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

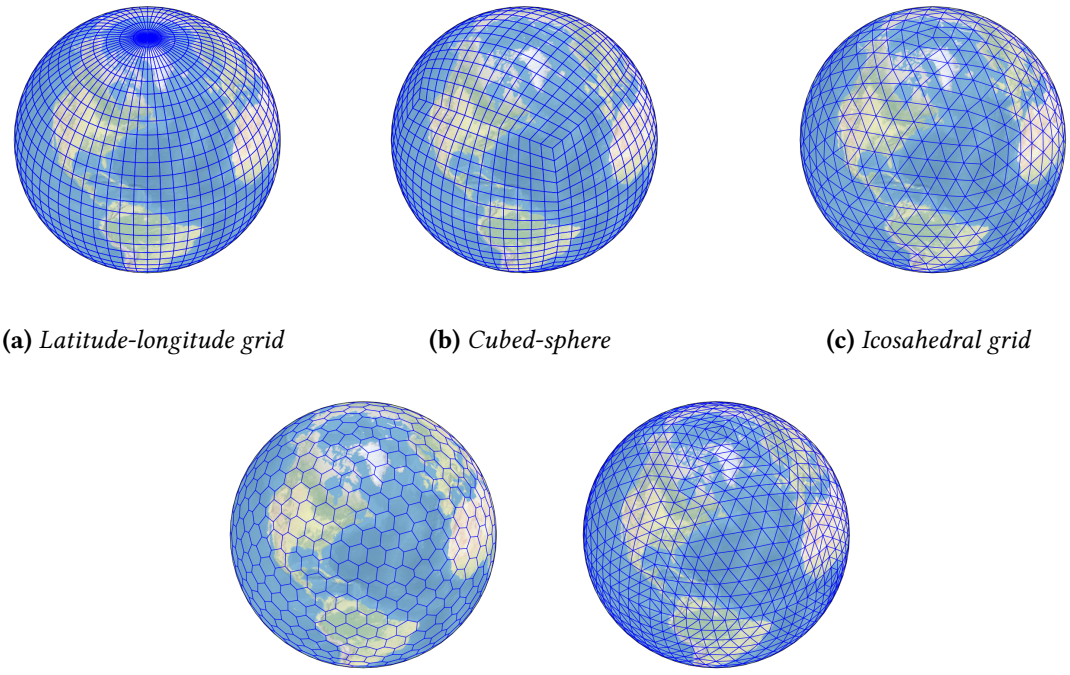
Weather and climate predictions are recognized as a good for mankind, due to the information they yield for diverse activities. For instance, short-range forecasts are useful for public use, while medium-range forecasts are helpful for industrial activities and agriculture. Seasonal forecasts (one up 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 addition 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 provide a detailed description of one-dimensional (1D) finite-volume (FV) schemes within a Semi-Lagrangian (SL) framework, specifically applied to the 1D advection equation with periodic boundary conditions. These schemes are also known as flux-form Semi-Lagrangian schemes, and they allow for time steps beyond the Courant-Friedrichs-Lewy (CFL) condition while preserving the total mass. FV-SL schemes have been explored in the literature since the work of LeVeque (1985), which extended the finite-volume schemes from Godunov (1959) to accommodate larger time steps. This approach has been further investigated in the literature (c.f, e.g. . Leonard et al. (1996) and Lin and Rood (1996)).

To introduce the FV-SL schemes, we begin by discretizing the spatial and temporal domains into uniform grids. Subsequently, the FV-SL schemes involve three steps. The first step involves computing the departure points of the spatial grid edges. The second step, known as reconstruction, utilizes the grid cell average values to determine a piecewise function within each cell. This piecewise function approximates the values of the advected quantity and ensures the preservation of its local mass within each grid cell. The third step entails updating the fluxes at the grid edges by integrating the reconstruction function over a domain that extends from the departure point of the grid edge to the grid edge itself.

The first step of FV-SL schemes can be accomplished by integrating an ordinary differential equation backward in time. The second step is performed using the Piecewise-Parabolic Method (PPM) proposed by Colella and Woodward (1984). As the name suggests, PPM employs piecewise-parabolic functions. The third and final step is computed easily, as the reconstruction functions consist of parabolas that preserve the local mass.

It is worth noting that the reconstruction function can be constructed using functions other than parabolas. In fact, the Piecewise-Parabolic Method (PPM) can be seen as an extension of the Piecewise-Linear method proposed by Van Leer (1977), which, in turn, was inspired by the Piecewise-Constant method introduced by Godunov (1959). Additionally, other schemes inspired by PPM have been proposed in the literature utilizing higher-order

polynomials, such as quartic polynomials (White & Adcroft, 2008). For a comprehensive review of general piecewise-polynomial reconstruction, we recommend referring to the technical report by Engwirda and Kelley (2016), Lauritzen et al. (2011), and the references therein.

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, PPM has been implemented in the FV3 dynamical core on its latitude-longitude grid (Lin, 2004) and cubed-sphere (Putman & Lin, 2007) versions. Although many other shapes for the basis functions and higher-order schemes are available in the literature, L. Harris et al. (2021) points out that the PPM scheme suits the needs of FV3 well. It is a flexible method that can be modified to ensure low diffusivity or shape preservation, for example. Additionally, a finite-volume numerical method usually requires monotonicity constraints, which, according to Godunov's theorem, limit the order of convergence to at most 1. Therefore, a higher-order scheme needs to strike a well-balanced trade-off between increasing computational cost and potential benefits.

This chapter begins with a basic review of one-dimensional advection equation in the integral form in Section 2.1. In Section 2.2, we establish the framework for general one-dimensional finite-volume Semi-Lagrangian schemes. Section 2.3 presents methods for computing the departure point. The PPM reconstruction is described in Section 2.4, while Subsection 2.4.2 introduces different approaches to ensure the monotonicity of parabolas. Section 2.5 focuses on the description and investigation of the PPM flux computation. Section 2.6 presents numerical results using the PPM scheme for the advection equation. Finally, Section 2.7 presents some concluding remarks. The application of PPM to solve two-dimensional problems will be addressed in Chapter 3.

2.1 One-dimensional advection equation in integral form

2.1.1 Notation

Before introducing the FV-SL schemes, let us establish some notation by introducing the concepts of a Δx-grid, a Δt-temporal grid, and the $(\Delta x, \Delta t, \lambda)$-discretization, as well as the concept of grid function/winds. In this chapter, we will use the notation $\Omega = [a, b]$ to represent the interval under consideration, and v to represent a non-negative integer indicating the number of ghost cell layers in each boundary. We also use the notations $\mathbb{R}_v^N := \mathbb{R}^{N+2v}$ and $\mathbb{R}_v^{N+1} := \mathbb{R}^{N+1+2v}$.

Definition 2.1 (Δx-grid). *For a given interval Ω and a positive real number Δx such that $\Delta x = (b-a)/N$ for some positive integer N, we say that $\Omega_{\Delta x} = \{X_i\}_{i=-v+1}^{N+v}$ is a Δx-grid for Ω if*

$$X_i = [x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}] = [a + (i-1)\Delta x, a + i\Delta x],$$

and $\Delta x = x_{i+\frac{1}{2}} - x_{i-\frac{1}{2}}$. Each X_i is referred to as a control volume or cell, and $x_{i-\frac{1}{2}}$ and $x_{i+\frac{1}{2}}$ are

the edges of the control volume X_i. The cell centroid is defined by

$$x_i = \frac{1}{2}(x_{i+\frac{1}{2}} + x_{i-\frac{1}{2}}), \quad \forall i = -\nu + 1, \dots, N + \nu,$$

and Δx is the cell length.

Remark 2.1. *If $1 \leq i \leq N$, we refer to i as an interior index; otherwise, i is considered a ghost cell index and we say the X_i is a ghost cell.*

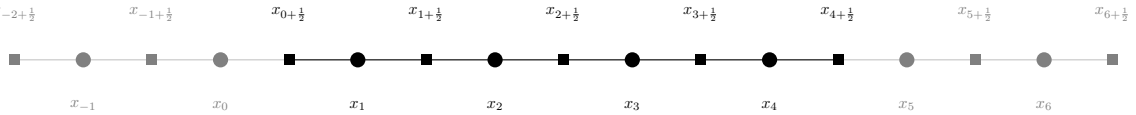


Figure 2.1: *Illustration of a Δx-grid with $N = 4$ cells in its interior (in black) and $\nu = 2$ ghost cell layers (in gray). The edges are denoted by squares and the cell centroids are denoted using circles.*

Definition 2.2 (Δt-temporal grid)**.** *For a given interval $[0, T]$ and a positive real number Δt such that $\Delta t = T/N_T$ for some positive integer N_T, we say that $T_{\Delta T} = \{T_n\}_{n=0}^{N_T}$ a Δt-temporal grid for $[0, T]$ if*

$$T_n = [t^n, t^{n+1}], \quad t^n = n\Delta t, \quad \Delta t = \frac{T}{N_T}, \quad \forall n = 0, \dots, N_T.$$

Definition 2.3 ($(\Delta x, \Delta t, \lambda)$-discretization)**.** *Given $\Omega \times [0, T]$ and positive real numbers Δx and Δt, we say that $(\Omega_{\Delta x}, T_{\Delta t})$ is a $(\Delta x, \Delta t, \lambda)$-discretization of $\Omega \times [0, T]$ if $\Omega_{\Delta x}$ is a Δx-grid for Ω, $T_{\Delta t}$ is a Δt-temporal grid for $[0, T]$, and $\frac{\Delta t}{\Delta x} = \lambda$.*

Remark 2.2. *Whenever we refer to a Δx-grid, a Δt-temporal grid, or a $(\Delta x, \Delta t, \lambda)$-discretization, X_i, N, t^n, and N_T are assumed to be implicitly defined.*

Next, we introduce the definitions of grid functions at cell centroids and edges.

Definition 2.4 (Δx-grid function)**.** *For a Δx-grid, we say that Q is a Δx-grid function if $Q = (Q_{-\nu+1}, \dots, Q_{N+\nu}) \in \mathbb{R}_\nu^N$.*

Definition 2.5 (Δx-grid wind)**.** *For a Δx-grid, we say that u is a Δx-grid wind if $u = (u_{-\nu+\frac{1}{2}}, \dots, u_{N+\nu+\frac{1}{2}}) \in \mathbb{R}_\nu^{N+1}$.*

The definition of a Δx-grid wind is based on the Arakawa grids (Arakawa & Lamb, 1977). Considering functions $q, u : \Omega \times [0, T] \to \mathbb{R}$ and a $(\Delta x, \Delta t, \lambda)$-discretization of $\Omega \times [0, T]$, we introduce the grid functions $q^n \in \mathbb{R}_\nu^N$ and $u^n \in \mathbb{R}_\nu^{N+1}$. Here, $q_i^n = q(x_i, t^n)$ and $u_{i+\frac{1}{2}}^n = u(x_{i+\frac{1}{2}}, t^n)$. These grid functions represent the discrete values of q and u at the cell centroids and edges, respectively, for each time level t^n (Figure 2.2).

In this chapter, our focus lies on periodic grid functions. We define a Δx-grid function Q as periodic if it satisfies the following conditions:

$$Q_i = Q_{N+i}, \quad i = -\nu + 1, \dots, 0,$$
$$Q_i = Q_{i-N}, \quad i = N + 1, \dots, N + \nu.$$

Similarly, we define a Δx-grid wind as periodic if it meets the following requirements:

$$u_{i-\frac{1}{2}} = u_{N+i+\frac{1}{2}}, \quad i = -\nu, \dots, -1,$$
$$u_{i+\frac{1}{2}} = u_{i+\frac{1}{2}-N}, \quad i = N+1, \dots, N+\nu.$$

We use the notation $\mathbb{P}_\nu^N$ and $\mathbb{P}_\nu^{N+1}$ to represent the spaces of periodic Δx-grid functions and winds, respectively.

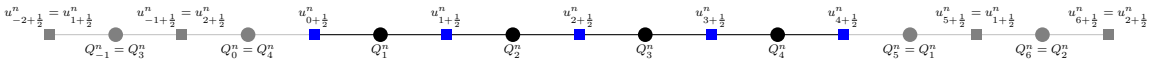


Figure 2.2: *Illustration of Δx-grid function Q (black circles) and a Δx-grid wind u (blue squares) and its ghost cell values (in gray) assuming periodicity.*

Given $Q \in \mathbb{P}_\nu^N$, we define the p-norm as

$$\|Q\|_{p,\Delta x} = \begin{cases} \left(\sum_{i=1}^N |Q_i|^p \right)^{\frac{1}{p}} & \text{if } 1 \le p < \infty, \\ \max_{i=1,\dots,N} |Q_i| & \text{otherwise}, \end{cases} \tag{2.1}$$

which is indeed a norm for periodic grid functions. Using a similar notation as in Engwirda and Kelley (2016), we define the stencil and a grid function evaluated on a stencil as follows.

Definition 2.6 (Stencil). *For a Δx-grid, and each $i = 0, \dots, N$, we define a stencil as a set of the form $\mathcal{S}_{i+\frac{1}{2}} = \{i - r + 1, \dots, i-1, i, i+1, \dots, i+s\} \subset \{-\nu+1, \dots, N+\nu\}$.*

Definition 2.7 (Grid function restricted to a stencil). *For a Δx-grid, a stencil $\mathcal{S}_{i+\frac{1}{2}}$, and a Δx-grid function Q, we define $Q(\mathcal{S}_{i+\frac{1}{2}}) = (Q_k)_{k \in \mathcal{S}_{i+\frac{1}{2}}}$.*

These definitions provide the necessary notation for describing grid functions and their evaluations on stencils. To achieve a more compact notation in some situations, we introduce the centered difference notation:

$$\delta_x g(x_i, t) = g(x_{i+\frac{1}{2}}, t) - g(x_{i-\frac{1}{2}}, t), \tag{2.2}$$

for any function $g : \Omega \times [0, T] \to \mathbb{R}$. Additionally, we introduce the average value of q in the i-th control volume at time t, denoted as $Q_i(t)$, defined by:

$$Q_i(t) = \frac{1}{\Delta x} \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} q(x, t)\, dx. \tag{2.3}$$

Moreover, we define the Δx-grid function of average values as $Q(t) = (Q_i(t))_{i=-\nu+1}^{N+\nu}$. Here, $Q_i(t)$ represents the average value of q in the i-th control volume at time t.

For the consideration of periodic boundary conditions, we can define spaces of periodic functions over the interval Ω as follows:

$$\mathcal{S}_P(\Omega) = \{q : \mathbb{R} \times [0, +\infty[\to \mathbb{R} \ : \ q(x + b - a, t) = q(x, t), \quad \forall x \in \mathbb{R}, \quad t \ge 0\}.$$

Similarly, the space of k-times periodically differentiable functions $C_P^k(\Omega)$ can be defined as:

$$C_P^k(\Omega) = S_P(\Omega) \cap C^k(\mathbb{R} \times [0, \infty[),$$

where $C^k(\mathbb{R}\times[0, +\infty[)$ denotes the space of functions that are k times continuously differentiable in both the spatial and temporal variables. In summary, $S_P(\Omega)$ represents the space of periodic functions, and $C_P^k(\Omega)$ represents the space of k-times periodically differentiable functions over the interval Ω subject to periodic boundary conditions.

2.1.2 The 1D advection equation

In this section, we will derive the integral form of the 1D advection equation with periodic boundary conditions over the interval Ω. What is going to be presented here follows LeVeque (1990, 2002) closely. The advection equation with periodic boundary conditions in its differential form is given by:

$$\begin{cases} [\partial_t q + \partial_x(uq)](x,t) = 0, & \forall (x,t) \in \mathbb{R}\times]0, +\infty[, \\ q(a,t) = q(b,t), & \forall t \geq 0, \\ q_0(x) = q(x,0), & \forall x \in \Omega. \end{cases} \tag{2.4}$$

Here, $q \in C_P^1(\Omega)$ represents the advected quantity, and $u \in C_P^1(\Omega)$ represents the velocity. We will focus on Equation (2.4) over the domain $D = \Omega \times [0, T]$, where $T > 0$ is a finite time. A strong or classical solution to the advection equation is defined as a function $q \in C_P^1(\Omega)$ and satisfies Equation (2.4). In order to deduce the integral form of Equation (2.4), we consider $[x_1, x_2] \times [t_1, t_2] \subset D$. Integrating Equation (2.5) over $[x_1, x_2]$, we obtain:

$$\frac{d}{dt} \int_{x_1}^{x_2} q(x,t)\, dx = -((uq)(x_2,t) - (uq)(x_1,t)), \tag{2.5}$$

and integrating Equation (2.5) over $[t_1, t_2]$, we get

$$\int_{x_1}^{x_2} q(x,t_2)\, dx = \int_{x_1}^{x_2} q(x,t_1) - \left(\int_{t_1}^{t_2} (uq)(x_2,t)\, dt - \int_{t_1}^{t_2} (uq)(x_1,t)\, dt \right). \tag{2.6}$$

The presented problem, Problem 2.1, aims to find a solution, called weak solution, to the advection equation in its integral form, considering the given initial condition q_0 and velocity function u.

Problem 2.1. *Given an initial condition q_0 and a velocity function u we would like to find a weak solution q of the advection equation in the integral form:*

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

$\forall [x_1, x_2] \times [t_1, t_2] \subset \Omega \times [0, T]$, *and* $q(x,0) = q_0(x)$, $\forall x \in \Omega$, $q(a,t) = q(b,t)$, $\forall t \in [0, T]$.

We point out that, for Problem 2.1, the total mass in Ω at time t defined by:

$$M_{[a,b]}(t) = \int_a^b q(x,t)\,dx,$$

remains constant over time, i.e.,

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

This conservation of total mass property is highly desirable for numerical schemes aiming to approximate general conservation law solutions accurately.

Applying the steps from Equation (2.4) to Equation (2.6) in reverse order, one can verify that if q is a weak solution and $q \in C_P^1(\Omega)$, then it satisfies Equation (2.4). Therefore, Equation (2.4) and Problem (2.1) are equivalent when $q \in C_P^1(\Omega)$. However, Problem (2.1) can be formulated for functions that are not C^1 and have discontinuities. In fact, Problem (2.1) only requires that q and uq are locally integrable.

It is worth noting that Equation (2.6) holds for all x_1, x_2, t_1, and t_2 such that $[x_1, x_2] \times [t_1, t_2] \subset D$. Therefore, let us consider a $(\Delta x, \Delta t, \lambda)$-discretization of D and rewrite Equation (2.6) in terms of this discretization. By replacing t_1, t_2, x_1, and x_2 with $t^n, t^{n+1}, x_{i-\frac{1}{2}}$, and $x_{i+\frac{1}{2}}$, respectively, in Equation (2.6), we obtain:

$$Q_i(t^{n+1}) = Q_i(t^n) - \frac{1}{\Delta x}\left(\int_{t^n}^{t^{n+1}} (uq)(x_{i+\frac{1}{2}}, t)\,dt - \int_{t^n}^{t^{n+1}} (uq)(x_{i-\frac{1}{2}}, t)\,dt \right), \quad \forall i = 1, \dots, N, \quad \forall n = 0, \dots, N_T - 1. \tag{2.7}$$

To achieve a more compact notation, we use the centered difference notation and then Equation (2.7) can be rewritten as:

$$Q_i(t^{n+1}) = Q_i(t^n) - \frac{1}{\Delta x}\delta_x\left(\int_{t^n}^{t^{n+1}} (uq)(x_i, t)\,dt \right), \quad \forall i = 1, \dots, N, \quad \forall n = 0, \dots, N_T - 1. \tag{2.8}$$

Now we can define a discretized version of Problem 2.1 as Problem 2.2.

Problem 2.2. *Let us consider the framework of Problem 2.1 and a $(\Delta x, \Delta t, \lambda)$-discretization of $\Omega \times [0,T]$. Since we are operating within the framework of Problem 2.1, the following relationship holds:*

$$Q_i(t^{n+1}) = Q_i(t^n) - \lambda\delta_x\left(\frac{1}{\Delta t} \int_{t^n}^{t^{n+1}} (uq)(x_i, t)\,dt \right), \quad \forall i = 1, \dots, N, \quad \forall n = 0, \dots, N_T - 1, \tag{2.9}$$

where $Q_i(t) = \frac{1}{\Delta x}\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} q(x,t)\,dx$. Our objective now is to determine the values $Q_i(t^n)$, $\forall i = 1, \dots, N$, $\forall n = 0, \dots, N_T - 1$, given the initial values $Q_i(0)$, $\forall i = 1, \dots N$. In other words, we aim to find the average values of q in each control volume X_i at the specified time instances.

It is important to note that no approximations have been made in problems (2.1) and (2.2). In Equation (2.9), we divided and multiplied by Δt to interpret $\frac{1}{\Delta t}\int_{t^n}^{t^{n+1}} (uq)(x_{i\pm\frac{1}{2}}, t)\,dt$ as a time-averaged flux. This interpretation is useful for deriving finite-volume schemes.

In Problem 2.2, we need to approximate the time-averaged flux at the cell edges $x_{i\pm\frac{1}{2}}$ to derive a finite-volume scheme. This flux, in principle, requires knowledge of q over the entire interval $[t^n, t^{n+1}]$. To overcome this, we can express the temporal integral as a spatial integral at time t^n. This approach avoids the need for information about q throughout the entire interval $[t^n, t^{n+1}]$. Furthermore, this spatial integral domain is closely related to the definition of the departure point.

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:

$$\begin{cases} \frac{\partial X}{\partial t}(t, s; x_{i+\frac{1}{2}}) = u(X(t, s; x_{i+\frac{1}{2}}), t), & t \in [t^n, s] \\ X(s, s; x_{i+\frac{1}{2}}) = x_{i+\frac{1}{2}}. \end{cases} \tag{2.10}$$

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. In Figure 2.3 we illustrate the departure point idea.

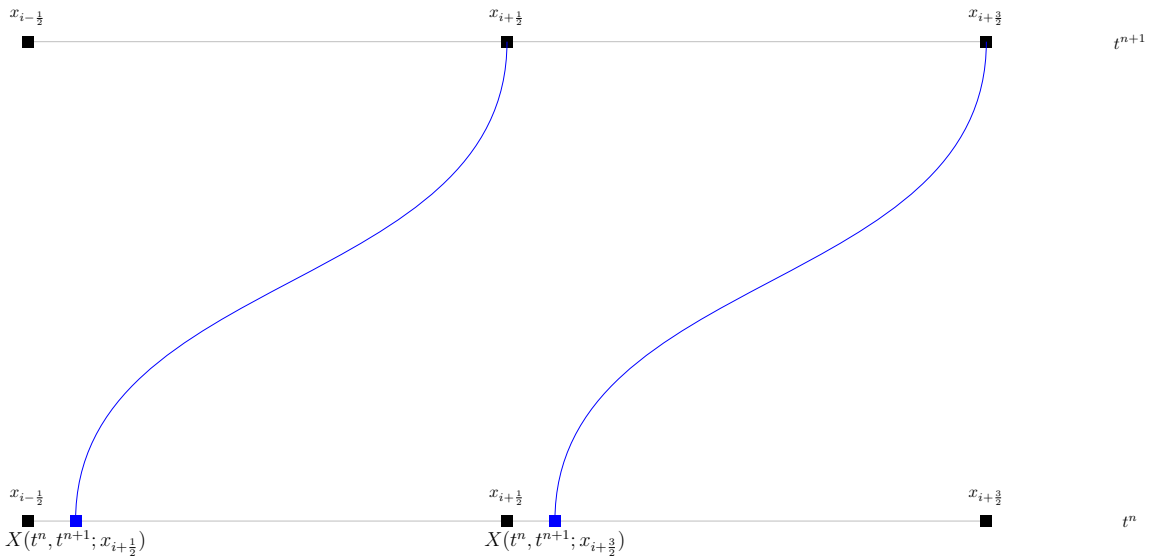


Figure 2.3: *Illustration of the departure point of the cell edges from time t^{n+1} to t^n.*

Integrating Equation (2.10) over the interval $[t, s]$, we get:

$$X(t, s; x_{i+\frac{1}{2}}) = x_{i+\frac{1}{2}} - \int_t^s u(X(\theta, s; x_{i+\frac{1}{2}}), \theta)\, d\theta. \tag{2.11}$$

In the following Proposition, we show how the time-averaged flux is related to a spatial integral over a interval depending on departure points.

