$$

where δQ_i is the average slope in the i-th control-volume:

$$\delta Q_i = \frac{1}{2}\left(Q_{i+1} - Q_{i-1}\right). \tag{2.40}$$

We notice that Formula (2.40) may be rewritten more explicitly as:

$$q_{i+\frac{1}{2}} = \frac{7}{12}\left(Q_{i+1} + Q_i\right) - \frac{1}{12}\left(Q_{i+2} + Q_{i-1}\right). \tag{2.41}$$

The Formula (2.41) is fourth-order accurate if q is at least C^4 (Colella & Woodward, 1984). Indeed, we prove this later in Proposition 2.2 by noticing that this Formula may be thought of as a finite-difference scheme. An explicit expression for the values of $q_{R,i}$ and $q_{L,i}$ are

given by:

$$q_{R,i} = q_{i+\frac{1}{2}} = \frac{7}{12}\Big(Q_{i+1} + Q_i\Big) - \frac{1}{12}\Big(Q_{i+2} + Q_{i-1}\Big), \tag{2.42}$$

$$q_{L,i} = q_{i-\frac{1}{2}} = \frac{7}{12}\Big(Q_i + Q_{i-1}\Big) - \frac{1}{12}\Big(Q_{i+1} + Q_{i-2}\Big). \tag{2.43}$$

We point out that a fifth-order accurate for the values of $q_{R,i}$ and $q_{L,i}$ is also possible, as it was developed by Putman and Lin (2007) based on the work Suresh and Huynh (1997). Indeed, we can use construct reconstruction with arbitrary orders using finite difference formulas since the PPM reconstruction may be thought of as a finite-differences scheme, This idea shall be clear soon. The fifth-order reconstruction formula reads:

$$q_{R,i} = \frac{1}{60}\Big(2Q_{i-2} - 13Q_{i-1} + 47Q_i + 27Q_{i+1} - 3Q_{i+2}\Big), \tag{2.44}$$

$$q_{L,i} = \frac{1}{60}\Big(-3Q_{i-2} + 27Q_{i-1} + 47Q_i - 13Q_{i+1} + 2Q_{i+2}\Big). \tag{2.45}$$

However, we notice that this reconstruction scheme allows discontinuity of the Piecewise-Parabolic function at the control volume edges.

PPM reconstruction numerical analysis

As we pointed out before, the approximation of q at the control volumes edges given by Equation (2.41) is fourth-order accurate when $q \in C^4([a,b])$. This is proved as a Corollary of the following Proposition 2.2.

Proposition 2.2. *Let $q \in C^4([a,b])$, $\overline{x} \in]a,b[$ and $h > 0$ such that $[\overline{x}-2h, \overline{x}+2h] \subset [a,b]$. Then, the following identity holds:*

$$q(\overline{x}) = \frac{7}{12}\left(\frac{1}{h}\int_{\overline{x}}^{\overline{x}+h} q(x)\,dx + \frac{1}{h}\int_{\overline{x}-h}^{\overline{x}} q(x)\,dx\right) - \frac{1}{12}\left(\frac{1}{h}\int_{\overline{x}+h}^{\overline{x}+2h} q(x)\,dx + \frac{1}{h}\int_{\overline{x}-2h}^{\overline{x}-h} q(x)\,dx\right) + C_1 h^4, \tag{2.46}$$

where C_1 is a constant that depends on q and h.

Proof. We define $Q(x) = \int_a^x q(\xi)\,d\xi$ for $x \in [a,b]$ as in Equation (2.37). It follows that:

$$\int_{\overline{x}}^{\overline{x}+h} q(\xi)\,d\xi + \int_{\overline{x}-h}^{\overline{x}} q(\xi)\,d\xi = Q(\overline{x}+h) - Q(\overline{x}-h),$$

$$\int_{\overline{x}+h}^{\overline{x}+2h} q(\xi)\,d\xi + \int_{\overline{x}-2h}^{\overline{x}-h} q(\xi)\,d\xi = Q(\overline{x}+2h) - Q(\overline{x}-2h) - (Q(\overline{x}+h) - Q(\overline{x}-h)).$$

Using these identities, Equation (2.46) may be rewritten as:

$$q(\overline{x}) = \frac{4}{3}\left(\frac{Q(\overline{x}+h) - Q(\overline{x}-h)}{2h}\right) - \frac{1}{3}\left(\frac{Q(\overline{x}+2h) - Q(\overline{x}-2h)}{4h}\right) + C_1 h^4, \tag{2.47}$$

which consists of finite-difference approximations. Thus, Equation (2.46) follows from Lemma A.1 with:

$$C_1 = C_1(\mu_1, \mu_2) = \frac{1}{720}\left(6q^{(4)}(\mu 1) - 32q^{(4)}(\mu_2)\right), \tag{2.48}$$

where $\mu_1, \mu_2 \in [\overline{x} - 2h, \overline{x} + 2h]$, which concludes the proof. □

Corollary 2.1. *It follows from Proposition 2.2 with* $\overline{x} = x_{i+\frac{1}{2}}$ *and* $h = \Delta x$ *that* $q_{i+\frac{1}{2}}$ *given by Equation* (2.41) *satisfies:*

$$q(x_{i+\frac{1}{2}}) - q_{i+\frac{1}{2}} = C_1 \Delta x^4, \tag{2.49}$$

with C_1 *given by Equation* (2.48).

Remark 2.4. *Similarly, one can show that the formulas are given by Equation* (2.44) *and Equation* (2.44) *are fifth-order accurate.*

The parabolic function from (2.46) given with coefficients specified before approximates q with order 3 when $q \in C^4([a, b])$. In order to check this, for $x \in X_i$ we rewrite Equation (2.33) as:

$$q_i(x) = q_{L,i} + \frac{(\Delta q_i + q_{6,i})}{\Delta x}(x - x_{i-\frac{1}{2}}) - \frac{q_{6,i}}{\Delta x^2}(x - x_{i-\frac{1}{2}})^2 \tag{2.50}$$

and we write q using its Taylor expansion assuming $q \in C^4([a, b])$:

$$q(x) = q(x_{i-\frac{1}{2}}) + q'(x_{i-\frac{1}{2}})(x - x_{i-\frac{1}{2}}) + \frac{q''(x_{i-\frac{1}{2}})}{2}(x - x_{i-\frac{1}{2}})^2 + \frac{q^{(3)}(\theta_i)}{6}(x - x_{i-\frac{1}{2}})^3, \tag{2.51}$$

where $\theta_i \in X_i$. Comparing Equation (2.50) with Equation (2.51), it is reasonable to seek to some bound to the expressions:

$$q'(x_{i-\frac{1}{2}}) - \frac{(\Delta q_i + q_{6,i})}{\Delta x}, \tag{2.52}$$

and:

$$\frac{q''(x_{i-\frac{1}{2}})}{2} - \left(-\frac{q_{6,i}}{\Delta x^2}\right). \tag{2.53}$$

We have seen that term $q_{L,i}$ gives a fourth-order approximation to $q(x_{i-\frac{1}{2}})$. The Corollary 2.2 shall prove that the term (2.52) has a bound proportional to Δx^2, and the Corollary 2.3 shall prove that the term (2.53) is bounded by a constant times Δx.

Before proving the desired bounds, it is useful to rewrite some terms explicitly as functions of the values Q_i's. Combining Equation (2.36) with Equations (2.42) and (2.43), we may write $q_{6,i}$ as:

$$q_{6,i} = \frac{1}{4}\left(Q_{i-2} - 6Q_{i-1} + 10Q_i - 6Q_{i+1} + Q_{i+2}\right). \tag{2.54}$$

Recalling the definition of Δq_i from Equation (2.34), and applying Equations (2.42) and

(2.43), we may express Δq_i as:

$$\Delta q_i = \frac{1}{12}\Big(Q_{i-2} - 8Q_{i-1} + 8Q_{i+1} - Q_{i+2} \Big). \tag{2.55}$$

Finally, we combine Equations (2.54) and (2.55) and write their sum as:

$$\frac{(\Delta q_i + q_{6,i})}{\Delta x} = \frac{2Q_{i-2} - 13Q_{i-1} + 15Q_i - 5Q_{i+1} + Q_{i+2}}{6\Delta x}. \tag{2.56}$$

The next Proposition 2.3 proves that Equation (2.56) approximates $q'(x_{i-\frac{1}{2}})$ with order 2.

Proposition 2.3. *Let $q \in C^3([a,b])$, $\overline{x} \in]a,b[$, and $h > 0$ such that $[\overline{x}-2h, \overline{x}+3h] \subset [a,b]$ Then, the following identity holds:*

$$\begin{aligned} q'(\overline{x}) = \frac{1}{6h}\Bigg(\frac{2}{h}\int_{\overline{x}-2h}^{\overline{x}-h} q(x)\,dx - \frac{13}{h}\int_{\overline{x}-h}^{\overline{x}} q(x)\,dx + \frac{15}{h}\int_{\overline{x}}^{\overline{x}+h} q(x)\,dx \\ - \frac{5}{h}\int_{\overline{x}+h}^{\overline{x}+2h} q(x)\,dx + \frac{1}{h}\int_{\overline{x}+2h}^{\overline{x}+3h} q(x)\,dx \Bigg) + C_2 h^2, \end{aligned} \tag{2.57}$$

where C_2 is a constant that depends on q and h.

Proof. We consider again $Q(x) = \int_a^x q(\xi)\,d\xi$ for $x \in [a,b]$ as in Equation (2.37). Like in Proposition 2.3, we have:

$$\begin{aligned} &\frac{1}{6h}\Bigg(\frac{2}{h}\int_{\overline{x}-2h}^{\overline{x}-h} q(x)\,dx - \frac{13}{h}\int_{\overline{x}-h}^{\overline{x}} q(x)\,dx + \frac{15}{h}\int_{\overline{x}}^{\overline{x}+h} q(x)\,dx - \frac{5}{h}\int_{\overline{x}+h}^{\overline{x}+2h} q(x)\,dx + \frac{1}{h}\int_{\overline{x}+2h}^{\overline{x}+3h} q(x)\,dx \Bigg) \\ &\quad = \frac{1}{6h}\Bigg(\frac{2}{h}\big(Q(\overline{x}-h) - Q(\overline{x}-2h)\big) - \frac{13}{h}\big(Q(\overline{x}) - Q(\overline{x}-h)\big) + \frac{15}{h}\big(Q(\overline{x}+h) - Q(\overline{x})\big) \\ &\qquad - \frac{5}{h}\big(Q(\overline{x}+2h) - Q(\overline{x}+h)\big) + \frac{1}{h}\big(Q(\overline{x}+3h) - Q(\overline{x}+2h)\big) \Bigg) \\ &\quad = \frac{1}{6h^2}\Big(-2Q(\overline{x}-2h) + 15Q(\overline{x}-h) - 28Q(\overline{x}) + 20Q(\overline{x}+h) - 6Q(\overline{x}+2h) + Q(\overline{x}+3h) \Big), \end{aligned}$$

which consists of the finite-difference scheme from Lemma A.2. Therefore, Equation (2.57) follows from Lemma A.2 with:

$$C_2 = C_2(\mu_1, \mu_2) = \frac{1}{24}\Big(128 q^{(3)}(\mu_1) - 116 q^{(3)}(\mu_2) \Big), \tag{2.58}$$

where $\mu_1, \mu_2 \in [x_0 - 2h, x_0 + 3h]$, which concludes the proof.

□

Corollary 2.2. *It follows from Proposition 2.3 with $\overline{x} = x_{i-\frac{1}{2}}$ and $h = \Delta x$ that Δq_i given by*

Equation (2.55) *and* $q_{6,i}$ *given by Equation* (2.54) *satisfy:*

$$q'(x_{i-\frac{1}{2}}) - \frac{(\Delta q_i + q_{6,i})}{\Delta x} = C_2 \Delta x^2, \tag{2.59}$$

with C_2 *given by Equation* (2.58).

Now, we analyse the following expression:

$$-\frac{2q_{6,i}}{\Delta x^2} = -\frac{1}{2\Delta x^2}\Big(Q_{i-2} - 6Q_{i-1} + 10Q_i - 6Q_{i+1} + Q_{i+2}\Big). \tag{2.60}$$

deduced from Equation (2.54) and we prove in Proposition 2.4 that Equation (2.60) approximates $q''(x_{i-\frac{1}{2}})$ with order 1.

Proposition 2.4. *Let* $q \in C^3([a,b])$, $\overline{x} \in]a,b[$ *and* $h > 0$ *such that* $[\overline{x}-2h, \overline{x}+3h] \subset [a,b]$. *Then, the following identity holds:*

$$\begin{aligned} q''(\overline{x}) = \frac{1}{2h^2}\Bigg(&-\frac{1}{h}\int_{\overline{x}-2h}^{\overline{x}-h} q(x)\,dx + \frac{6}{h}\int_{\overline{x}-h}^{\overline{x}} q(x)\,dx - \frac{10}{h}\int_{\overline{x}}^{\overline{x}+h} q(x)\,dx \\ &+\frac{6}{h}\int_{\overline{x}+h}^{\overline{x}+2h} q(x)\,dx - \frac{1}{h}\int_{\overline{x}+2h}^{\overline{x}+3h} q(x)\,dx\Bigg) + C_3 h, \end{aligned} \tag{2.61}$$

where C_3 *is a constant that depends on* q *and* h.

Proof. Similarly to Proposition 2.3 using the same function Q, we have:

$$\begin{aligned} &\frac{1}{2h^2}\Bigg(-\frac{1}{h}\int_{\overline{x}-2h}^{\overline{x}-h} q(x)\,dx + \frac{6}{h}\int_{\overline{x}-h}^{\overline{x}} q(x)\,dx - \frac{10}{h}\int_{\overline{x}}^{\overline{x}+h} q(x)\,dx + \frac{6}{h}\int_{\overline{x}+h}^{\overline{x}+2h} q(x)\,dx - \frac{1}{h}\int_{\overline{x}+2h}^{\overline{x}+3h} q(x)\,dx\Bigg) \\ &\quad= \frac{1}{2h^2}\Bigg(-\frac{1}{h}\big(Q(\overline{x}-h) - Q(\overline{x}-2h)\big) + \frac{6}{h}\big(Q(\overline{x}) - Q(\overline{x}-h)\big) - \frac{10}{h}\big(Q(\overline{x}+h) - Q(\overline{x})\big) \\ &\qquad\qquad +\frac{6}{h}\big(Q(\overline{x}+2h) - Q(\overline{x}+h)\big) - \frac{1}{h}\big(Q(\overline{x}+3h) - Q(\overline{x}+2h)\big)\Bigg) \\ &\quad= \frac{1}{2h^3}\Big(Q(\overline{x}-2h) - 7Q(\overline{x}-h) + 16Q(\overline{x}) - 16Q(\overline{x}+h) + 7Q(\overline{x}+2h) - Q(\overline{x}+3h)\Big), \end{aligned}$$

which consists of the finite-difference scheme from Lemma A.3. Therefore, Equation (2.61) follows from Lemma A.3 with:

$$C_3 = C_3(\mu_1,\mu_2) = \frac{1}{48}\Big(104q^{(3)}(\mu_1) - 128q^{(3)}(\mu_2)\Big), \tag{2.62}$$

where $\mu_1, \mu_2 \in [x_0 - 2h, x_0 + 3h]$, which concludes the proof. □

Corollary 2.3. *It follows from Proposition 2.4 with* $\overline{x} = x_{i-\frac{1}{2}}$ *and* $h = \Delta x$ *that* $q_{6,i}$ *given by Equation* (2.41) *satisfies:*

$$q''(x_{i-\frac{1}{2}}) - \Big(-\frac{2q_{6,i}}{\Delta x^2}\Big) = C_3 \Delta x, \tag{2.63}$$

with C_3 given by Equation (2.62).

With the aid of Corollaries 2.1, 2.2, and 2.3, we are able to prove that the PPM reconstruction approximates q with order 3. Indeed, we prove this on the follow up Proposition 2.5.

Proposition 2.5. *Let $q \in C^4([a,b])$. Then, the Piecewise-Parabolic function given by Equation (2.33) with the parameters $q_{R,i}$ and $q_{L,i}$ obeying Equations (2.42) and (2.43) gives a third-order approximation to q on the control volume X_i. Namely, there exist constants M_1 and M_2 such that*

$$|q(x) - q_i(x)| \leq M_1 \Delta x^4 + M_2 \Delta x^3, \quad \forall x \in X_i.$$

Proof. For $x \in X_i$, from Equations (2.51) and (2.50), we have:

$$\begin{aligned} q(x) - q_i(x) = (q'(x_{i-\frac{i}{2}}) - q_{L,i}) + \left(q'(x_{i-\frac{1}{2}}) - \frac{(\Delta q_i + q_{6,i})}{\Delta x} \right)(x - x_{i-\frac{1}{2}}) \\ + \left(\frac{q''(x_{i-\frac{1}{2}})}{2} + \frac{q_{6,i}}{\Delta x^2} \right)(x - x_{i-\frac{1}{2}})^2 + \frac{q^{(3)}(\theta_i)}{6}(x - x_{i-\frac{1}{2}})^3. \end{aligned}$$

Using this fact with Corollaries 2.1, 2.2, and 2.3, we have:

$$q(x) - q_i(x) = C_1 \Delta x^4 + C_2 \Delta x^2 (x - x_{i-\frac{1}{2}}) + \frac{C_3}{2} \Delta x (x - x_{i-\frac{1}{2}})^2 + C_4 (x - x_{i-\frac{1}{2}})^3$$

where C_1, C_2 and C_3 are given by Equations (2.48), (2.58) and (2.62), respectively, and

$$C_4 = C_4(\theta_i) = \frac{q^{(3)}(\theta_i)}{6}. \tag{2.64}$$

For $x \in X_i$, we have $|x - x_{i-\frac{i}{2}}| \leq \Delta x$, thus:

$$|q(x) - q_i(x)| \leq M_1 \Delta x^4 + M_2 \Delta x^3,$$

where

$$M_1 = \frac{38}{720} \sup_{\xi \in [a,b]} |q^{(4)}(\xi)|,$$

$$M_2 = \left(\frac{244}{24} + \frac{232}{96} + \frac{1}{6} \right) \sup_{\xi \in [a,b]} |q^{(3)}(\xi)| = \frac{143}{12} \sup_{\xi \in [a,b]} |q^{(3)}(\xi)|,$$

which concludes the proof. □

Remark 2.5. *Replacing the formulas for $q_{R,i}$ and $q_{L,i}$ given by Equations (2.42)and (2.43) by the formulas given by Equations (2.44) and (2.45), does not change the order of convergence of the parabolic approximation.*

2.3.2 Monotonization

This section is dedicated to presenting possible ways of ensuring the creation of new extrema values in the PPM reconstruction. We are going to present the original monotonic scheme from Colella and Woodward (1984) and an alternative scheme from Lin (2004), which was an attempt to reduce the diffusion of the original scheme Colella and Woodward (1984) and is currently employed in the FV3 dynamical core (L. Harris et al., 2021).

Limiter from Colella and Woodward (1984)

To avoid numerical oscillations in the parabolas, especially when discontinuities are present, Colella and Woodward (1984) ensures that the reconstructed value at cell edges (namely, $q_{i+\frac{1}{2}}$) does not stay outside of the range of its neighbors average values (Q_i and Q_{i+1}). This can be achieved by replacing the term δQ_i in Equation (2.39) by the values $\delta_m Q_i$ given by:

$$\delta_m Q_i = \begin{cases} \max(|\delta Q_i|, 2|Q_{i+1} - Q_i|, 2|Q_i - Q_{i-1}|) \cdot \operatorname{sgn}(\delta Q_i) & \text{if } (Q_{i+1} - Q_i)(Q_i - Q_{i-1}) > 0, \\ 0 & \text{otherwise,} \end{cases} \tag{2.65}$$

where sgn denotes the sign function. To ensure, monotonicity we also must ensure that the parabola has values between $q_{R,i}$ and $q_{L,i}$. This step will introduce a discontinuity on the edges of the PPM approximation. If Q_i is the local maximum/minimum, then we make the parabola constant. This is expressed as:

$$q_{L,i} \leftarrow Q_i, \quad q_{R,i} \leftarrow Q_i, \quad \text{if} \quad (Q_{R,i} - Q_i)(Q_i - Q_{L,i}) \geq 0 \tag{2.66}$$

This step eliminates the introduction of new extremes when we already have an extremum. The other case where we need to modify the values $q_{L,i}$ and $q_{R,i}$ is when the extrema of the parabola falls in $[x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}]$. It is easy to see from Equation (2.50) that, the extrema of the parabola falling in $[x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}]$ is equivalent to $|\Delta q_i| \leq |q_{6,i}|$. In this case, the values are updated as follows:

$$\begin{cases} q_{L,i} \leftarrow 3Q_i - 2q_{R,i} & \text{if } \Delta q_i \cdot q_{6,i} > (\Delta q_i)^2, \\ q_{R,i} \leftarrow 3Q_i - 2q_{L,i} & \text{if } -(\Delta q_i)^2 > \Delta q_i \cdot q_{6,i} \end{cases} \tag{2.67}$$

In this step, we are changing the value at the edge where the extreme is closer and ensuring again that no new extreme is created.