Proposition 2.1. *Assume the framework of Problem 2.2. If q and u are C^1 functions, then:*

$$\int_{t^n}^{t^{n+1}} (uq)(x_{i+\frac{1}{2}}, s)\, ds = \int_{X(t^n, t^{n+1}; x_{i+\frac{1}{2}})}^{x_{i+\frac{1}{2}}} q(x, t^n)\, dx \tag{2.12}$$

Proof. Using the Leibniz rule for integration in Equation (2.11), it follows that:

$$\begin{aligned}\frac{\partial X}{\partial s}(t,s;x_{i+\frac{1}{2}}) &= -\left(u(x_{i+\frac{1}{2}},s) + \int_t^s \frac{du}{ds}(X(\theta,s;x_{i+\frac{1}{2}}),\theta)\,d\theta\right)\\ &= -u(x_{i+\frac{1}{2}},s) - \int_t^s \frac{\partial u}{\partial x}(X(\theta,s;x_{i+\frac{1}{2}}),\theta)\frac{\partial X}{\partial s}(\theta,s;x_{i+\frac{1}{2}})\,d\theta.\end{aligned} \tag{2.13}$$

Taking the derivative with respect to t of Equation (2.13), we have:

$$\frac{\partial}{\partial t}\left(\frac{\partial X}{\partial s}\right)(t,s;x_{i+\frac{1}{2}}) = \frac{\partial u}{\partial x}(X(t,s;x_{i+\frac{1}{2}}),t)\frac{\partial X}{\partial s}(t,s;x_{i+\frac{1}{2}}). \tag{2.14}$$

Using standard ordinary differential equations techniques (ODE), we get that X that solves Equations (2.13) and (2.14) is given by:

$$\frac{\partial X}{\partial s}(t,s;x_{i+\frac{1}{2}}) = -\exp\left(\int_t^s \frac{\partial u}{\partial x}(X(\theta,s;x_{i+\frac{1}{2}}),\theta)\,d\theta\right)u(x_{i+\frac{1}{2}},s). \tag{2.15}$$

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

$$\begin{aligned}\frac{dq}{dt}(X(t,s;x_{i+\frac{1}{2}}),t) &= \frac{\partial q}{\partial t}(X(t,s;x_{i+\frac{1}{2}}),t) + u(X(t,s;x_{i+\frac{1}{2}}),t)\frac{\partial q}{\partial x}(X(t,s;x_{i+\frac{1}{2}}),t)\\ &= -\frac{\partial u}{\partial x}(X(t,s;x_{i+\frac{1}{2}}),t)q(X(t,s;x_{i+\frac{1}{2}}),t),\end{aligned} \tag{2.16}$$

where we used that q satisfies the linear advection equation on its differential form and that $X(t,s;x_{i+\frac{1}{2}})$ solves Equation (2.10). Using again standard ODE techniques, we get that q that solves Equation (2.16) is given by:

$$q(X(t,s;x_{i+\frac{1}{2}}),t) = \exp\left(-\int_t^s \frac{\partial u}{\partial x}(X(\theta,s;x_{i+\frac{1}{2}}),\theta)\,d\theta\right)q(x_{i+\frac{1}{2}},s). \tag{2.17}$$

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

$$\int_{X(t^n,t^n;x_{i+\frac{1}{2}})=x_{i+\frac{1}{2}}}^{X(t^n,t^{n+1};x_{i+\frac{1}{2}})} q(x,t^n)\,dx = \int_{t^n}^{t^{n+1}} q(X(t^n,s;x_{i+\frac{1}{2}}),t^n)\frac{\partial X}{\partial s}(t^n,s;x_{i+\frac{1}{2}})\,ds, \tag{2.18}$$

where we are just using the variable change integration formula. Then, it follows from Equations (2.15) and (2.17) with $t = t^n$ that:

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

which is the desired formula. □

With the aid of Proposition 2.1, we can rewrite Problem 2.2 in terms of the departure

point, avoiding the need for knowledge about q over the entire interval $[t^n, t^{n+1}]$. This is described in Problem 2.3:

Problem 2.3. *Assume the framework of Problem 2.1 and a $(\Delta x, \Delta t, \lambda)$-discretization of $\Omega \times [0,T]$. Since we are in the framework of Problem 2.1, it follows that:*

$$Q_i(t^{n+1}) = Q_i(t^n) - \lambda\left(\frac{1}{\Delta t}\int_{X(t^n,t^{n+1};x_{i+\frac{1}{2}})}^{x_{i+\frac{1}{2}}} q(x,t^n)\,dx - \frac{1}{\Delta t}\int_{X(t^n,t^{n+1};x_{i-\frac{1}{2}})}^{x_{i-\frac{1}{2}}} q(x,t^n)\,dx\right), \quad \forall i = 1,\dots,N, \quad \forall n = 0,\dots,N_T-1, \tag{2.19}$$

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

At each time step t^n, we compute the values of $Q_i(t^{n+1})$ based on $Q_i(t^n)$ and the integrals of $q(x,t^n)$ over specific intervals. These intervals are defined by the departure points $X(t^n,t^{n+1};x_{i+\frac{1}{2}})$ and $X(t^n,t^{n+1};x_{i-\frac{1}{2}})$. To perform the computations, we need to determine the departure points from the edges of all control volumes and calculate the required integrals. This idea serves as the motivation for defining finite-volume Semi-Lagrangian schemes. These schemes involve estimating the departure points and reconstructing the function q at time t^n using its average values $Q_i(t^n)$, which enables us to compute the necessary integrals.

2.2 The finite-volume Semi-Lagrangian approach

Finally, we define the 1D FV-SL scheme problem as follows in Problem 2.3.

Problem 2.4 (1D FV-SL scheme)**.** *Assume the framework defined in Problem 2.3. The finite-volume Semi-Lagrangian approach of Problem 2.3 consists of finding a scheme of the form:*

$$Q_i^{n+1} = Q_i^n - \lambda(F_{i+\frac{1}{2}}^n - F_{i-\frac{1}{2}}^n), \quad \forall i = 1,\dots,N, \quad \forall n = 0,\dots,N_T-1, \tag{2.20}$$

where $Q^n \in \mathbb{P}_v^N$ is intended to be an approximation of $Q(t^n) \in \mathbb{P}_v^N$ in some sense. We define $Q_i^0 = Q_i(0)$ or $Q_i^0 = q_i^0$. The terms $F_{i\pm\frac{1}{2}}^n$ are known as numerical flux and are given by

$$F_{i\pm\frac{1}{2}}^n = \frac{1}{\Delta t}\int_{\tilde{x}_{i\pm\frac{1}{2}}^n}^{x_{i\pm\frac{1}{2}}} \tilde{q}(x;Q^n)\,dx, \tag{2.21}$$

where $\tilde{x}_{i\pm\frac{1}{2}}^n$ is an estimate of the departure point $X(t^n,t^{n+1};x_{i\pm\frac{1}{2}})$, and $\tilde{q}$ is a reconstruction function for q built with the values Q^n. Thus, $F_{i\pm\frac{1}{2}}^n$ approximates $\frac{1}{\Delta t}\int_{X(t^n,t^{n+1};x_{i\pm\frac{1}{2}})}^{x_{i\pm\frac{1}{2}}} q(x,t^n)\,dx$.

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

$$M^n = \Delta x \sum_{i=1}^{N} Q_i^n. \tag{2.22}$$

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

$$M^{n+1} = \Delta x \sum_{i=1}^{N} Q_i^{n+1} = M^n - \Delta t \sum_{i=1}^{N} (F^n_{i+\frac{1}{2}} - F^n_{i-\frac{1}{2}}) = M^n - \Delta t (F^n_{N+\frac{1}{2}} - F^n_{\frac{1}{2}}) = M^n,$$

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

We would like to highlight an important relationship between the average values of q and its values at the cell centroids. In Problem 2.4, we mentioned that the initial condition can be represented as q_i^0 instead of $Q_i(0)$. Moreover, when analyzing the convergence of a FV-SL scheme, it is useful to compare Q_i^n with q_i^n since computing $Q_i(t^n)$ requires evaluating an analytical integral, which can be challenging in certain cases. In Proposition 2.2, we provide a simple proof that q_i^n approximates $Q_i(t^n)$ with second-order error when q is twice continuously differentiable.

Proposition 2.2. *If $q \in C^2_P(\Omega)$, then $Q_i(t^n) - q_i^n = C_1 \Delta x^2$, where $C_1 = \frac{1}{24}\frac{\partial^2 q}{\partial x^2}(\eta, t^n)$, $\eta \in X_i$.*

Proof. Just apply Theorem A.3 for the function $q(x, t^n)$. □

The Problem of the convergence of 1D FV-SL schemes is addressed in Section A.5. Now we are going to address the problem of the departure point estimation and the reconstruction problem.

2.3 Departure point computation

2.3.1 RK1 scheme

Equation (2.11) enables us to compute or estimate the departure point. For instance, if u is constant, the departure point at time t^n for the point $x_{i+\frac{1}{2}}$ at time t^{n+1} is given by:

$$X(t^n, t^{n+1}; x_{i+\frac{1}{2}}) = x_{i+\frac{1}{2}} - u\Delta t. \tag{2.23}$$

In general, the estimated departure point, denoted by $\tilde{x}^n_{i+\frac{1}{2}}$, takes the form:

$$\tilde{x}^n_{i+\frac{1}{2}} = x_{i+\frac{1}{2}} - \tilde{u}^n_{i+\frac{1}{2}} \Delta t, \tag{2.24}$$

where $\tilde{u}^n_{i+\frac{1}{2}}$ represents the time-averaged wind and approximates:

$$\frac{1}{\Delta t} \int_{t^n}^{t^{n+1}} u(X(\theta, t^{n+1}; x_{i+\frac{1}{2}}), \theta)\, d\theta. \tag{2.25}$$

The departure point $\tilde{x}^n_{i+\frac{1}{2}}$ is said to be p-order accurate if there exists a constant C that does not depend on Δt, such that:

$$X(t^n, t^{n+1}; x_{i+\frac{1}{2}}) - \tilde{x}^n_{i+\frac{1}{2}} = C\Delta t^p. \tag{2.26}$$

One possible way of estimating the time-averaged wind is by using:

$$\tilde{u}^n_{i+\frac{1}{2}} = u^n_{i+\frac{1}{2}}, \tag{2.27}$$

as in the original PPM scheme by Colella and Woodward (1984). This scheme will be referred to as **RK1**. Our objective now is to determine the constant C and the value of p in Equation (2.26) in the following proposition. It is useful to introduce the concept of a material derivative beforehand:

$$\frac{Dh}{Dt} = \frac{\partial h}{\partial t} + u\frac{\partial h}{\partial x},$$

where h is a function belonging to $\mathcal{C}^1$.

Proposition 2.3. *If $u \in \mathcal{C}^1$ and the time-averaged wind is computed using Equation (2.27), then the departure point from Equation (2.24) satisfies:*

$$X(t^n, t^{n+1}; x_{i+\frac{1}{2}}) - \tilde{x}^n_{i+\frac{1}{2}} = C\Delta t^2, \tag{2.28}$$

for a constant C that depends on u.

Proof. Using Corollary A.1 for the function $f(t) = u(X(t, t^{n+1}; x_{i+\frac{1}{2}}), t)$ in Equation (2.11), we obtain:

$$X(t^n, t^{n+1}; x_{i+\frac{1}{2}}) = x_{i+\frac{1}{2}} - u(X(t^n, t^{n+1}; x_{i+\frac{1}{2}}), t^n)\Delta t - \frac{1}{2}\frac{Du}{Dt}(X(\theta_1, t^{n+1}; x_{i+\frac{1}{2}}), \theta_1)\Delta t^2, \tag{2.29}$$

for $\theta_1 \in [t^n, t^{n+1}]$. Using Taylor's expansion of $u(X(t, t^{n+1}; x_{i+\frac{1}{2}}), t^n)$, we have:

$$u(X(t^n, t^{n+1}; x_{i+\frac{1}{2}}), t^n) = u^n_{i+\frac{1}{2}} - \left(u\frac{\partial u}{\partial x}\right)(X(\theta_2, t^{n+1}; x_{i+\frac{1}{2}}), t^n))\Delta t, \tag{2.30}$$

for $\theta_2 \in [t^n, t^{n+1}]$. Substituting Equation (2.30) into Equation (2.29), we obtain the desired constant C given by:

$$C = C(\theta_1, \theta_2) = -\frac{1}{2}\frac{Du}{Dt}(X(\theta_1, t^{n+1}; x_{i+\frac{1}{2}}), \theta_1) - \left(u\frac{\partial u}{\partial x}\right)(X(\theta_2, t^{n+1}; x_{i+\frac{1}{2}}), t^n). \tag{2.31}$$

□

2.3.2 RK2 scheme

Before presenting a higher-order estimate for the departure point, let us recall the definition of the CFL number.

Definition 2.8. *For Problem 2.4, the CFL number at an edge $x_{i+\frac{1}{2}}$ and at a time level t^n is defined by*

$$c^n_{i+\frac{1}{2}} = \frac{\Delta t}{\Delta x}u^n_{i+\frac{1}{2}}. \tag{2.32}$$

The CFL number is the maximum of the values $c^n_{i+\frac{1}{2}}$. The problem of estimating the

departure point is very common in Semi-Lagrangian schemes, which are quite popular in atmospheric modeling. For a review of departure point calculation methods, we refer to Tumolo (2011, Chapter 3) and the references therein. There are different approaches to compute the departure point, such as integrating the ODE from Equation 2.1 using different time integrators (D. Durran, 2011) backward in time. The Runge-Kutta methods are a possible choice to compute the departure point (*cf. e.g.* Guo et al. (2014), Lu et al. (2022)). In this work, we shall consider a second-order Runge-Kutta method to compute the departure point, which we express in terms of $\tilde{u}^n_{i+\frac{1}{2}}$ using the following equations (D. R. Durran, 2010):

$$\begin{aligned} \tilde{x}^{n+\frac{1}{2}}_{i+\frac{1}{2}} &= x_{i+\frac{1}{2}} - u^n_{i+\frac{1}{2}} \frac{\Delta t}{2} = x_{i+\frac{1}{2}} - c^n_{i+\frac{1}{2}} \frac{\Delta x}{2}, \\ \tilde{u}^n_{i+\frac{1}{2}} &= u\left(\tilde{x}^{n+\frac{1}{2}}_{i+\frac{1}{2}}, t^n + \frac{\Delta t}{2}\right). \end{aligned} \tag{2.33}$$

Notice that this scheme requires values of u at points that are not grid points, both in time and space. This problem is addressed firstly using a second-order extrapolation in time:

$$u^{n+\frac{1}{2}}_{i+\frac{1}{2}} = \frac{3}{2} u^n_{i+\frac{1}{2}} - \frac{1}{2} u^{n-1}_{i+\frac{1}{2}}. \tag{2.34}$$

Then we use linear interpolation in space:

$$\tilde{u}^n_{i+\frac{1}{2}} = \begin{cases} \left(1 - \alpha^n_{i+\frac{1}{2}}\right) u^{n+\frac{1}{2}}_{i+\frac{1}{2}-k} + \alpha^n_{i+\frac{1}{2}} u^{n+\frac{1}{2}}_{i-\frac{1}{2}-k} & \text{if } u^n_{i+\frac{1}{2}} \geq 0, \\ \alpha^n_{i+\frac{1}{2}} u^{n+\frac{1}{2}}_{i+\frac{3}{2}-k} + \left(1 - \alpha^n_{i+\frac{1}{2}}\right) u^{n+\frac{1}{2}}_{i+\frac{1}{2}-k} & \text{if } u^n_{i+\frac{1}{2}} < 0, \end{cases} \tag{2.35}$$

where $\frac{c^n_{i+\frac{1}{2}}}{2} = \alpha^n_{i+\frac{1}{2}} + k$, $k = \lfloor \frac{c^n_{i+\frac{1}{2}}}{2} \rfloor$, $\alpha^n_{i+\frac{1}{2}} \in [0, 1[$, and $\lfloor \cdot \rfloor$ is the floor function. This scheme leads to a third-order error in the departure point estimate (see *e.g.* D. R. Durran (2010, Section 7.1.2)). This scheme shall be referred to as **RK2**. Notice that for this scheme, we need ghost values for the velocity, depending on how large the CFL number is. In particular, if the CFL number is less than 2, then $k = 0$ and we need the ghost values $u^n_{-1+\frac{1}{2}}$ and $u^n_{N+\frac{3}{2}}$. Finally, the impact of the departure point approximation on the time-averaged flux is investigated in Section A.5.3.

2.4 Reconstruction

In this section, we will review the Piecewise-Parabolic Method (PPM). The analysis of its accuracy will be presented in Section A.4. PPM was originally proposed by Colella and Woodward (1984) for gas dynamic simulations, and its applicability to atmospheric simulations has been demonstrated by Carpenter et al. (1990). This method is based on utilizing parabolas to reconstruct the function using its average values, ensuring both mass conservation and monotonicity. PPM is an extension of the Piecewise-Linear Method introduced by Van Leer (1977), and it is implemented in the FV3 model using the dimension splitting method developed by Lin and Rood (1996).

Let's consider a function q defined in $\Omega = [a, b]$ and a Δx-grid covering Ω. We assume

that we are given the average values $Q_i = \frac{1}{\Delta x}\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} q(x)\,dx$ for each control volume X_i, where $i = 1, \dots, N$. In this context, it is convenient to define the Δx-grid function $Q \in \mathbb{P}_\nu^N$ with the entries given by Q_i. To facilitate the discussion, we introduce the indicator function $\chi_i(x)$ for each control volume X_i, defined as:

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

Drawing inspiration from Stoer and Bulirsch (2002, Chapter 1), we consider a family of functions $\Phi(\xi;\mu)$ defined for $\xi \in [0,1]$, depending on a parameter $\mu = (\mu_0, \mu_1, \dots, \mu_d) \in \mathbb{R}^{d+1}$. The reconstruction problem involves finding a piecewise function:

$$\tilde{q}(x;Q) = \sum_{i=1}^{N} \chi_i(x) q_i(x;Q), \tag{2.36}$$

where $q_i(x;Q) = \Phi\left(\frac{x-x_{i-\frac{1}{2}}}{\Delta x};\alpha_i\right)$ and $\alpha_i = (\alpha_{i0}, \alpha_{i1}, \dots \alpha_{id}) \in \mathbb{R}^{d+1}$. It is required that:

$$\frac{1}{\Delta x}\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} \tilde{q}(x;Q)\,dx = \frac{1}{\Delta x}\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} q_i(x;Q)\,dx = \int_0^1 \Phi(\xi;\alpha_i)\,d\xi = Q_i,$$

which means that $q_i(x;Q)$ preserves the mass within each control volume X_i.

Notice that, given $q_i(x;Q) = \Phi\left(\frac{x-x_{i-\frac{1}{2}}}{\Delta x};\alpha_i\right)$, it is reasonable to expect that $\Phi(0;\alpha_i)$ approximates $q_i(x_{i-\frac{1}{2}})$ and $\Phi(1;\alpha_i)$ approximates $q_i(x_{i+\frac{1}{2}})$. Additionally, if both q and Φ are sufficiently differentiable, $\Phi^{(l)}(0;\alpha_i)$ should approximate $(\Delta x)^l q^{(l)}(x_{i-\frac{1}{2}})$ and $\Phi^{(l)}(1;\alpha_i)$ should approximate $(\Delta x)^l q^{(l)}(x_{i+\frac{1}{2}})$, provided these derivatives exist.

One approach to estimating these values at the edges $x_{i+\frac{1}{2}}$ using the average values Q is by employing a reconstruction method based on primitive functions (LeVeque, 2002, Chapter 17). It is worth noting that if we define:

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

we have $Q^{(l)}(x) = q^{(l-1)}(x)$. Specifically, $Q^{(l)}(x_{i+\frac{1}{2}}) = q^{(l-1)}(x_{i+\frac{1}{2}})$ and $Q(x_{i+\frac{1}{2}}) = \Delta x \sum_{k=1}^{i} Q_k$, for all $i = 0, \dots, N$. Therefore, we can employ finite-difference schemes to estimate $q^{(l-1)}(x_{i+\frac{1}{2}})$ using the Δx-grid function Q, given that it is assumed to be known.

Let us assume that the l-th derivative of Q at $x_{i+\frac{1}{2}}$ is approximated using a stencil $S_{i+\frac{1}{2}}^{(l)}$ and weights $\beta_{k,i}^{(l)}$, where $k \in S_{i+\frac{1}{2}}^{(l)}$. When d is odd, we can seek a parameter $\alpha_i \in \mathbb{R}^{d+1}$ that ensures mass conservation and approximates q and its derivatives at the edges by solving the following system:

$$\begin{cases} \int_0^1 \Phi(\xi;\alpha_i)\,d\xi & = Q_i, \\ \Phi^{(l)}(0;\alpha_i) & = (\Delta x)^l \sum_{k\in S_{i-\frac{1}{2}}^{(l)}} \beta_{k,i}^{(l)} Q_k, \quad \text{for } l = 0, \dots, d-1. \end{cases} \tag{2.38}$$

If d is even, similarly we look for a parameter $\alpha_i \in \mathbb{R}^{d+1}$ that solves:

$$\begin{cases} \int_0^1 \Phi(\xi;\alpha_i)\,d\xi & = Q_i, \\ \Phi^{(l)}(0;\alpha_i) & = (\Delta x)^l \sum_{k\in S^{(l)}_{i-\frac{1}{2}}} \beta^{(l)}_{k,i} Q_k, \quad \text{for } l = 0,\dots,\frac{d}{2}-1, \\ \Phi^{(l)}(1;\alpha_i) & = (\Delta x)^l \sum_{k\in S^{(l)}_{i+\frac{1}{2}}} \beta^{(l)}_{k,i} Q_k, \quad \text{for } l = 0,\dots,\frac{d}{2}-1. \end{cases} \tag{2.39}$$

The reconstruction problem becomes linear when $\Phi(\xi;\mu)$ can be expressed as:

$$\Phi(\xi;\mu) = \sum_{k=0}^{d} \mu_k \Phi_k(\xi),$$

where Φ_k are functions defined on $[0,1]$. In this case, Equation (2.38) and Equation (2.39) form $(d+1)\times(d+1)$ linear systems. It is common to assume that the Φ_k's are linearly independent. Therefore, we have described a method that allows us to reconstruct a function from its average values, preserving its mass in each control volume, and approximating q at the edges. This method works for functions Φ_k as long as they are sufficiently differentiable. For example, choosing $d = 0$ and $\Phi_0(\xi) = 1$ gives us piecewise constant functions, as used in Godunov (1959). If we choose $d = 1$, $\Phi_0(\xi) = 1$, and $\Phi_1(\xi) = \xi$, we obtain a piecewise linear reconstruction, similar to Van Leer (1977). For polynomial reconstruction schemes, we refer to Engwirda and Kelley (2016) and the references therein.