Limiter from Lin (2004)

Similarly to Colella and Woodward (1984), Lin (2004) reduces numerical oscilations in the parabolas replacing the term δQ_i in Equation (2.39) by the values $\delta_m Q_i$ given by:

$$\delta_m Q_i = \max(|\delta Q_i|, 2\delta Q_{\min,i}, 2\delta Q_{\max,i}) \cdot \operatorname{sgn}(\delta Q_i), \tag{2.68}$$

where $\delta Q_{\min,i} = Q_i - \min(Q_{i+1}, Q_i, Q_{i-1})$ and $\delta Q_{\max,i} = \max(Q_{i+1}, Q_i, Q_{i-1}) - Q_i$. The monotonicity is achieved by the following scheme:

$$q_{L,i} \leftarrow Q_i - \max(|\delta_m Q_i|, |q_{L,i} - Q_i|) \cdot \operatorname{sgn}(\delta_m Q_i), \tag{2.69}$$

$$q_{R,i} \leftarrow Q_i - \max(|\delta_m Q_i|, |q_{R,i} - Q_i|) \cdot \operatorname{sgn}(\delta_m Q_i). \tag{2.70}$$

This scheme may be further improved to reduce the diffusion even more as described by Lin (2004), but we are not going to assess this approach here.

2.3.3 Flux

Let us assume the framework from Problem 2.3 for the linear advection where the velocity function u is given. Supposing that the average grid values $Q^n = (Q_1^n, \cdots, Q_N^n)$ are known, we would like to compute the values Q^{n+1}. This is achieved using a scheme of the type given in Problem 2.3. Therefore, we need to estimate the time-average flux. For each control volume edge $i = 0, \cdots, N$ and $y > 0$ we define the following average of the Piecewise-Parabolic approximation defined in Equation (2.32) for the data Q^n (Colella & Woodward, 1984):

$$F_{L,i+\frac{1}{2}}(y) = \frac{1}{y} \int_{x_{i+\frac{1}{2}}-y}^{x_{i+\frac{1}{2}}} q_{PP}(\xi)\, d\xi, \tag{2.71}$$

and

$$F_{R,i+\frac{1}{2}}(y) = \frac{1}{y} \int_{x_{i+\frac{1}{2}}}^{x_{i+\frac{1}{2}}+y} q_{PP}(\xi)\, d\xi, \tag{2.72}$$

If $y \leq \Delta x$, then both of the above integral domains are constrained to a single control volume. Thus, it follows from a straightforward computation using Equation (2.33) that:

$$F_{L,i+\frac{1}{2}}(y) = \frac{1}{y} \int_{x_{i+\frac{1}{2}}-y}^{x_{i+\frac{1}{2}}} q_i(\xi)\, d\xi = q_{R,i} + \frac{(q_{6,i} - \Delta q_i)}{2\Delta x} y - \frac{q_{6,i}}{3\Delta x^2} y^2, \tag{2.73}$$

and

$$F_{R,i+\frac{1}{2}}(y) = \frac{1}{y} \int_{x_{i+\frac{1}{2}}}^{x_{i+\frac{1}{2}}+y} q_{i+1}(\xi)\, d\xi = q_{L,i+1} + \frac{(q_{6,i+1} + \Delta q_{i+1})}{2\Delta x} y - \frac{q_{6,i+1}}{3\Delta x^2} y^2. \tag{2.74}$$

The numerical flux function is then defined by:

$$\mathcal{F}(Q; u^n, i) = F^n_{i+\frac{1}{2}} = \begin{cases} u^n_{i+\frac{1}{2}} F_{L,i+\frac{1}{2}}(u^n_{i+\frac{1}{2}} \Delta t) & \text{if } u^n_{i+\frac{1}{2}} \geq 0, \\ u^n_{i+\frac{1}{2}} F_{R,i+\frac{1}{2}}(-u^n_{i+\frac{1}{2}} \Delta t) & \text{if } u^n_{i+\frac{1}{2}} < 0, \end{cases} \tag{2.75}$$

The numerical flux function may be also expressed as:

$$\mathcal{F}(Q; u^n, i) = \frac{1}{\Delta t} \int_{x_{i+\frac{1}{2}} - u^n_{i+\frac{1}{2}} \delta t}^{x_{i+\frac{1}{2}}} q_{PP}(x)\, dx. \tag{2.76}$$

Notice that if we define:

$$c^n_{i+\frac{1}{2}} = u^n_{i+\frac{1}{2}} \frac{\Delta t}{\Delta x},$$

the requirement $y \leq \Delta x$ for Equation (2.75) is equivalent to require that $|c^n_{i+\frac{1}{2}}| \leq 1$ for all i, which is the CFL condition. In the absence of monotonization, it follows from Equations (2.42), (2.43), (2.54) and (2.55) that the numerical flux may be expressed as the following stencil:

$$\mathcal{F}(Q; u^n, i) = u^n_{i+\frac{1}{2}} \sum_{k=-2}^{3} \alpha_{i,k} Q^n_{i+k},$$

where the coefficients are satisfies:

$$12\alpha_{i,-2} = \begin{cases} c_{i+\frac{1}{2}} - c^2_{i+\frac{1}{2}} & \text{if } u^n_{i+\frac{1}{2}} \geq 0, \\ 0 & \text{if } u^n_{i+\frac{1}{2}} < 0, \end{cases}$$

$$12\alpha_{i,-1} = \begin{cases} -1 - 5c_{i+\frac{1}{2}} + 6c^2_{i+\frac{1}{2}} & \text{if } u^n_{i+\frac{1}{2}} \geq 0, \\ -1 + 2c_{i+\frac{1}{2}} - c^2_{i+\frac{1}{2}} & \text{if } u^n_{i+\frac{1}{2}} < 0, \end{cases}$$

$$12\alpha_{i,0} = \begin{cases} 7 + 15c_{i+\frac{1}{2}} - 10c^2_{i+\frac{1}{2}} & \text{if } u^n_{i+\frac{1}{2}} \geq 0, \\ 7 - 13c_{i+\frac{1}{2}} + 6c^2_{i+\frac{1}{2}} & \text{if } u^n_{i+\frac{1}{2}} < 0, \end{cases}$$

$$12\alpha_{i,1} = \begin{cases} 7 - 13c_{i+\frac{1}{2}} + 6c^2_{i+\frac{1}{2}} & \text{if } u^n_{i+\frac{1}{2}} \geq 0, \\ 7 + 15c_{i+\frac{1}{2}} - 10c^2_{i+\frac{1}{2}} & \text{if } u^n_{i+\frac{1}{2}} < 0, \end{cases}$$

$$12\alpha_{i,2} = \begin{cases} -1 + 2c_{i+\frac{1}{2}} - c^2_{i+\frac{1}{2}} & \text{if } u^n_{i+\frac{1}{2}} \geq 0, \\ -1 - 5c_{i+\frac{1}{2}} + 6c^2_{i+\frac{1}{2}} & \text{if } u^n_{i+\frac{1}{2}} < 0, \end{cases}$$

$$12\alpha_{i,3} = \begin{cases} 0 & \text{if } u^n_{i+\frac{1}{2}} \geq 0, \\ c_{i+\frac{1}{2}} - c^2_{i+\frac{1}{2}} & \text{if } u^n_{i+\frac{1}{2}} < 0, \end{cases}$$

If the reconstruction at the edges is calculated using Equation (2.44) and (2.45), which is leads to a scheme called hybrid PPM (Putman & Lin, 2007), the flux stencil coefficients may be written as:

$$60\alpha_{i,-2} = \begin{cases} 2 - c_{i+\frac{1}{2}} - c^2_{i+\frac{1}{2}} & \text{if } u^n_{i+\frac{1}{2}} \geq 0, \\ 0 & \text{if } u^n_{i+\frac{1}{2}} < 0, \end{cases}$$

$$60\alpha_{i,-1} = \begin{cases} -13 - c_{i+\frac{1}{2}} + 14c^2_{i+\frac{1}{2}} & \text{if } u^n_{i+\frac{1}{2}} \geq 0, \\ -3 + 4c_{i+\frac{1}{2}} - c^2_{i+\frac{1}{2}} & \text{if } u^n_{i+\frac{1}{2}} < 0, \end{cases}$$

$$60\alpha_{i,0} = \begin{cases} 47 + 39c_{i+\frac{1}{2}} - 26c_{i+\frac{1}{2}}^2 & \text{if } u_{i+\frac{1}{2}}^n \geq 0, \\ 27 - 41c_{i+\frac{1}{2}} + 14c_{i+\frac{1}{2}}^2 & \text{if } u_{i+\frac{1}{2}}^n < 0, \end{cases}$$

$$60\alpha_{i,1} = \begin{cases} 27 - 41c_{i+\frac{1}{2}} + 14c_{i+\frac{1}{2}}^2 & \text{if } u_{i+\frac{1}{2}}^n \geq 0, \\ 47 + 39c_{i+\frac{1}{2}} - 26c_{i+\frac{1}{2}}^2 & \text{if } u_{i+\frac{1}{2}}^n < 0, \end{cases}$$

$$60\alpha_{i,2} = \begin{cases} -3 + 4c_{i+\frac{1}{2}} - c_{i+\frac{1}{2}}^2 & \text{if } u_{i+\frac{1}{2}}^n \geq 0, \\ -13 - c_{i+\frac{1}{2}} + 14c_{i+\frac{1}{2}}^2 & \text{if } u_{i+\frac{1}{2}}^n < 0, \end{cases}$$

$$60\alpha_{i,3} = \begin{cases} 0 & \text{if } u_{i+\frac{1}{2}}^n \geq 0, \\ 2 - c_{i+\frac{1}{2}} - c_{i+\frac{1}{2}}^2 & \text{if } u_{i+\frac{1}{2}}^n < 0, \end{cases}$$

Flux numerical analysis

With the stencil coefficients, we can compute the amplification factor (Equation (2.31)) for the PPM and the hybrid PPM schemes, both without monotonization. We assume a constant velocity equal to one and $N = 100$ (number of control volumes). In Figure 2.1 we show the amplification factor for both PPM and hybrid PPM schemes considering different CFL numbers. We can observe that both schemes damp most of the Fourier modes for larger k, regardless of the CFL number. Besides that, the hybrid scheme is more effective when reducing the Fourier modes amplitude. We point out that both schemes are exact when the CFL number is equal to 1. From this analysis, we can conclude that the PPM and hybrid PPM schemes satisfy the Von Neumann stability criteria when the CFL restriction is respected.

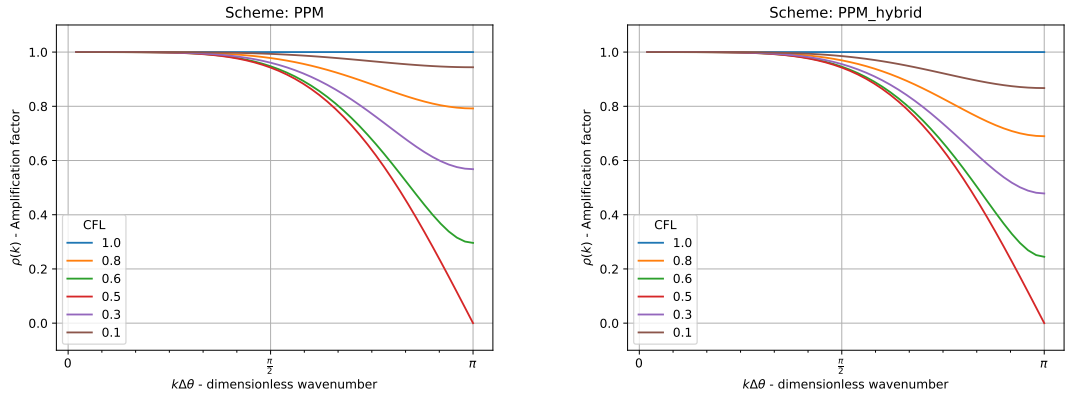


Figure 2.1: *Amplification factor for the PPM (left) and hybrid PPM (right) schemes for different CFL numbers.*

In order to investigate the consistency of the PPM scheme, we notice that when we are

deducing the time average flux, we are making some approximations of the form:

$$\int_{t^n}^{t^{n+1}} (uq)(x_{i+\frac{1}{2}}, t)\, dt \approx \int_{x_{i+\frac{1}{2}} - u^n_{i+\frac{1}{2}}\Delta t}^{x_{i+\frac{1}{2}}} q(x, t^n)\, dx, \tag{2.77}$$

as we can see from Equations (2.71) and (2.72), which basically replace q by q_{PP} on the right-hand side of Equation (2.77). As we shall see, the approximation (2.77) is in fact exact if we assume that u is constant. Therefore, in this case, the only error in the flux computation is due to the approximation made by q_{PP}. In the case where the velocity is not constant, this approximation will have an unique source of error related to a computation of a departure point which shall be second-order accurate.

To introduce the definition of departure point, for each $s \in [t^n, t^{n+1}]$, we consider the following Cauchy problem backward in time:

$$\begin{cases} \frac{\partial X}{\partial t}(t, s; x_{i+\frac{1}{2}}) = u(X(t, s; x_{i+\frac{1}{2}}), t), \quad t \in [t^n, s] \\ X(s, s; x_{i+\frac{1}{2}}) = x_{i+\frac{1}{2}}. \end{cases} \tag{2.78}$$

The point $X(t^n, s; x_{i+\frac{1}{2}})$ is called departure point at time t^n of the point $x_{i+\frac{1}{2}}$ at time s. Integrating Equation (2.78) over the interval $[t, s]$, we get:

$$X(t, s; x_{i+\frac{1}{2}}) = x_{i+\frac{1}{2}} - \int_t^s u(X(\theta, s; x_{i+\frac{1}{2}}), \theta)\, d\theta. \tag{2.79}$$

Equation (2.79) is useful to show that the trajectory $X(t, s; x_{i+\frac{1}{2}})$ remains on a singe control volume, under some basic assumptions, as we show on the next proposition.

Proposition 2.6. *If $u \in C^1$, the CFL condition is satisfied and if Δt and Δx are small enough, then for any $s \in [t^n, t^{n+1}]$, $t \in [t^n, s]$, we have that $X(t, s; x_{i+\frac{1}{2}}) \in X_i$ if $u^n_{i+\frac{1}{2}} > 0$ and $X(t, s; x_{i+\frac{1}{2}}) \in X_{i+1}$ if $u^n_{i+\frac{1}{2}} < 0$.*

Proof. Let us assume $u^n_{i+\frac{1}{2}} > 0$. Since u is continuous and the CFL condition is satisfied, there exist Δt and Δx such that $|u(X(t, s; x_{i+\frac{1}{2}}), t)|\frac{\Delta t}{\Delta x} \leq 1$ and $u(X(t, s; x_{i+\frac{1}{2}}), t) > 0$, $\forall s \in [t^n, t^{n+1}]$, $t \in [t^n, s]$. Hence, it follows from Equation (2.79) and the mean value theorem for integrals that:

$$X(t, s; x_{i+\frac{1}{2}}) = x_{i+\frac{1}{2}} - (t - s)u(X(\theta_1, s; x_{i+\frac{1}{2}}), \theta_1),$$

for some $\theta_1 \in [t, s]$. Therefore:

$$X(t, s; x_{i+\frac{1}{2}}) \geq x_{i+\frac{1}{2}} - \Delta t u(X(\theta_1, s; x_{i+\frac{1}{2}}), \theta_1) \geq x_{i+\frac{1}{2}} - \Delta x = x_{i-\frac{1}{2}},$$

from which the claim follows. The case $u^n_{i+\frac{1}{2}} < 0$ is very similar to the case $u^n_{i+\frac{1}{2}} > 0$. □

Equation (2.79) allow us to compute or estimate the departure point. For instance, if u is constant, then the departure point at time t^n of the point $x_{i+\frac{1}{2}}$ at time t^{n+1} is given by:

$$X(t^n, t^{n+1}; x_{i+\frac{1}{2}}) = x_{i+\frac{1}{2}} - u\Delta t. \tag{2.80}$$

If the velocity is not constant, it follows from Equation (2.79) that we can write a second-order approximation to the departure point:

$$X(t^n, t^{n+1}; x_{i+\frac{1}{2}}) = x_{i+\frac{1}{2}} - u^n_{i+\frac{1}{2}}\Delta t + O(\Delta t^2). \tag{2.81}$$

This approximation is exactly what is used in the PPM flux and we give a bound to this expression on the next proposition.

Proposition 2.7. *Assuming that $u \in C^1$ and X satisfies Equation (2.79), then:*

$$X(t^n, t^{n+1}; x_{i+\frac{1}{2}}) - (x_{i+\frac{1}{2}} - u^n_{i+\frac{1}{2}}\Delta t) = M_1 \Delta t^2, \tag{2.82}$$

where M_1 is a constant that depends only on u.

Proof. We basically need to estimate the integral of the right-hand side of (2.79). Defining $f(t) = u(X(t, t^{n+1}; x_{i+\frac{1}{2}}), t)$, similarly to Proposition 2.1 it is easy to show that:

$$\int_{t^n}^{t^{n+1}} f(t)\,dt - \Delta t f(t^n) = \frac{\Delta t^2}{2} f'(\theta)$$

where $\theta \in [t^n, t^{n+1}]$. Writing this expression in terms of u, we have:

$$\begin{aligned}\int_{t^n}^{t^{n+1}} u(X(t, t^{n+1}; x_{i+\frac{1}{2}}), t)\,dt - u^n_{i+\frac{1}{2}}\Delta t &= \frac{\Delta t^2}{2}\left(\frac{\partial u}{\partial t}(X(\theta, t^{n+1}; x_{i+\frac{1}{2}}), \theta) + \frac{\partial X}{\partial t}(\theta, t^{n+1}; x_{i+\frac{1}{2}})\frac{\partial u}{\partial x}(X(\theta, t^{n+1}; x_{i+\frac{1}{2}}), \theta)\right) \\ &= \frac{\Delta t^2}{2}\left(\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x}\right)(X(\theta, t^{n+1}; x_{i+\frac{1}{2}}), \theta)\end{aligned}$$

from which the proposition follows with

$$M_1 = M_1(\mu_1, \mu_2) = \frac{1}{2}\left(\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x}\right)(\mu_1, \mu_2), \tag{2.83}$$

for some $\mu_2 \in [t^n, t^{n+1}]$ and $\mu_1 \in X_i \cup X_{i+1}$, assuming that Δt and Δx are small enough as in Proposition 2.6. □

We point out that higher-order departure points estimates may be obtained integrating the ordinary differential equation (2.78) using higher-order time integration schemes. On the next Proposition 2.8, we prove that approximation (2.77) is exact if we use the exact departure point.