2.4.1 The Piecewise-Parabolic Method

Hereafter, we are going the focus on the piecewise parabolic method from Colella and Woodward (1984) that uses $d = 2$, $\Phi_0(\xi) = 1$, $\Phi_1(\xi) = \xi$, $\Phi_1(\xi) = (1-\xi)\xi$. In order to follow the notation from Colella and Woodward (1984), we write $\alpha_{0i} = q_{L,i}$, $\alpha_{1i} = \Delta q_i$ and $\alpha_{2i} = q_{6,i}$. Therefore, each q_i may be expressed as:

$$q_i(x;Q) = q_{L,i} + \Delta q_i z_i(x) + q_{6,i} z_i(x)(1-z_i(x)), \quad \text{where } z_i(x) = \frac{x - x_{i-\frac{1}{2}}}{\Delta x}, \quad x \in X_i, \tag{2.40}$$

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 $[0,1]$. It is easy to see that $\lim_{x\to x^+_{i-\frac{1}{2}}} q_i(x;Q) = q_{L,i}$. If we define $q_{R,i} = \lim_{x\to x^-_{i+\frac{1}{2}}} q_i(x;Q)$, then we have:

$$\Delta q_i = q_{R,i} - q_{L,i}. \tag{2.41}$$

The average value of q_i is given by:

$$\frac{1}{\Delta x}\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} q_i(x;Q)\,dx = \frac{(q_{L,i}+q_{R,i})}{2} + \frac{q_{6,i}}{6}. \tag{2.42}$$

Under the hypothesis of mass conservation, we have:

$$q_{6,i} = 6\left(Q_i - \frac{(q_{L,i}+q_{R,i})}{2}\right). \tag{2.43}$$

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 parabola q_i from (2.36) guarantees mass conservation. To completely determine the parabola q_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 $q(x_{i+\frac{1}{2}})$. As we mentioned before in after introducing Equation (2.37), this is achieved using finite-differences. An explicit expression for the approximation of $q(x_{i-\frac{1}{2}})$, denoted by $q_{i+\frac{1}{2}}$, is given by (Colella & Woodward, 1984):

$$q_{i+\frac{1}{2}} = \frac{1}{2}\Big(Q_{i+1} + Q_i\Big) - \frac{1}{6}\Big(\delta Q_{i+1} - \delta Q_i\Big), \tag{2.44}$$

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

$$\delta Q_i = \frac{1}{2}\Big(Q_{i+1} - Q_{i-1}\Big). \tag{2.45}$$

We notice that Formula (2.45) may be rewritten more explicitly as:

$$q_{i+\frac{1}{2}} = \frac{7}{12}\Big(Q_{i+1} + Q_i\Big) - \frac{1}{12}\Big(Q_{i+2} + Q_{i-1}\Big). \tag{2.46}$$

The Formula (2.46) is fourth-order accurate if q is at least C^4 (Colella & Woodward, 1984). Indeed, we prove this later in Proposition A.1. An explicit expression for the values of $q_{R,i}$ and $q_{L,i}$ are given by:

$$q_{R,i} = q_{i+\frac{1}{2}} = \frac{7}{12}\Big(Q_{i+1} + Q_i\Big) - \frac{1}{12}\Big(Q_{i+2} + Q_{i-1}\Big), \tag{2.47}$$

$$q_{L,i} = q_{i-\frac{1}{2}} = \frac{7}{12}\Big(Q_i + Q_{i-1}\Big) - \frac{1}{12}\Big(Q_{i+1} + Q_{i-2}\Big). \tag{2.48}$$

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). The fifth-order reconstruction formula reads:

$$q_{R,i} = \frac{1}{60}\Big(2Q_{i-2} - 13Q_{i-1} + 47Q_i + 27Q_{i+1} - 3Q_{i+2}\Big), \tag{2.49}$$

$$q_{L,i} = \frac{1}{60}\Big(-3Q_{i-2} + 27Q_{i-1} + 47Q_i - 13Q_{i+1} + 2Q_{i+2}\Big). \tag{2.50}$$

However, we notice that this reconstruction scheme allows discontinuity of the Piecewise-Parabolic function at the control volume edges since the stencil it is not symmetric. The PPM scheme, utilizing Equations (2.47) and (2.48), will be referred to as **PPM-0**. On the other hand, the PPM scheme, employing Equations (2.49) and (2.50), will be denoted as **PPM-PL07**.

2.4.2 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) - PPM-CW84

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 (2.44) by the values $\delta_m Q_i$ given by:

$$\delta_m Q_i = \begin{cases} \max(|\delta Q_i|, 2|Q_{i+1} - Q_i|, 2|Q_i - Q_{i-1}|) \cdot \operatorname{sgn}(\delta Q_i) & \text{if } (Q_{i+1} - Q_i)(Q_i - Q_{i-1}) > 0, \\ 0 & \text{otherwise,} \end{cases} \tag{2.51}$$

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:

$$q_{L,i} \leftarrow Q_i, \quad q_{R,i} \leftarrow Q_i, \quad \text{if} \quad (Q_{R,i} - Q_i)(Q_i - Q_{L,i}) \geq 0 \tag{2.52}$$

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 (A.33) 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:

$$\begin{cases} q_{L,i} \leftarrow 3Q_i - 2q_{R,i} & \text{if } \Delta q_i \cdot q_{6,i} > (\Delta q_i)^2, \\ q_{R,i} \leftarrow 3Q_i - 2q_{L,i} & \text{if } -(\Delta q_i)^2 > \Delta q_i \cdot q_{6,i} \end{cases} \tag{2.53}$$

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. This scheme is referred to as **PPM-CW84**.

Limiter from Lin (2004) - PPM-L04

Similarly to Colella and Woodward (1984), Lin (2004) reduces numerical oscillations in the parabolas by replacing the term δQ_i in Equation (2.44) with the values $\delta_m Q_i$ given by:

$$\delta_m Q_i = \max(|\delta Q_i|, 2\delta Q_{\min,i}, 2\delta Q_{\max,i}) \cdot \operatorname{sgn}(\delta Q_i), \tag{2.54}$$

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:

$$q_{L,i} \leftarrow Q_i - \max(|\delta_m Q_i|, |q_{L,i} - Q_i|) \cdot \operatorname{sgn}(\delta_m Q_i), \tag{2.55}$$
$$q_{R,i} \leftarrow Q_i - \max(|\delta_m Q_i|, |q_{R,i} - Q_i|) \cdot \operatorname{sgn}(\delta_m Q_i). \tag{2.56}$$

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. This scheme is referred to as **PPM-L04**.

2.5 Flux

Let's consider the framework outlined in Problem 2.4. Assuming that $Q^n \in \mathbb{P}_\nu^N$ is known, our objective is to compute the values Q^{n+1}. To accomplish this, we utilize a scheme similar to the one presented in Problem 2.4, taking into account the presence of a reconstruction function $\tilde{q}(x; Q^n)$ as discussed in Section 2.4, and an initial estimation $\tilde{x}^n_{i+\frac{1}{2}} = x_{i+\frac{1}{2}} - \tilde{u}^n_{i+\frac{1}{2}} \Delta t$ for a time-averaged wind $\tilde{u}^n_{i+\frac{1}{2}}$ as explained in Section 2.3. The numerical flux function is then expressed as:

$$F^n_{i+\frac{1}{2}}(Q^n, \tilde{u}^n_{i+\frac{1}{2}}) = \frac{1}{\Delta t} \int_{x_{i+\frac{1}{2}} - \tilde{u}^n_{i+\frac{1}{2}} \Delta t}^{x_{i+\frac{1}{2}}} \tilde{q}(x; Q^n)\, dx. \tag{2.57}$$

Notice that if we define the averaged CFL number,

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

where $\tilde{c}^n_{i+\frac{1}{2}} = k + \alpha^n_{i+\frac{1}{2}}$, $k = \lfloor \tilde{c}^n_{i+\frac{1}{2}} \rfloor$, $\alpha^n_{i+\frac{1}{2}} \in [0, 1[$, we can express the numerical flux as (Y. Chen et al., 2017; Lin & Rood, 1996):

$$F^n_{i+\frac{1}{2}}(Q^n, \tilde{u}^n_{i+\frac{1}{2}}) = \frac{1}{\Delta t} \begin{cases} \Delta x \sum_{l=0}^{k-1} Q_{i-l} + \int_{x_{i-k+\frac{1}{2}} - \alpha^n_{i+\frac{1}{2}} \Delta x}^{x_{i-k+\frac{1}{2}}} \tilde{q}(x; Q^n)\, dx, & \text{if } \tilde{u}^n_{i+\frac{1}{2}} \geq 0, \\ \Delta x \sum_{l=0}^{k-1} Q_{i-l} - \int_{x_{i-k+\frac{1}{2}}}^{x_{i-k+\frac{1}{2}} - \alpha^n_{i+\frac{1}{2}} \Delta x} \tilde{q}(x; Q^n)\, dx, & \text{if } \tilde{u}^n_{i+\frac{1}{2}} < 0. \end{cases} \tag{2.58}$$

where we used that $\tilde{q}$ preserves the local mass.

We will provide explicit expressions for the integrals in Equation (2.58) when using the PPM method. For each control volume edge, denoted by $i = 0, \ldots, N$, and $y > 0$, we define the following averages of the Piecewise-Parabolic approximation, as defined in Equation (2.36) for Q^n (Colella & Woodward, 1984):

$$F_{L,i+\frac{1}{2}}(Q^n, y) = \frac{1}{y} \int_{x_{i+\frac{1}{2}} - y}^{x_{i+\frac{1}{2}}} \tilde{q}(x; Q^n)\, dx, \tag{2.59}$$

and

$$F_{R,i+\frac{1}{2}}(Q^n, y) = \frac{1}{y} \int_{x_{i+\frac{1}{2}}}^{x_{i+\frac{1}{2}} + y} \tilde{q}(x; Q^n)\, dx. \tag{2.60}$$

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 (2.40) that:

$$F_{L,i+\frac{1}{2}}(Q^n, y) = \frac{1}{y}\int_{x_{i+\frac{1}{2}}-y}^{x_{i+\frac{1}{2}}} q_i(x; Q^n)\,dx = q_{R,i} + \frac{(q_{6,i} - \Delta q_i)}{2\Delta x}y - \frac{q_{6,i}}{3\Delta x^2}y^2, \tag{2.61}$$

and

$$F_{R,i+\frac{1}{2}}(Q^n, y) = \frac{1}{y}\int_{x_{i+\frac{1}{2}}}^{x_{i+\frac{1}{2}}+y} q_{i+1}(x; Q^n)\,dx = q_{L,i+1} + \frac{(q_{6,i+1} + \Delta q_{i+1})}{2\Delta x}y - \frac{q_{6,i+1}}{3\Delta x^2}y^2. \tag{2.62}$$

The numerical flux function for PPM is then defined by:

$$F^n_{i+\frac{1}{2}}(Q^n, \tilde{u}^n_{i+\frac{1}{2}}) = \frac{1}{\Delta t}\begin{cases} \Delta x \sum_{l=0}^{k-1} Q_{i-l} + \Delta x \alpha^n_{i+\frac{1}{2}} F_{L,i+\frac{1}{2}}(Q^n, \alpha^n_{i+\frac{1}{2}}\Delta x) & \text{if } \tilde{u}^n_{i+\frac{1}{2}} \geq 0, \\ \Delta x \sum_{l=0}^{k-1} Q_{i-l} + \Delta x \alpha^n_{i+\frac{1}{2}} F_{R,i+\frac{1}{2}}(Q^n, -\alpha^n_{i+\frac{1}{2}}\Delta x) & \text{if } \tilde{u}^n_{i+\frac{1}{2}} < 0, \end{cases} \tag{2.63}$$

where $\tilde{u}^n_{i+\frac{1}{2}}$ is the velocity used in the departure point estimation. In particular, if the CFL number is less than one, then:

$$F^n_{i+\frac{1}{2}}(Q^n, \tilde{u}^n_{i+\frac{1}{2}}) = \begin{cases} \tilde{u}^n_{i+\frac{1}{2}} F_{L,i+\frac{1}{2}}(Q^n, \tilde{u}^n_{i+\frac{1}{2}}\Delta t) & \text{if } \tilde{u}^n_{i+\frac{1}{2}} \geq 0, \\ \tilde{u}^n_{i+\frac{1}{2}} F_{R,i+\frac{1}{2}}(Q^n, -\tilde{u}^n_{i+\frac{1}{2}}\Delta t) & \text{if } \tilde{u}^n_{i+\frac{1}{2}} < 0, \end{cases} \tag{2.64}$$

If the monotonic scheme from Lin (2004) is employed, then $F_{i+\frac{1}{2}}$ uses the stencil $\mathcal{S}_{i+\frac{1}{2}} = \{i-3, i-2, i-1, i, i+1, i+2, i+3\}$ and we need $\nu = 4$ layers of ghost cells. Otherwise the stencil used by $F_{i+\frac{1}{2}}$ is given by $\mathcal{S}_{i+\frac{1}{2}} = \{i-2, i-1, i, i+1, i+2\}$ for all the other schemes that we presented and $\nu = 3$. The accuracy of this flux formulation is described in Section A.5.3.

2.6 Numerical experiments

This section is dedicated to presenting the numerical results of the PPM and its variations discussed here. We will consider the following reconstruction schemes: PPM-0, PPM-PL07, PPM-CW84, and PPM-L04, which were presented in Section 2.4, as well as the departure point schemes RK1 and RK2 described in Section 2.3. The code used in this section can be found in Appendix B.

For all the simulations presented here, we will consider the spatial domain $[0, 1]$ and the time interval $[0, 5]$. The relative change at time step n in the mass is computed as:

$$\frac{|M^n - M^0|}{|M_0|},$$

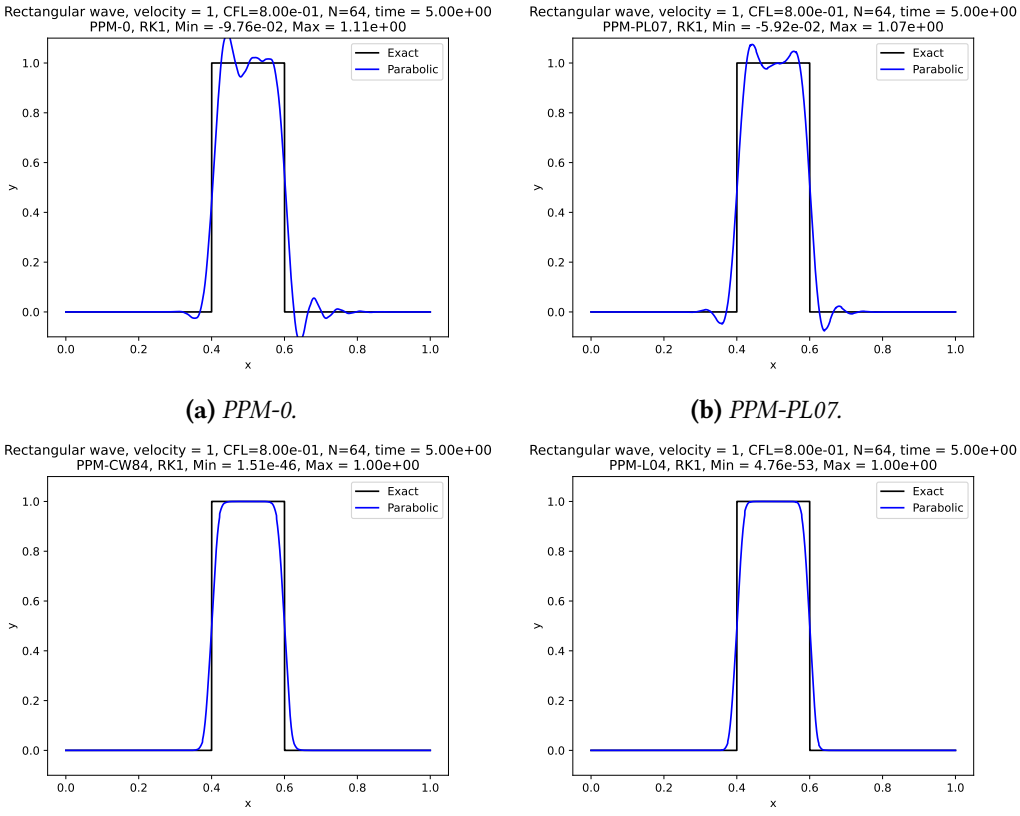
where M^n is given by Equation (2.22). For all the simulations, the mass is preserved with machine precision. Furthermore, we compute the initial average values $Q_i(0)$ using the initial values of q_i^0 at the control volume centroids for all simulations, which is second-order accurate by Proposition 2.2. In the error calculation, only when q_0 is given by Equation (2.67), we replace $Q_i(t^n)$ by its centroid value $q_i(t^n)$, which again gives a second-order

approximation by Proposition 2.2.

As a first numerical experiment, we consider a discontinuous initial condition given by:

$$q_0(x) = \begin{cases} 1 & \text{if } x \in [0.4, 0.6], \\ 0 & \text{otherwise.} \end{cases} \tag{2.65}$$

for the linear advection equation with constant velocity, which we adopt as $u = 0.2$. It is easy to check that the exact solution of Problem 2.1 is given by $q_0(x-ut)$. We will use a CFL number equal to 0.8. The spatial domain will be $[0, 1]$, and the time integration interval will be $[0, 5]$. Since we are assuming periodic boundary conditions, the period is equal to 5. Therefore, the simulations presented here will advect an initial profile for one time period. The departure schemes RK1 and RK2 compute the departure point exactly in this case, so we will only use the RK1 scheme. In Figure 2.4, we present the obtained results.



(a) *PPM-0.* **(b)** *PPM-PL07.*

(c) *PPM-CW84.* **(d)** *PPM-L04.*

Figure 2.4: *Linear advection experiment using a constant velocity equal to* 0.2*, a CFL number equal to* 0.8*,* $N = 64$ *cells, and the initial condition is given by Equation* (2.65)*. These figures show the advected profile after 5 time units (one time period). Reconstruction schemes employed: PPM-0 (a), PPM-PL07 (b), PPM-CW84 (c) and PPM-L04 (d).*

It is evident that the monotonic schemes exhibit a significant advantage. These schemes effectively prevent the strong oscillations observed in the non-monotonic schemes, as well as the generation of new extrema, which aligns with our expectations. Besides that, the scheme PPM-PL07 is less dispersive than the scheme PPM-0

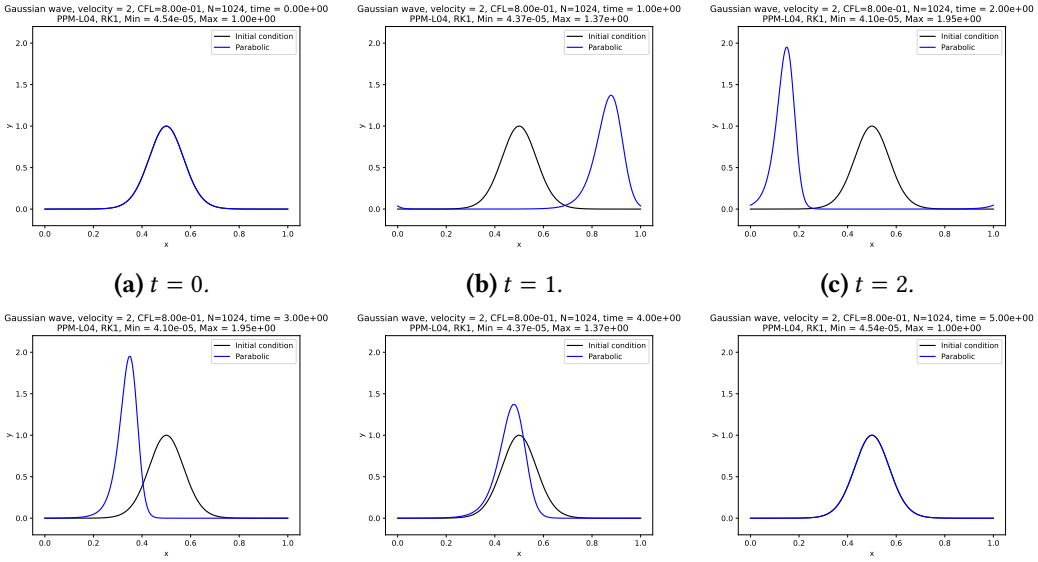
As a second experiment, we shall investigate the how the PPM schemes behave when the velocity is variable. We are going to consider the velocity

$$u(x,t) = u_0 \cos\left(\frac{\pi t}{T}\right) \sin^2\left(\pi\left(x - \frac{t}{T}\right)\right) + u_1. \tag{2.66}$$

We adopt the parameters $u_0 = u_1 = 0.2$, and $T = 5$. Following the approach in Trefethen (2000), we initialize the periodic Gaussian profile defined as:

$$q(x) = \exp(-10\cos^2(\pi x)), \quad x \in [0,1]. \tag{2.67}$$

The velocity function given by Equation (2.66) is based on the deformational flow test case in Nair and Lauritzen (2010). As the velocity is variable, we utilize the departure point schemes RK1 and RK2. In this case, the solution exhibits a period of 5 time units, meaning that the profile deforms and returns to its initial shape and position after 5 time units, allowing us to compute the error. Indeed, in Figure 2.5, we show how the solution behaves using a high-resolution ($N = 1024$), the PPM-L04 scheme and the RK1 departure point scheme.



(a) $t = 0$. **(b)** $t = 1$. **(c)** $t = 2$.

(d) $t = 3$. **(e)** $t = 4$. **(f)** $t = 5$.