Proposition 2.8. *Assume the framework of Problem 2.2. If q and u are C^1 functions, then:*

$$\int_{X(t^n, t^{n+1}; x_{i+\frac{1}{2}})}^{x_{i+\frac{1}{2}}} q(x, t^n)\,dx = \int_{t^n}^{t^{n+1}} (uq)(x_{i+\frac{1}{2}}, s)\,ds \tag{2.84}$$

Proof. From Equation (2.79), using the Leibniz rule for integration it follows that:

$$\begin{aligned}\frac{\partial X}{\partial s}(t,s;x_{i+\frac{1}{2}}) &= -\left(u(x_{i+\frac{1}{2}},s) + \int_t^s \frac{du}{ds}(X(\theta,s;x_{i+\frac{1}{2}}),\theta)\,d\theta\right)\\ &= -u(x_{i+\frac{1}{2}},s) - \int_t^s \frac{\partial u}{\partial x}(X(\theta,s;x_{i+\frac{1}{2}}),\theta)\frac{\partial X}{\partial s}(\theta,s;x_{i+\frac{1}{2}})\,d\theta.\end{aligned} \tag{2.85}$$

Taking the derivative with respect to t of Equation (2.85), we have:

$$\frac{\partial}{\partial t}\left(\frac{\partial X}{\partial s}\right)(t,s;x_{i+\frac{1}{2}}) = \frac{\partial u}{\partial x}(X(t,s;x_{i+\frac{1}{2}}),t)\frac{\partial X}{\partial s}(t,s;x_{i+\frac{1}{2}}). \tag{2.86}$$

Using standard ODE techniques, we get that X that solves Equation (2.86) is given by:

$$\frac{\partial X}{\partial s}(t,s;x_{i+\frac{1}{2}}) = -\exp\left(\int_s^t \frac{\partial u}{\partial x}(X(\theta,s;x_{i+\frac{1}{2}}),\theta)\,d\theta\right)u(x_{i+\frac{1}{2}},s). \tag{2.87}$$

Computing q on the trajectory give by $X(t,s;x_{i+\frac{1}{2}})$ and taking its time derivative, we obtain:

$$\begin{aligned}\frac{dq}{dt}(X(t,s;x_{i+\frac{1}{2}}),t) &= \frac{\partial q}{\partial t}(X(t,s;x_{i+\frac{1}{2}}),t) + u(X(t,s;x_{i+\frac{1}{2}}),t)\frac{\partial q}{\partial x}(X(t,s;x_{i+\frac{1}{2}}),t)\\ &= -\frac{\partial u}{\partial x}(X(t,s;x_{i+\frac{1}{2}}),t)q(X(t,s;x_{i+\frac{1}{2}}),t),\end{aligned} \tag{2.88}$$

where we used that q satisfies the linear advection equation and that $X(t,s;x_{i+\frac{1}{2}})$ solves Equation (2.78). Using again standard ODE techniques, we get that q that solves Equation (2.88) is given by:

$$q(X(t,s;x_{i+\frac{1}{2}}),t) = \exp\left(-\int_s^t \frac{\partial u}{\partial x}(X(\theta,s;x_{i+\frac{1}{2}}),\theta)\,d\theta\right)q(x_{i+\frac{1}{2}},s). \tag{2.89}$$

Notice that if u does not depend on x, then q is constant along the trajectory $X(t,s;x_{i+\frac{1}{2}})$.

Let us consider the mapping $s \in [t^n, t^{n+1}] \to X(t^n,s,x_{i+\frac{1}{2}})$. Integrating q over all departure points at time t^n from $x_{i+\frac{1}{2}}$ at time s, we have

$$\int_{X(t^n,t^{n+1};x_{i+\frac{1}{2}})}^{X(t^n,t^n;x_{i+\frac{1}{2}})=x_{i+\frac{1}{2}}} q(x,t^n)\,dx = -\int_{t^n}^{t^{n+1}} q(X(t^n,s;x_{i+\frac{1}{2}}),t^n)\frac{\partial X}{\partial s}(t^n,s;x_{i+\frac{1}{2}})\,ds, \tag{2.90}$$

where we are just using the variable change integration formula. Then, it follows from Equations (2.87) and (2.89) with $t = t^n$ that:

$$\int_{X(t^n,t^{n+1};x_{i+\frac{1}{2}})}^{x_{i+\frac{1}{2}}} q(x,t^n)\,dx = \int_{t^n}^{t^{n+1}} (uq)(x_{i+\frac{1}{2}},s)\,ds \tag{2.91}$$

□

The next Proposition (2.9) gives an estimate to the approximation (2.77), but now we are going to use the estimate from Equation (2.81) to compute the departure point in

Equation (2.84).

Proposition 2.9. *Assume the framework of Problem 2.2 and also assume the CFL condition and that Δx and Δt are small enough as in Proposition 2.6. If $q \in C^1$ and $u \in C^2$, then:*

$$\left| \int_{t^n}^{t^{n+1}} (uq)(x_{i+\frac{1}{2}}, s)\, ds - \int_{x_{i+\frac{1}{2}} - u^n_{i+\frac{1}{2}}\Delta t}^{x_{i+\frac{1}{2}}} q(x, t^n)\, dx - \left(\int_{t^n}^{t^{n+1}} (uq)(x_{i-\frac{1}{2}}, s)\, ds - \int_{x_{i-\frac{1}{2}} - u^n_{i-\frac{1}{2}}\Delta t}^{x_{i-\frac{1}{2}}} q(x, t^n)\, dx \right) \right| \leq K_1 \Delta t^3,$$

where K_1 depends on q and u.

Proof. Using Equations (2.82), the mean value theorem for integrals and Equation (2.84), we get:

$$\int_{t^n}^{t^{n+1}} (uq)(x_{i+\frac{1}{2}}, s)\, ds - \int_{x_{i+\frac{1}{2}} - u^n_{i+\frac{1}{2}}\Delta t}^{x_{i+\frac{1}{2}}} q(x, t^n)\, dx = \int_{X(t^n, t^{n+1}; x_{i+\frac{1}{2}})}^{x_{i+\frac{1}{2}}} q(x, t^n)\, dx - \int_{x_{i+\frac{1}{2}} - u^n_{i+\frac{1}{2}}\Delta t}^{x_{i+\frac{1}{2}}} q(x, t^n)\, dx$$
$$= \int_{X(t^n, t^{n+1}; x_{i+\frac{1}{2}})}^{x_{i+\frac{1}{2}} - u^n_{i+\frac{1}{2}}\Delta t} q(x, t^n)\, dx = \left(X(t^n, t^{n+1}; x_{i+\frac{1}{2}}) - x_{i+\frac{1}{2}} - u^n_{i+\frac{1}{2}}\Delta t \right) q(\mu_3, t^n) = M_1(\mu_1, \mu_2)\Delta t^2 q(\mu_3, t^n),$$

for some $\mu_1, \mu_2 \in X_i \cup X_{i+1}$, $\mu_3 \in [t^n, t^{n+1}]$, where $M_1(\mu_1, \mu_2)$ is given by (2.83). Similarly, we have:

$$\int_{t^n}^{t^{n+1}} (uq)(x_{i-\frac{1}{2}}, s)\, ds - \int_{x_{i-\frac{1}{2}} - u^n_{i-\frac{1}{2}}\Delta t}^{x_{i-\frac{1}{2}}} q(x, t^n)\, dx = M_1(\lambda_1, \lambda_2)\Delta t^2 q(\lambda_3, t^n),$$

and again, $\lambda_1, \lambda_3 \in X_{i-1} \cup X_i$, $\lambda_2 \in [t^n, t^{n+1}]$, $M_1(\lambda_1, \lambda_2)$ is given by (2.83). We introduce the following auxiliary C^1 function:

$$F(v) = M_1(v_1, v_2) q(v_3, t^n).$$

where $v = (v_1, v_2, v_3)$, $v_1, v_3 \in [a, b]$, $v_2 \in [0, T]$. Using the mean value theorem, we have:

$$|F(\lambda) - F(\mu)| \leq \left(\sup_{v \in [a,b]\times[0,T]\times[a,b]} \|\nabla F(v)\|_2 \right) \|\lambda - \mu\|_2 \leq \left(\sqrt{1 + \frac{18}{\sigma^2}} \sup_{v \in [a,b]\times[0,T]\times[a,b]} \|\nabla F(v)\|_2 \right) \Delta t,$$

where $\|\cdot\|_2$ is the 2-norm in $\mathbb{R}^3$, and we used that $|\lambda_1 - \mu_1| \leq 3\Delta x$, $|\lambda_3 - \mu_3| \leq 3\Delta x$ and $|\lambda_2 - \mu_2| \leq \Delta t$. Finally, we have the desired bound:

$$\left| \int_{t^n}^{t^{n+1}} (uq)(x_{i+\frac{1}{2}}, s)\, ds - \int_{x_{i+\frac{1}{2}} - u^n_{i+\frac{1}{2}}\Delta t}^{x_{i+\frac{1}{2}}} q(x, t^n)\, dx - \left(\int_{t^n}^{t^{n+1}} (uq)(x_{i-\frac{1}{2}}, s)\, ds - \int_{x_{i-\frac{1}{2}} - u^n_{i-\frac{1}{2}}\Delta t}^{x_{i-\frac{1}{2}}} q(x, t^n)\, dx \right) \right|$$
$$= |(F(\mu) - F(\lambda))|\Delta t^2 \leq K_1 \Delta t^3,$$

where $K_1 = \sqrt{1 + \frac{18}{\sigma^2}} \sup_{v \in [a,b]\times[0,T]\times[a,b]} \|\nabla F(v)\|_2$. □

The next proposition gives a measure of the impact of the piecewise parabolic ap-

proximation on the time average flux, considering this last computed using an estimated departure point.

Proposition 2.10. *Assume the framework of Problem 2.2. If $q \in C^5$ and $u \in C^1$, then:*

$$\left| \frac{1}{\Delta t} \int_{x_{i+\frac{1}{2}} - u^n_{i+\frac{1}{2}} \Delta t}^{x_{i+\frac{1}{2}}} q(x, t^n)\, dx - \mathcal{F}(Q(t_n); u^n, i) - \left(\frac{1}{\Delta t} \int_{x_{i-\frac{1}{2}} - u^n_{i-\frac{1}{2}} \Delta t}^{x_{i-\frac{1}{2}}} q(x, t^n)\, dx - \mathcal{F}(Q(t_n); u^n, i-1) \right) \right| \leq K_2 \Delta x^4,$$

where K_2 depends on q and u.

Proof. We denote by q_{PP} the piecewise-parabolic approximation of $Q(t^n)$. Then:

$$\frac{1}{\Delta t} \int_{x_{i+\frac{1}{2}} - u^n_{i+\frac{1}{2}} \Delta t}^{x_{i+\frac{1}{2}}} q(x, t^n)\, dx - \mathcal{F}(Q(t_n); u^n, i) = \frac{1}{\Delta t} \int_{x_{i+\frac{1}{2}} - u^n_{i+\frac{1}{2}} \Delta t}^{x_{i+\frac{1}{2}}} q(x, t^n)\, dx - \frac{1}{\Delta t} \int_{x_{i+\frac{1}{2}} - u^n_{i+\frac{1}{2}} \Delta t}^{x_{i+\frac{1}{2}}} q_{PP}(x)\, dx$$

and

$$\frac{1}{\Delta t} \int_{x_{i-\frac{1}{2}} - u^n_{i-\frac{1}{2}} \Delta t}^{x_{i-\frac{1}{2}}} q(x, t^n)\, dx - \mathcal{F}(Q(t_n); u^n, i-1) = \frac{1}{\Delta t} \int_{x_{i-\frac{1}{2}} - u^n_{i-\frac{1}{2}} \Delta t}^{x_{i-\frac{1}{2}}} q(x, t^n)\, dx - \frac{1}{\Delta t} \int_{x_{i-\frac{1}{2}} - u^n_{i-\frac{1}{2}} \Delta t}^{x_{i-\frac{1}{2}}} q_{PP}(x)\, dx.$$

Similarly to Proposition 2.5, we can write:

$$\begin{aligned} q(x, t^n) - q_{PP}(x) =& C_1(\mu_1, \mu_2)\Delta x^4 + C_2(\mu_3, \mu_4)\Delta x^2 (x - x_L) + \frac{C_3}{2}(\mu_5, \mu_6)\Delta x (x - x_L)^2 \\ &+ C_4(\mu_7)(x - x_L)^3, \end{aligned}$$

where C_1, C_2, C_3 and C_4 are given by Equations (2.48), (2.58), (2.62) and (2.64) respectively, x_L is the left boundary of the control volume that contains x (X_i or X_{i+1}) and $\mu_k \in [x_{i+\frac{1}{2}} - 3\Delta x, x_{i+\frac{1}{2}} + 3\Delta x]$, $\forall k = 1, \cdots, 7$. Similarly to Proposition 2.9, using the mean value theorem for integrals, one can write:

$$\int_{x_{i+\frac{1}{2}} - u^n_{i+\frac{1}{2}} \Delta t}^{x_{i+\frac{1}{2}}} (q(x, t^n) - q_{PP}(x))\, dx = F(\lambda)\Delta x^4, \tag{2.92}$$

and

$$\int_{x_{i-\frac{1}{2}} - u^n_{i-\frac{1}{2}} \Delta t}^{x_{i-\frac{1}{2}}} (q(x, t^n) - q_{PP}(x))\, dx = F(\mu)\Delta x^4, \tag{2.93}$$

for an auxiliary function $F : [a, b]^8 \to \mathbb{R}$, $F \in C^1$, where F depends on q, u, and C_1, C_2, C_3 and C_4. Subtracting Equation (2.93) from Equation (2.92) and using the mean value theorem, we get the desired inequality. □

Now we are able to tackle the consistency problem on the next proposition.

Proposition 2.11. *Assume the framework of Problem 2.2 and also assume the CFL condition and that Δx and Δt are small enough as in Proposition 2.6. Denote by q_{PP} the Piecewise-Parabolic approximation of $q(x, t^n)$. Then, the LTE given by Equation (2.25) satisfies:*

$$|\tau_i^n| \leq M_1 \Delta t + M_2 \Delta x^3, \tag{2.94}$$

where M_1 and M_2 are constants depending only on q and u.

Proof. We have:

$$\Delta x \tau_i^n = \left| \frac{1}{\Delta t} \int_{t^n}^{t^{n+1}} (uq)(x_{i+\frac{1}{2}}, s)\, ds - \mathcal{F}(Q(t_n); u^n, i) - \left(\frac{1}{\Delta t} \int_{t^n}^{t^{n+1}} (uq)(x_{i-\frac{1}{2}}, s)\, ds - \mathcal{F}(Q(t_n); u^n, i-1) \right) \right| =$$

$$\frac{1}{\Delta t} \left| \int_{t^n}^{t^{n+1}} (uq)(x_{i+\frac{1}{2}}, s)\, ds - \int_{x_{i+\frac{1}{2}} - u^n_{i+\frac{1}{2}} \Delta t}^{x_{i+\frac{1}{2}}} q_{PP}(x)\, dx - \left(\int_{t^n}^{t^{n+1}} (uq)(x_{i-\frac{1}{2}}, s)\, ds - \int_{x_{i-\frac{1}{2}} - u^n_{i-\frac{1}{2}} \Delta t}^{x_{i-\frac{1}{2}}} q_{PP}(x)\, dx \right) \right| =$$

$$\frac{1}{\Delta t} \left| \int_{t^n}^{t^{n+1}} (uq)(x_{i+\frac{1}{2}}, s)\, ds - \int_{x_{i+\frac{1}{2}} - u^n_{i+\frac{1}{2}} \Delta t}^{x_{i+\frac{1}{2}}} q(x, t^n)\, dx + \int_{x_{i+\frac{1}{2}} - u^n_{i+\frac{1}{2}} \Delta t}^{x_{i+\frac{1}{2}}} q(x, t^n)\, dx - \int_{x_{i+\frac{1}{2}} - u^n_{i+\frac{1}{2}} \Delta t}^{x_{i+\frac{1}{2}}} q_{PP}(x)\, dx \right.$$

$$\left. - \left(\int_{t^n}^{t^{n+1}} (uq)(x_{i-\frac{1}{2}}, s)\, ds - \int_{x_{i-\frac{1}{2}} - u^n_{i-\frac{1}{2}} \Delta t}^{x_{i-\frac{1}{2}}} q(x, t^n)\, dx + \int_{x_{i-\frac{1}{2}} - u^n_{i-\frac{1}{2}} \Delta t}^{x_{i-\frac{1}{2}}} q(x, t^n)\, dx - \int_{x_{i-\frac{1}{2}} - u^n_{i-\frac{1}{2}} \Delta t}^{x_{i-\frac{1}{2}}} q_{PP}(x)\, dx \right) \right| \leq$$

$$\frac{1}{\Delta t} \left| \int_{t^n}^{t^{n+1}} (uq)(x_{i+\frac{1}{2}}, s)\, ds - \int_{x_{i+\frac{1}{2}} - u^n_{i+\frac{1}{2}} \Delta t}^{x_{i+\frac{1}{2}}} q(x, t^n)\, dx - \left(\int_{t^n}^{t^{n+1}} (uq)(x_{i-\frac{1}{2}}, s)\, ds - \int_{x_{i-\frac{1}{2}} - u^n_{i-\frac{1}{2}} \Delta t}^{x_{i-\frac{1}{2}}} q(x, t^n)\, dx \right) \right| +$$

$$\frac{1}{\Delta t} \left| \int_{x_{i+\frac{1}{2}} - u^n_{i+\frac{1}{2}} \Delta t}^{x_{i+\frac{1}{2}}} q(x, t^n)\, dx - \int_{x_{i+\frac{1}{2}} - u^n_{i+\frac{1}{2}} \Delta t}^{x_{i+\frac{1}{2}}} q_{PP}(x)\, dx - \left(\int_{x_{i-\frac{1}{2}} - u^n_{i-\frac{1}{2}} \Delta t}^{x_{i-\frac{1}{2}}} q(x, t^n)\, dx - \int_{x_{i-\frac{1}{2}} - u^n_{i-\frac{1}{2}} \Delta t}^{x_{i-\frac{1}{2}}} q_{PP}(x)\, dx \right) \right|$$

Therefore, it follows from Propositions 2.8 and 2.9 and than

$$|\tau_i^n| \leq \frac{1}{\Delta x \Delta t} K_1 \Delta t^3 + \frac{1}{\Delta x} K_2 \Delta x^4 = K_1 \sigma \Delta t + K_2 \Delta x^3$$

from which the proposition follows. □

Thus, in the PPM flux estimation, we have a first-order error related to the departure point computation. We point out that, if the velocity is constant, then no error is obtained using the PPM flux, except for the approximation q_{PP} to q.

2.4 Numerical experiements

This Section is dedicated to presenting the numerical results of the PPM and its variations discussed here. For non-monotonic schemes, we are going to consider the original PPM from Colella and Woodward (1984) and the hybrid PPM from Putman and Lin (2007). For monotonic schemes, we are going to consider the monotonization schemes from Colella and Woodward (1984) and Lin (2004), which are referred to as CW84 monotonization and L04 monotonization hereafter. In Subsection 2.4.1 we present results using the linear advection equation with constant velocity and in Subsection 2.4.2 the results are based on the linear advection equation with variable velocity. The code used in this Section may be found in Appendix C.

2.4.1 Linear advection equation with constant velocity simulations

For the linear advection equation with the constant velocity we shall adopt the $u = 0.2$ and a CFL number equal to 0.8. The spatial domain will be given by $[0, 1]$ and the time integration interval will be $[0, 5]$. Since we are going to assume periodic boundary conditions, the period is equal to 5. Hence, the simulations presented here shall advect an initial profile for one time period. This shall be the general setup for all simulations presented in this subsection. What will distinguish the simulations is the initial condition. The first q_0 is given by:

$$q_0(x) = \sin(2\pi k x) + 1. \tag{2.95}$$

Here k denotes the wavenumber and we adopt $k = 5$. Inspired by Trefethen (2000), we adopted the following periodic Gaussian profile.

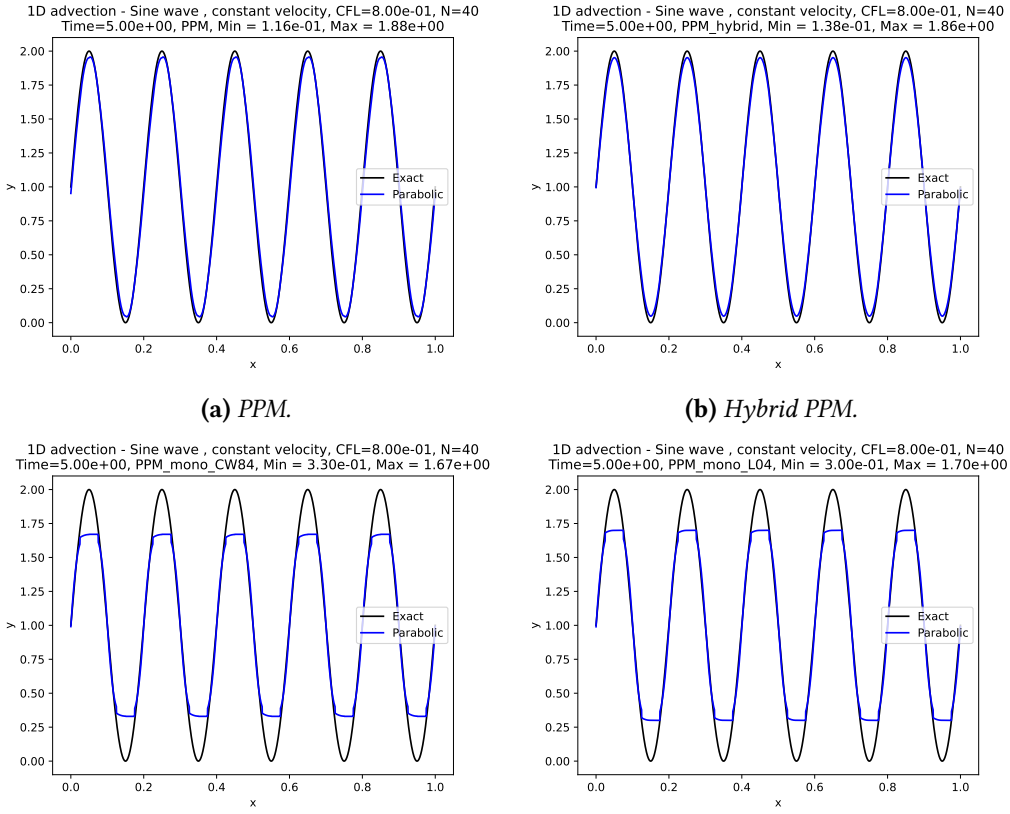
$$q_0(x) = \exp(-10\cos^2(2\pi x)). \tag{2.96}$$

Both functions from Equations (2.96) and (2.96) are smooth. We also consider a discontinuous initial condition given by:

$$q_0(x) = \begin{cases} 1 & \text{if } x \in [0.4, 0.6], \\ 0 & \text{otherwise.} \end{cases} \tag{2.97}$$

It is easy to check that the exact solution of Problem 2.1 is given by $q_0(x - ut)$ for all q_0 presented here.

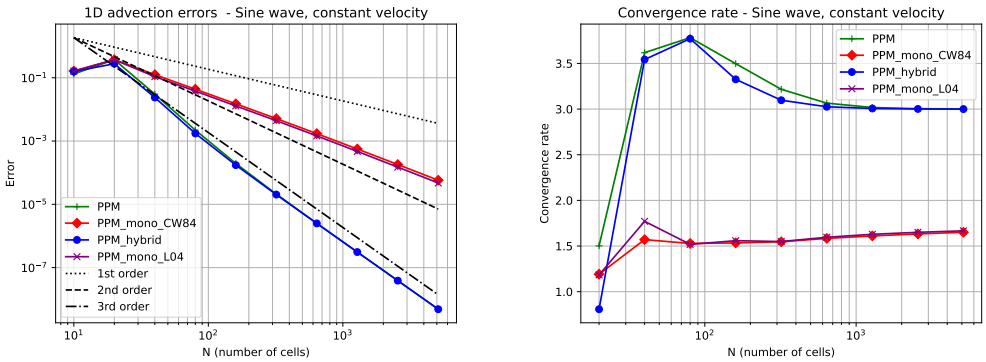
As pointed in Subsection 2.2.2, when q_0 is given by Equation (2.96), we are going to compute the initial average values $Q_i(0)$ using the initial values of q_i^0 at the control volume centroids, which is second-order accurate by Proposition 2.1. In the error calculation, only when q_0 is given by Equation (2.96), we replace $Q_i(t^n)$ by its centroid value $q_i(t^n)$, which again gives a second-order approximation by Proposition 2.1. Therefore, since q_0 is smooth, we expect that the error convergence shall be at least second-order accurate. Finally, the error norm is normalized by dividing the error norm by the exact solution norm. The norm adopted here is the maximum norm.



(a) *PPM.* **(b)** *Hybrid PPM.*

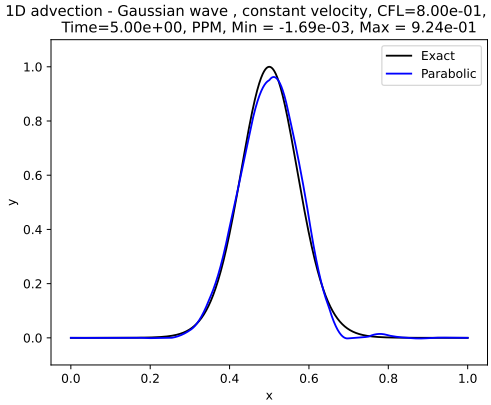
(c) *PPM + CW84 monotonization.* **(d)** *PPM + L04 monotonization.*

Figure 2.2: *Linear advection experiement using a constant velocity equal to* 0.1*, a CFL number equal to* 0.8*,* $N = 40$ *cells, and the initial condition is given by Equation* (2.95)*. These figures show the advected profile after 5 seconds (one time period). Schemes employed: PPM (a), hybrid PPM (b), PPM with the CW84 monotonization (c) and PPM with L04 monotonization (d).*

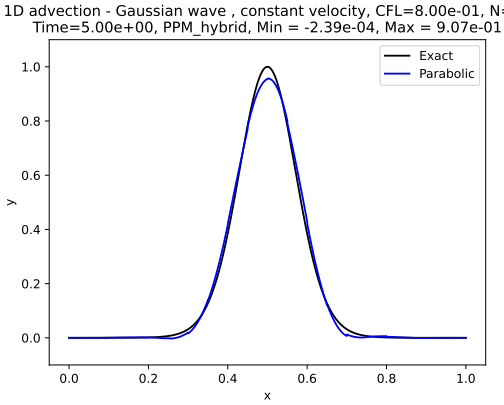


(a) *Error.* **(b)** *Convergence rate.*

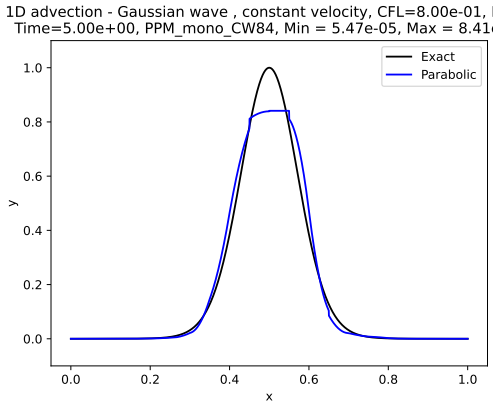
Figure 2.3: *Convergence of the error (a) and convergence rate (b) for the schemes PPM, hybrid PPM, PPM with the CW84 monotonization and PPM with L04 monotonization applied to the linear advection problem using a constant velocity equal to* 0.1*, a CFL number equal to* 0.8*, a final time of integration equal to 5 seconds and the initial condition given by Equation* (2.95)*.*



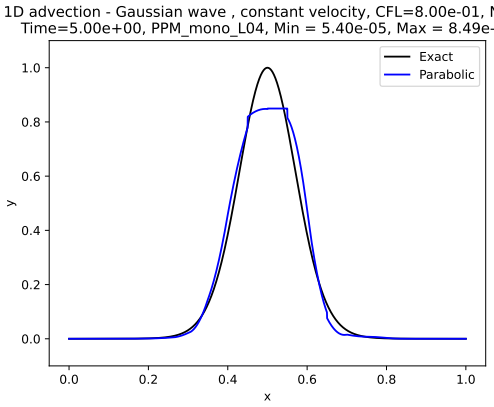
(a) *PPM.*



(b) *Hybrid PPM.*

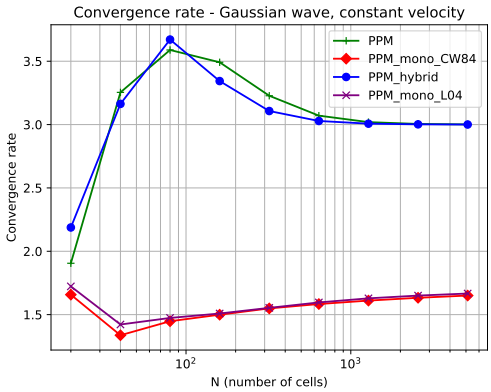


(c) *PPM + CW84 monotonization.*

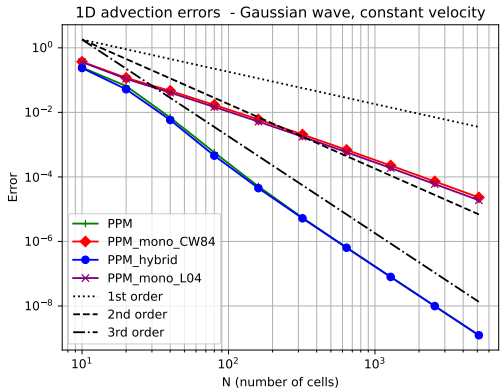


(d) *PPM + L04 monotonization.*

Figure 2.4: *Similar to Figure 2.2 but using $N = 20$ and the initial condition given by Equation (2.96).*

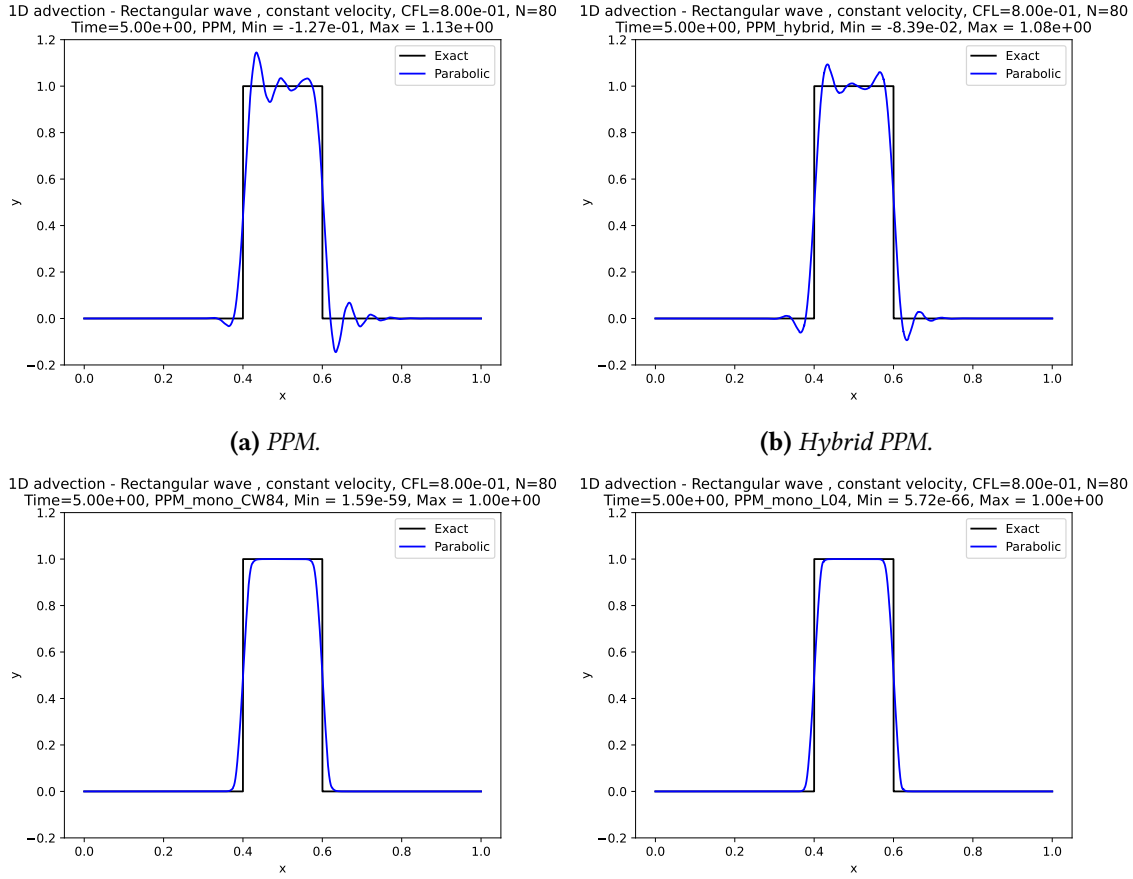


(a) *Error.*



(b) *Convergence rate.*

Figure 2.5: *Similar to Figure 2.3 but using the initial condition given by Equation (2.96).*



(a) *PPM.*

(b) *Hybrid PPM.*

(c) *PPM + CW84 monotonization.*

(d) *PPM + L04 monotonization.*

Figure 2.6: *Similar to Figure 2.2 but using* $N = 80$ *and the initial condition given by Equation (2.97).*

OKOK

2.4.2 Linear advection equation with variable velocity simulations

In this Subsection, we shall investigate the how the PPM schemes behaves when the velocity is variable. The initial condition is always given by Equation (2.96). The relative errors are computed using the centroid values of q as described in Subsection 2.4.2. We are going to consider the velocities

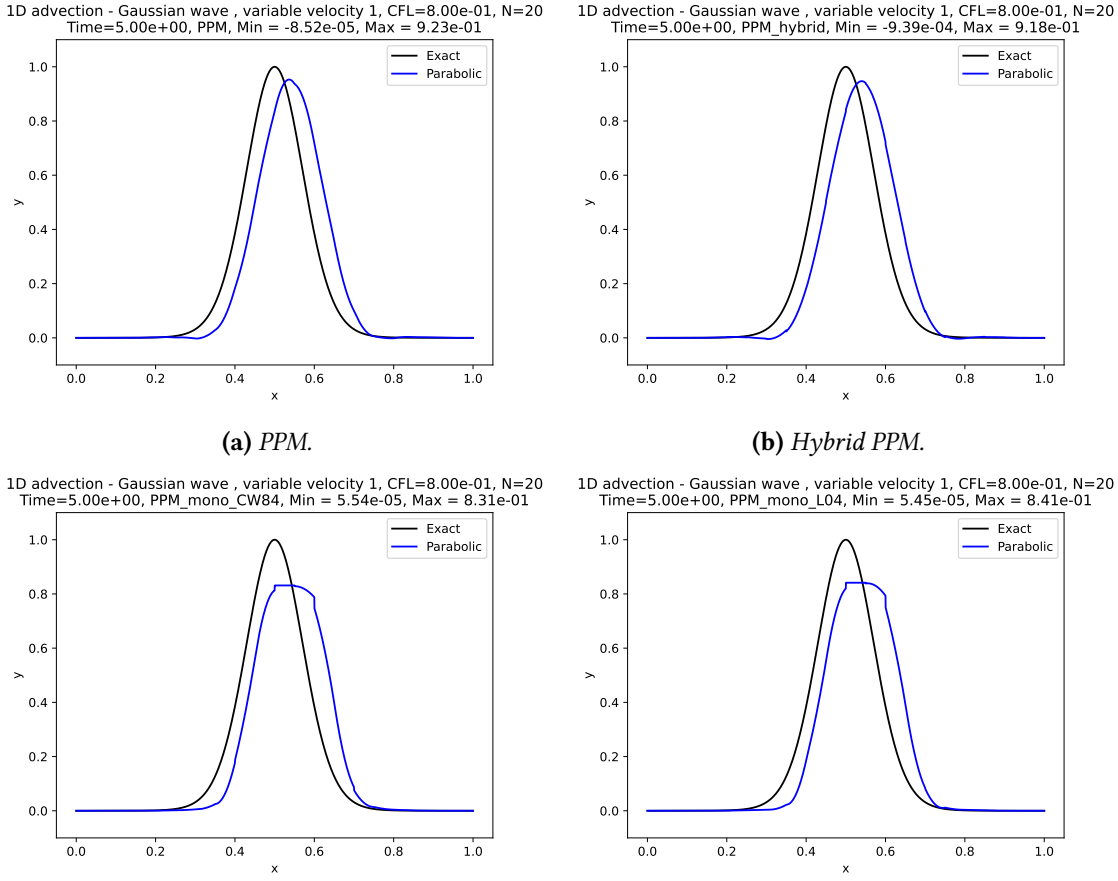
$$u(x,t) = u_0 \cos\left(\frac{\pi t}{T}\right), \tag{2.98}$$

and

$$u(x,t) = u_0 \cos\left(\frac{\pi t}{T}\right) \sin^2(\pi x). \tag{2.99}$$

We adopt the parameters $u_0 = 0.2$ and $T = 5$. In both cases, the solution has a period equal to 5. Therefore, the profile after 5 seconds is equal to the initial profile and we can compute the error. We remark that the velocity from Equation (2.99) is based on the deformational flow test case from onNair and Lauritzen (2010).

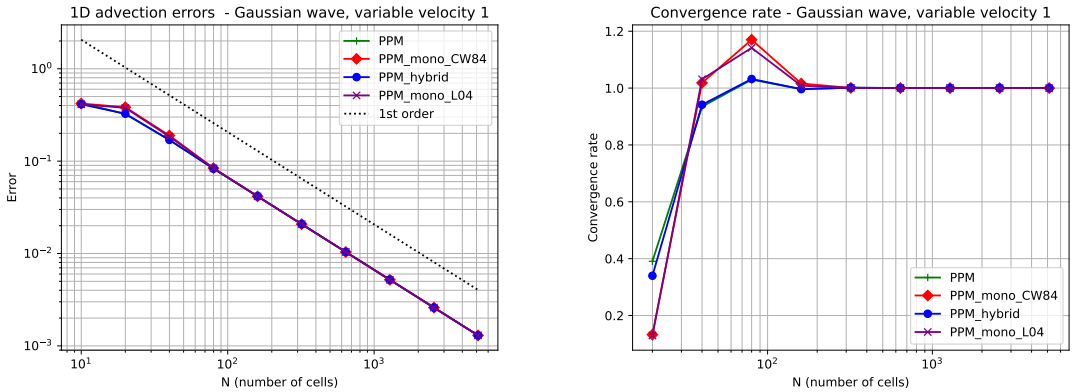
The velocity from Equation (2.98) varies only with time while the velocity from Equation (2.99) varies with both time and space.



(a) *PPM.* **(b)** *Hybrid PPM.*

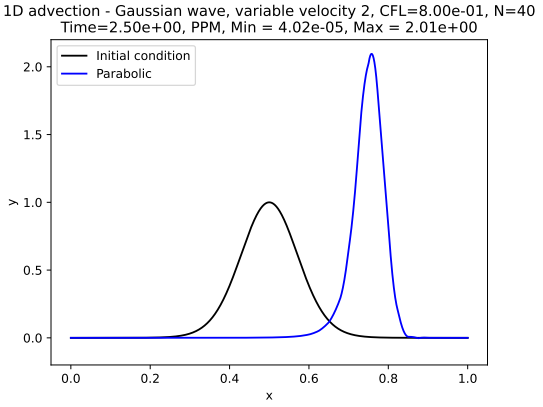
(c) *PPM + CW84 monotonization.* **(d)** *PPM + L04 monotonization.*

Figure 2.7: *Similar to Figure 2.2 but using $N = 20$, the initial condition given by Equation (2.96) and the variable velocity given by Equation*

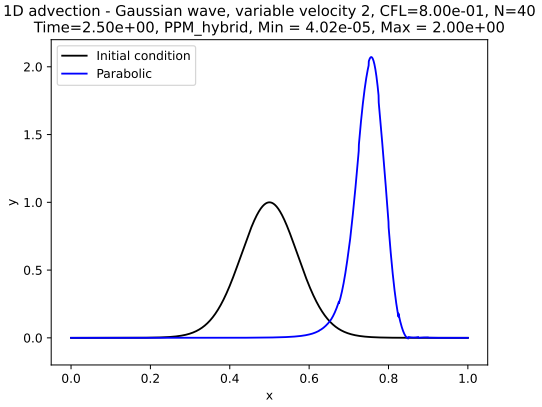


(a) *Error.* **(b)** *Convergence rate.*

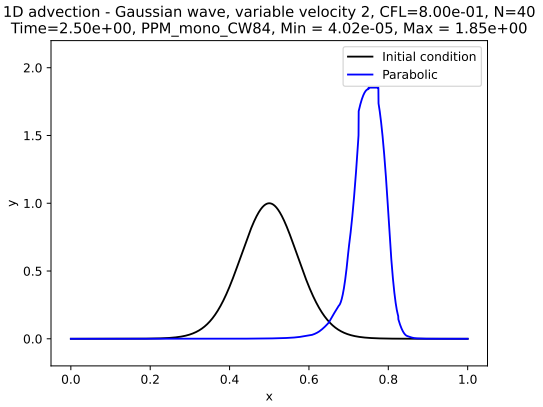
Figure 2.8: *Similar to Figure 2.3 but using the initial condition given by Equation (2.96) and the variable velocity given by Equation (2.98).*