Figure 2.5: *Linear advection experiment using the velocity from Equation (2.67), a CFL number equal to* 0.8, $N = 1024$ *cells, and the initial condition is given by Equation (2.67) (2.5a). These figures show the advected profile at 1 (2.5b), 2 (2.5c), 3 (2.5d), 4 (2.5e), and 5 (2.5f) time units. We are using the PPM-L04 scheme with the RK1 departure point scheme.*

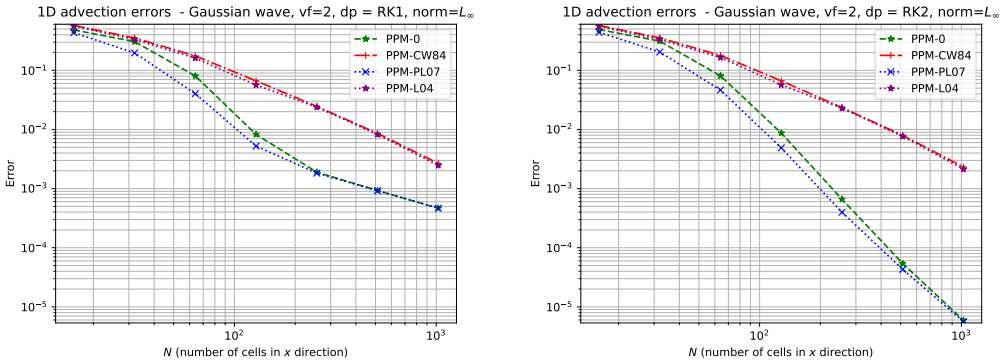
To investigate the error convergence, we employ $\Delta x^{(k)}$-grids with $\Delta x^{(k)} = 1/2^k$ for $k = 4, \ldots, 10$. To measure the accuracy, we consider the relative error in the maximum norm as follows:

$$E_k = \max_{n=0,\ldots,N_T} \frac{\|Q^n - Q(t^n)\|_{\infty,\Delta x}}{\|Q(t^n)\|_{\infty,\Delta x}}.$$

The convergence rate is defined by

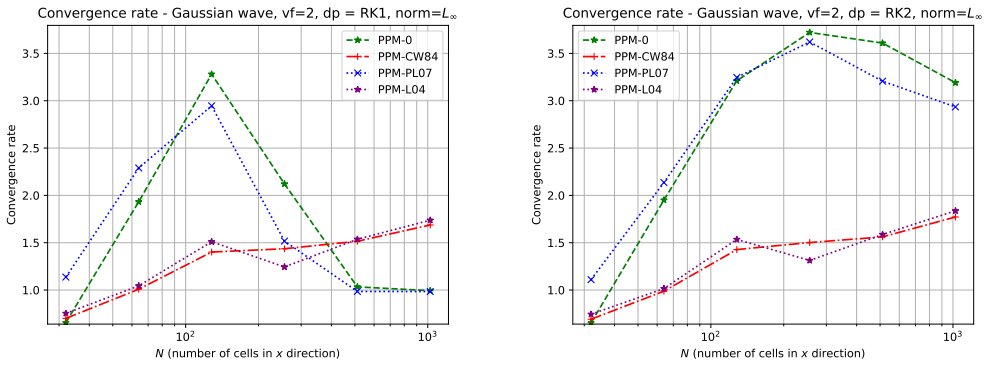
$$CR_k = \frac{\ln\left(\frac{E_k}{E_{k-1}}\right)}{\ln 2}, \quad \text{for} \quad k = 5, \ldots 10.$$

The difference between the RK1 and RK2 schemes becomes clear when observing the relative error in Figure 2.6 and the convergence rate in Figure 2.7. The RK1 scheme results in a first-order error in the departure point, which dominates the total error for all PPM schemes. This observation is in agreement with the discussion in Section 2.3. On the other hand, when employing the RK2 scheme, we can achieve a third-order accuracy for the PPM-0 and PPM-PL07 schemes, surpassing the expected second-order accuracy of the departure point scheme. This experiment illustrates the impact of the departure point calculation error on the overall error. Furthermore, regardless of the departure point scheme, the PPM-PL07 scheme exhibits a smaller error. Among the monotonic schemes, the PPM-L04 has a slightly smaller error compared to PPM-CW84.



(a) *RK1.* **(b)** *RK2.*

Figure 2.6: *Relative error for different PPM schemes using the RK1 (left) and RK2 (right) departure point scheme for the initial condition given by Equation (2.67) and the variable velocity given by Equation (2.66).*



(a) *RK1.* **(b)** *RK2.*

Figure 2.7: *As in Figure 2.6 but considering the convergence rate.*

2.7 Concluding remarks

In this chapter, we provided a general overview of 1D FV-SL schemes for the advection equation. We discussed the three essential tasks involved in these schemes. The first task is the reconstruction of a function from its average values. We employed the PPM method introduced by Colella and Woodward (1984) and its variants. We were able to achieve third-order accuracy in the reconstruction process, even without imposing monotonicity constraints. The second task involves computing the departure point of the control volume edges. For this purpose, we utilized the first-order departure point calculation method from Colella and Woodward (1984) known as RK1. Additionally, we explored a second-order approach by employing a two-stages Runge-Kutta scheme (RK2) to integrate the departure point ordinary differential equation (ODE). Lastly, the third task entails computing the flux, which involves integrating the reconstructed function over a domain determined by the departure point.

From the numerical experiments, we observed that the PPM-PL07 (Putman & Lin, 2007), which uses fifth-order reconstruction at the edges, leads to third-order accuracy but is more accurate than the PPM-0 (Colella & Woodward, 1984) scheme, which uses fourth-order reconstruction at the edges. Regarding the monotonic schemes, we observed that both schemes were able to avoid overshoots, with PPM-L04 (Lin, 2004) being more accurate than the scheme PPM-CW84 proposed by Colella and Woodward (1984).

The difference between the departure point schemes became apparent when we performed a test with variable velocity. The simulation using the RK1 scheme resulted in a final first-order error, despite the scheme having third-order accuracy in space. However, the RK2 scheme preserved third-order accuracy despite being only second-order accurate. We expect that, in general, combining PPM with the RK2 scheme should result in at least second-order accuracy.

Clearly, the RK2 scheme is more computationally expensive since it requires time extrapolation and linear interpolation of the velocity field. One possible way to reduce its cost would be to use larger CFL numbers allowed by the FV-SL schemes, as discussed in Section 2.5.

Chapter 3

Two-dimensional finite-volume methods

In Chapter 2, we addressed the problem of solving the one-dimensional linear advection equation using the finite-volume method based on PPM. In this chapter, our focus shifts to solving the two-dimensional linear advection equation using the finite-volume method. This step is crucial in our work since, as we will explore in Chapter 5, solving the linear advection equation on the cubed-sphere relies on solving two-dimensional linear advection equations at each cube face, with interpolation between adjacent panels.

A natural approach to develop a finite-volume method for the two-dimensional linear advection equation would involve extending PPM to two dimensions. Indeed, Rančić (1992) proposed a piecewise bi-parabolic extension of PPM using a semi-Lagrangian temporal discretization. However, this method suffers from a significant drawback—its computationally expensive nature. As a popular alternative, dimension-splitting methods are often used, which replace the two-dimensional problem with a sequence of one-dimensional problems. For example, we can solve the two-dimensional linear advection equation by solving a series of one-dimensional linear advection equations using the PPM from Chapter 2. Moreover, in principle, we can employ any numerical method that solves the one-dimensional linear advection equation.

A comparison between two-dimensional and dimension-splitting semi-Lagrangian schemes on a plane was investigated by Y. Chen et al. (2017), utilizing the PPM as the one-dimensional solver and distorted two-dimensional grids. Their main conclusion was that dimension-splitting schemes are more sensitive to grid distortions, but they are computationally cheaper and more accurate than two-dimensional methods, particularly when dealing with large CFL numbers.

The primary objective of this chapter is to provide a comprehensive explanation of the dimension splitting method proposed by Lin and Rood (1996). This method is currently utilized in the FV3 dynamical core and is applied to the two-dimensional linear advection equation using the one-dimensional finite-volume schemes described in Chapter 2. Similar to Chapter 2, we start this chapter with a review of the integral form of the two-dimensional advection equation in Section 3.1. In Section 3.2, we establish the framework for general

two-dimensional finite-volume schemes. The dimension splitting method is presented in Section 3.3, and we showcase numerical experiments in Section 3.4.

3.1 Two-dimensional advection equation in integral form

3.1.1 Notation

This section is dedicated to extending the notation of Section 2.1.1. Based on definitions 2.1 and 2.3, we introduce the concepts of a $(\Delta x, \Delta y)$-grid and $(\Delta x, \Delta y, \Delta t, \lambda)$ discretization. Throughout this chapter, we will use the notation $\Omega = [a, b] \times [c, d]$ and v to represent a non-negative integer indicating the number of ghost cell layers in each boundary. We also use the notations $\mathbb{R}_v^{N\times M} := \mathbb{R}^{(N+2v)\times(M+2v)}$ and $\mathbb{R}_v^{(N+1)\times M} := \mathbb{R}^{(N+1+2v)\times(M+2v)}$, $\mathbb{R}_v^{N\times(M+1)} := \mathbb{R}^{(N+2v)\times(M+1+2v)}$.

Definition 3.1 ($(\Delta x, \Delta y)$-grid). *Given Ω and positive real numbers Δx and Δy such that $\Delta x = (b-a)/N$, $\Delta y = (d-c)/M$, for positive integers N and M, we say that $\Omega_{\Delta x,\Delta y} = (\Omega_{ij})_{i=-v+1,\dots,N+v}^{j=-v+1,\dots,M+v}$ is a $(\Delta x, \Delta y)$-grid for Ω if*

$$\Omega_{ij} = [x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}] \times [y_{j-\frac{1}{2}}, y_{j+\frac{1}{2}}] = [a + (i-1)\Delta x, a + i\Delta x] \times [c + (j-1)\Delta x, c + j\Delta x],$$

$\Delta x = x_{i+\frac{1}{2}} - x_{i-\frac{1}{2}}$, $\Delta y = y_{j+\frac{1}{2}} - y_{j-\frac{1}{2}}$. Each Ω_{ij} is called control volume or cell. The cell centroids (x_i, y_j) are defined by

$$x_i = \frac{1}{2}(x_{i+\frac{1}{2}} + x_{i-\frac{1}{2}}), y_j = \frac{1}{2}(y_{j+\frac{1}{2}} + y_{j-\frac{1}{2}}).$$

Remark 3.1. *If $1 \le i \le N$,$1 \le j \le M$, we refer to (i, j) as an interior index; otherwise, (i, j) is considered a ghost cell index and we say the Ω_{ij} is a ghost cell.*

Definition 3.2 ($(\Delta x, \Delta y, \Delta t, \lambda)$-discretization). *Given $\Omega \times [0, T]$, and positive real numbers Δx Δy and Δt, we say that $(\Omega_{\Delta x,\Delta y}, T_{\Delta t})$ is a $(\Delta x, \Delta y, \Delta t, \lambda)-$discretization of $\Omega \times [0, T]$ if $\Omega_{\Delta x,\Delta y}$ is a $(\Delta x, \Delta y)$ grid for Ω and $T_{\Delta t}$ is a Δt-temporal grid for $[0, T]$, $\frac{\Delta t}{\Delta x} = \lambda$ and $\frac{\Delta t}{\Delta y} = \lambda$.*

Remark 3.2. *Whenever we mention a $(\Delta x, \Delta y)-$grid, or a $(\Delta x, \Delta y, \Delta t, \lambda)$-discretization, then Ω_{ij}, N and M are implicitly defined.*

Next, we introduce the definitions of grid functions at cell centroids and C-grid functions.

Definition 3.3 ($(\Delta x, \Delta y)$-grid function). *For a $(\Delta x, \Delta y)$-grid, we say that $Q = (Q_{ij})_{i=-v+1,\dots,N+v}^{j=-v+1,\dots,M+v} \in \mathbb{R}_v^{N\times M}$ is a $(\Delta x, \Delta y)$-grid function.*

Definition 3.4 ($(\Delta x, \Delta y)$-C grid wind). *For a $(\Delta x, \Delta y)$-grid, we say that (u, v) is a $(\Delta x, \Delta y)$-C grid wind if $u = (u_{i+\frac{1}{2},j})_{i=-v,\dots,N+v}^{j=-v+1,\dots,M+v} \in \mathbb{R}_v^{(N+1)\times M}$, $v = (v_{i,j+\frac{1}{2}})_{i=-v+1,\dots,N+v}^{j=-v,\dots,M+v} \in \mathbb{R}_v^{N\times(M+1)}$.*

Considering a function $q : \Omega \times [0, T] \to \mathbb{R}$, a vector field $\boldsymbol{u} : \Omega \times [0, T] \to \mathbb{R}$, $\boldsymbol{u} = (u, v)$, a $(\Delta x, \Delta y, \Delta t, \lambda)$-discretization of $\Omega \times [0, T]$, we introduce the grid functions $q^n \in \mathbb{R}_v^{N\times M}$, $u^n \in \mathbb{R}_v^{(N+1)\times M}$, $v^n \in \mathbb{R}_v^{N\times(M+1)}$. Here, $q_{ij}^n = q(x_i, y_j, t^n)$, $u_{i+\frac{1}{2},j}^n = u(x_{i+\frac{1}{2}}, y_j, t^n)$,

$v^n_{i,j+\frac{1}{2}} = u(x_i, y_{j+\frac{1}{2}}, t^n)$. These grid functions represent the discrete values of q and $\boldsymbol{u}$ at the cell centroids and edges, respectively, for each time level t^n (Figure 2.2). We shall also use the notations $q^n_{i+\frac{1}{2},j} = q(x_{i+\frac{1}{2}}, y_j, t^n)$ and $q^n_{i,j+\frac{1}{2}} = q(x_i, y_{j+\frac{1}{2}}, t^n)$.

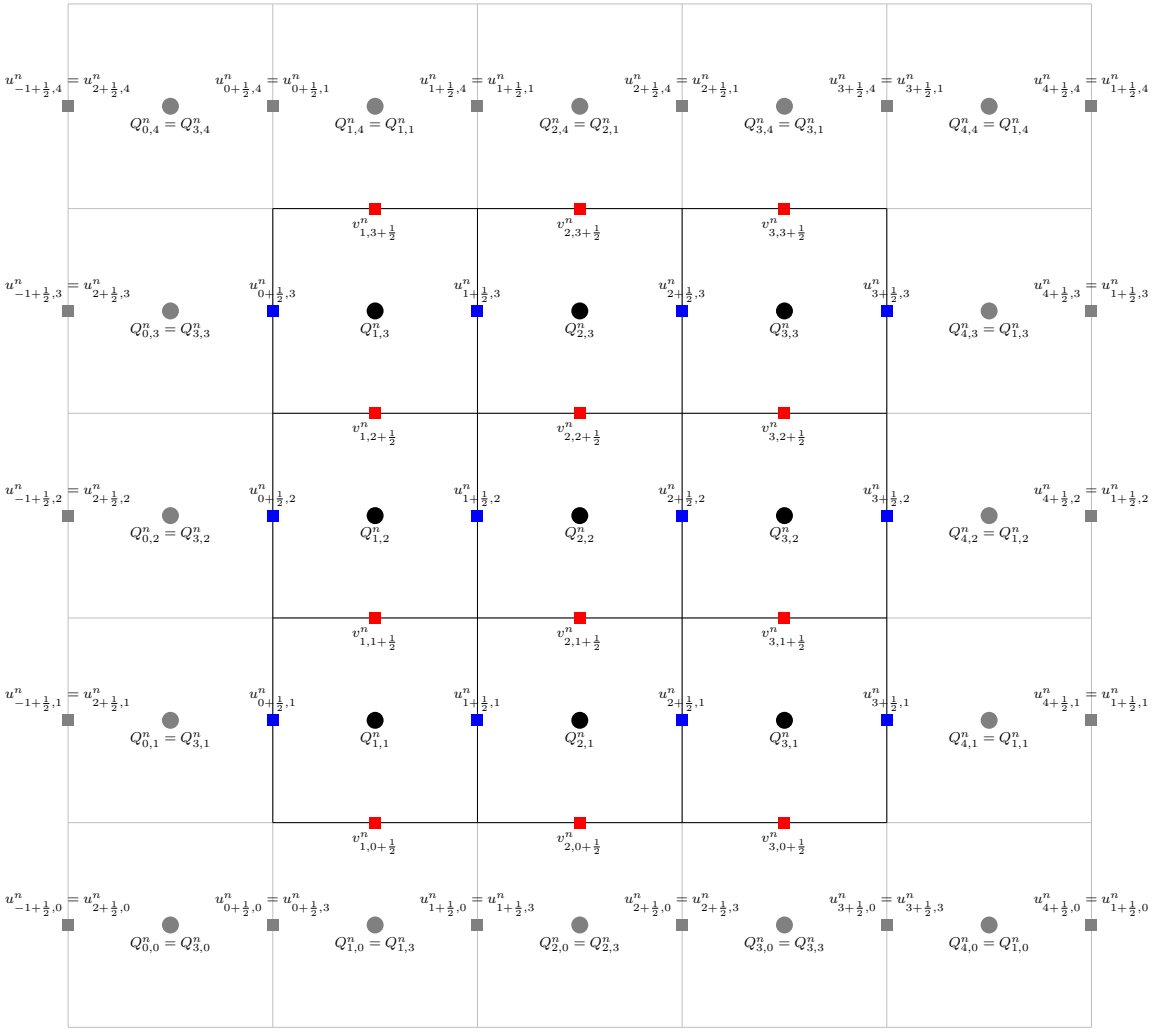


Figure 3.1: *Illustration of $(\Delta x, \Delta y)$-grid function Q (black circles) and a $(\Delta x \Delta y)$-C grid wind u (blue squares) and v (red squares) and its ghost cell values (in gray) assuming biperiodicity.*

We recall that we say the $\boldsymbol{u}$ is **non-divergent** if $\nabla \cdot \boldsymbol{u} = 0$. We denote by $\nabla \cdot (q\boldsymbol{u})$ the divergence operator:

$$\nabla \cdot (q\boldsymbol{u})(x, y, t) = [\partial_x(uq) + \partial_y(vq)](x, y, t). \tag{3.1}$$

We define the $(\Delta x, \Delta y)$-grid function δ^n as the exact divergence of $q\boldsymbol{u}$ at the cell centers, namely

$$\delta^n_{ij} = \nabla \cdot (\boldsymbol{u}q)(x_i, y_j, t^n). \tag{3.2}$$

In this chapter, our focus also lies on periodic grid functions. We define a $(\Delta x, \Delta y)$-grid

function Q as periodic if it satisfies the following conditions:

$$\begin{aligned}
Q_{i,j} &= Q_{N+i,j}, \quad i = -\nu+1,\dots,0, && j = -\nu+1,\dots,M+\nu,\\
Q_{i,j} &= Q_{i-N,j}, \quad i = N+1,\dots,N+\nu, && j = -\nu+1,\dots,M+\nu,\\
Q_{i,j} &= Q_{i,M+j}, \quad j = -\nu+1,\dots,0, && i = -\nu+1,\dots,N+\nu,\\
Q_{i,j} &= Q_{i,j-M}, \quad j = M+1,\dots,M+\nu, && i = -\nu+1,\dots,N+\nu.
\end{aligned}$$

We use the notation $\mathbb{P}_\nu^{N\times M}$ represent the spaces of periodic $(\Delta x, \Delta y)$-grid functions. Similarly, we define a $(\Delta x, \Delta y)$-grid wind (u, v) as periodic if it meets the following requirements:

$$\begin{aligned}
u_{i-\frac{1}{2},j} &= u_{N+i+\frac{1}{2},j}, \quad i = -\nu,\dots,-1, && j = -\nu+1,\dots,M+\nu,\\
u_{i+\frac{1}{2},j} &= u_{i+\frac{1}{2}-N,j}, \quad i = N+1,\dots,N+\nu, && j = -\nu+1,\dots,M+\nu,\\
u_{i+\frac{1}{2},j} &= u_{i+\frac{1}{2},M+j}, \quad i = -\nu,\dots,N+1+\nu, && j = -\nu+1,\dots,0,\\
u_{i+\frac{1}{2},j} &= u_{i+\frac{1}{2},j-M}, \quad i = -\nu,\dots,N+1+\nu, && j = M+1,\dots,M+\nu,\\
v_{i,j-\frac{1}{2}} &= v_{i,M+j+\frac{1}{2}}, \quad j = -\nu,\dots,-1, && i = -\nu+1,\dots,N+\nu,\\
v_{i,j+\frac{1}{2}} &= v_{i,j+\frac{1}{2}-M}, \quad j = M+1,\dots,M+\nu, && i = -\nu+1,\dots,N+\nu,\\
v_{i,j+\frac{1}{2}} &= v_{N+i,j+\frac{1}{2}}, \quad j = -\nu,\dots,M+1+\nu, && i = -\nu+1,\dots,0,\\
v_{i,j+\frac{1}{2}} &= c_{i-N,j+\frac{1}{2}}, \quad j = -\nu,\dots,N+1+\nu, && i = N+1,\dots,N+\nu.
\end{aligned}$$

In this case, we use the notation $u \in \mathbb{P}_\nu^{(N+1)\times M}$, $v \in \mathbb{P}_\nu^{N\times(M+1)}$.

For a grid function Q we also use the notations:

$$\begin{aligned}
Q_{\times,j} &:= (Q_{-\nu+1,j},\dots,Q_{N+\nu,j}) \in \mathbb{R}_\nu^N,\\
Q_{i,\times} &:= (Q_{i,-\nu+1},\dots,Q_{i,M+\nu}) \in \mathbb{R}_\nu^M.
\end{aligned}$$

Given $Q = (Q_{ij}) \in \mathbb{P}_{\nu,P}^{N\times M}$, we define the p-norm by

$$\|Q\|_{p,\Delta x\times\Delta y} = \begin{cases} \left(\sum_{i=1}^N \sum_{j=1}^M |Q_{ij}|^p\right)^{\frac{1}{p}} & \text{if } 1 \le p < \infty,\\ \max_{i=1,\dots,N,j=1,\dots,M} |Q_{ij}| & \text{otherwise .}\end{cases} \tag{3.3}$$

We also introduce the centered difference notation:

$$\delta_x h(x_i, y, t) = h(x_{i+\frac{1}{2}}, y, t) - h(x_{i-\frac{1}{2}}, y, t), \tag{3.4}$$

$$\delta_y h(x, y_j, t) = h(x, y_{j+\frac{1}{2}}, t) - h(x, y_{j-\frac{1}{2}}, t), \tag{3.5}$$

for any function $h : \Omega \times [0, T] \to \mathbb{R}$. Additionally, we introduce the average value of q in the control volume Ω_{ij} at time t, denoted as $Q_{ij}(t)$, defined by:

$$Q_{ij}(t) = \frac{1}{\Delta x \Delta y} \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} \int_{y_{j-\frac{1}{2}}}^{y_{j-\frac{1}{2}}} q(x, y, t)\, dx. \tag{3.6}$$

Moreover, we define the $(\Delta x, \Delta y)$-grid function of average values as $Q(t) = (Q_{ij}(t))_{i=-\nu+1,\dots,N+\nu}^{j=-\nu+1,\dots,M+\nu}$.

For the consideration of periodic boundary conditions, we can define spaces of periodic functions over the interval Ω as follows:

$$S_P(\Omega) = \{q : \mathbb{R}^2 \times [0, +\infty[\to \mathbb{R} : q(x + b - a, y + d - c, t) = q(x, y, t), \quad \forall x, y \in \mathbb{R}, \quad t \geq 0\}.$$

Similarly, the space of k-times periodically differentiable functions $C_P^k(\Omega)$ can be defined as:

$$C_P^k(\Omega) = S_P(\Omega) \cap C^k(\mathbb{R}^2 \times [0, \infty[),$$

where $C^k(\mathbb{R}^2 \times [0, +\infty[)$ denotes the space of functions that are k times continuously differentiable in both the spatial and temporal variables. In summary, $S_P(\Omega)$ represents the space of periodic functions, and $C_P^k(\Omega)$ represents the space of k-times periodically differentiable functions over Ω subject to periodic boundary conditions.