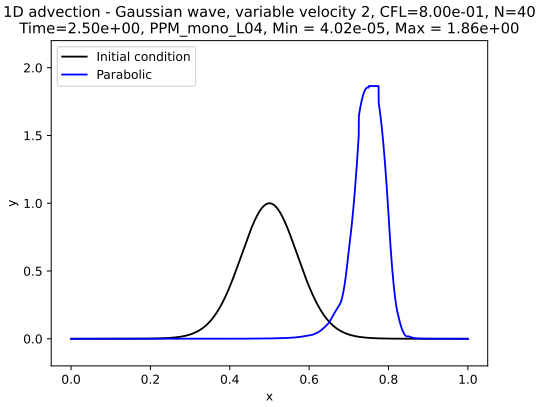
(a) *PPM.*



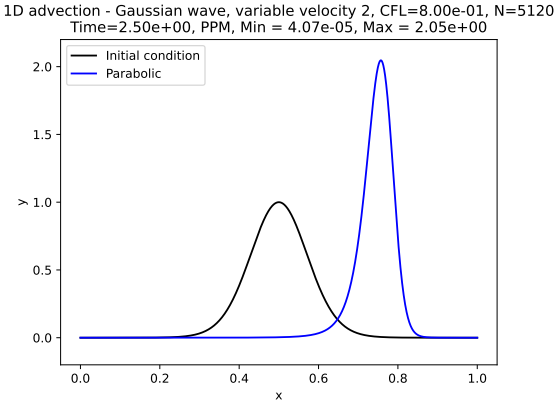
(b) *Hybrid PPM.*



(c) *PPM + CW84 monotonization.*

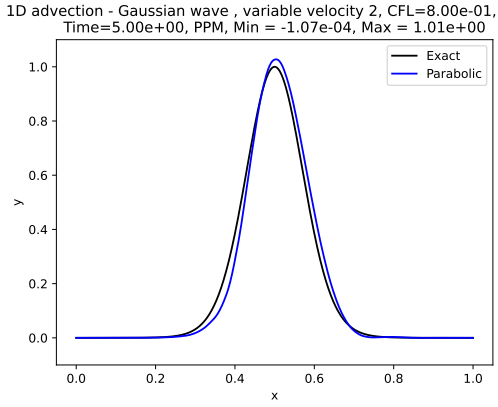


(d) *PPM + L04 monotonization.*

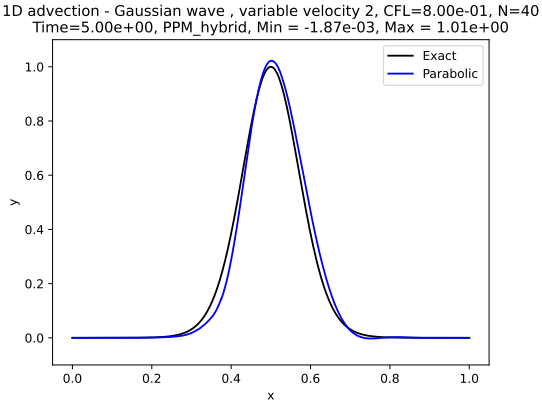


(e) *Reference solution.*

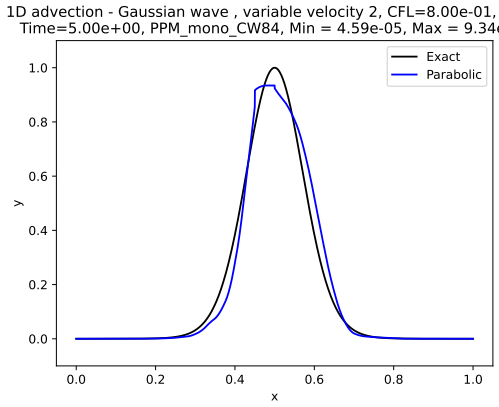
Figure 2.9: *Similar to Figure 2.2 but using $N = 40$, the initial condition given by Equation (2.96), the variable velocity given by Equation (2.99) and the final time is 2.5 (half a period). In (e) we show a reference solution, using the PPM scheme with 5120 cells.*



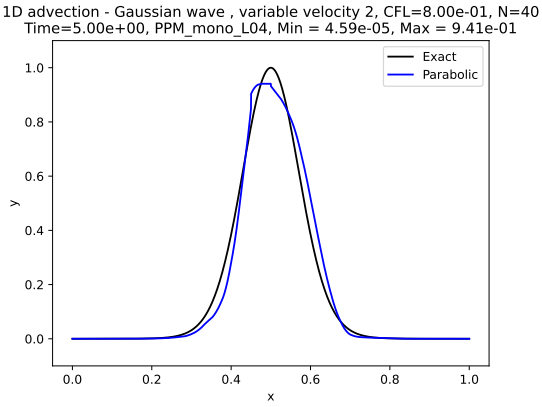
(a) *PPM.*



(b) *Hybrid PPM.*

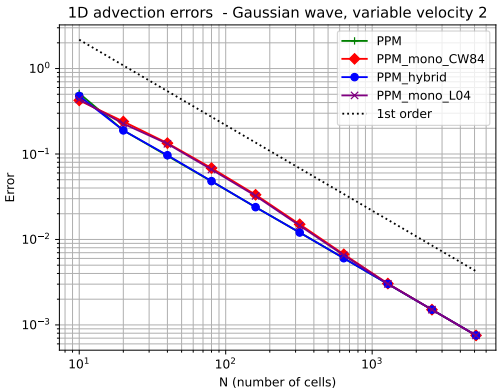


(c) *PPM + CW84 monotonization.*

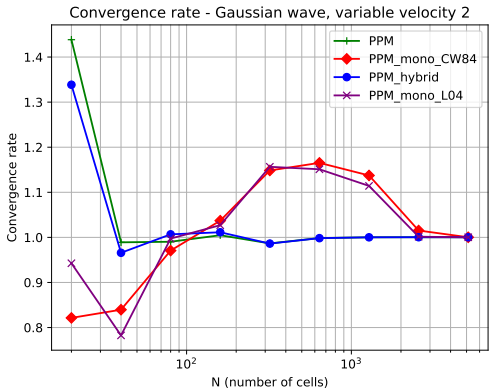


(d) *PPM + L04 monotonization.*

Figure 2.10: *Similar to Figure 2.4 but using $N = 40$, the initial condition given by Equation (2.96) and the variable velocity given by Equation (2.99) .*



(a) *Error.*



(b) *Convergence rate.*

Figure 2.11: *Similar to Figure 2.3 but using the initial condition given by Equation (2.96) and the variable velocity given by Equation (2.99).*

2.5 Concluding remarks

Chapter 3

Two-dimensional finite-volume methods

The main aim of this chapter is to give a detailed description of the dimensional splitting method proposed by Lin and Rood (1996). Similarly to Chapter 2, we start this chapter start with a review of two-dimensional conservation laws in the integral form in Section 3.1, and in Section 3.2 we set the framework of general two-dimensional finite-volumes schemes. Section 3.3 presents the dimension splitting method and numerical simulation are shown in Section 3.4.

3.1 Two-dimensional system of conservation laws in integral form

Let us consider $\mathcal{C}^1$ flux functions $f : \mathbb{R}^m \to \mathbb{R}^m$ and $g : \mathbb{R}^m \to \mathbb{R}^m$ in x and y direction, respectively. A two-dimensional system of conservation laws in the differential form in a domain $\Omega = [a,b] \times [c,d] \subset \mathbb{R}^2$ associated to the fluxes f and g is given by:

$$\frac{\partial}{\partial t} q(x,y,t) + \frac{\partial}{\partial x} f(q(x,y,t)) + \frac{\partial}{\partial y} g(q(x,y,t)) = 0, \quad \forall (x,y,t) \in \Omega^{o} \times]0,+\infty[.^{1} \tag{3.1}$$

The solution q is interpreted as the vector of state variable densities. A classical or strong solution to this system of conservation laws is a $\mathcal{C}^1$ function q satisfying Equation (3.1). As we did in Section 2.1, our goal is to deduce an integral form of Equation (3.1). To do so, let us consider $[x_1,x_2] \times [y_1,y_2] \subset \Omega^{o}$ and $[t_1,t_2] \subset [0,+\infty[$. Integrating Equation (3.1)

[1] Ω^{o} denotes the interior of Ω. Namely, $\Omega^{o} =]a,b[\times]c,d[$.

over $[x_1, x_2] \times [y_1, y_2]$ yields:

$$\frac{d}{dt}\left(\int_{x_1}^{x_2} \int_{y_1}^{y_2} q(x,y,t)\, dx\, dy \right) = -\int_{y_1}^{y_2} \Big(f(q(x_2,y,t)) - f(q(x_1,y,t)) \Big)\, dy - \int_{x_1}^{x_2} \Big(g(q(x,y_2,t)) - g(q(x,y_1,t)) \Big)\, dx. \tag{3.2}$$

Integrating Equation (3.2) over the time interval $[t_1, t_2]$, we have:

$$\begin{aligned} \int_{x_1}^{x_2} \int_{y_1}^{y_2} q(x,y,t_{n+1})\, dx\, dy = & \int_{x_1}^{x_2} \int_{y_1}^{y_2} q(x,y,t_n)\, dx\, dy \\ & - \int_{t_1}^{t_2} \int_{y_1}^{y_2} \Big(f(q(x_2,y,t)) - f(q(x_1,y,t)) \Big)\, dy\, dt \\ & - \int_{t_1}^{t_2} \int_{x_1}^{x_2} \Big(g(q(x,y_2,t)) - g(q(x,y_1,t)) \Big)\, dx\, dt. \end{aligned} \tag{3.3}$$

Equation (3.3) is the integral form of Equation (3.1). We say that $q \in L^\infty(\Omega \times [0, +\infty[, \mathbb{R}^m)$ is a weak solution to the system of conservation laws (3.1) if q satisfies the integral form (3.3), $\forall [x_1, x_2] \times [y_1, y_2] \subset \Omega^o$ and $\forall [t_1, t_2] \subset [0, +\infty[$. Similarly to Section 2.1, these problems are equivalent when q is a C^1 function.

We consider an initial condition $q_0 \in L^\infty(\Omega)$, $q(x, y, 0) = q_0(x, y)$, $\forall (x, y) \in \Omega$. Boundary conditions will be assumed bi-periodic. At last, the matrix $\alpha Df(q) + \beta Dg(q)$ is assumed to have real eigenvalues and be diagonalizable $\forall q \in \mathbb{R}^m, \forall \alpha, \beta \in \mathbb{R}$ (LeVeque, 1990), so that we have a hyperbolic conservation law. Therefore, we are again dealing with a Cauchy problem.

To move in the direction of a discrete version of Equation (3.3), let us discretize the domain $D = \Omega \times [0, T]$ following the notations of Section 2.1. Given a positive integer N_T, we define the time step $\Delta t = \frac{T}{N_T}$, $t_n = n\Delta t$, for $n = 0, 1, \cdots, N_T$. The spatial discretization is constructed through an uniformly spaced partition of Ω given by:

$$[a, b] = \bigcup_{i=1}^{N} X_i, \text{ where } X_i = [x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}] \text{ and } a = x_{\frac{1}{2}} < x_{\frac{3}{2}} < \cdots < x_{N-\frac{1}{2}} < x_{N+\frac{1}{2}} = b, \tag{3.4}$$

$$[c, d] = \bigcup_{j=1}^{M} Y_j, \text{ where } Y_j = [y_{j-\frac{1}{2}}, y_{j+\frac{1}{2}}] \text{ and } c = y_{\frac{1}{2}} < y_{\frac{3}{2}} < \cdots < y_{M-\frac{1}{2}} < y_{M+\frac{1}{2}} = d, \tag{3.5}$$

$$\Omega = \bigcup_{i=1}^{N} \bigcup_{j=1}^{M} \Omega_{ij}, \text{ where } \Omega_{ij} = X_i \times Y_j. \tag{3.6}$$

The regions Ω_{ij} are known as control volumes. Similarly to Chapter 2 we employ the notations $\Delta x = x_{i+\frac{1}{2}} - x_{i-\frac{1}{2}}$, $\Delta y = y_{j+\frac{1}{2}} - y_{j-\frac{1}{2}}$ and $x_i = \frac{1}{2}(x_{i+\frac{1}{2}} + x_{i-\frac{1}{2}})$ $y_j = \frac{1}{2}(y_{j+\frac{1}{2}} + y_{j-\frac{1}{2}})$, $\forall i = 1, \cdots, N, \forall j = 1, \cdots, M$, to define the control volume lengths and midpoints, respectively. Finally, we denote by $Q_{ij}(t) \in \mathbb{R}^m$ as the vector of average values of state variable vector at

time t in the control volume Ω_{ij}, that is:

$$Q_{ij}(t) = \frac{1}{\Delta x \Delta y} \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} \int_{y_{j-\frac{1}{2}}}^{y_{j-\frac{1}{2}}} q(x,t)\,dx \in \mathbb{R}^m. \tag{3.7}$$

Substituting t_1, t_2, x_1, x_2, y_1 and y_2 by $t_n, t_{n+1}, x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}, y_{j-\frac{1}{2}}, y_{j+\frac{1}{2}}$, respectively, in Equation (3.3), we obtain:

$$\begin{aligned} Q_{ij}(t_{n+1}) = Q_{ij}(t_n) &- \frac{\Delta t}{\Delta x \Delta y}\delta_x \left(\frac{1}{\Delta t} \int_{t_1}^{t_2} \int_{y_{j-\frac{1}{2}}}^{y_{j+\frac{1}{2}}} f(q(x_i, y, t))\,dy\,dt \right) \\ &- \frac{\Delta t}{\Delta x \Delta y}\delta_y \left(\frac{1}{\Delta t} \int_{t_1}^{t_2} \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} g(q(x, y_j, t))\,dx\,dt \right), \end{aligned} \tag{3.8}$$

where we are using the centered finite-difference notation:

$$\delta_x h(x_i, y, t) = h(x_{i+\frac{1}{2}}, y, t) - h(x_{i-\frac{1}{2}}, y, t), \tag{3.9}$$

$$\delta_y h(x, y_j, t) = h(x, y_{j+\frac{1}{2}}, t) - h(x, y_{j-\frac{1}{2}}, t), \tag{3.10}$$

for any function h. The Equation (3.8) is useful to motivate two-dimensional finite-volume schemes, as we shall see in the next section.

3.2 The finite-volume approach

This Section is basically an extension to two dimensions of the concepts presented in Section 2.2. The problem of two-dimensional system of conservation laws in the integral form presented Section 3.1 is written in a concise way in Problem 3.1.

Problem 3.1. *Given $\Omega = [a,b] \times [c,d]$, $D = \Omega \times [0,T]$, C^1 flux functions $f, g : \mathbb{R}^m \to \mathbb{R}^m$, $m \geq 1$, we would like to find the weak solution $q \in L^\infty(D, \mathbb{R}^m)$ of the two-dimensional system of conservation laws in the integral form:*

$$\begin{aligned} \int_{x_1}^{x_2} \int_{y_1}^{y_2} q(x,y,t)\,dx\,dy = &\int_{x_1}^{x_2} \int_{y_1}^{y_2} q(x,y,t)\,dx\,dy \\ &- \int_{t_1}^{t_2} \int_{y_1}^{y_2} \Big(f(q(x_2,y,t)) - f(q(x_1,y,t)) \Big)\,dy\,dt \\ &- \int_{t_1}^{t_2} \int_{x_1}^{x_2} \Big(g(q(x,y_2,t)) - g(q(x,y_1,t)) \Big)\,dx\,dt. \end{aligned}$$

$\forall [x_1,x_2]\times[y_1,y_2]\times[t_1,t_2] \subset D$, given the initial condition $q(x,y,0) = q_0(x,y)$, $\forall (x,y) \in \Omega$, and assuming bi-periodic boundary conditions, i.e., $q(a,y,t) = q(b,y,t)$, $\forall t \in [0,T]$, $\forall y \in [c,d]$, and $q(x,c,t) = q(x,d,t)$, $\forall t \in [0,T]$, $\forall x \in [a,b]$.

For Problem 3.1, the total mass in Ω is defined by:

$$M_\Omega(t) = \int_\Omega q(x,y,t)\,dx\,dy \in \mathbb{R}^m, \quad \forall t \in [0,T], \tag{3.11}$$

and is conserved within time:

$$M_\Omega(t) = M_\Omega(0), \quad \forall t \in [0,T]. \tag{3.12}$$

Section 3.1 introduced a version of Problem 3.1 considering a discretization of the domain D. This version is also summarized in Problem 3.2.

Problem 3.2. *Assume the framework of Problem 3.1. We consider positive integers N and N_T, a spatial discretization of [a,b] given by $X_i = [x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}]$, $\forall i = 1, \cdots, N$, $a = x_{\frac{1}{2}} < x_{\frac{3}{2}} < \cdots < x_{N-\frac{1}{2}} < x_{N+\frac{1}{2}} = b$, $\Delta x = x_{i+\frac{1}{2}} - x_{i-\frac{1}{2}}$, $Y_j = [y_{j-\frac{1}{2}}, y_{j+\frac{1}{2}}]$, $\forall j = 1, \cdots, M$, $c = y_{\frac{1}{2}} < y_{\frac{3}{2}} < \cdots < y_{M-\frac{1}{2}} < y_{M+\frac{1}{2}} = d$, $\Delta y = y_{j+\frac{1}{2}} - y_{j-\frac{1}{2}}$, $\Omega_{ij} = X_i \times Y_j$, a time discretization $t_n = n\Delta t$, $\Delta t = \frac{T}{N_T}$, $\forall n = 1, \cdots, N_T$. Since we are in the framework of Problem 3.2, it follows that:*

$$\begin{aligned} Q_{ij}(t_{n+1}) = Q_{ij}(t_n) &- \frac{\Delta t}{\Delta x \Delta y}\delta_x\left(\frac{1}{\Delta t}\int_{t_1}^{t_2}\int_{y_{j-\frac{1}{2}}}^{y_{j+\frac{1}{2}}} f(q(x_i, y, t))\, dy\, dt\right) \\ &- \frac{\Delta t}{\Delta x \Delta y}\delta_y\left(\frac{1}{\Delta t}\int_{t_1}^{t_2}\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} g(q(x, y_j, t))\, dx\, dt\right), \end{aligned}$$

where $Q_{ij}(t) = \frac{1}{\Delta x \Delta y}\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}}\int_{y_{j-\frac{1}{2}}}^{y_{j+\frac{1}{2}}} q(x, y, t)\, dx\, dy$.

Our problem now consists of finding the values $Q_{ij}(t_n)$, $\forall i = 1, \cdots, N$, $\forall j = 1, \cdots, M$, $\forall n = 1, \cdots, N_T$, given the initial values $Q_{ij}(0)$, $\forall i = 1, \cdots N$, $\forall j = 1, \cdots, M$. In other words, we would like to find the average values of q in each control volume Ω_{ij} at the considered time instants.

Finally, we define the one-dimensional (2D) finite-volume (FV) scheme problem as follows in Problem 3.3. We use the notation $q_{ij}^n = q(x_i, y_j, t_n)$ to represent the values of q in the discrete domain D and $u_{i+\frac{1}{2},j}^n = u(x_{i+\frac{1}{2}}, y_j, t_n)$ to represent the velocity in x direction at control volume edges midpoints in the x direction, and $v_{i,j+\frac{1}{2}}^n = v(x_i, y_{j+\frac{1}{2}}, t_n)$ to represent the velocity in y direction at control volume edges midpoints in the y direction,

Problem 3.3 (2D-FV scheme)**.** *Assume the framework defined in Problem 3.2. The finite-volume approach of Problem 3.1 consists of a finding a scheme of the form:*

$$\begin{gathered} Q_{ij}^{n+1} = Q_{ij}^n - \frac{\Delta t}{\Delta x \Delta y}\delta_i F_{i,j}^n - \frac{\Delta t}{\Delta x \Delta y}\delta_j G_{i,j}^n, \\ \forall i = 1, \cdots, N, \quad \forall j = 1, \cdots, M, \quad \forall n = 1, \cdots, N_T - 1, \end{gathered}$$

where $\delta_i F_{ij}^n = F_{i+\frac{1}{2},j}^n - F_{i-\frac{1}{2},j}^n$, $\delta_j G_{ij}^n = G_{i,j+\frac{1}{2}}^n - G_{i,j-\frac{1}{2}}^n$ and $Q_{ij}^n \in \mathbb{R}^m$ is intended to be an approximation of $Q_{ij}(t_n)$ in some sense. We define by $Q_{ij}^0 = Q_{ij}(0)$ or $Q_{ij}^0 = q_{i,j}^0$.

The term $F_{i+\frac{1}{2},j}^n = \mathcal{F}(Q^n; u^n; v^n, i, j)$ is known as numerical flux in the x direction, where $\mathcal{F}$ is the numerical flux function, and it approximates $\frac{1}{\Delta t}\int_{t_n}^{t_{n+1}}\int_{y_{j-\frac{1}{2}}}^{y_{j+\frac{1}{2}}} f(q(x_{i+\frac{1}{2}}, y, t))\, dy\, dt$, $\forall i = 0, 1, \cdots, N$, and $G_{i,j+\frac{1}{2}}^n = \mathcal{G}(Q^n, u^n, v^n; i, j)$ is known as numerical flux in the y direction, where $\mathcal{G}$ is the numerical flux function, and it approximates $\frac{1}{\Delta t}\int_{t_n}^{t_{n+1}}\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} g(q(x, y_{j+\frac{1}{2}}, t))\, dx\, dt$, $\forall j = 0, 1, \cdots, M$, or, in other words, they estimate the time-averaged fluxes at the control

volume Ω_{ij} boundaries.

Notice that we are using the notations $Q^n = (Q^n_{ij})_{i=1,\cdots,N,j=1,\cdots,M}$, $u^n = (u^n_{i+\frac{1}{2},j})_{i=0,\cdots,N,j=1,\cdots,M}$ and $v^n = (v^n_{i,j+\frac{1}{2}})_{i=1,\cdots,N,j=0,\cdots,M}$.

3.3 Dimension splitting

In this section we shall describe how we can combine one-dimensional to solve the bidimensional advection equation on the plane. As we shall later, the solution of the advection equation on the cubed-sphere consists of applying the scheme described here in each panel combined with interpolation on ghost cells. The dimension splitting method presented here was developed by Lin and Rood (1996) and is currently employed in the FV3 dynamical core.

As we mentioned, we consired the two-dimensional advection linear equation on the plane with biperiodic boundary conditions. The flux functions are then given by $f(q) = u(x,y,t)q(x,y,t)$ and $g(q) = v(x,y,t)q(x,y,t)$.

The idea of the dimension splitting consists of solving the advection equation firstly in the x direction using $F^n_{i+\frac{1}{2},j}$ as the PPM flux:

$$Q^{x,n+1}_{ij} = Q^n_{ij} - \frac{\Delta t}{\Delta x}(F_{i+\frac{1}{2},j}(Q^n) - F_{i-\frac{1}{2},j}(Q^n)) \tag{3.13}$$

Then, we solve the advection equation in the y direction using $G^n_{i,j+\frac{1}{2}}$ as the PPM flux in the y direction:

$$Q^{xy,n+1}_{ij} = Q^{x,n+1}_{ij} - \frac{\Delta t}{\Delta y}(G_{i,j+\frac{1}{2}}(Q^{x,n+1}) - G_{i,j-\frac{1}{2}}(Q^{x,n+1}))$$

Notice that we can repeat this process in a reverse order. Indeed, we can solve the advection equation in y direction using $G^n_{i,j+\frac{1}{2}}$ as the PPM flux in the y direction:

$$Q^{y,n+1}_{ij} = Q^n_{ij} - \frac{\Delta t}{\Delta y}(G_{i,j+\frac{1}{2}}(Q^n) - G_{i,j-\frac{1}{2}}(Q^n))$$

Then, we can solve the advection equation in the x direction using $F^n_{i+\frac{1}{2},j}$ as the PPM flux in the x direction:

$$Q^{yx,n+1}_{ij} = Q^{x,n+1}_{ij} - \frac{\Delta t}{\Delta x}(F_{i+\frac{1}{2},j}(Q^{y,n+1}) - F_{i-\frac{1}{2},j}(Q^{y,n+1}))$$

Finally, we can eliminate the direction bias by computing the updated solution using an averaged solution: $Q^{n+1} = \frac{(Q^{xy,n+1}+Q^{yx,n+1})}{2}$

3.4 Numerical experiments

Chapter 4

Cubed-sphere grids

The cubed-sphere grid was originally proposed by Sadourny (1972) and was reinvestigated by Ronchi et al. (1996) and Rančić et al. (1996). As it is usual to Planotic grids, we start with a cube circumscribed in a sphere and project its faces on the sphere. This chapter aims to review and investigated geometrical properties of the cubed-spheres available in the literature. This chapter is under development.

4.1 Cubed-sphere mappings

4.1.1 Equidistant cubed-sphere

We consider a sphere radius $R > 0$ and $a = \frac{R}{\sqrt{3}}$ representing the half-length of the cube, and the family of maps $\Psi_p : [-a, a] \times [-a, a] \to \mathbb{S}^2_R$, $p = 1, \cdots, 6$, where:

$$\Psi_1(x, y) = \frac{R}{\sqrt{a^2 + x^2 + y^2}}(a, x, y), \tag{4.1}$$

$$\Psi_2(x, y) = \frac{R}{\sqrt{a^2 + x^2 + y^2}}(-x, a, y), \tag{4.2}$$

$$\Psi_3(x, y) = \frac{R}{\sqrt{a^2 + x^2 + y^2}}(-a, -x, y), \tag{4.3}$$

$$\Psi_4(x, y) = \frac{R}{\sqrt{a^2 + x^2 + y^2}}(x, -a, y), \tag{4.4}$$

$$\Psi_5(x, y) = \frac{R}{\sqrt{a^2 + x^2 + y^2}}(-y, x, a), \tag{4.5}$$

$$\Psi_6(x, y) = \frac{R}{\sqrt{a^2 + x^2 + y^2}}(y, x, -a). \tag{4.6}$$

The family of maps $\{\Psi_p, p = 1, \cdots, 6\}$ allow us to cover the sphere, and by creating an uniform partition of the square $[-a, a] \times [-a, a]$, we generate the equidistant cubed-sphere grid. we show an example of an equidistante cubed-sphere grid.

The derivative of the maps Ψ_p are given by:

$$D\Psi_1(x,y) = \frac{R}{(a^2+x^2+y^2)^{3/2}} \begin{bmatrix} -ax & -ay \\ a^2+y^2 & -xy \\ -xy & a^2+x^2 \end{bmatrix}, \tag{4.7}$$

$$D\Psi_2(x,y) = \frac{R}{(a^2+x^2+y^2)^{3/2}} \begin{bmatrix} -(a^2+y^2) & xy \\ -ax & -ay \\ -xy & a^2+x^2 \end{bmatrix}, \tag{4.8}$$

$$D\Psi_3(x,y) = \frac{R}{(a^2+x^2+y^2)^{3/2}} \begin{bmatrix} ax & ay \\ -(a^2+y^2) & xy \\ -xy & a^2+x^2 \end{bmatrix}, \tag{4.9}$$

$$D\Psi_4(x,y) = \frac{R}{(a^2+x^2+y^2)^{3/2}} \begin{bmatrix} a^2+y^2 & -xy \\ ax & ay \\ -xy & a^2+x^2 \end{bmatrix}, \tag{4.10}$$

$$D\Psi_5(x,y) = \frac{R}{(a^2+x^2+y^2)^{3/2}} \begin{bmatrix} xy & -(a^2+x^2) \\ a^2+y^2 & -xy \\ -ax & -ay \end{bmatrix}, \tag{4.11}$$

$$D\Psi_6(x,y) = \frac{R}{(a^2+x^2+y^2)^{3/2}} \begin{bmatrix} -xy & a^2+x^2 \\ a^2+y^2 & -xy \\ ax & ay \end{bmatrix}. \tag{4.12}$$

With the aid of the derivative, we may define a basis of tangent vectors $\{\boldsymbol{g}_1, \boldsymbol{g}_2\}$ on each point on the sphere by:

$$\boldsymbol{g}_1(x,y;p) = D\Psi_p(x,y)\begin{bmatrix}1\\0\end{bmatrix}, \boldsymbol{g}_2(x,y;p) = D\Psi_p(x,y)\begin{bmatrix}0\\1\end{bmatrix}, \tag{4.13}$$

In other words, we have $\{\boldsymbol{g}_1(x,y;p), \boldsymbol{g}_2(x,y;p)\} \subset T_{\Psi_p(x,y)}\mathbb{S}^2_R$, $\forall (x,y) \in [-a,a] \times [-a,a]$. Notice that

$$[D\Psi_p(x,y)]^T D\Psi_p(x,y) = \frac{R^2}{(a^2+x^2+y^2)^2} \begin{bmatrix} a^2+x^2 & -xy \\ -xy & a^2+y^2 \end{bmatrix}, \tag{4.14}$$

does not depend on p. Hence, it makes sense to define the matrix $G_\Psi(x,y) =$

$[D\Psi_p(x,y)]^T D\Psi_p(x,y)$ which is known as metric tensor. It is easy to see that:

$$G_\Psi(x,y) = \begin{bmatrix} \langle \boldsymbol{g}_1(x,y;p), \boldsymbol{g}_1(x,y;p)\rangle & \langle \boldsymbol{g}_1(x,y;p), \boldsymbol{g}_2(x,y;p)\rangle \\ \langle \boldsymbol{g}_1(x,y;p), \boldsymbol{g}_2(x,y;p)\rangle & \langle \boldsymbol{g}_2(x,y;p), \boldsymbol{g}_2(x,y;p)\rangle \end{bmatrix} \tag{4.15}$$

and that $G_\Psi(x,y)$ is positive-definite, $\forall (x,y) \in [-a,a]\times[-a,a]$. The Jacobian of the metric tensor $G_\Psi(x,y)$ is then given by:

$$\sqrt{|\det G_\Psi(x,y)|} = \frac{R^2}{(a^2+x^2+y^2)^{3/2}} a \tag{4.16}$$

4.1.2 Equiangular cubed-sphere

We consider again $a = \frac{R}{\sqrt{3}}$ and we define the family of maps $\Phi_p : [-\frac{\pi}{4},\frac{\pi}{4}]\times[-\frac{\pi}{4},\frac{\pi}{4}] \to \mathbb{S}^2_R$, $p = 1,\cdots,6$, given by $\Phi_p(x,y) = \Psi_p(a\tan x, a\tan y)$. The coordinates $(a\tan x, a\tan y)$ are called as angular coordinates. By the chain rule:

$$D\Phi_p(x,y) = aD\Psi_p(a\tan x, a\tan y)\begin{bmatrix} \frac{1}{\cos^2 x} & 0 \\ 0 & \frac{1}{\cos^2 y} \end{bmatrix} \tag{4.17}$$

and therefore we can define the following tangent vectors

$$\boldsymbol{r}_1(x,y;p) = D\Phi_p(x,y)\begin{bmatrix}1\\0\end{bmatrix} = \frac{a}{\cos^2 x}\boldsymbol{g}_1(\tan x, \tan y; p), \tag{4.18}$$

$$\boldsymbol{r}_2(x,y;p) = D\Phi_p(x,y)\begin{bmatrix}0\\1\end{bmatrix} = \frac{a}{\cos^2 y}\boldsymbol{g}_2(\tan x, \tan y; p), \tag{4.19}$$

that is, $\{\boldsymbol{r}_1(x,y;p), \boldsymbol{r}_{2,p}(x,y;p)\} \subset T_{\Phi_p(x,y)}\mathbb{S}^2_R$, $\forall (x,y) \in [-\frac{\pi}{4},\frac{\pi}{4}]\times[-\frac{\pi}{4},\frac{\pi}{4}]$.

Again, it makes sense to define the matrix

$$G_\Phi(x,y) = [D\Phi_p(x,y)]^T D\Phi_p(x,y) \tag{4.20}$$

$$= a^2[D\Psi_p(a\tan x, a\tan y)]^T \begin{bmatrix} \frac{1}{\cos^4 x} & 0 \\ 0 & \frac{1}{\cos^4 y} \end{bmatrix} D\Psi_p(a\tan x, a\tan y) \tag{4.21}$$

that does not depend on p and is the metric tensor. It is easy to see that:

$$G_\Phi(x,y) = \begin{bmatrix} \langle \boldsymbol{r}_1(x,y;p), \boldsymbol{r}_1(x,y;p)\rangle & \langle \boldsymbol{r}_1(x,y;p), \boldsymbol{r}_2(x,y;p)\rangle \\ \langle \boldsymbol{r}_1(x,y;p), \boldsymbol{r}_2(x,y;p)\rangle & \langle \boldsymbol{r}_2(x,y;p), \boldsymbol{r}_2(x,y;p)\rangle \end{bmatrix} \tag{4.22}$$

and that $G_\Phi(x,y)$ is positive-definite, $\forall (x,y) \in [-\frac{\pi}{4},\frac{\pi}{4}]\times[-\frac{\pi}{4},\frac{\pi}{4}]$. The Jacobian of the metric tensor $G_\Phi(x,y)$ is then given by:

$$\sqrt{|\det G_\Phi(x,y)|} = \frac{a}{\cos^2 x\cos^2 y}\frac{R^2}{(a^2+a^2\tan^2 x + a^2\tan^2 y)^{3/2}} a \tag{4.23}$$

$$= \frac{R^2}{\cos^2 x\cos^2 y}\frac{1}{(1+\tan^2 x+\tan^2 y)^{3/2}} \tag{4.24}$$

Chapter 5

Cubed-sphere finite-volume methods

5.1 Advection finite-volume scheme

In this Chapter, we show how we can use the dimension splitting method presented in Chapter 3 to solve the advection equation on the cubed-sphere with base on Putman and Lin (2007).

We denote by $\Psi_p : [-a,a]\times[-a,a] \to \mathbb{S}^2_R$, $p = 1,\cdots,6$, as a cubed-sphere mapping introduce in Chapter 4. We introduce the notations:

- $(x, y; p)$ represents a point on the cubed-sphere using a cubed-sphere mapping;
- $[-a,a]^2 = \bigcup_{i,j=1}^{N} [x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}] \times [y_{j-\frac{1}{2}}, y_{j+\frac{1}{2}}]$;
- $\Delta x = x_{i+\frac{1}{2}} - x_{i-\frac{1}{2}}, \Delta y = y_{j+\frac{1}{2}} - y_{j-\frac{1}{2}}$;
- $\Omega_{ijp} = \Psi_p([x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}] \times [y_{j-\frac{1}{2}}, y_{j+\frac{1}{2}}])$ are the cubed-sphere control-volumes;
- $\boldsymbol{g}_1(x,y;p) = D\Psi_p(x,y)\begin{bmatrix}1\\0\end{bmatrix}$ and $\boldsymbol{g}_2(x,y;p) = D\Psi_p(x,y)\begin{bmatrix}0\\1\end{bmatrix}$ are the tangent vectors;
- $g_\Psi(x,y) = \begin{bmatrix}\langle \boldsymbol{g}_1(x,y;p), \boldsymbol{g}_1(x,y;p)\rangle & \langle \boldsymbol{g}_1(x,y;p), \boldsymbol{g}_2(x,y;p)\rangle \\ \langle \boldsymbol{g}_1(x,y;p), \boldsymbol{g}_2(x,y;p)\rangle & \langle \boldsymbol{g}_2(x,y;p), \boldsymbol{g}_2(x,y;p)\rangle\end{bmatrix}$ is the metric tensor;
- $\sqrt{\det g_\Psi(x,y)}$ is the metric tensor Jacobian;
- $$|\Omega_{ijp}| = \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} \int_{y_{i-\frac{1}{2}}}^{y_{i+\frac{1}{2}}} \sqrt{\det g_\Psi(x,y)}\, dx\, dy$$
  are the control-volume areas
- $$Q_{ijp}(t) = \frac{1}{|\Omega_{ijp}|} \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} \int_{y_{j-\frac{1}{2}}}^{y_{j+\frac{1}{2}}} q(x,y,t;p)\sqrt{\det g_\Psi(x,y)}\, dx\, dy$$

are the averages of q on the control-volumes;

- $u^n_{i+\frac{1}{2},j,p} = u(x_{i+\frac{1}{2}}, y_j, t_n; p)$;
- $v^n_{i,j+\frac{1}{2},p} = v(x_i, y_{j+\frac{1}{2}}, t_n; p)$.

Given a tangent velocity field $\boldsymbol{u}$ on the sphere, we denote its contravariant components by $\tilde{u}$ and $\tilde{v}$. For a give a detailed discussion on contravariant representations in Appendix B. The advection equation on panel the p of the cubed-sphere is given by:

$$\frac{\partial}{\partial t}q + \frac{1}{\sqrt{\det g_\Psi}}\left(\frac{\partial}{\partial x}(\tilde{u}\sqrt{\det g_\Psi}q) + \frac{\partial}{\partial y}(\tilde{v}\sqrt{\det g_\Psi}q)\right) = 0,$$

$\forall (x, y, t) \in [-a, a]^2 \times [0, T]$, $q = q(x, y, t; p)$. Its integral form is given by:

$$\begin{aligned} Q_{ijp}(t_{n+1}) = Q_{ijp}(t_n) &- \frac{\Delta t}{|\Omega_{ijp}|}\delta_x\left(\frac{1}{\Delta t}\int_{t_1}^{t_2}\int_{y_{j-\frac{1}{2}}}^{y_{j+\frac{1}{2}}} (\tilde{u}\sqrt{\det g_\Psi}q)(x_i, y, t; p)\, dy\, dt\right) \\ &- \frac{\Delta t}{|\Omega_{ijp}|}\delta_y\left(\frac{1}{\Delta t}\int_{t_1}^{t_2}\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} (\tilde{v}\sqrt{\det g_\Psi}q)(x, y_j, t; p)\, dx\, dt\right), \end{aligned}$$

Hence, we can use the dimension splitting presented in Chapter 3 to the variable $\sqrt{\det g_\Psi}q$. However, when computing the stencils near to the cube edges, we need to approximate the values of q in the ghost cells in order to compute the stencils.

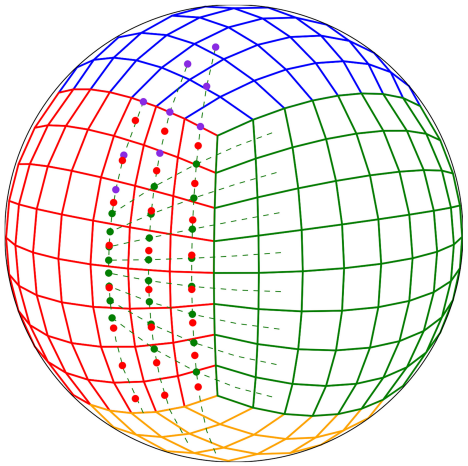


Figure 5.1: *Ghost cells on the equiangular cubed-sphere. Figure from Chen (2021)*

We adopted the strategy proposed by Chen (2021), that works only in the equiangular cubed-sphere. This approach exploits that the center of the ghost cells lye on a geodesic that connects the center of cells of the neighbor panel, as it is shown in Figure 5.1. Therefore, we can employ a Lagrange interpolation over a geodesic to approximate the values at the ghost cells. We decided to use a third-order Lagrange interpolation.

5.2 Numerical experiments

We are going to show some results using the velocity vector field from the deformational flow test case on the proposed by Nair and Lauritzen (2010). In these simulation, we do not use the monotonicity constraints.

For the first test case, we assume that the scalar field q is constant and equal to 1. Since the velocity field is non-divergente, the scalar field should remain constant.

Appendix A

Finite-difference estimatives

This appendix aims to prove all finite-difference error estimations used throughout this text. All the proves are very simple and consist of applying Taylor's expansions, as it is usual when showing the order of accuracy of many numerical schemes.

Lemma A.1. *Let $F \in C^5([a,b])$, $x_0 \in]a,b[$ and $h > 0$ such that $[x_0 - 2h, x_0 + 2h] \subset [a,b]$. Then, the following identity holds:*

$$F'(x_0) = \frac{4}{3}\left(\frac{F(x_0+h) - F(x_0-h)}{2h}\right) - \frac{1}{3}\left(\frac{F(x_0+2h) - F(x_0-2h)}{4h}\right) + C_1 h^4, \qquad \text{(A.1)}$$

where C_1 is a constant that depends only on F and h.