3.1.2 The 2D advection equation

Let us consider a velocity field given by $\boldsymbol{u} = (u, v)$, where u is the velocity in x-direction and v is the velocity in x and y direction and $u, v \in C_P^1(\Omega)$. The two-dimensional advection equation in its differential form in a domain Ω associated to the velocity field $\boldsymbol{u}$ and assuming biperiodic boundary conditions is given by:

$$\begin{cases} [\partial_t q + \partial_x(uq) + \partial_y(vq)](x, y, t) = 0, \quad \forall (x, y, t) \in \mathbb{R}^2 \times]0, +\infty[, \\ q(a, y, t) = q(b, y, t), \quad \forall y \in [c, d], \quad \forall t \geq 0, \\ q(x, c, t) = q(x, d, t), \quad \forall x \in [a, b], \quad \forall t \geq 0, \\ q_0(x) = q(x, y, 0), \quad \forall (x, y) \in \Omega. \end{cases} \tag{3.7}$$

A classical or strong solution to the two-dimensional advection equation is a $C_P^1(\Omega)$ function q satisfying Equation (3.7). As we did in Section 2.1, our goal is to deduce an integral form of Equation (3.7). Thus, let us consider $[x_1, x_2] \times [y_1, y_2] \subset \Omega$ and $[t_1, t_2] \subset [0, +\infty[$. Integrating Equation (3.7) over $[x_1, x_2] \times [y_1, y_2]$ yields:

$$\begin{aligned} \frac{d}{dt}\left(\int_{x_1}^{x_2} \int_{y_1}^{y_2} q(x, y, t)\, dx\, dy \right) = &- \int_{y_1}^{y_2} \Big((uq)(x_2, y, t) - (uq)(x_1, y, t) \Big)\, dy \\ &- \int_{x_1}^{x_2} \Big((vq)(x, y_2, t) - (vq)(x, y_1, t) \Big)\, dx. \end{aligned} \tag{3.8}$$

Integrating Equation (3.8) over the time interval $[t_1, t_2]$, we have:

$$\begin{aligned}\int_{x_1}^{x_2}\int_{y_1}^{y_2} q(x,y,t_{n+1})\,dx\,dy = &\int_{x_1}^{x_2}\int_{y_1}^{y_2} q(x,y,t_n)\,dx\,dy \\ &- \int_{t_1}^{t_2}\int_{y_1}^{y_2}\Big((uq)(x_2,y,t) - (uq)(x_1,y,t)\Big)\,dy\,dt \\ &- \int_{t_1}^{t_2}\int_{x_1}^{x_2}\Big((vq)(x,y_2,t) - (vq)(x,y_1,t)\Big)\,dx\,dt.\end{aligned} \tag{3.9}$$

Equation (3.9) is the integral form of Equation (3.7). We say that q is a weak solution to the advection equation (3.7) if q satisfies the integral form (3.9), $\forall [x_1, x_2] \times [y_1, y_2] \subset \Omega^o$ and $\forall [t_1, t_2] \subset [0, +\infty[$. We summarize the weak version of Equation (3.7) in Problem (3.1).

Problem 3.1. *Given an initial condition q_0 and a velocity function $\boldsymbol{u} = (u, v)$ we would like to find a weak solution q of the two-dimensional advection equation in its integral form:*

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

$\forall [x_1, x_2] \times [y_1, y_2] \times [t_1, t_2] \subset \Omega \times [0, T]$, *and* $q(x, y, 0) = q_0(x, y)$, $\forall (x, y) \in \Omega$, $q(a, y, t) = q(b, y, t)$, $\quad \forall y \in [c, d], \quad \forall t \geq 0$, $q(x, c, t) = q(x, d, t), \quad \forall x \in [a, b], \quad \forall t \geq 0$.

Similarly to Section 2.1, Equation (3.7) and Problem (3.1). when $q \in C_P^1(\Omega)$ function. For Problem 3.1, the total mass in Ω is defined by:

$$M_\Omega(t) = \int_\Omega q(x,y,t)\,dx\,dy, \quad \forall t \in [0,T], \tag{3.10}$$

and is conserved within time:

$$M_\Omega(t) = M_\Omega(0), \quad \forall t \in [0,T]. \tag{3.11}$$

Considering a $(\Delta x, \Delta y, \Delta t, \lambda)$ discretization of $D = \Omega \times [0, T]$ and 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 (3.9), we obtain:

$$\begin{aligned}Q_{ij}(t_{n+1}) = Q_{ij}(t_n) &- \frac{\Delta t}{\Delta x \Delta y}\delta_x\left(\frac{1}{\Delta t}\int_{t_1}^{t_2}\int_{y_{j-\frac{1}{2}}}^{y_{j+\frac{1}{2}}}(uq)(x_i,y,t)\,dy\,dt\right) \\ &- \frac{\Delta t}{\Delta x \Delta y}\delta_y\left(\frac{1}{\Delta t}\int_{t_1}^{t_2}\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}}(vq)(x,y_j,t)\,dx\,dt\right),\end{aligned} \tag{3.12}$$

where we are using the centered finite-difference notation. Now we can define a discretized version of Problem 3.1 as Problem 3.2.

Problem 3.2. *Assume the framework of Problem 3.1 and consider a $(\Delta x, \Delta y, \Delta t, \lambda)$-discretization of $\Omega \times [0,T]$. Since we are in the framework of Problem 3.1, it follows that:*

$$Q_{ij}(t_{n+1}) = Q_{ij}(t_n) - \lambda\delta_x\left(\frac{1}{\Delta t\Delta y}\int_{t^n}^{t^{n+1}}\int_{y_{j-\frac{1}{2}}}^{y_{j+\frac{1}{2}}}(uq)(x_i,y,t)\,dy\,dt\right) - \lambda\delta_y\left(\frac{1}{\Delta t\Delta x}\int_{t^n}^{t^{n+1}}\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}}(vq)(x,y_j,t)\,dx\,dt\right),$$

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 $Q_{ij}(t_n)$, $\forall i = 1,\dots,N$, $\forall j = 1,\dots,M$, $\forall n = 0,\dots,N_T-1$, given the initial values $Q_{ij}(0)$, $\forall i = 1,\dots N$, $\forall j = 1,\dots,M$. In other words, we aim to find the average values of q in each control volume Ω_{ij} at the specified time instances.

It is important to note that no approximations have been made in Problems (3.1) and (3.2).

3.2 The finite-volume approach

Finally, we define the 2D-FV scheme problem as follows in Problem 3.3.

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:*

$$Q_{ij}^{n+1} = Q_{ij}^n - \lambda\delta_i F_{ij}^n - \lambda\delta_j G_{ij}^n, \qquad \forall i = 1,\dots,N, \quad \forall j = 1,\dots,M, \quad \forall n = 0,\dots,N_T-1, \tag{3.13}$$

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 $Q^n \in \mathbb{P}_v^{N\times M}$ is intended to be an approximation of $Q(t_n) \in \mathbb{P}_v^{N\times M}$ in some sense. We define $Q_{ij}^0 = Q_{ij}(0)$ or $Q_{ij}^0 = q_{i,j}^0$.

The term $F_{i+\frac{1}{2},j}^n$ is known as numerical flux in the x direction and it approximates $\frac{1}{\Delta t\Delta y}\int_{t_n}^{t_{n+1}}\int_{y_{j-\frac{1}{2}}}^{y_{j+\frac{1}{2}}}(uq)(x_{i+\frac{1}{2}},y,t)\,dy\,dt$, $\forall i = 0,1,\dots,N$, and $G_{i,j+\frac{1}{2}}^n$ is known as numerical flux in the y direction and it approximates $\frac{1}{\Delta t\Delta x}\int_{t_n}^{t_{n+1}}\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}}(vq)(x,y_{j+\frac{1}{2}},t)\,dx\,dt$, $\forall j = 0,1,\dots,M$, or, in other words, they estimate the time-averaged fluxes at the control volume Ω_{ij} boundaries.

Remark 3.3. *For Problem 3.3, we define the CFL number in the x and y direction by $\max\{|u_{i+\frac{1}{2},j}^n|\}$ and $\max\{|v_{i,j+\frac{1}{2}}^n|\}$, respectively. The CFL number is maximum between these numbers and we say that the CFL condition is satisfied if the CFL number is less than one.*

For a 2D-FV the discrete total mass at the time-step n is given by

$$M^n = \Delta x\Delta y\sum_{i=1}^{N}\sum_{j=1}^{M} Q_{ij}^n.$$

Therefore, the discrete total mass is constant for a 2D-FV scheme, which follows from a

straightforward computation:

$$M^{n+1} = \Delta x \sum_{i=1}^{N}\sum_{j=1}^{M} Q_{ij}^{n+1} = M^n - \Delta t \sum_{i=1}^{N}\sum_{j=1}^{M}(F^n_{i+\frac{1}{2},j} - F^n_{i-\frac{1}{2},j}) - \Delta t \sum_{i=1}^{N}\sum_{j=1}^{M}(G^n_{i,j+\frac{1}{2}} - G^n_{i,j-\frac{1}{2}})$$
$$= M^n - \Delta t \sum_{j=1}^{M}(F^n_{N+\frac{1}{2},j} - F^n_{\frac{1}{2},j}) - \Delta t \sum_{i=1}^{N}(G^n_{i,M+\frac{1}{2}} - G^n_{i,\frac{1}{2}}) = M^n,$$

where we are using that $F^n_{N+\frac{1}{2},j} = F^n_{\frac{1}{2},j}$, $G^n_{i,M+\frac{1}{2}} = G^n_{i,\frac{1}{2}}$ since we are assuming bi-periodic boundary conditions.

As we mentioned in Problem 3.3, the initial condition may be assumed as q^0_{ij} or $Q_{ij}(0)$. For two-dimensional simulations, we are going to assume q^0_{ij} as initial data to avoid the computation of integrals. Furthermore, the errors will be calculated using the values q^n_{ij} instead of $Q_{ij}(t_n)$. Similarly to Proposition 2.2, we have that the centroid value approximates the average value with second order, as Proposition 3.1 shows.

Proposition 3.1. *If $q \in C^2$, then $|Q_{ij}(t^n) - q^n_{ij}| \leq C_1\Delta x^2 + C_2\Delta x\Delta y + C_3\Delta y^2$, where C_1, C_2 and C_3 are constants.*

Proof. Just apply Theorem A.4 for the function $q(x, y, t^n)$. □

In order to check the consistency of 2D-FV, it is useful to use the notion of discrete divergence.

Definition 3.5 (Discrete divergence)**.** *For Problem 3.3, we define the discrete divergence as a $(\Delta x, \Delta y)$-grid function $\mathbb{D}^n(Q^n, u^n, v^n) \in \mathbb{P}_v^{N\times M}$ given by:*

$$\mathbb{D}^n_{ij}(Q^n, u^n, v^n) = \frac{1}{\Delta t}\left(\frac{\delta_i F^n_{ij}}{\Delta x} + \frac{\delta_j G^n_{ij}}{\Delta y}\right), \quad i = 1, \dots, N, \quad j = 1, \dots, M. \tag{3.14}$$

With the aid of the discrete divergence, we may rewrite Equation (3.13) as:

$$Q^{n+1} = Q^n - \Delta t\mathbb{D}^n(Q^n, u^n, v^n), \tag{3.15}$$

Notice that if we replace Q^n by the exact solution $Q(t^n)$ in Equation (3.15), we have

$$Q(t^{n+1}) = Q(t^n) - \Delta t\mathbb{D}^n(Q(t^n), u^n, v^n) - \Delta t\tau^n, \tag{3.16}$$

where $\tau^n \in \mathbb{P}_v^{N\times M}$ is the local truncation error (LTE). Rearranging the terms of Equation (3.16), we obtain:

$$\tau^n = \frac{Q(t^{n+1}) - Q(t^n)}{\Delta t} - \mathbb{D}^n(Q(t^n), u^n, v^n). \tag{3.17}$$

We define the consistency of the 2D-FV scheme as follows.

Definition 3.6 (Consistency)**.** *Let us consider the framework of Problem 3.3. A 2D-FV scheme is said to be consist in the p-norm if for any sequence of $(\Delta x^{(k)}, \Delta y^{(k)}, \Delta t^{(k)}, \lambda)$-discretizations,*

$k \in \mathbb{N}$, *with* $\lim_{k\to\infty} \Delta x^{(k)} = \lim_{k\to\infty} \Delta y^{(k)} = \lim_{k\to\infty} \Delta t^{(k)} = 0$, *we have:*

$$\lim_{k\to\infty} \left[\max_{1\le n\le N_T^{(k)}} \|\tau^n\|_{p,\Delta x^{(k)}\times\Delta y^{(k)}} \right] = 0,$$

and it is said to be consistent with order d *in the* $p-$*norm if*

$$\max_{1\le n\le N_T^{(k)}} \|\tau^n\|_{p,\Delta x^{(k)}\times\Delta y^{(k)}} = O(\Delta x^d).$$

The relationship between consistency and convergence is explained in Section A.6. If q satisfies Equation (3.7), it can be observed that consistency is equivalent to the following:

$$\max_{1\le n\le N_T^{(k)}} \|\delta^n - \mathbb{D}^n(Q^n, u^n, v^n)\|_{p,\Delta x^{(k)}\times\Delta y^{(k)}} = O(\Delta x^d),$$

where $\delta^n \in \mathbb{P}_v^{N\times M}$ is defined in Equation (3.2). Therefore, we can determine whether a 2D-FV scheme is consistent by comparing the discrete divergence to the exact divergence.

3.3 Dimension splitting

This section aims to demonstrate how a 2D-FV scheme, such as the one presented in Problem 3.3, can be constructed using 1D-FV schemes through a technique known as dimension splitting. Before introducing the dimension splitting scheme proposed by Lin and Rood (1996), it is helpful to examine general operator splitting schemes, as the dimension splitting technique is a specific instance of operator splitting methods.

For a given time interval $[0, T]$, we utilize a Δt-temporal grid. Let us consider the abstract Cauchy problem.

$$\begin{cases} \frac{dq}{dt}(t) &= Aq(t), \quad t \in [t^n, t^{n+1}], \\ q(t^n) &= q_n, \end{cases}$$

for $n = 0, \dots, N_T - 1$, where $q(t) \in \mathcal{B}$ for some Banach space $\mathcal{B}$, and $A : \mathcal{B} \to \mathcal{B}$ is a linear operator following the framework of Richtmyer and Morton (1968, Chapter 3). We are interested in finding $q(t^{n+1})$ given q_n. Assuming that $A = A_1 + A_2$ for two linear operators $A_1, A_2 : \mathcal{B} \to \mathcal{B}$, we consider the following abstract Cauchy sub-problems:

$$\begin{cases} \frac{dq^1}{dt}(t) &= A_1 q(t), \quad t \in [t^n, t^{n+1}], \\ q^1(t^n) &= q_n, \end{cases}$$

and

$$\begin{cases} \frac{dq^{21}}{dt}(t) &= A_2 q(t), \quad t \in [t^n, t^{n+1}], \\ q^{21}(t^n) &= q^1(t^{n+1}). \end{cases}$$

Then we can approximate $q(t_0 + \Delta t)$ as $q^{21}(t^n + \Delta t)$ with an error of $O(\Delta t)$ if A_1 and A_2

do not commute. Otherwise, this method is exact. This approach is known as Lie-Trotter splitting. It's worth noting that the Lie-Trotter splitting can also be performed in reverse order when solving the sub-problems:

$$\begin{cases} \frac{dq^2}{dt}(t) & = A_2 q(t), \quad t \in [t^n, t^{n+1}], \\ q^2(t^n) & = q_n, \end{cases}$$

and

$$\begin{cases} \frac{dq^{21}}{dt}(t) & = A_1 q(t), \quad t \in [t^n, t^{n+1}], \\ q^{12}(t^n) & = q^1(t^{n+1}), \end{cases}$$

and again we estimate $q(t^{n+1})$ by $q^{12}(t^{n+1})$ with error $O(\Delta t)$. As noted by Strang (1968), we can consider the following equation to approximate $q(t^{n+1})$ using a second-order ($O(\Delta t^2)$) symmetric scheme:

$$q^*(t^{n+1}) = \frac{q^{21}(t^{n+1}) + q^{12}(t^{n+1})}{2}, \tag{3.18}$$

This scheme is referred to as the average Lie-Trotter splitting (Holden et al., 2010). The process of averaging two Lie-Trotter splittings is a specific case of methods known as weighted sequential splitting methods in the literature. Furthermore, this scheme averaging process can be extended to achieve higher-order schemes (Jia & Li, 2011). For an analysis of the accuracy of weighted sequential splitting methods, we recommend referring to Csomós et al. (2005).

It is worth noting that one of the most commonly used second-order splitting schemes in the literature is the Strang splitting (Strang, 1968). This scheme requires solving three sub-problems per time-step, with one of them at time $t_n + \frac{\Delta t}{2}$. In contrast, the average Lie-Trotter splitting requires solving four sub-problems per time-step. Consequently, the Strang splitting is computationally more efficient. However, as we will observe in this chapter, when applied to the linear advection equation, the average Lie-Trotter splitting allows for a modification that eliminates a splitting error arising from considering a constant scalar field and non-divergent velocity (Lin & Rood, 1996).

To move towards the scheme from (Lin & Rood, 1996), let us consider Problem 3.1 in its differential form (Equation (3.7)). We are going to consider $N + 2\nu$ one-dimensional advection equations in the x-direction:

$$[\partial_t q^x + \partial_x (u q^x)](x, y_j, t) = 0,$$

and the $N + 2\nu$ one-dimensional advection equations in the y-direction

$$[\partial_t q^y + \partial_y (v q^y)](x_i, y, t) = 0.$$

We shall assume that these problems are solved using a 1D-FV scheme as in Problem 2.4 with numerical flux functions $F^n_{i+\frac{1}{2},j}(Q^n_{\times,j}, \tilde{u}^n_{i+\frac{1}{2},j})$ and $G^n_{i,j+\frac{1}{2}}(Q^n_{i,\times}, \tilde{v}^n_{i,j+\frac{1}{2}})$, respectively where $\tilde{u}^n_{i+\frac{1}{2},j}$ is the time-averaged used in the departure point estimation in the x direction and $\tilde{v}^n_{i,j+\frac{1}{2}}$) is the time-averaged used in the departure point estimation in the y direction. In this

work, we assume that the fluxes $F^n_{i+\frac{1}{2},j}(Q^n_{\times,j}, \tilde{u}^n_{i+\frac{1}{2},j})$ and $G^n_{i,j+\frac{1}{2}}(Q^n_{i,\times}, \tilde{v}^n_{i,j+\frac{1}{2}})$ are computed using the PPM flux assuming that the CFL number is less than one (see Equation (2.64)).

More explicitily, as in Section 2.4.1, we have a piecewise-parabolic approximation in the x direction:

$$\begin{cases} q^x_{ij}(x; Q^n_{\times,j}) = q^x_{L,i,j} + \Delta q^x_{ij} z_i(x) + q^x_{6,i,j} z_i(x)(1 - z_i(x)), \\ z_i(x) = \frac{x - x_{i-\frac{1}{2}}}{\Delta x}, \quad x \in X_i, \\ q^x_{L,i,j} = q^n_{i-\frac{i}{2},j} + O(\Delta x^2), \\ q^x_{R,i,j} = q^n_{i+\frac{i}{2},j} + O(\Delta x^2), \\ \Delta q^x_{ij} = q^x_{R,i,j} - q^x_{L,i,j}, \\ q^x_{6,i,j} = 6\left(Q^n_{ij} - \frac{(q^x_{L,i,j} + q^x_{R,i,j})}{2} \right), \end{cases} \tag{3.19}$$

for $i = 1, \dots, N$, $j = -\nu + 1, \dots, M + \nu$, and we also construct a piecewise-parabolic approximation in the y direction:

$$\begin{cases} q^y_{ij}(y; Q^n_{i,\times}) = q^y_{L,i,j} + \Delta q^y_{ij} z_j(y) + q^y_{6,i,j} z_j(y)(1 - z_j(y)), \\ z_j(y) = \frac{y - y_{j-\frac{1}{2}}}{\Delta y}, \quad y \in Y_j, \\ q^y_{L,i,j} = q^n_{i,j-\frac{1}{2}} + O(\Delta y^2), \\ q^y_{R,i,j} = q^n_{i,j+\frac{1}{2}} + O(\Delta y^2), \\ \Delta q^y_{ij} = q^y_{R,i,j} - q^y_{L,i,j}, \\ q^y_{6,i,j} = 6\left(Q^n_{ij} - \frac{(q^y_{L,i,j} + q^y_{R,i,j})}{2} \right), \end{cases} \tag{3.20}$$

for $i = -\nu+1, \dots, N+\nu$, $j = 1, \dots, M$. The values $q^x_{L,i,j}$, $q^x_{R,i,j}$, $q^y_{L,i,j}$, and $q^y_{R,i,j}$, which approximate the values of q at C-grid wind positions, are computed using one of the schemes PPM-0, PPM-PL07, PPM-CW84, or PPM-L04, as described in Sections 2.4.1 and 2.4.2. These approximations are expected to be second-order accurate because the given average values are computed on the 2D control volume Ω_{ij} instead of the 1D control volumes X_i or Y_j. Then, we may express the fluxes as in Equation (2.64), namely:

$$F^n_{i+\frac{1}{2},j}(Q^n_{\times,j}, \tilde{u}^n_{i+\frac{1}{2},j}) = \tilde{u}^n_{i+\frac{1}{2},j} \times \begin{cases} q^x_{R,i,j} + \frac{1}{2}(q^x_{6,i,j} - \Delta q^x_{ij})\tilde{c}^{x,n}_{i+\frac{1}{2},j} + \frac{1}{3} q^x_{6,i,j} (\tilde{c}^{x,n}_{i+\frac{1}{2},j})^2, & \text{if} \quad \tilde{u}^n_{i+\frac{1}{2},j} > 0, \\ q^x_{L,i+1,j} - \frac{1}{2}(q^x_{6,i+1,j} + \Delta q^x_{i+1,j})\tilde{c}^{x,n}_{i+\frac{1}{2},j} - \frac{1}{3} q^x_{6,i+1,j} (\tilde{c}^{x,n}_{i+\frac{1}{2},j})^2, & \text{if} \quad \tilde{u}^n_{i+\frac{1}{2},j} \le 0, \end{cases} \tag{3.21}$$

for $i = 1, \dots, N$, $j = -\nu + 1, \dots, M + \nu$, and

$$G^n_{i,j+\frac{1}{2}}(Q^n_{i,\times}, \tilde{v}^n_{i,j+\frac{1}{2}}) = \tilde{v}^n_{i,j+\frac{1}{2}} \times \begin{cases} q^y_{R,i,j} + \frac{1}{2}(q^y_{6,i,j} - \Delta q^y_{ij})\tilde{c}^{y,n}_{i,j+\frac{1}{2}} + \frac{1}{3} q^y_{6,i,j} (\tilde{c}^{y,n}_{i,j+\frac{1}{2}})^2, & \text{if} \quad \tilde{v}^n_{i+\frac{1}{2},j} > 0, \\ q^y_{L,i,j+1} - \frac{1}{2}(q^y_{6,i,j+1} + \Delta q^y_{i,j+1})\tilde{c}^{y,n}_{i,j+\frac{1}{2}} - \frac{1}{3} q^y_{6,i,j+1} (\tilde{c}^{y,n}_{i,j+\frac{1}{2}})^2, & \text{if} \quad \tilde{v}^n_{i,j+\frac{1}{2}} \le 0, \end{cases} \tag{3.22}$$

for $i = -\nu+1, \dots, N+\nu$, $j = 1, \dots, M$, and

$$\tilde{c}^{x,n}_{i+\frac{1}{2},j} = \tilde{u}^{x,n}_{i+\frac{1}{2},j}\frac{\Delta t}{\Delta x},$$
$$\tilde{c}^{y,n}_{i,j+\frac{1}{2}} = \tilde{v}^{y,n}_{i,j+\frac{1}{2}}\frac{\Delta t}{\Delta y},$$

are the time-averaged CFL numbers in the x and y directions, respectively, which are assumed to have absolute value less than one. The time-averaged winds are computed using the RK1 and RK2 schemes from Section 2.3. When we use the PPM-L04 scheme, we set $\nu = 4$; otherwise, $\nu = 3$.