Proof. Given $\delta \in]0, 2h]$, then $x_0 + \delta \in]x_0, x_0 + 2h]$ and $x_0 - \delta \in]x_0 - 2h, x_0]$. Then, we get using the Taylor expansion of F:

$$F(x_0+\delta) = F(x_0) + F'(x_0)\delta + F^{(2)}(x_0)\frac{\delta^2}{2} + F^{(3)}(x_0)\frac{\delta^3}{3!} + F^{(4)}(x_0)\frac{\delta^3}{4!} + F^{(5)}(\theta_\delta)\frac{\delta^5}{5!} \quad \theta_\delta \in [x_0, x_0+\delta],$$
$$F(x_0-\delta) = F(x_0) - F'(x_0)\delta + F^{(2)}(x_0)\frac{\delta^2}{2} - F^{(3)}(x_0)\frac{\delta^3}{3!} + F^{(4)}(x_0)\frac{\delta^4}{4!} - F^{(5)}(\theta_{-\delta})\frac{\delta^5}{5!}, \quad \theta_{-\delta} \in [x_0-\delta, x_0].$$

Thus:

$$\frac{F(x_0+\delta) - F(x_0-\delta)}{2\delta} = F'(x_0) + F^{(3)}(x_0)\frac{\delta^2}{3!} + \left(F^{(5)}(\theta_\delta) + F^5(\theta_{-\delta})\right)\frac{\delta^4}{2\cdot 5!}, \qquad \text{(A.2)}$$

Applying Equation (A.2) for $\delta = h$ and $\delta = 2h$, we get, respectively:

$$\frac{F(x_0+h) - F(x_0-h)}{2h} = F'(x_0) + F^{(3)}(x_0)\frac{h^2}{3!} + \left(F^{(5)}(\theta_h) + F^{5)}(\theta_{-h})\right)\frac{h^4}{2\cdot 5!}, \quad \theta_h \in [x_0, x_0+h], \quad \theta_{-h} \in [x_0-h, x_0], \qquad \text{(A.3)}$$

and

$$\frac{F(x_0+2h)-F(x_0-2h)}{4h} = F'(x_0) + F^{(3)}(x_0)\frac{4h^2}{3!} + \left(F^{(5)}(\theta_{2h}) + F^{(5)}(\theta_{-2h})\right)\frac{16h^4}{2\cdot 5!}, \quad \theta_{2h} \in [x_0, x_0+2h], \quad \theta_{-2h} \in [x_0-2h, x_0]. \tag{A.4}$$

Using Equations (A.3) and (A.4), we obtain:

$$\frac{4}{3}\left(\frac{F(x_0+h)-F(x_0-h)}{2h}\right) = \frac{4}{3}F'(x_0) + F^{(3)}(x_0)\frac{4h^2}{3\cdot 3!} + \left(F^{(5)}(\theta_h) + F^{(5)}(\theta_{-h})\right)\frac{h^4}{2\cdot 5!}, \tag{A.5}$$

$$\frac{1}{3}\left(\frac{F(x_0+2h)-F(x_0-2h)}{4h}\right) = \frac{1}{3}F'(x_0) + F^{(3)}(x_0)\frac{4h^2}{3\cdot 3!} + \left(F^{(5)}(\theta_{2h}) + F^{(5)}(\theta_{-2h})\right)\frac{16h^4}{3\cdot 2\cdot 5!} \tag{A.6}$$

Subtracting Equation (A.6) from Equation (A.5) we get the desired Equation (A.1) with

$$C_1 = \frac{1}{720}\left(3F^{(5)}(\theta_h) + 3F^{(5)}(\theta_{-h}) - 16F^{(5)}(\theta_{2h}) - 16F^{(5)}(\theta_{-2h})\right), \tag{A.7}$$

where $\theta_h \in [x_0, x_0+h], \theta_{-h} \in [x_0-h, x_0]$, $\theta_{2h} \in [x_0, x_0+2h], \theta_{-2h} \in [x_0-2h, x_0]$. Using the intermediate value theorem, we can express C_1 in a more compact way as

$$C_1 = \frac{1}{720}\left(6F^{(5)}(\eta_1) - 32F^{(5)}(\eta_2)\right), \tag{A.8}$$

where $\eta_1, \eta_2 \in [x_0-2h, x_0+2h]$, which concludes the proof. □

Lemma A.2. *Let $F \in C^4([a,b])$, $x_0 \in]a,b[$ and $h > 0$ such that $[x_0-2h, x_0+3h] \subset [a,b]$. Then, the following identity holds:*

$$F''(x_0) = \frac{-2F(x_0-2h) + 15F(x_0-h) - 28F(x_0) + 20F(x_0+h) - 6F(x_0+2h) + F(x_0+3h)}{6h^2} + C_2h^2, \tag{A.9}$$

where C_2 is a constant that depends only on F and h.

Proof. From the Taylor's expansion, we have:

$$\begin{aligned}
F(x_0-2h) &= F(x_0) - 2F'(x_0)h + 2F^{(2)}(x_0)h^2 - \frac{8}{6}F^{(3)}(x_0)h^3 + \frac{16}{24}F^{(4)}(\theta_{-2h})h^4,\\
F(x_0-h) &= F(x_0) - F'(x_0)h + \frac{1}{2}F^{(2)}(x_0)h^2 - \frac{1}{6}F^{(3)}(x_0)h^3 + \frac{1}{24}F^{(4)}(\theta_{-h})h^4,\\
F(x_0+h) &= F(x_0) + F'(x_0)h + \frac{1}{2}F^{(2)}(x_0)h^2 + \frac{1}{6}F^{(3)}(x_0)h^3 + \frac{1}{24}F^{(4)}(\theta_{h})h^4,\\
F(x_0+2h) &= F(x_0) + 2F'(x_0)h + 2F^{(2)}(x_0)h^2 + \frac{8}{6}F^{(3)}(x_0)h^3 + \frac{16}{24}F^{(4)}(\theta_{2h})h^4,\\
F(x_0+3h) &= F(x_0) + 3F'(x_0)h + \frac{9}{2}F^{(2)}(x_0)h^2 + \frac{27}{6}F^{(3)}(x_0)h^3 + \frac{81}{24}F^{(4)}(\theta_{3h})h^4,
\end{aligned}$$

where $\theta_{-2h} \in [x_0-2h, x_0-h], \theta_{-h} \in [x_0-h, x_0], \theta_h \in [x_0, x_0+h], \theta_{2h} \in [x_0+h, x_0+2h], \theta_{3h} \in [x_0+2h, x_0+3h]$. Multiplying these equations by their respective coefficients given in Equation (A.9), one get:

$$\begin{aligned}
-2F(x_0-2h) &= -2F(x_0) + 4F'(x_0)h - 4F^{(2)}(x_0)h^2 + \frac{16}{6}F^{(3)}(x_0)h^3 - \frac{32}{24}F^{(4)}(\theta_{-2h})h^4,\\
15F(x_0-h) &= 15F(x_0) - 15F'(x_0)h + \frac{15}{2}F^{(2)}(x_0)h^2 - \frac{15}{6}F^{(3)}(x_0)h^3 + \frac{15}{24}F^{(4)}(\theta_{-h})h^4,\\
-28F(x_0) &= -28F(x_0),\\
20F(x_0+h) &= 20F(x_0) + 20F'(x_0)h + 10F^{(2)}(x_0)h^2 + \frac{20}{6}F^{(3)}(x_0)h^3 + \frac{20}{24}F^{(4)}(\theta_{h})h^4,\\
-6F(x_0+2h) &= -6F(x_0) - 12F'(x_0)h - 12F^{(2)}(x_0)h^2 - 8F^{(3)}(x_0)h^3 - \frac{96}{24}F^{(4)}(\theta_{2h})h^4,\\
F(x_0+3h) &= F(x_0) + 3F'(x_0)h + \frac{9}{2}F^{(2)}(x_0)h^2 + \frac{27}{6}F^{(3)}(x_0)h^3 + \frac{81}{24}F^{(4)}(\theta_{3h})h^4.
\end{aligned}$$

Summing all these equations, we get the desired Formula (A.9) with C given by:

$$C_2 = \frac{1}{24}\left(32F^{(4)}(\theta_{-2h}) - 15F^{(4)}(\theta_{-h}) - 20F^{(4)}(\theta_h) + 96F^{(4)}(\theta_{2h}) - 81F^{(4)}(\theta_{3h})\right). \tag{A.10}$$

Using the intermediate value theorem, we can express C_2 in a more compact way as

$$C_2 = \frac{1}{24}\left(128F^{(5)}(\eta_1) - 116F^{(5)}(\eta_2)\right), \tag{A.11}$$

where $\eta_1, \eta_2 \in [x_0-2h, x_0+3h]$, which concludes the proof. □

Lemma A.3. *Let $F \in C^4([a,b])$, $x_0 \in]a,b[$ and $h > 0$ such that $[x_0-2h, x_0+3h] \subset [a,b]$.*

Then, the following identity holds:

$$F^{(3)}(x_0) = \frac{F(x_0-2h) - 7F(x_0-h) + 16F(x_0) - 16F(x_0+h) + 7F(x_0+2h) - F(x_0+3h)}{2h^3} + C_3 h, \tag{A.12}$$

where C_3 is a constant that depends only on F and h.

Proof. From the Taylor's expansion, we have:

$$\begin{aligned}
F(x_0-2h) &= F(x_0) - 2F'(x_0)h + 2F^{(2)}(x_0)h^2 - \frac{8}{6}F^{(3)}(x_0)h^3 + \frac{16}{24}F^{(4)}(\theta_{-2h})h^4,\\
F(x_0-h) &= F(x_0) - F'(x_0)h + \frac{1}{2}F^{(2)}(x_0)h^2 - \frac{1}{6}F^{(3)}(x_0)h^3 + \frac{1}{24}F^{(4)}(\theta_{-h})h^4,\\
F(x_0+h) &= F(x_0) + F'(x_0)h + \frac{1}{2}F^{(2)}(x_0)h^2 + \frac{1}{6}F^{(3)}(x_0)h^3 + \frac{1}{24}F^{(4)}(\theta_{h})h^4,\\
F(x_0+2h) &= F(x_0) + 2F'(x_0)h + 2F^{(2)}(x_0)h^2 + \frac{8}{6}F^{(3)}(x_0)h^3 + \frac{16}{24}F^{(4)}(\theta_{2h})h^4,\\
F(x_0+3h) &= F(x_0) + 3F'(x_0)h + \frac{9}{2}F^{(2)}(x_0)h^2 + \frac{27}{6}F^{(3)}(x_0)h^3 + \frac{81}{24}F^{(4)}(\theta_{3h})h^4,
\end{aligned}$$

where $\theta_{-2h} \in [x_0-2h, x_0-h], \theta_{-h} \in [x_0-h, x_0], \theta_h \in [x_0, x_0+h], \theta_{2h} \in [x_0+h, x_0+2h], \theta_{3h} \in [x_0+2h, x_0+3h]$. Multiplying these equations by their respective coefficients given in Equation (A.12), one get:

$$\begin{aligned}
F(x_0-2h) &= F(x_0) - 2F'(x_0)h + \frac{4}{2}F^{(2)}(x_0)h^2 - \frac{8}{6}F^{(3)}(x_0)h^3 + \frac{16}{24}F^{(4)}(\theta_{-2h})h^4,\\
-7F(x_0-h) &= -7F(x_0) + 7F'(x_0)h - \frac{7}{2}F^{(2)}(x_0)h^2 + \frac{7}{6}F^{(3)}(x_0)h^3 - \frac{7}{24}F^{(4)}(\theta_{-h})h^4,\\
16F(x_0) &= 16F(x_0),\\
-16F(x_0+h) &= -16F(x_0) - 16F'(x_0)h - \frac{16}{2}F^{(2)}(x_0)h^2 - \frac{16}{6}F^{(3)}(x_0)h^3 - \frac{16}{24}F^{(4)}(\theta_{h})h^4,\\
7F(x_0+2h) &= 7F(x_0) + 14F'(x_0)h + \frac{28}{2}F^{(2)}(x_0)h^2 + \frac{56}{6}F^{(3)}(x_0)h^3 + \frac{112}{24}F^{(4)}(\theta_{2h})h^4,\\
-F(x_0+3h) &= -F(x_0) - 3F'(x_0)h - \frac{9}{2}F^{(2)}(x_0)h^2 - \frac{27}{6}F^{(3)}(x_0)h^3 - \frac{81}{24}F^{(4)}(\theta_{3h})h^4.
\end{aligned}$$

Summing all these equations, we have:

$$F(x_0-2h) - 7F(x_0-h) + 16F(x_0) - 16F(x_0+h) + 7F(x_0+2h) - F(x_0+3h) = 2F^{(3)}(x_0)h^3 - 2Ch^4,$$

we get the derised Formula (A.12) with C given by:

$$C_3 = \frac{1}{48}\left(-16F^{(4)}(\theta_{-2h}) + 7F^{(4)}(\theta_{-h}) + 16F^{(4)}(\theta_h) - 112F^{(4)}(\theta_{2h}) + 81F^{(4)}(\theta_{3h}) \right). \tag{A.13}$$

Using the intermediate value theorem, we can express C_3 in a more compact way as

$$C_3 = \frac{1}{48}\left(104F^{(5)}(\eta_1) - 128F^{(5)}(\eta_2)\right), \tag{A.14}$$

where $\eta_1, \eta_2 \in [x_0 - 2h, x_0 + 3h]$, which concludes the proof. □

Appendix B

Spherical coordinates and geometry

Given $R > 0$, we denote the sphere of radius R centered at the origin of $\mathbb{R}^3$:

$$\mathbb{S}^2_R = \{(X, Y, Z) \in \mathbb{R}^3 \ : \ X^2 + Y^2 + Z^2 = R^2\}.$$

The tangent space at $P \in \mathbb{S}^2_R$ by $T_P\mathbb{S}^2$. It is easy to see that:

$$T_P\mathbb{S}^2_R = \{Q \in \mathbb{R}^3 \ : \ \langle P, Q\rangle = 0\},$$

where $\langle \cdot, \cdot \rangle$ denotes the standard inner product of $\mathbb{R}^3$. The tangent bundle is denoted by:

$$T\mathbb{S}^2_R = \bigcup_{P \in \mathbb{S}^R_2} T_P\mathbb{S}^2_R.$$

We are going to consider three ways to represent an element of $\mathbb{S}^2_R$: using (X, Y, Z) coordinates, or using (λ, ϕ) latitude-longitude coordinates, or, at last, using the cubed-sphere coordinates (x, y, p), where (x, y) are the cube face coordinates and $p \in \{1, 2, \cdots, 6\}$ stands for a cube panel, as presented in Chapter 4.

B.1 Tangent vectors

Let us consider $P \in \mathbb{S}^2_R$, we consider the projection on the tangent space $T_P\mathbb{S}^2_R$ which is given by:

$$\Pi_P(Q) = \frac{\langle P, Q\rangle}{R} P - Q$$

We introduce the tangent vectors at $\Psi_p(x, y)$ given by: $\boldsymbol{g}_x(x, y; p) = \Pi_P(\Psi_p(x + \Delta x, y))$, $\boldsymbol{g}_y(x, y; p) = \Pi_P(\Psi_p(x, y+\Delta y))$, where $P = \Psi_p(x, y+\Delta y)$. We then introduce the normalized vectors:

$$\boldsymbol{e}_x(x, y; p) = \frac{\boldsymbol{g}_x(x, y; p)}{\|\boldsymbol{g}_x(x, y; p)\|}, \quad \boldsymbol{e}_y(x, y; p) = \frac{\boldsymbol{g}_y(x, y; p)}{\|\boldsymbol{g}_y(x, y; p)\|},$$

where $\|\cdot\|$ is the Euclidean norm. It can be shown that the tangent vector of the geodesic from $\Psi_p(x,y)$ from $\Psi_p(x+\Delta x,y)$ is a multiple of $\boldsymbol{e}_x$ and the tangent vector of the geodesic from $\Psi_p(x,y)$ from $\Psi_p(x,y+\Delta y)$ is a multiple of $\boldsymbol{e}_y$. Thus, this process using the projection operator on the tangent space allow us to compute the unit tangent vectors for any cubed-sphere mapping gridlines at a given point.

B.2 Conversions between latitude-longitude and contravariant coordinates

We consider the latitude-longitude mapping $\Psi_{ll} : [0,2\pi]\times[-\frac{\pi}{2},\frac{\pi}{2}] \to \mathbb{S}^2_R$, given by:

$$X(\lambda,\phi) = R\cos\phi\cos\lambda, \tag{B.1}$$
$$Y(\lambda,\phi) = R\cos\phi\sin\lambda, \tag{B.2}$$
$$Z(\lambda,\phi) = R\sin\phi, \tag{B.3}$$

The derivative or Jacobian matrix of the mapping Ψ_{ll} is given by:

$$D\Psi_{ll}(\lambda,\phi) = R\begin{bmatrix} -\cos\phi\sin\lambda & -\sin\phi\cos\lambda \\ \cos\phi\cos\lambda & \sin\phi\sin\lambda \\ 0 & \cos\phi \end{bmatrix} \tag{B.4}$$

Using this matrix columns, we can define the tangent vectors:

$$\boldsymbol{g}_\lambda(\lambda,\phi) = D\Psi_{ll}(\lambda,\phi)\begin{bmatrix}1\\0\end{bmatrix}, \quad \boldsymbol{g}_\phi(\lambda,\phi) = D\Psi_{ll}(\lambda,\phi)\begin{bmatrix}0\\1\end{bmatrix}, \tag{B.5}$$

We normalize the vectors $\boldsymbol{g}_\lambda$ and $\boldsymbol{g}_\phi$ and we obtain unit tangent vectors on the sphere at $\Phi_{ll}(\lambda,\phi)$:

$$\boldsymbol{e}_\lambda(\lambda,\phi) = \begin{bmatrix}-\sin\lambda\\ \cos\lambda\\ 0\end{bmatrix}, \quad \boldsymbol{e}_\phi(\lambda,\phi) = \begin{bmatrix}-\sin\phi\cos\lambda\\ -\sin\phi\sin\lambda\\ \cos\phi\end{bmatrix}, \tag{B.6}$$

Let us consider a tangent vector field $\boldsymbol{u} : \mathbb{S}^2_R \to T\mathbb{S}^2_R$ on the sphere, *i.e.*, $\boldsymbol{u}(P)\in T_P\mathbb{S}^2_R$, $\forall P\in\mathbb{S}^2_R$. We may express this vector fields in latitude-longitude coordinates as:

$$\boldsymbol{u}(\lambda,\phi) = u_\lambda(\lambda,\phi)\boldsymbol{e}_\lambda(\lambda,\phi) + v_\phi(\lambda,\phi)\boldsymbol{e}_\phi(\lambda,\phi). \tag{B.7}$$

Or, we may also represent this vector field using the basis obtained by cubed-sphere coordinates:

$$\boldsymbol{u}(x,y;p) = \tilde{u}(x,y;p)\boldsymbol{e}_x(x,y;p) + \tilde{v}(x,y;p)\boldsymbol{e}_y(x,y;p). \tag{B.8}$$

This representation is known as contravariant representation. In order to relate the latitude-

longitude representation with the contravariant representation, we notice that:

$$\boldsymbol{e_x}(x,y;p) = \langle \boldsymbol{e}_x, \boldsymbol{e}_\lambda \rangle \boldsymbol{e}_\lambda(\lambda,\phi) + \langle \boldsymbol{e}_x, \boldsymbol{e}_\phi \rangle \boldsymbol{e}_\phi(\lambda,\phi), \tag{B.9}$$

$$\boldsymbol{e_y}(x,y;p) = \langle \boldsymbol{e}_y, \boldsymbol{e}_\lambda \rangle \boldsymbol{e}_\lambda(\lambda,\phi) + \langle \boldsymbol{e}_y, \boldsymbol{e}_\phi \rangle \boldsymbol{e}_\phi(\lambda,\phi), \tag{B.10}$$

which holds since the vectors $\boldsymbol{e}_\lambda(\lambda,\phi)$ and $\boldsymbol{e}_\phi(\lambda,\phi)$ are orthogonal. Replacing Equations (B.9) and (B.10) in Equation (B.8), we obtain the values (u_λ, v_ϕ) in terms of the contravariant components $(\tilde{u},\tilde{v})$ as the following matrix equation:

$$\begin{bmatrix} u_\lambda(\lambda,\phi) \\ v_\phi(\lambda,\phi) \end{bmatrix} = \begin{bmatrix} \langle \boldsymbol{e}_x, \boldsymbol{e}_\lambda \rangle & \langle \boldsymbol{e}_y, \boldsymbol{e}_\lambda \rangle \\ \langle \boldsymbol{e}_x, \boldsymbol{e}_\phi \rangle & \langle \boldsymbol{e}_y, \boldsymbol{e}_\phi \rangle \end{bmatrix} \begin{bmatrix} \tilde{u}(x,y;p) \\ \tilde{v}(x,y;p) \end{bmatrix}. \tag{B.11}$$

Conversely, we may express the contravariant components in terms of latitude-longitude components by inverting Equation (B.11):

$$\begin{bmatrix} \tilde{u}(x,y;p) \\ \tilde{v}(x,y;p) \end{bmatrix} = \frac{1}{\langle \boldsymbol{e}_x, \boldsymbol{e}_\lambda \rangle \langle \boldsymbol{e}_y, \boldsymbol{e}_\phi \rangle - \langle \boldsymbol{e}_y, \boldsymbol{e}_\lambda \rangle \langle \boldsymbol{e}_x, \boldsymbol{e}_\phi \rangle} \begin{bmatrix} \langle \boldsymbol{e}_y, \boldsymbol{e}_\phi \rangle & -\langle \boldsymbol{e}_y, \boldsymbol{e}_\lambda \rangle \\ -\langle \boldsymbol{e}_x, \boldsymbol{e}_\phi \rangle & \langle \boldsymbol{e}_x, \boldsymbol{e}_\lambda \rangle \end{bmatrix} \begin{bmatrix} u_\lambda(\lambda,\phi) \\ v_\phi(\lambda,\phi) \end{bmatrix}. \tag{B.12}$$

B.3 Covariant/contravariant conversion

Given Equation Let us consider again a tangent vector field $\boldsymbol{u} : \mathbb{S}^2_R \to T\mathbb{S}^2_R$ on the sphere, the contravariant representation of $\boldsymbol{u}$ is given by Equation (B.8). The covariant components (u,v) are given by:

$$u(x,y;p) = \langle \boldsymbol{u}(x,y;p), e_x(x,y;p) \rangle, \tag{B.13}$$

$$v(x,y;p) = \langle \boldsymbol{u}(x,y;p), e_y(x,y;p) \rangle. \tag{B.14}$$

Replacing Equation (B.8) in Equations (B.13) and (B.14) we obtain the relation covariant components in terms of the contravariant terms:

$$\begin{bmatrix} u(x,y;p) \\ v(x,y;p) \end{bmatrix} = \begin{bmatrix} 1 & \langle \boldsymbol{e}_x, \boldsymbol{e}_y \rangle \\ \langle \boldsymbol{e}_x, \boldsymbol{e}_y \rangle & 1 \end{bmatrix} \begin{bmatrix} \tilde{u}(x,y;p) \\ \tilde{v}(x,y;p) \end{bmatrix}. \tag{B.15}$$

Denoting the angle between $\boldsymbol{e}_x$ and $\boldsymbol{e}_y$ by α, we have $\langle \boldsymbol{e}_x, \boldsymbol{e}_y \rangle = \cos\alpha$. Thus, we may express the contravariant components in terms of the covariant terms inverting Equation (B.15):

$$\begin{bmatrix} \tilde{u}(x,y;p) \\ \tilde{v}(x,y;p) \end{bmatrix} = \frac{1}{\sin^2\alpha} \begin{bmatrix} 1 & -\cos\alpha \\ -\cos\alpha & 1 \end{bmatrix} \begin{bmatrix} u(x,y;p) \\ v(x,y;p) \end{bmatrix}. \tag{B.16}$$

Notice that combining Equations (B.15) and (B.16) with Equations (B.11) and (B.12) one may get relations between the latitude-longitude components and the covariant components.

Appendix C

Code avaibility

The codes needed for this work have been built openly at GitHub. The PPM implementation for the one-dimensional advection equation used in Chapter 2 is available at https://github.com/luanfs/py-ppm. The dimension-splitting implementation for the advection equation on the plane used in Chapter 3 is available at https://github.com/luanfs/py-operator-splitting. At last, all the grid tools for the cubed sphere used Chapters 4 and 5, including the finite volume model on this grid, is available in a Python version at https://github.com/luanfs/py-cubed-sphere and in a Fortran 90 version at https://github.com/luanfs/cubed-sphere.

Finally, the report for the qualification exam has also been built at GitHub in the following repository: https://github.com/luanfs/doc-qualification.

All the contents in GitHub's repositories shown previously are under constant development.

References

Arakawa, A., & Lamb, V. R. (1977). Computational design of the basic dynamical processes of the ucla general circulation model. In *General circulation models of the atmosphere* (pp. 173–265). Elsevier. https://doi.org/https://doi.org/10.1016/B978-0-12-460817-7.50009-4. (Cit. on p. 4)

Barros, S., Dent, D., Isaksen, L., Robinson, G., Mozdzynski, G., & Wollenweber, F. (1995). The ifs model: A parallel production weather code. *Parallel Computing*, *21*(10), 1621–1638. https://doi.org/https://doi.org/10.1016/0167-8191(96)80002-0 (cit. on p. 3)

Benacchio, T., & Wood, N. (2016). Semi-implicit semi-lagrangian modelling of the atmosphere: A met office perspective. *Communications in Applied and Industrial Mathematics*, *7*(3), 4–25. https://doi.org/doi:10.1515/caim-2016-0020 (cit. on p. 1)

Carpenter, R. L., Droegemeier, K. K., Woodward, P. R., & Hane, C. E. (1990). Application of the piecewise parabolic method (ppm) to meteorological modeling. *Monthly Weather Review*, *118*(3), 586–612. https://doi.org/10.1175/1520-0493(1990)118<0586:AOTPPM>2.0.CO;2 (cit. on pp. 4, 7, 19)

Chen, X. (2021). The lmars based shallow-water dynamical core on generic gnomonic cubed-sphere geometry [e2020MS002280 2020MS002280]. *Journal of Advances in Modeling Earth Systems*, *13*(1), e2020MS002280. https://doi.org/https://doi.org/10.1029/2020MS002280 (cit. on p. 56)

Colella, P., & Woodward, P. R. (1984). The piecewise parabolic method (ppm) for gas-dynamical simulations. *Journal of Computational Physics*, *54*(1), 174–201. https://doi.org/https://doi.org/10.1016/0021-9991(84)90143-8 (cit. on pp. 4, 7, 19, 20, 26, 27, 36)

Cooley, J. W., & Tukey, J. W. (1965). An algorithm for the machine calculation of complex fourier series. *Mathematics of Computation*, *19*(90), 297–301. http://www.jstor.org/stable/2003354 (cit. on p. 2)

Dennis, J., Edwards, J., Evans, K., Guba, O., Lauritzen, P., Mirin, A., St-Cyr, A., Taylor, M., & Worley, P. (2012). Cam-se: A scalable spectral element dynamical core for the community atmosphere model. *Internat. J. High Perf. Comput. Appl.*, *26*, 74–89. https://doi.org/10.1177/1094342011428142 (cit. on p. 4)

Eliasen, E., Machenhauer, B., & Rasmussen, E. (1970). On a numerical method for integration of the hydrodynamical equations wirh a spectral representation of the horizontal fields. https://doi.org/10.13140/RG.2.2.13894.88645 (cit. on p. 2)

Engwirda, D., & Kelley, M. (2016). A weno-type slope-limiter for a family of piecewise polynomial methods. https://doi.org/10.48550/ARXIV.1606.08188. (Cit. on p. 7)

Figueroa, S., Bonatti, J., Kubota, P., Grell, G., Morrison, H., R. M. Barros, S., Fernandez, J., Ramirez-Gutierrez, E., Siqueira, L., Luzia, G., Silva, J., Silva, J., Pendharkar, J., Capistrano, V., Alvim, D., Enore, D., Diniz, F., Satyamurty, P., Cavalcanti, I., & Panetta, J. (2016). The brazilian global atmospheric model (bam): Performance for tropical rainfall forecasting and sensitivity to convective scheme and horizontal resolution. *Weather Forecast.*, *31(5)*, 1547–1572. https://doi.org/10.1175/WAF-D-16-0062.1 (cit. on p. 3)

Giraldo, F. X., Kelly, J. F., & Constantinescu, E. M. (2013). Implicit-explicit formulations of a three-dimensional nonhydrostatic unified model of the atmosphere (numa). *SIAM Journal on Scientific Computing*, *35*(5), B1162–B1194. https://doi.org/10.1137/120876034 (cit. on p. 4)

Godunov, S. (1959). A difference method for numerical calculation of discontinuous solutions of the equations of hydrodynamics. *Mat. Sb.*, *47(89):3*, 271–306 (cit. on p. 7).

Harris, L., Chen, X., Putman, W., Zhou, L., & Chen, J.-H. (2021). A scientific description of the gfdl finite-volume cubed-sphere dynamical core. *Series : NOAA technical memorandum OAR GFDL ; 2021-001*. https://doi.org/https://doi.org/10.25923/6nhs-5897 (cit. on pp. 7, 26)

Harris, L. M., & Lin, S.-J. (2013). A two-way nested global-regional dynamical core on the cubed-sphere grid. *Monthly Weather Review*, *141*(1), 283–306. https://doi.org/10.1175/MWR-D-11-00201.1 (cit. on pp. 4, 5)

Kent, J., Melvin, T., & Wimmer, G. A. (2022). A mixed finite element discretisation of the shallow water equations. *Geoscientific Model Development Discussions*, *2022*, 1–17. https://doi.org/10.5194/gmd-2022-225 (cit. on p. 4)

Krishnamurti, T., Hardiker, V., Bedi, H., & Ramaswamy, L. (2006). *An introduction to global spectral modeling* (Vol. 35). https://doi.org/10.1007/0-387-32962-5. (Cit. on p. 2)

Lauritzen, P. H., Ullrich, P. A., & Nair, R. D. (2011). Atmospheric transport schemes: Desirable properties and a semi-lagrangian view on finite-volume discretizations. In P. Lauritzen, C. Jablonowski, M. Taylor, & R. Nair (Eds.), *Numerical techniques for global atmospheric models* (pp. 185–250). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-11640-7_8. (Cit. on p. 7)

LeVeque, R. J. (1990). *Numerical methods for conservation laws*. Birkhäuser Basel. https://doi.org/10.1007/978-3-0348-5116-9. (Cit. on pp. 8–10, 46)

LeVeque, R. J. (2002). *Finite volume methods for hyperbolic problems*. Cambridge University Press. https://doi.org/10.1017/CBO9780511791253. (Cit. on pp. 8, 12, 14, 16, 17)

Lin, S.-J. (2004). A "vertically lagrangian" finite-volume dynamical core for global models. *Monthly Weather Review*, *132*(10), 2293–2307. https://doi.org/10.1175/1520-0493(2004)132<2293:AVLFDC>2.0.CO;2 (cit. on pp. 4, 7, 26, 27, 36)

Lin, S.-J., Chao, W. C., Sud, Y. C., & Walker, G. K. (1994). A class of the van leer-type transport schemes and its application to the moisture transport in a general circulation model. *Monthly Weather Review*, *122*(7), 1575–1593. https://doi.org/10.1175/1520-0493(1994)122<1575:ACOTVL>2.0.CO;2 (cit. on p. 4)

Lin, S.-J., & Rood, R. B. (1996). Multidimensional flux-form semi-lagrangian transport schemes. *Monthly Weather Review*, *124*(9), 2046–2070. https://doi.org/10.1175/1520-0493(1996)124<2046:MFFSLT>2.0.CO;2 (cit. on pp. 4, 19, 45, 49)

Lin, S.-J., & Rood, R. B. (1997). An explicit flux-form semi-lagrangian shallow-water model on the sphere. *Quarterly Journal of the Royal Meteorological Society*, *123*(544), 2477–2498. https://doi.org/https://doi.org/10.1002/qj.49712354416 (cit. on p. 4)

Müller, A., Deconinck, W., Kühnlein, C., Mengaldo, G., Lange, M., Wedi, N., Bauer, P., Smolarkiewicz, P. K., Diamantakis, M., Lock, S.-J., Hamrud, M., Saarinen, S., Mozdzynski, G., Thiemert, D., Glinton, M., Bénard, P., Voitus, F., Colavolpe, C., Marguinaud, P., … New, N. (2019). The escape project: Energy-efficient scalable algorithms for weather prediction at exascale. *Geoscientific Model Development*, *12*(10), 4425–4441. https://doi.org/10.5194/gmd-12-4425-2019 (cit. on p. 3)

Nair, R. D., & Lauritzen, P. H. (2010). A class of deformational flow test cases for linear transport problems on the sphere. *Journal of Computational Physics*, *229*(23), 8868–8887. https://doi.org/https://doi.org/10.1016/j.jcp.2010.08.014 (cit. on pp. 40, 57)

Nair, R. D., Levy, M. N., & Lauritzen, P. H. (2011). Emerging numerical methods for atmospheric modeling. In *Numerical techniques for global atmospheric models* (pp. 251–311). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-11640-7_9. (Cit. on p. 10)

Orszag, S. A. (1970). Transform method for the calculation of vector-coupled sums: Application to the spectral form of the vorticity equation. *Journal of Atmospheric Sciences*, *27*(6), 890–895. https://doi.org/10.1175/1520-0469(1970)027<0890:TMFTCO>2.0.CO;2 (cit. on p. 2)

Peixoto, P. (2016). Accuracy analysis of mimetic finite volume operators on geodesic grids and a consistent alternative. *J. Comput. Phys.*, *310*, 127–160. https://doi.org/10.1016/j.jcp.2015.12.058 (cit. on p. 5)

Peixoto, P., & Barros, S. R. M. (2013). Analysis of grid imprinting on geodesic spherical icosahedral grids. *J. Comput. Phys.*, *237*, 61–78. https://doi.org/10.1016/j.jcp.2012.11.041 (cit. on p. 5)

Putman, W. M. (2007). *Development of the finite-volume dynamical core on the cubed-sphere* (Doctoral dissertation). Florida State University. Florida, US. http://purl.flvc.org/fsu/fd/FSU_migr_etd-0511. (Cit. on p. 4)

Putman, W. M., & Lin, S.-J. (2007). Finite-volume transport on various cubed-sphere grids. *Journal of Computational Physics*, *227*(1), 55–78. https://doi.org/https://doi.org/10.1016/j.jcp.2007.07.022 (cit. on pp. 4, 5, 7, 21, 28, 36, 55)

Rančić, M., Purser, R. J., & Mesinger, F. (1996). A global shallow-water model using an expanded spherical cube: Gnomonic versus conformal coordinates. *Quarterly Journal of the Royal Meteorological Society*, *122*(532), 959–982. https://doi.org/https://doi.org/10.1002/qj.49712253209 (cit. on p. 51)

Rančić, M., Purser, R. J., Jović, D., Vasic, R., & Black, T. (2017). A nonhydrostatic multiscale model on the uniform jacobian cubed sphere. *Monthly Weather Review*, *145*(3), 1083–1105. https://doi.org/10.1175/MWR-D-16-0178.1 (cit. on p. 4)

Randall, D. A., Bitz, C. M., Danabasoglu, G., Denning, A. S., Gent, P. R., Gettelman, A., Griffies, S. M., Lynch, P., Morrison, H., Pincus, R., & Thuburn, J. (2018). 100 years of earth system model development. *Meteorological Monographs*, *59*, 12.1–12.66. https://doi.org/10.1175/AMSMONOGRAPHS-D-18-0018.1 (cit. on pp. 1, 3)

Ringler, T., Thuburn, J., Klemp, J., & Skamarock, W. (2010). A unified approach to energy conservation and potential vorticity dynamics on arbitrarily structured C-grids. *J. Comput. Phys.*, *229*, 3065–3090. https://doi.org/10.1016/j.jcp.2009.12.007 (cit. on p. 5)

Ronchi, C., Iacono, R., & Paolucci, P. (1996). The "cubed sphere": A new method for the solution of partial differential equations in spherical geometry. *Journal of Computational Physics*, *124*(1), 93–114. https://doi.org/https://doi.org/10.1006/jcph.1996.0047 (cit. on pp. 4, 51)

Sadourny, R. (1972). Conservative finite-difference approximations of the primitive equations on quasi-uniform spherical grids. *Monthly Weather Review*, *100*(2), 136–144. https://doi.org/10.1175/1520-0493(1972)100<0136:CFAOTP>2.3.CO;2 (cit. on pp. 4, 51)

Samenow, J. (2019). *National weather service launches upgraded, improved global forecast model*. Retrieved July 29, 2022, from https://www.washingtonpost.com/weather/2019/06/12/national-weather-service-launches-upgraded-improved-global-forecast-model/. (Cit. on p. 4)

Santos, L. F., & Peixoto, P. S. (2021). Topography-based local spherical voronoi grid refinement on classical and moist shallow-water finite-volume models. *Geoscientific Model Development*, *14*(11), 6919–6944. https://doi.org/10.5194/gmd-14-6919-2021 (cit. on p. 5)

Skamarock, W., Klemp, J., Duda, M., Fowler, L., Park, S.-H., & Ringler, T. (2012). A multiscale nonhydrostatic atmospheric model using centroidal Voronoi tesselations and C-grid staggering. *Mon. Weather. Rev.*, *140(09)*, 3090–3105. https://doi.org/10.1175/MWR-D-11-00215.1 (cit. on p. 5)

Staniforth, A., & Thuburn, J. (2012). Horizontal grids for global weather and climate prediction models: A review. *Q. J. Roy. Meteor. Soc.*, *138*, 1–26. https://doi.org/10.1002/qj.958 (cit. on p. 3)

Strikwerda, J. C. (2004). *Finite difference schemes and partial differential equations, second edition*. Society for Industrial; Applied Mathematics. https://doi.org/10.1137/1.9780898717938. (Cit. on p. 18)

Suresh, A., & Huynh, H. (1997). Accurate monotonicity-preserving schemes with runge–kutta time stepping. *Journal of Computational Physics*, *136*(1), 83–99. https://doi.org/https://doi.org/10.1006/jcph.1997.5745 (cit. on pp. 12, 21)

Thuburn, J. (2011). Conservation in dynamical cores: What, how and why? In *Numerical techniques for global atmospheric models* (pp. 345–355). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-11640-7_11. (Cit. on p. 3)

Thuburn, J., Ringler, T., Skamarock, W., & Klemp, J. (2009). Numerical representation of geostrophic modes on arbitrarily structured C-grids. *J. Comput. Phys.*, *228*, 8321–8335. https://doi.org/10.1016/j.jcp.2009.08.006 (cit. on p. 5)

Trefethen, L. N. (2000). *Spectral methods in matlab*. Society for Industrial; Applied Mathematics. https://doi.org/10.1137/1.9780898719598. (Cit. on pp. 17, 36)

Ullrich, P. A., Jablonowski, C., Kent, J., Lauritzen, P. H., Nair, R., Reed, K. A., Zarzycki, C. M., Hall, D. M., Dazlich, D., Heikes, R., Konor, C., Randall, D., Dubos, T., Meurdesoif, Y., Chen, X., Harris, L., Kühnlein, C., Lee, V., Qaddouri, A., … Viner, K. (2017). Dcmip2016: A review of non-hydrostatic dynamical core design and intercomparison of participating models. *Geoscientific Model Development*, *10*(12), 4477–4509. https://doi.org/10.5194/gmd-10-4477-2017 (cit. on p. 3)

Van Leer, B. (1977). Towards the ultimate conservative difference scheme. iv. a new approach to numerical convection. *Journal of Computational Physics*, *23*(3), 276–299. https://doi.org/https://doi.org/10.1016/0021-9991(77)90095-X (cit. on pp. 4, 7, 19)

Weller, H. (2012). Controlling the computational modes of the arbitrarily structured c grid, *Mon. Weather. Rev.*, *140(10)*, 3220–3234. https://doi.org/doi.org/10.1175/MWR-D-11-00221.1 (cit. on p. 5)

Whitaker, J. (2015). *Hiwpp non-hydrostatic dynamical core tests: Results from idealized test cases*. Retrieved November 5, 2022, from https://www.weather.gov/media/sti/nggps/HIWPP_idealized_tests-v8%20revised%2005212015.pdf/. (Cit. on p. 4)

White, L., & Adcroft, A. (2008). A high-order finite volume remapping scheme for nonuniform grids: The piecewise quartic method (pqm). *Journal of Computational Physics*, *227*(15), 7394–7422. https://doi.org/https://doi.org/10.1016/j.jcp.2008.04.026 (cit. on p. 7)

Williamson, D., Drake, J., Hack, J., Jakob, R., & Swarztrauber, P. (1992). A standard test set for numerical approximations to the shallow water equations in spherical geometry. *J. Comput. Phys.*, *102*, 211–224. https://doi.org/10.1016/S0021-9991(05)80016-6 (cit. on p. 5)

Williamson, D. L. (2007). The evolution of dynamical cores for global atmospheric models. *Journal of the Meteorological Society of Japan. Ser. II*, *85B*, 241–269. https://doi.org/10.2151/jmsj.85B.241 (cit. on pp. 1, 2)

Wood, N., Staniforth, A., White, A., Allen, T., Diamantakis, M., Gross, M., Melvin, T., Smith, C., Vosper, S., Zerroukat, M., & Thuburn, J. (2014). An inherently mass-conserving semi-implicit semi-lagrangian discretization of the deep-atmosphere global non-hydrostatic equations. *Quarterly Journal of the Royal Meteorological Society*, *140*(682), 1505–1520. https://doi.org/https://doi.org/10.1002/qj.2235 (cit. on p. 1)

Woodward, P. R. (1986). Piecewise-parabolic methods for astrophysical fluid dynamics. In K.-H. A. Winkler & M. L. Norman (Eds.), *Astrophysical radiation hydrodynamics* (pp. 245–326). Springer Netherlands. https://doi.org/10.1007/978-94-009-4754-2_8. (Cit. on p. 7)

Zheng, Y., & Marguinaud, P. (2018). Simulation of the performance and scalability of message passing interface (mpi) communications of atmospheric models running on exascale supercomputers. *Geoscientific Model Development*, *11*(8), 3409–3426. https://doi.org/10.5194/gmd-11-3409-2018 (cit. on p. 3)