We introduce the auxiliary grid functions $\mathbf{F}$ and $\mathbf{G}$, both belonging to $\mathbb{R}^{N\times M}_{\nu}$, given by:

$$\mathbf{F}_{ij}(Q^n, \tilde{u}^n) = -\lambda\left(F^n_{i+\frac{1}{2},j}(Q^n_{\times,j}, \tilde{u}^n_{i+\frac{1}{2},j}) - F^n_{i-\frac{1}{2},j}(Q^n_{\times,j}, \tilde{u}^n_{i-\frac{1}{2},j})\right),$$

for $i = 1, \dots, N$, $j = -\nu+1, \dots, M+\nu$, and

$$\mathbf{G}_{ij}(Q^n, \tilde{v}^n) = -\lambda\left(G^n_{i,j+\frac{1}{2}}(Q^n_{i,\times}, \tilde{v}^n_{i,j+\frac{1}{2}}) - G^n_{i,j-\frac{1}{2}}(Q^n_{i,\times}, \tilde{v}^n_{i,j-\frac{1}{2}})\right),$$

for $i = -\nu+1, \dots, N+\nu$ $j = 1, \dots,$, which are the numerical flux update of the 1D-FV schemes in the x and y direction, respectively. The Lie-Trotter splitting is obtained by solving the advection in the x direction

$$Q^{x,n} = Q^n + \mathbf{F}(Q^n, \tilde{u}^n),$$

for $j = \nu+1, \dots, M+\nu$ (Figure 3.2a), and then we advect in the y direction with initial data $Q^{x,n}$

$$Q^{yx,n} = Q^{x,n} + \mathbf{G}(Q^{x,n}, \tilde{v}^n),$$

for $i = -\nu+1, \dots, N+\nu$ (Figure 3.2b). To get the average Lie-Trotter splitting we repeat the process in the reverse order by solving the advection equation in the y direction

$$Q^{y,n} = Q^n + \mathbf{G}(Q^n, \tilde{v}^n),$$

for $i = -\nu+1, \dots, N+\nu$ (Figure 3.3a), and then we advect in the x-direction with initial data $Q^{y,n+1}$

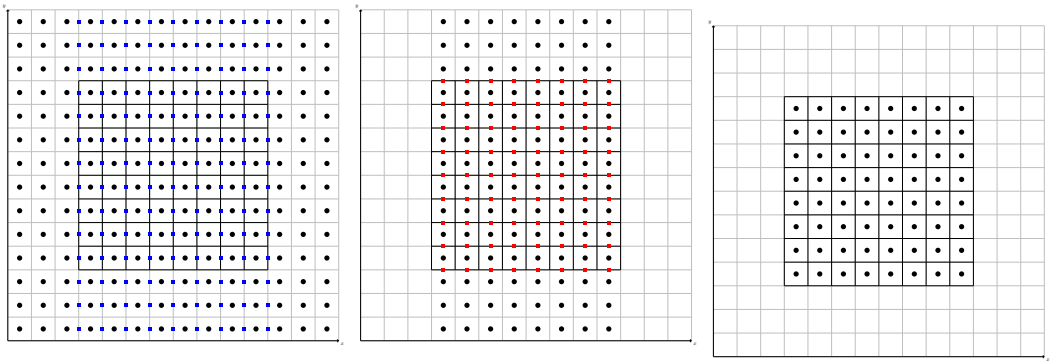
$$Q^{xy,n} = Q^{y,n} + \mathbf{F}(Q^{y,n}, \tilde{u}^n),$$

for $j = -\nu+1, \dots, M+\nu$, (Figure 3.3b) and thus we have the average Lie-Trotter solution:

$$\begin{aligned} Q^{n+1} &= \frac{(Q^{xy,n} + Q^{yx,n})}{2} \\ &= Q^n + \frac{1}{2}\mathbf{F}(Q^n, \tilde{u}^n) + \frac{1}{2}\mathbf{G}(Q^n, \tilde{v}^n) + \frac{1}{2}\mathbf{F}\left(Q^n + \frac{1}{2}\mathbf{G}(Q^n, \tilde{v}^n), \tilde{u}^n\right) + \frac{1}{2}\mathbf{G}\left(Q^n + \frac{1}{2}\mathbf{F}(Q^n, \tilde{u}^n), \tilde{v}^n\right), \end{aligned}$$

assuming that the numerical flux functions are linear in the input Q, we may rewrite a computationally cheaper version of the average Lie-Trotter splitting as (Lin & Rood, 1996):

$$Q^{n+1} = Q^n + \mathbf{F}\left(Q^n + \frac{1}{2}\mathbf{G}(Q^n, \tilde{v}^n), \tilde{u}^n\right) + \mathbf{G}\left(Q^n + \frac{1}{2}\mathbf{F}(Q^n, \tilde{u}^n), \tilde{v}^n\right). \tag{3.23}$$

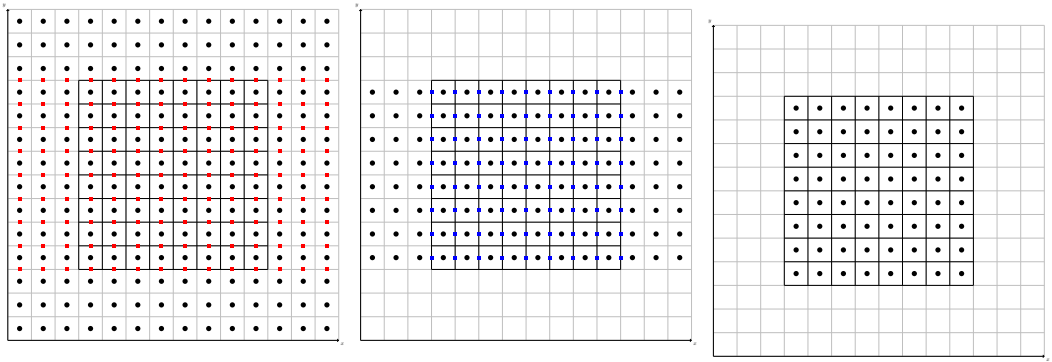


(a) Q^n *(black circles) and* u *at edges (blue squares).*

(b) $Q^{x,n}$ *(black circles) and* v *at edges (red squares).*

(c) $Q^{yx,n}$ *(black circles).*

Figure 3.2: *Illustration of the Lie-Trotter splitting applied in the* x *direction (operator* $\mathbf{F}$*) and then in the* y *direction (operator* $\mathbf{G}$*). Interior cells are depicted using black lines, while ghost cells are depicted using gray lines. All the winds shown are the ones used in the RK1 departure point scheme. If the RK2 scheme is used, an additional layer of wind ghost values should be added at each boundary in (a) and (b).*



(a) Q^n *(black circles) and* v *at edges (red squares).*

(b) $Q^{y,n}$ *(black circles) and* u *at edges (blue squares).*

(c) $Q^{xy,n}$ *(black circles).*

Figure 3.3: *Similar to Figure 3.2 but considering the Lie-Trotter splitting in reverse order.*

The numerical flux functions defined in Chapter 2 are indeed linear the input Q if there are no monotonic constrain, but we are going to consider this scheme even when there are monotonic constraints since it requires fewer operations. Further, if we if assume that $q = \overline{q}$ is constant and $\nabla \cdot \boldsymbol{u} = 0$ then the solution remains constant and then, assuming also

that $\boldsymbol{u}$ does not depend on t, then $\mathbf{F}$ and $\mathbf{G}$ are given by

$$\mathbf{F}_{ij}(Q^n, \tilde{u}^n) = -\overline{q}\lambda\delta_x u(x_i, y_j),$$
$$\mathbf{G}_{ij}(Q^n, \tilde{v}^n) = -\overline{q}\lambda\delta_y v(x_i, y_j).$$

However, if we compute the updated solution using Equation (3.23), we have that the error is given by

$$Q_{ij}^{n+1} - \overline{q} = -\Delta t\left(\frac{\delta_x u(x_i, y_j)}{\Delta x} + \frac{\delta_y v(x_i, y_j)}{\Delta y}\right) - \Delta t^2\overline{q}\left(\frac{\delta_y v\delta_x u(x_i, y_j) + \delta_x u\delta_y v(x_i, y_j)}{2\Delta x\Delta y}\right) \tag{3.24}$$

$$= \Delta t(O(\Delta x^2) + O(\Delta y^2)) - \Delta t^2\overline{q}\left(\frac{\delta_y v\delta_x u(x_i, y_j) + \delta_x u\delta_y v(x_i, y_j)}{2\Delta x\Delta y}\right). \tag{3.25}$$

Thus, the terms in the equation above multiplied by Δt^2 are related to an splitting error, even if we consider the exact fluxes. Aiming to eliminate the error from, Lin and Rood (1996) proposes to consider a modification of the average Lie-Trotter splitting as

$$Q^{n+1} = Q^n + \mathbf{F}\left(Q^n + \frac{1}{2}\mathbf{g}(Q^n, \tilde{v}^n), \tilde{u}^n\right) + \mathbf{G}\left(Q^n + \frac{1}{2}\mathbf{f}(Q^n, \tilde{u}^n), \tilde{v}^n\right), \tag{3.26}$$

where $\mathbf{f}$ and $\mathbf{g}$ are called inner advective operators and approximate $-\Delta t u\frac{\partial q}{\partial x}$ and $-\Delta t v\frac{\partial q}{\partial y}$.

In this work, we shall consider the following inner advective operator proposed by Lin (2004) (hereafter, **L04**) and the one proposed by Putman and Lin (2007) (hereafter, **PL07**). The PL07 scheme is currently used in the FV3 dynamical core. We also shall consider the average Lie-Trotter splitting (hereafter, **AVLT**). All the expressions of each inner advective operator mentioned are shown in Table 3.1. It is easy to see that both operators L04 and PL07 eliminate the term multiplied by Δt^2 that appeared in Equation (3.24) when we apply these operators for a constant grid function Q^n and a non-divergent velocity field in Equation (3.26). Therefore, these inner advective operators eliminate the splitting error for a constant field and a non-divergent velocity field.

Scheme	$\mathbf{f}_{ij}(Q^n, \tilde{u}^n)$	$\mathbf{g}_{ij}(Q^n, \tilde{v}^n)$
AVLT	$\mathbf{F}_{ij}(Q^n, \tilde{u}^n)$	$\mathbf{G}_{ij}(Q^n, \tilde{v}^n)$
L04	$\mathbf{F}_{ij}(Q^n, \tilde{u}^n) + Q_{ij}^n\frac{\Delta t}{\Delta x}(\tilde{u}^n_{i+\frac{1}{2},j} - \tilde{u}^n_{i-\frac{1}{2},j})$	$\mathbf{G}_{ij}(Q^n, \tilde{v}^n) + Q_{ij}^n\frac{\Delta t}{\Delta y}(\tilde{v}^n_{i,j+\frac{1}{2}} - \tilde{v}^n_{i,j-\frac{1}{2}})$
PL07	$\frac{1}{2}\left(-Q_{ij}^n + \frac{Q_{ij}^n + \mathbf{F}_{ij}(Q^n, \tilde{u}^n)}{1-\frac{\Delta t}{\Delta x}\left(\tilde{u}^n_{i+\frac{1}{2},j} - \tilde{u}^n_{i-\frac{1}{2},j}\right)}\right)$	$\frac{1}{2}\left(-Q_{ij}^n + \frac{Q_{ij}^n + \mathbf{G}_{ij}(Q^n, \tilde{v}^n)}{1-\frac{\Delta t}{\Delta y}\left(\tilde{v}^n_{i,j+\frac{1}{2}} - \tilde{v}^n_{i,j-\frac{1}{2}}\right)}\right)$

Table 3.1: *Expression of the inner advective operators considered in this work. AVLT stands for the average Lie-Trotter scheme, while L04 and PL07 stands for the inner advective operators from Lin (2004) and from Putman and Lin (2007), respectively.*

Recalling the definition of discrete divergence (Definition 3.5) we have:

$$\mathbb{D}^n = -\frac{1}{\Delta t}\left[\mathbf{F}\left(Q^n + \frac{1}{2}\mathbf{g}(Q^n, \tilde{v}^n), \tilde{u}^n\right) + \mathbf{G}\left(Q^n + \frac{1}{2}\mathbf{f}(Q^n, \tilde{u}^n), \tilde{v}^n\right)\right], \tag{3.27}$$

and as pointed out in Section A.6, we may use the discrete divergence to check the scheme consistency.

3.4 Numerical experiments

To assess the dimension splitting schemes introduced previously, we are going to consider the linear advection equation on $[0, 1] \times [0, 1]$ with bi-biperiodic boundary conditions. For the 1D schemes, we consider the FV-SL schemes PPM-0, PPM-PL07, PPM-CW84, PPM-L04 with the departure point schemes RK1 and RK2 as in Section 2.6. For all simulations, the CFL number is set to 0.8, and the time integration interval is $[0, 5]$, which represents the period of the exact solution for all tests. We employ $(\Delta x^{(k)}, \Delta y^{(k)})$-grids with $\Delta x^{(k)} = \Delta y^{(k)} = 1/2^k$ for $k = 4, \ldots, 10$. We introduce the relative error in the maximum norm:

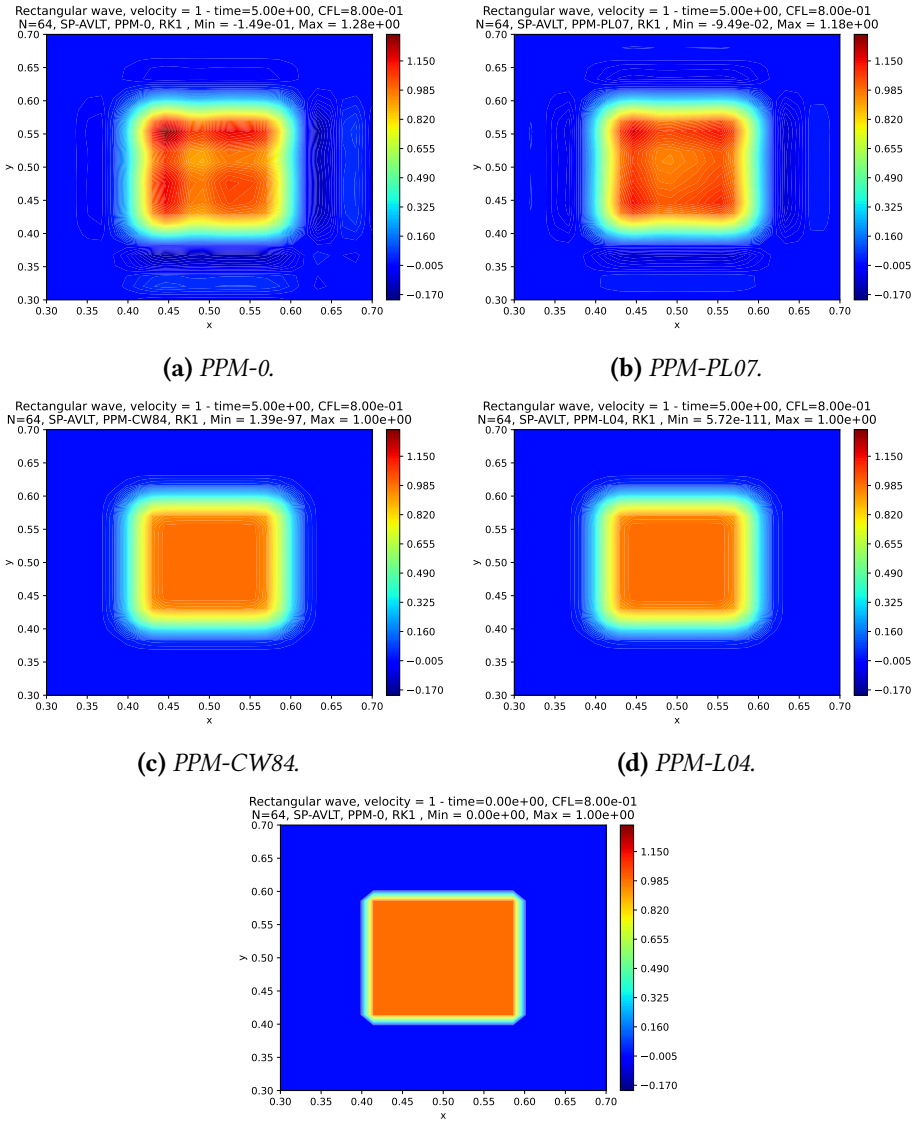
$$E_k = \max_{n=0,\ldots,N_T} \frac{\|Q^n - q(t^n)\|_{\infty,\Delta x\times\Delta y}}{\|q(t^n)\|_{\infty,\Delta x\times\Delta y}}.$$

The convergence rate is defined as in Section 2.6, as well as the total mass variation, which is preserved with machine precision in all experiments presented here. Notice that in the error computation, we use the centroid values instead of the exact average values to avoid computations of analytical integrals. As we pointed out in Proposition 3.1, this approximation introduces a second-order error. As a first test, we consider a constant velocity $\boldsymbol{u} = (0.2, -0.2)$. We are going to consider as initial condition a rectangular profile (Figure 3.4e) given by:

$$q_0(x, y) = \begin{cases} 1 & \text{if } (x, y) \in [0.4, 0.6] \times [0.4, 0.6], \\ 0 & \text{otherwise.} \end{cases} \tag{3.28}$$

The exact solution of Problem 3.1 in this case is $q_0((x, y) - \boldsymbol{u}t)$. Since the velocity field is constant, all the splitting schemes introduced in Section 3.3 are the same, so we will only consider the AVLT splitting. Additionally, it can be easily seen that the Lie-Trotter splitting is exact in this case (cf. eg. LeVeque, 1990, p. 202-203), meaning no splitting error is introduced. For the 1D schemes, we use RK1 to compute the departure point since this scheme is exact when the velocity is constant.

The conclusions for this test are very similar to the first 1D test from Section 2.6. This behavior is due to the fact that no splitting error is introduced when the velocity is constant. From Figure 3.4, it can be observed that the AVLT splitting preserves monotonicity when we use the monotonic 1D schemes PPM-CW84 and PPM-L04. For the non-monotonic schemes, as seen in Figure 2.4, PPM-PL07 produces less numerical dispersion than PPM-0, similar to the first 1D test.



(a) *PPM-0.* **(b)** *PPM-PL07.*

(c) *PPM-CW84.* **(d)** *PPM-L04.*

(e) *Initial condition.*

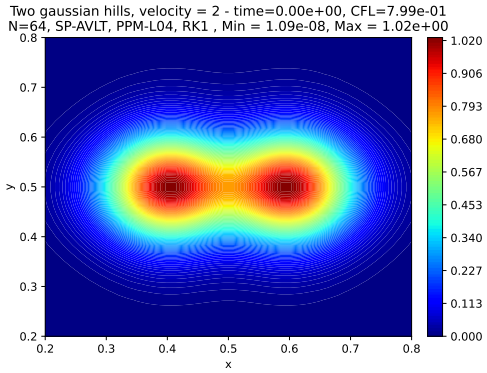
Figure 3.4: *Linear advection experiment using a constant velocity equal to* $\boldsymbol{u} = (0.2, -0.2)$*, a CFL number equal to* 0.8*,* $N = M = 64$*, and the initial condition is given by Equation* (3.28)*. We use the schemes PPM-0, PPM-PL07, PPM-CW84 and PPM-L04 with AVLT splitting. These figures show the advected profile after 5 time units (one time period). The initial condition is shown in (e).*

For variable velocity testing, we consider two Gaussian hills given by:

$$\begin{aligned} q_0(x,y) = &\exp(-10\cos^2(\pi(x-0.1))\exp(-10\cos^2(\pi y))+ \\ &\exp(-10\cos^2(\pi(x+0.1)))\exp(-10\cos^2(\pi y)), \end{aligned} \tag{3.29}$$

defined in $[0,1]\times[0,1]$, whose graph is shown in Figure 3.5a.

We consider the Cartesian version of the deformational flow test case on the sphere from Nair and Lauritzen (2010) proposed by Y. Chen et al. (2017). The velocity is given



(a) *Initial condition from Equation (3.29).*

by:

$$\begin{cases} u(x,y,t) &= c\frac{\pi}{L_y}\sin^2(\alpha_1)(2\cos(\alpha_2)\sin(\alpha_2))(\cos(\alpha_3)) - \frac{L_x}{T}, \\ v(x,y,t) &= \frac{-c}{\pi}\frac{2\pi}{L_x}(2\sin(\alpha_1)\cos(\alpha_1)\cos^2(\alpha_2))\cos(\alpha_3), \end{cases} \tag{3.30}$$

where $L_x = 2\pi, L_y = \pi, T = 5, c = \frac{10}{T}(\frac{L_x}{2\pi})^2, \alpha_1 = 2\pi\left(\frac{X}{L_x} - \frac{t}{T}\right), \alpha_2 = \frac{\pi Y}{L_y}, \alpha_3 = \frac{\pi t}{T}, X = -\pi + 2\pi x$, $Y = -\frac{\pi}{2} + \pi y$. Y. Chen et al. (2017) uses periodic boundary conditions in the $x-$direction and zero-gradient in the $y-$direction. However, we will employ biperiodic boundary conditions to simplify the problem. Both velocity fields are divergence-free, and they deform the initial condition. After 5 time units, the scalar field returns to its initial position and shape, allowing us to compute the error.

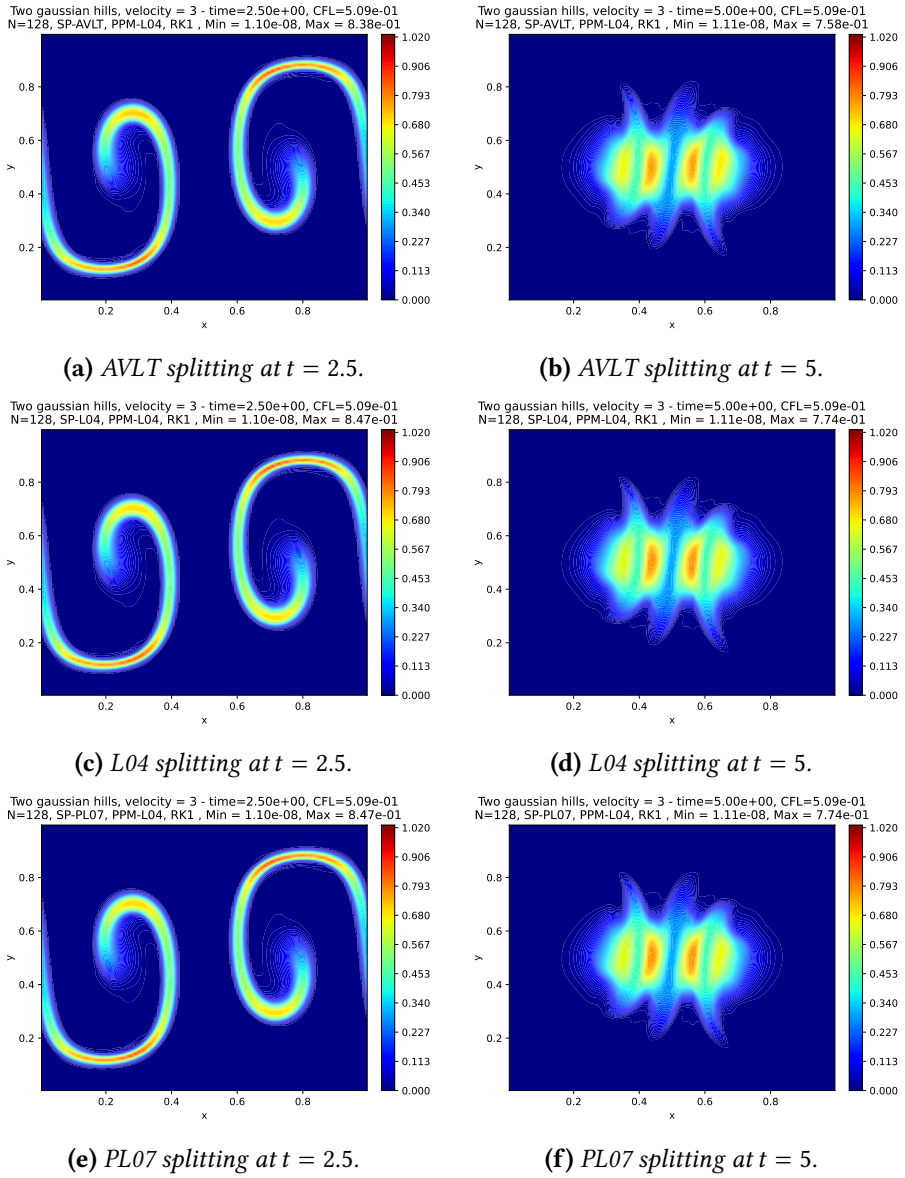
In Figure 3.6, the results obtained using two Gaussian hills and the velocity field from Equation (3.30) are depicted. The PPM-L04 scheme with AVLT, L04, and PL07 splitting, along with RK1 as the departure point scheme, is used. It can be observed how the scalar field deforms and eventually returns to its initial position. Comparing the schemes, it is noticeable that the PL07 and L04 schemes produce almost identical results and are less diffusive than AVLT.

From Figure 3.7a, it can be observed that when using the RK1 schemes, the PL07 and L04 schemes are slightly more accurate than AVLT for the 1D PPM-PL07 schemes. All these schemes exhibit a convergence order greater than two (Figure 3.8). When the 1D scheme PPM-L04 is used, all the splitting methods yield very similar results, regardless of the departure point scheme. However, when the RK2 schemes are used, AVLT is the most accurate and achieves a third-order convergence (Figure 3.7b).

3.5 Concluding remarks

In this chapter, we introduced the dimension-splitting method, which replaces the solution of the 2D advection equation with the solution of multiple 1D advection equations, resulting in more cost-effective 2D-FV schemes. For our simulations, we adopted the 1D FV-SL scheme based on PPM to solve the 1D equations.

We modified the average of two Lie-Trotter splittings, which is second-order accurate, to ensure the preservation of a constant scalar field with a divergence-free velocity. This



(a) *AVLT splitting at* $t = 2.5$.

(b) *AVLT splitting at* $t = 5$.

(c) *L04 splitting at* $t = 2.5$.

(d) *L04 splitting at* $t = 5$.

(e) *PL07 splitting at* $t = 2.5$.

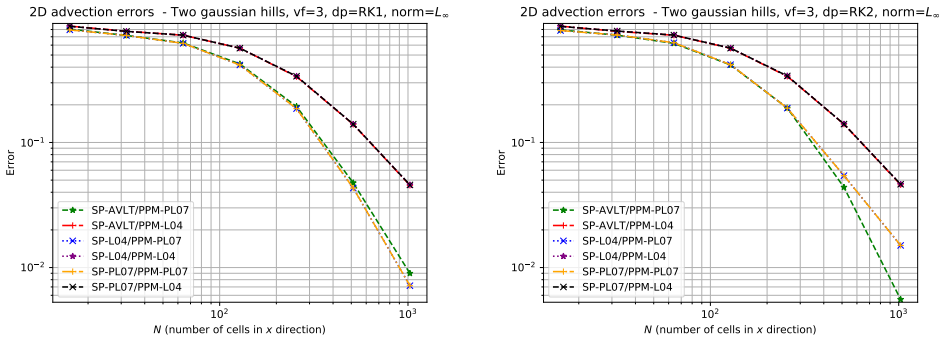
(f) *PL07 splitting at* $t = 5$.

Figure 3.6: *Linear advection experiment using the velocity from Equation* (3.30)*, a CFL number equal to* 0.8*,* $N = M = 128$*, and the initial condition is given by Equation* (3.29)*. We use the scheme PPM-L04 with AVLT (3.6a and 3.6b), L04 (3.6c and 3.6d) and PL07 (3.6e and 3.6f) splitting and RK1 departure point scheme. These figures show the advected profile after 2.5 (left) and 5 (right) time units (one time period).*

modification addresses the limitation of the classical averaging Lie-Trotter splitting and follows the methodology used in FV3.

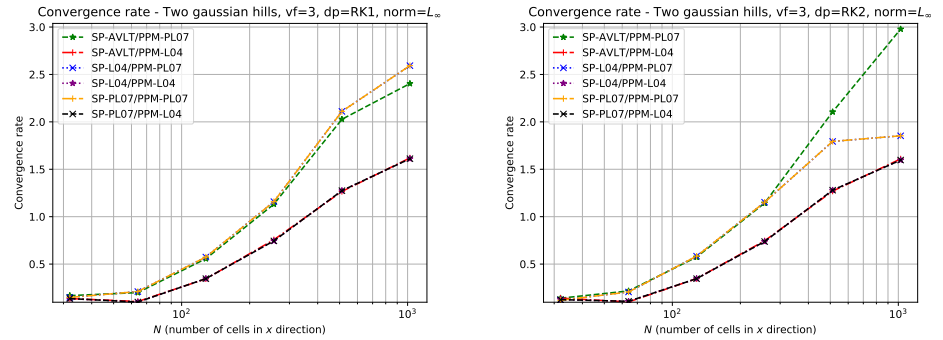
Based on the simulation with constant velocity, we concluded that all the splitting schemes are equivalent and do not introduce any splitting errors. In fact, the splittings are exact in this case. We observed that all the conclusions from the 1D simulations hold true in the 2D case as well, with mass conservation and monotonicity being preserved when using the monotonic limiters in the 1D subproblems.

In the simulation with variable velocity, we conducted a flow deformation test case. We



(a) *RK1.* **(b)** *RK2.*

Figure 3.7: *Convergence of the error for the schemes PPM-PL07and PPM-L04 with AVLT/L04/PL07/ splitting applied to the linear advection problem using a velocity from Equation* (3.30)*, a CFL number equal to* 0.8*, a final time of integration equal to 5 time units and the initial condition given by Equation* (3.28)*. The departure points are computed using the RK1 (left) and RK2 (right) schemes.*



(a) *RK1.* **(b)** *RK2.*

Figure 3.8: *Similar to Figure* 3.7 *but considering the convergence rate.*

observed that all splitting schemes preserved monotonicity in all simulations. The PL07 and L04 schemes yielded very similar results and introduced a first-order error, which was observed when using the RK2 scheme to compute the departure point. Surprisingly, the AVLT scheme achieved third-order accuracy, surpassing its expected second-order accuracy. This indicates that a more accurate departure point calculation benefits the AVLT splitting. However, when using a first-order departure point computation, the splitting schemes PL07 and L04 produced slightly smaller errors.

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 is usual for Planotic grids, we start with a Platonic solid, in this case, a cube, which is circumscribed in a sphere. We then project its faces onto the sphere. The original cubed-sphere, called the equidistant cubed-sphere, was proposed by Sadourny (1972) but resulted in a non-uniform grid. To address this issue, a solution was proposed by introducing angular coordinates, leading to a quasi-uniform grid known as the equiangular cubed-sphere. The cubed sphere consists of six panels, each one having a local Cartesian coordinate system. This makes it easier to extend methods from the plane to the sphere. In fact, Putman and Lin (2007) extends the dimension splitting technique from Lin and Rood (1996), as presented in Chapter 3, to the cubed-sphere.

There are essentially two major challenges when working with the cubed-sphere grid:

1. The non-orthogonal grid system: This challenge is primarily related to the appearance of metric terms in the equations. It adds computational cost and often requires conversions between contravariant and covariant components of a velocity field.

2. The discontinuity of the coordinate system at the cube edges: This is perhaps the most problematic challenge. Computing stencils along the cube edges becomes challenging due to the discontinuous nature of the coordinate system.

One possible approach to compute stencils at the edges is to extend the local coordinate of each panel to its neighboring panels, adding ghost cells in the halo region. In the case of the equiangular cubed-sphere, ghost cell values lie on the same geodesics containing the data from the neighboring panels. This allows for the use of one-dimensional high-order Lagrange interpolation to compute the stencils at the edges. This approach has been extensively used in the literature (X. Chen, 2021; Croisille, 2013; Katta et al., 2015a, 2015b) and was initially introduced by Ronchi et al. (1996). Alternatively, Putman and Lin (2007) uses extrapolation for grid values near the cube edges. Another approach that avoids the need for interpolation or extrapolation near the edges is the conformal cubed-sphere developed by Rančić et al. (1996). While this grid leads to an orthogonal and continuous coordinate system near the edges, it generates grid singularities near the cube corners,

similar to the pole problem. An improved and more uniform conformal grid, called the Uniform Jacobian cubed sphere, was later proposed by Rančić et al. (2017). Each approach is likely to generate grid imprinting, and one of the goals of this work is to investigate the amount of grid imprinting produced by each method.

This chapter aims to review and investigated the geometrical properties of the cubed-sphere. Besides that, we also aim to investigate the process of interpolating/extrapolating near the cube edges. We start with a basic review of the cubed-sphere mappings in Section 4.1, while Section 4.3 investigates the interpolation/extrapolation near the cube edges with some numerical experiments.

4.1 Cubed-sphere mappings

4.1.1 Equidistant cubed-sphere

We start this chapter by introducing the equidistant cubed-sphere proposed by Sadourny (1972). Given $R > 0$, we denote the sphere of radius R centered at the origin of $\mathbb{R}^3$ as:

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

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

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

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

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

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

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

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

The set of 6 maps $\{\Psi_p, p = 1, \dots, 6\}$ allow us to cover the sphere. Here p denotes a panel, and they are defined and orientated as Figure 4.1 shows. Then, we can represent a point on the sphere using the cubed-sphere coordinates (x, y, p).

The derivative of the maps Ψ_p are given by:

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

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

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

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

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

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

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

$$\boldsymbol{g}_x(x,y,p) = D\Psi_p(x,y)\begin{bmatrix}1\\0\end{bmatrix}, \boldsymbol{g}_y(x,y,p) = D\Psi_p(x,y)\begin{bmatrix}0\\1\end{bmatrix}.$$

Notice that

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

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:

$$G_\Psi(x,y) = \begin{bmatrix} \langle \boldsymbol{g}_x(x,y,p), \boldsymbol{g}_x(x,y,p)\rangle & \langle \boldsymbol{g}_x(x,y,p), \boldsymbol{g}_y(x,y,p)\rangle \\ \langle \boldsymbol{g}_x(x,y,p), \boldsymbol{g}_y(x,y,p)\rangle & \langle \boldsymbol{g}_y(x,y,p), \boldsymbol{g}_y(x,y,p)\rangle \end{bmatrix},$$

where $\langle\cdot,\cdot\rangle$ denotes the standard inner product of $\mathbb{R}^3$, 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:

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

4.1.2 Equiangular cubed-sphere

Another cubed-sphere mapping is the equiangular mapping (Ronchi et al., 1996), which leads to a more uniform grid. This mapping is a composition of equidistant mapping with angular coordinates. We consider again $a = \frac{R}{\sqrt{3}}$ and we define the family of maps $\Phi_p : [-\frac{\pi}{4}, \frac{\pi}{4}]\times[-\frac{\pi}{4}, \frac{\pi}{4}] \to \mathbb{S}^2_R$, $p = 1,\dots,6$, given by $\Phi_p(x,y) = \Psi_p(a\tan x, a\tan y)$. The

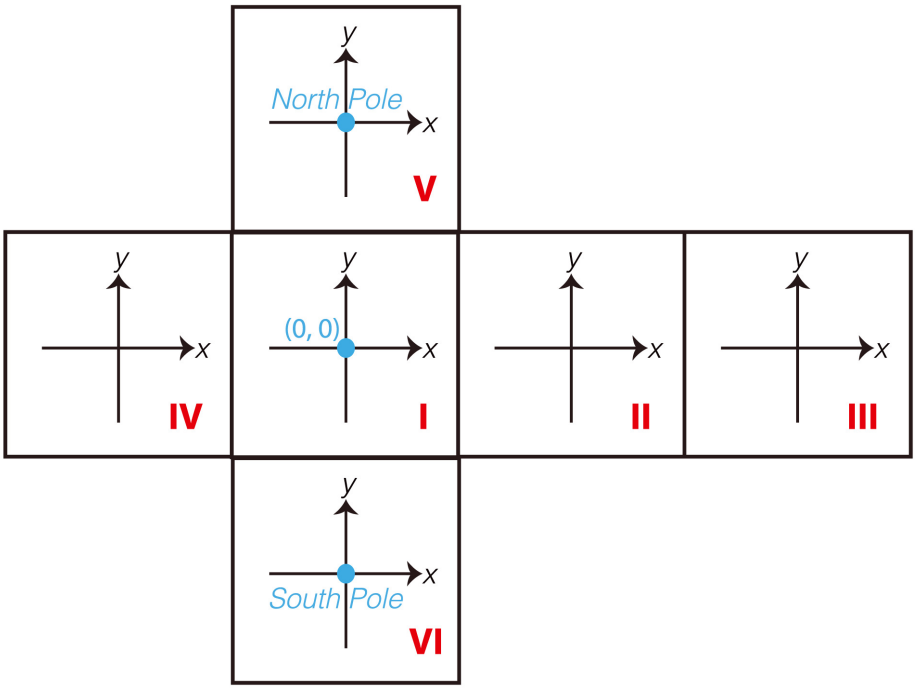


Figure 4.1: *Cubed-sphere panels definition and orientation. Figure taken from Jung et al. (2019).*

coordinates $(a \tan x, a \tan y)$ are called angular coordinates. By the chain rule:

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

and therefore we can define the following tangent vectors

$$\boldsymbol{r}_x(x,y,p) = D\Phi_p(x,y)\begin{bmatrix}1\\0\end{bmatrix} = \frac{a}{\cos^2 x}\boldsymbol{g}_x(\tan x, \tan y, p), \tag{4.1}$$

$$\boldsymbol{r}_y(x,y,p) = D\Phi_p(x,y)\begin{bmatrix}0\\1\end{bmatrix} = \frac{a}{\cos^2 y}\boldsymbol{g}_y(\tan x, \tan y, p), \tag{4.2}$$

Again, it makes sense to define the matrix

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

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

$$G_\Phi(x,y) = \begin{bmatrix} \langle \boldsymbol{r}_x(x,y,p), \boldsymbol{r}_x(x,y,p)\rangle & \langle \boldsymbol{r}_x(x,y,p), \boldsymbol{r}_y(x,y,p)\rangle \\ \langle \boldsymbol{r}_x(x,y,p), \boldsymbol{r}_y(x,y,p)\rangle & \langle \boldsymbol{r}_y(x,y,p), \boldsymbol{r}_y(x,y,p)\rangle \end{bmatrix}, \tag{4.3}$$

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:

$$\begin{aligned}\sqrt{|\det G_\Phi(x,y)|} &= \frac{a}{\cos^2 x \cos^2 y}\frac{R^2}{(a^2+a^2\tan^2 x + a^2\tan^2 y)^{3/2}}a \\ &= \frac{R^2}{\cos^2 x\cos^2 y}\frac{1}{(1+\tan^2 x+\tan^2 y)^{3/2}}.\end{aligned} \tag{4.4}$$

4.1.3 Tangent vectors on the sphere

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

$$T_P\mathbb{S}^2_R = \{P_0 \in \mathbb{R}^3 \,:\, \langle P, P_0\rangle = 0\}.$$

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. We say that a vector field $\boldsymbol{u} : \mathbb{S}^2_R \to \mathbb{R}^3$ is tangent on the sphere if $\boldsymbol{u}(P) \in T_P\mathbb{S}^2_R$, $\forall P \in \mathbb{S}^2_R$.

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:

$$X(\lambda,\phi) = R\cos\phi\cos\lambda, \tag{4.5}$$
$$Y(\lambda,\phi) = R\cos\phi\sin\lambda, \tag{4.6}$$
$$Z(\lambda,\phi) = R\sin\phi. \tag{4.7}$$

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

$$D\Psi_{ll}(\lambda,\phi) = R\begin{bmatrix}-\cos\phi\sin\lambda & -\sin\phi\cos\lambda \\ \cos\phi\cos\lambda & \sin\phi\sin\lambda \\ 0 & \cos\phi\end{bmatrix}. \tag{4.8}$$

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

$$\boldsymbol{r}_\lambda(\lambda,\phi) = D\Psi_{ll}(\lambda,\phi)\begin{bmatrix}1\\0\end{bmatrix}, \quad \boldsymbol{r}_\phi(\lambda,\phi) = D\Psi_{ll}(\lambda,\phi)\begin{bmatrix}0\\1\end{bmatrix}. \tag{4.9}$$

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

$$\boldsymbol{e}_\lambda(\lambda,\phi) = \begin{bmatrix}-\sin\lambda\\ \cos\lambda\\ 0\end{bmatrix}, \quad \boldsymbol{e}_\phi(\lambda,\phi) = \begin{bmatrix}-\sin\phi\cos\lambda\\ -\sin\phi\sin\lambda\\ \cos\phi\end{bmatrix}. \tag{4.10}$$

Let us consider a tangent vector field $\boldsymbol{u} : \mathbb{S}^2_R \to \mathbb{R}^3$ on the sphere, represented as

$$\boldsymbol{u}(\lambda, \phi) = u_\lambda(\lambda, \phi)\boldsymbol{e}_\lambda(\lambda, \phi) + v_\phi(\lambda, \phi)\boldsymbol{e}_\phi(\lambda, \phi). \tag{4.11}$$

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

$$\boldsymbol{u}(x, y, p) = u(x, y, p)\boldsymbol{r}_x(x, y, p) + v(x, y, p)\boldsymbol{r}_y(x, y, p). \tag{4.12}$$

This representation is known as contravariant representation. In order to relate the latitude-longitude representation with the contravariant representation, we notice that:

$$\boldsymbol{r}_x(x, y, p) = \langle \boldsymbol{r}_x, \boldsymbol{e}_\lambda \rangle \boldsymbol{e}_\lambda(\lambda, \phi) + \langle \boldsymbol{r}_x, \boldsymbol{e}_\phi \rangle \boldsymbol{e}_\phi(\lambda, \phi), \tag{4.13}$$

$$\boldsymbol{r}_y(x, y, p) = \langle \boldsymbol{r}_y, \boldsymbol{e}_\lambda \rangle \boldsymbol{e}_\lambda(\lambda, \phi) + \langle \boldsymbol{r}_y, \boldsymbol{e}_\phi \rangle \boldsymbol{e}_\phi(\lambda, \phi), \tag{4.14}$$

which holds since the vectors $\boldsymbol{e}_\lambda(\lambda, \phi)$ and $\boldsymbol{e}_\phi(\lambda, \phi)$ are orthogonal. Replacing Equations (4.13) and (4.14) in Equation (4.12), we obtain the values (u_λ, v_ϕ) in terms of the contravariant components (u, v) as the following matrix equation:

$$\begin{bmatrix} u_\lambda(\lambda, \phi) \\ v_\phi(\lambda, \phi) \end{bmatrix} = \begin{bmatrix} \langle \boldsymbol{r}_x, \boldsymbol{e}_\lambda \rangle & \langle \boldsymbol{r}_y, \boldsymbol{e}_\lambda \rangle \\ \langle \boldsymbol{r}_x, \boldsymbol{e}_\phi \rangle & \langle \boldsymbol{r}_y, \boldsymbol{e}_\phi \rangle \end{bmatrix} \begin{bmatrix} u(x, y, p) \\ v(x, y, p) \end{bmatrix}. \tag{4.15}$$

Conversely, we may express the contravariant components in terms of latitude-longitude components by inverting Equation (4.15):

$$\begin{bmatrix} u(x, y, p) \\ v(x, y, p) \end{bmatrix} = \frac{1}{\langle \boldsymbol{r}_x, \boldsymbol{e}_\lambda \rangle \langle \boldsymbol{r}_y, \boldsymbol{e}_\phi \rangle - \langle \boldsymbol{r}_y, \boldsymbol{e}_\lambda \rangle \langle \boldsymbol{r}_x, \boldsymbol{e}_\phi \rangle} \begin{bmatrix} \langle \boldsymbol{r}_y, \boldsymbol{e}_\phi \rangle & -\langle \boldsymbol{r}_y, \boldsymbol{e}_\lambda \rangle \\ -\langle \boldsymbol{r}_x, \boldsymbol{e}_\phi \rangle & \langle \boldsymbol{r}_x, \boldsymbol{e}_\lambda \rangle \end{bmatrix} \begin{bmatrix} u_\lambda(\lambda, \phi) \\ v_\phi(\lambda, \phi) \end{bmatrix}. \tag{4.16}$$

Covariant/contravariant conversion

Let us consider again a tangent vector field $\boldsymbol{u} : \mathbb{S}^2_R \to \mathbb{R}^3$ on the sphere. Its contravariant representation is given by Equation (4.12). The covariant components (U, V) are given by:

$$U(x, y, p) = \langle \boldsymbol{u}(x, y, p), \boldsymbol{r}_x(x, y, p) \rangle, \tag{4.17}$$

$$V(x, y, p) = \langle \boldsymbol{u}(x, y, p), \boldsymbol{r}_y(x, y, p) \rangle. \tag{4.18}$$

Replacing Equation (4.12) in Equations (4.17) and (4.18) we obtain the relation covariant components in terms of the contravariant terms:

$$\begin{bmatrix} U(x, y, p) \\ V(x, y, p) \end{bmatrix} = \begin{bmatrix} \langle \boldsymbol{r}_x, \boldsymbol{r}_x \rangle & \langle \boldsymbol{r}_x, \boldsymbol{r}_y \rangle \\ \langle \boldsymbol{r}_x, \boldsymbol{r}_y \rangle & \langle \boldsymbol{r}_y, \boldsymbol{r}_y \rangle \end{bmatrix} \begin{bmatrix} u(x, y, p) \\ v(x, y, p) \end{bmatrix}. \tag{4.19}$$

We may express the contravariant components in terms of the covariant terms inverting Equation (4.19):

$$\begin{bmatrix} u(x, y, p) \\ v(x, y, p) \end{bmatrix} = \frac{1}{\langle \boldsymbol{r}_x, \boldsymbol{r}_x \rangle \langle \boldsymbol{r}_y, \boldsymbol{r}_y \rangle - \langle \boldsymbol{r}_x, \boldsymbol{r}_y \rangle^2} \begin{bmatrix} \langle \boldsymbol{r}_y, \boldsymbol{r}_y \rangle & -\langle \boldsymbol{r}_x, \boldsymbol{r}_y \rangle \\ -\langle \boldsymbol{r}_x, \boldsymbol{r}_y \rangle & \langle \boldsymbol{r}_x, \boldsymbol{r}_x \rangle \end{bmatrix} \begin{bmatrix} U(x, y, p) \\ V(x, y, p) \end{bmatrix}. \tag{4.20}$$

Notice that combining Equations (4.19) and (4.20) with Equations (4.15) and (4.16) one may get relations between the latitude-longitude components and the covariant components.

4.2 Notation

Let us denote by $\Sigma_p : \Omega \to \mathbb{S}^2_R$, $p = 1, \dots, 6$, $\Omega = [-a, a]^2$, a cubed-sphere mapping like the equiangular or the equidistant mappings introduced in Section 4.1. With these mappings, we denote by $\boldsymbol{r}_x(x, y, p), \boldsymbol{r}_y(x, y, p)$ the basis of tangent vectors at (x, y, p). The metric tensor is denoted by $\sigma(x, y) = \sqrt{|\det G_\Sigma(x, y, p)|}$, where

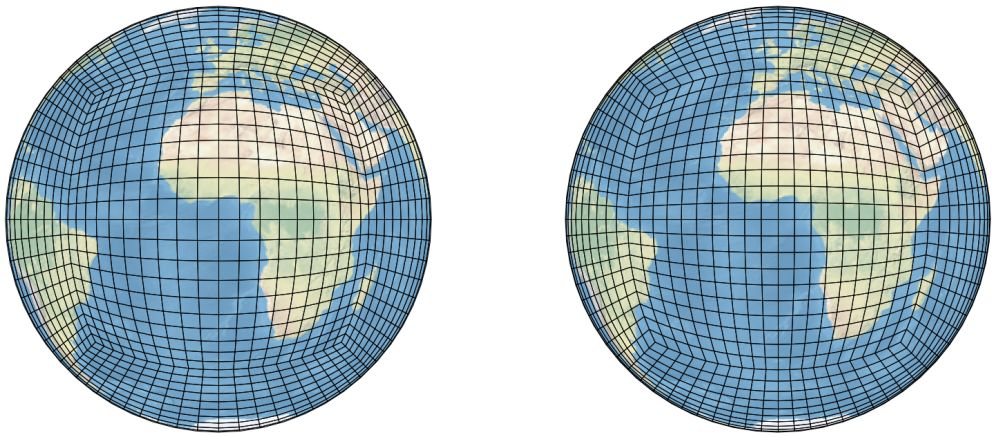
$$G_\Sigma(x, y, p) = \begin{bmatrix} \langle \boldsymbol{r}_x(x, y, p), \boldsymbol{r}_x(x, y, p) \rangle & \langle \boldsymbol{r}_x(x, y, p), \boldsymbol{r}_y(x, y, p) \rangle \\ \langle \boldsymbol{r}_x(x, y, p), \boldsymbol{r}_y(x, y, p) \rangle & \langle \boldsymbol{r}_y(x, y, p), \boldsymbol{r}_y(x, y, p) \rangle \end{bmatrix}. \tag{4.21}$$

We will utilize the notation introduced in Section 3.1.1 throughout this chapter. The parameter ν represents a non-negative integer indicating the number of ghost cell layers in each panel boundary, called halo size.

To introduce the cubed-sphere, we consider a $(\Delta x, \Delta y)$-grid denoted by $\Omega_{\Delta x, \Delta y} = (\Omega_{ij})_{i,j=-\nu+1,\dots,N+\nu}$, where $\Delta x = \Delta y$, and it covers the domain Ω. A control volume of the cubed-sphere is denoted by Ω_{ijp}, defined as follows:

$$\Omega_{ijp} = \Sigma_p(\Omega_{ij}) \quad -\nu + 1 \leq i, j \leq N + \nu, \quad 1 \leq p \leq 6.$$

The cubed-sphere grid refers to the collection of control volumes $(\Omega_{ijp})^{p=1,\dots,6}_{i,j=-\nu+1,\dots,N+\nu}$. In Figure 4.2, an example of the cubed-sphere grid is depicted, excluding the ghost cells. This grid is generated using the equidistant and equiangular mappings for $N = 20$.



(a) *Equidistant* **(b)** *Equiangular*

Figure 4.2: *Equidistant (a) and equiangular (b) cubed-spheres generated with $N = 20$.*

From Figure 4.2, it is evident that the equiangular cubed-sphere exhibits a higher

degree of uniformity compared to the equidistant cubed-sphere. As noted by Rančić et al. (1996), the ratio between the maximum and minimum cell areas on the equiangular cubed-sphere is approximately 1.3, whereas the same ratio is approximately 5.2 on the equidistant cubed-sphere.

We also utilize the notation $\mathcal{CS}_N = \mathbb{R}^{(N+\nu)\times(N+\nu)\times 6}$ to represent grid functions on the cubed-sphere at cell centers. Let's assume we have a function $q : \mathbb{S}^2_R \times [0,T] \to \mathbb{R}$, and we have a $(\Delta x, \Delta y, \Delta t, \lambda)$-discretization of $\Omega \times [0,T]$. We introduce $q^n \in \mathcal{CS}^N$, which represents the grid function q evaluated at the discrete points. In other words, $q^n_{ijp} = q(x_i, y_j, p, t^n)$, where $i, j = -\nu+1, \ldots, N+\nu$, and $p = 1, \ldots, 6$. Furthermore, we use the notations $q^n_{i+\frac{1}{2},j,p} = q(x_{i+\frac{1}{2}}, y_j, t^n)$ for $i = -\nu, \ldots, N+\nu$ and $j = -\nu+1, \ldots, N+\nu$ to represent q at the midpoint of edges in the x direction. Similarly, we use $q^n_{i,j+\frac{1}{2},p} = q(x_i, y_{j+\frac{1}{2}}, t^n)$ for $i = -\nu+1, \ldots, N+\nu$ and $j = -\nu, \ldots, N+\nu$ to represent q at the midpoint of edges in the y direction. When q does not depend on the time variable t, we can omit the index n. In Figure 4.3, we depict a grid function at centers, edge midpoints in the x direction and edge midpoints in the y direction for the equiangular cubed-sphere considering the halo size equal to three.

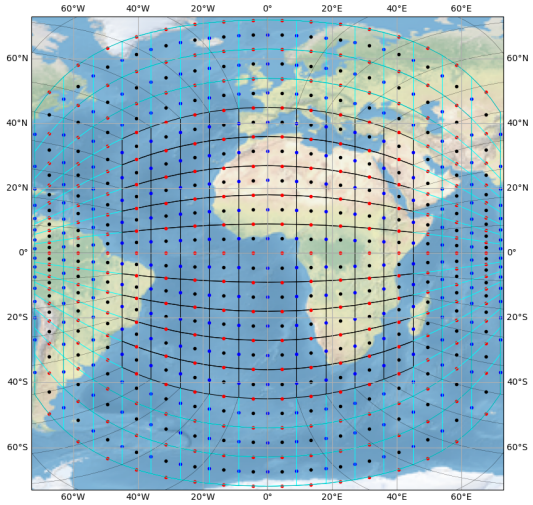


Figure 4.3: *Grid function at centers (black dots), edge midpoints in the x direction (blue dots) and edge midpoints in the y direction (red dots) at panel 1 for the equiangular cubed-sphere using three layers of ghost cells and $N = 10$. Ghost cells of panel 1 are drawn using cyan colors.*

We define the average values of a function q with the aid of the metric tensor $\sigma(x,y)$ at time t:

$$Q_{ijp}(t) = \frac{1}{|\Omega_{ijp}|} \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} \int_{y_{j-\frac{1}{2}}}^{y_{j+\frac{1}{2}}} q(x,y,p,t)\sigma(x,y)\,dx\,dy,$$

where $|\Omega_{ijp}|$ is the control volume area given by:

$$|\Omega_{ijp}| = \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} \int_{y_{j-\frac{1}{2}}}^{y_{j+\frac{1}{2}}} \sigma(x,y)\,dx\,dy.$$

In this context, we define $Q(t) \in \mathcal{CS}_N$ given by $Q(t) = (Q_{ijp}(t))_{i,j=-\nu+1,\dots,N+\nu}^{p=1,\dots,6}$. Similar to Proposition 3.1, we may approximate the average value using the centroid value, that is

$$Q_{ijp}^n - q_{ijp}^n = O(\Delta x^2),$$

In this work, we shall always approximate the average values since our schemes are expected to be at most second-order, this approximation does not deteriorate the convergence order.

4.3 Edges treatment

4.3.1 Ghost cells scalar field interpolation

Let's consider a function $q : \mathbb{S}_R^2 \to \mathbb{R}$ given at the cell centroids, denoted by q_{ijp}, where $i, j = 1 \dots, N$ and $p = 1, \dots, 6$. Our objective is to estimate these values at positions outside the range $1, \dots, N$, specifically at ghost cell positions.

To solve this problem, we will employ the strategy outlined in Zerroukat and Allen (2022). As previously mentioned, the ghost cells in the local Cartesian systems are mapped onto the geodesics of adjacent panels, which enables us to use Lagrange interpolation to obtain the values of ghost cells.

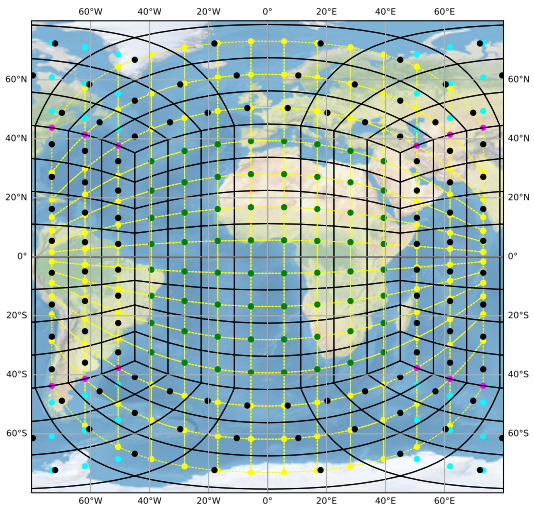


Figure 4.4: *Equiangular cubed-sphere panel 1 with $N = 8$: centroid at panel 1 (green circles) and others panel centroids (black circles), ghost cell points at panel 1 (yellow and magenta circles) and others panels (cyan circles).*

To illustrate this process in Panel 1, we depict the values of q_{ijp} in Figure 4.4. The green circles represent the values in Panel 1, while the black circles represent the values in the other panels, for a given $N = 8$. Assuming a halo size of 3, we also indicate the target values at the ghost cell positions using yellow and magenta circles. It is worth noting that the dashed yellow lines in Figure 4.4 illustrate how the ghost cell points lie on geodesics

containing grid positions from adjacent panels. With the exception of the magenta circles, all the ghost cell values can be obtained using 1D Lagrange interpolation, utilizing the surrounding black circles on the geodesic. This interpolation procedure can be performed for all panels. Subsequently, the magenta circles can be interpolated using the values obtained in the first step of interpolation (depicted as cyan circles in Figure 4.4), while preserving the order of accuracy of the interpolation, assuming it is fixed.

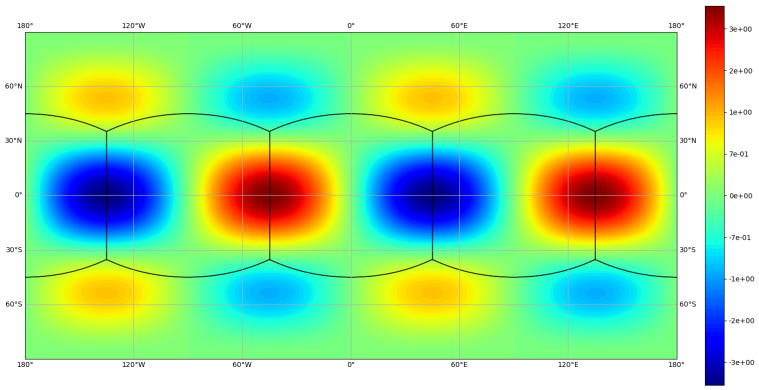
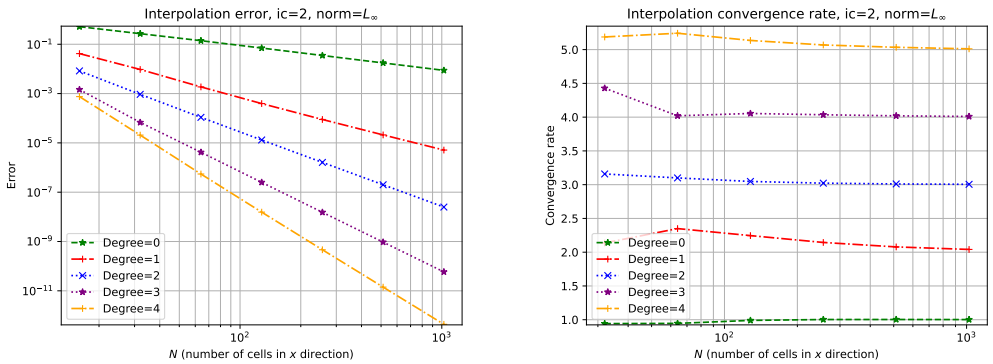


Figure 4.5: *Trigonometric function defined by Equation* (4.22).

We are going to show a numerical example of this interpolation process using a halo region of size 4 and assuming the radius of the sphere is equal to one. We shall consider the following trigonometric function, which is the divergence of a velocity field, as in Peixoto and Barros (2013) in our tests:

$$q(\lambda,\phi) = \frac{1}{\cos(\phi)}\Big(-2\cos^3(\phi)\sin(\lambda)\cos(\lambda) + 16\sin^2(\lambda)\cos(\lambda)\cos^3(\phi)\sin(\phi)\Big), \quad (4.22)$$

whose graph is depicted in Figure 4.5.



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

Figure 4.6: *Relative error convergence (a) and convergence rate (b) for the ghost cell interpolation process for different polynomial degrees, using the trigonometric function given by Equation* (4.22).

We are going to consider the relative error in the maximum norm and the convergence rate at the ghost cell positions defined analogously as in Section 3.4. We also consider values os N given by 2^k, for $k = 4, \dots, 10$, in order to compute the error and convergence rate. In Figures 4.6 we show the error and convergence rate for the trigonometric function (Equation (4.22)), respectively, considering polynomials of degrees from 0 up to 4. As both graphs show, we were able to achieve the expected order of convergence.

4.3.2 Ghost cells wind interpolation

4.3.3 Edges reconstruction

Let us consider the following problem: given the values q_{ijp} we wish to find approximations of the function q at the control volume edge midpoints denoted by $q^{L,x}_{ijp} \approx q_{i-\frac{1}{2},j,p}$, $q^{R,x}_{ijp} \approx q_{i+\frac{1}{2},j,p}$, $q^{L,y}_{ijp} \approx q_{i,j-\frac{1}{2},p}$, $q^{R,y}_{ijp} \approx q_{i,j+\frac{1}{2},p}$, where we also using the notations $q_{i+\frac{1}{2},j,p} \approx q(x_{i+\frac{1}{2}}, y_j; p)$, $q_{i,j+\frac{1}{2},p} \approx q(x_i, y_{j+\frac{1}{2}}; p)$. These points are illustrated in Figure 4.7 for panel 1.

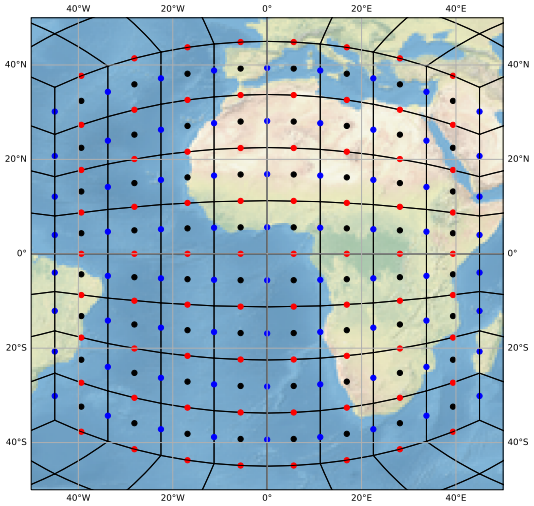


Figure 4.7: *The reconstruction problem on the cubed-sphere panel 1: we are given the centroid values of a function (black circles) and we wish to estimate these values at the edges midpoints in the x direction (blue circles) and y direction (red circles). This figure uses an equiangular cubed-sphere with* $N = 8$.

We can estimate the desired values using the one-dimensional reconstruction schemes from Sections 2.4 and 2.4.2 by performing the PPM reconstruction in the x and y directions independently. Notice that all the schemes from Sections 2.4 and 2.4.2 are expected to be second-order accurate due to centroid point approximation. The major difference here is when we compute the stencil near the cube edges. Unlikely the previous chapters where we assumed periodic boundary conditions, the boundary conditions are related to the adjacent panels. One way to overcome this problem is just to add ghost cell layers and use the process described in Section 4.3.1, hence all the stencils may be computed. This approach shall be referred to as **ET-R96** (ET stands for edge treatment) since this approach of extending the grid lines and using Lagrange interpolation was originally proposed in Ronchi et al. (1996).

Another way to fill the ghost cell values used for instance in Sadourny (1972), is just to ignore the discontinuity of the coordinate system and use the values of the cells in the adjacent panels as the ghost cell values. This scheme is labeled **ET-S72**. An approach that avoids the use of ghost cells has been developed by Putman and Lin (2007) using extrapolation at the cells surrounding to the cube edge. We are going to describe this

scheme that we shall name **ET-PL07**. This scheme uses the following extrapolations:

$$
\begin{aligned}
q^{L,x}_{1,j,p} &= \frac{1}{2}\Big(3Q_{1,j,p} - Q_{2,j,p}\Big),\\
q^{R,x}_{N,j,p} &= \frac{1}{2}\Big(3Q_{N,j,p} - Q_{N-1,j,p}\Big),\\
q^{L,y}_{i,1,p} &= \frac{1}{2}\Big(3Q_{i,1,p} - Q_{i,2,p}\Big),\\
q^{R,y}_{i,N,p} &= \frac{1}{2}\Big(3Q_{i,N,p} - Q_{i,N-1,p}\Big),
\end{aligned}
$$

at the points that are located on the cube edges. The other edge values are estimated as:

$$
\begin{aligned}
q^{R,x}_{1,j,p} &= \frac{1}{14}\Big(3Q_{1,j,p} + 11Q_{2,j,p} - 2(Q_{3,j,p} - Q_{1,j,p})\Big),\\
q^{L,x}_{2,j,p} &= q^{R,x}_{1,j,p},\\
q^{L,x}_{N,j,p} &= \frac{1}{14}\Big(3Q_{N,j,p} + 11Q_{N-1,j,p} - 2(Q_{N-2,j,p} - Q_{N,j,p})\Big),\\
q^{R,x}_{N-1,j,p} &= q^{L,x}_{N,j,p},
\end{aligned}
$$

in the x direction and in the y direction we use the formulas

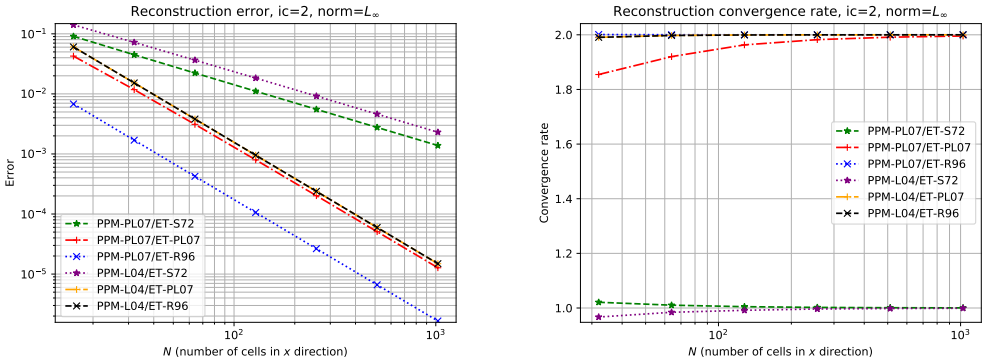
$$
\begin{aligned}
q^{R,y}_{i,1,p} &= \frac{1}{14}\Big(3Q_{i,1,p} + 11Q_{i,2,p} - 2(Q_{i,3,p} - Q_{i,1,p})\Big),\\
q^{L,y}_{i,2,p} &= q^{R,x}_{i,1,p},\\
q^{L,y}_{i,N,p} &= \frac{1}{14}\Big(3Q_{i,N,p} + 11Q_{i,N-1,p} - 2(Q_{i,N-2,p} - Q_{i,N,p})\Big),\\
q^{R,y}_{i,N-1,p} &= q^{L,y}_{i,N,p}.
\end{aligned}
$$

We are going to use the trigonometric function (Equation (4.22)) as before on the unit sphere to compare the schemes ET-R96, ET-S72 and ET-PL07. The scheme ET-R96 uses

cubic polynomials. We introduce the relative errors:

$$
\begin{aligned}
e_{i-\frac{1}{2},j,p} &= (|q_{i-\frac{1}{2},j,p} - q^{L,x}_{ijp}|)/|q_{i-\frac{1}{2},j,p}|,\\
e_{i+\frac{1}{2},j,p} &= (|q_{i+\frac{1}{2},j,p} - q^{R,x}_{ijp}|)/|q_{i+\frac{1}{2},j,p}|,\\
e_{i,j-\frac{1}{2},p} &= (|q_{i,j-\frac{1}{2},p} - q^{L,y}_{ijp}|)/|q_{i,j-\frac{1}{2},p}|,\\
e_{i,j+\frac{1}{2},p} &= (|q_{i,j+\frac{1}{2},p} - q^{R,y}_{ijp}|)/|q_{i,j+\frac{1}{2},p}|,\\
e_{ijp} &= \max\{e_{i-\frac{1}{2},j,p}, e_{i+\frac{1}{2},j,p}, e_{i,j-\frac{1}{2},p}, e_{i,j+\frac{1}{2},p}\},\\
E &= \max\{e_{ijp}\}.
\end{aligned}
$$

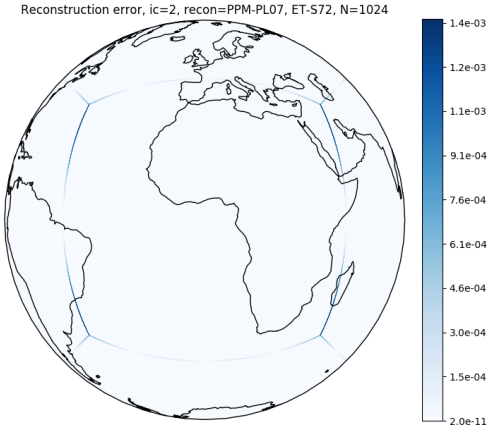
We are going to compute E for different values of N as in the numerical experiments of Section 4.3.1.



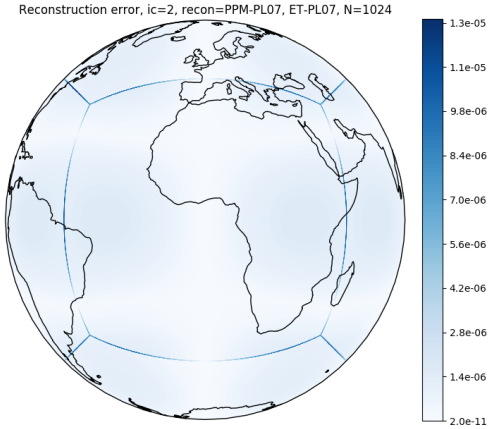
(a) *Error.* **(b)** *Convergence rate.*

Figure 4.8: *Relative error convergence (a) and convergence rate (b) for the reconstruction problem for different edge treatments (ET), using the trigonometric function given by Equation (4.22).*

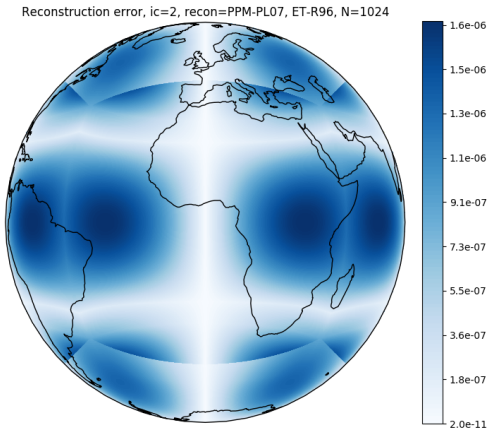
In Figure 4.8 we show the errors and the convergence rate using the PPM-PL07 and PPM-L04 reconstruction schemes. We can observe that all the schemes converge to zero with second-order. The difference between the edge treatments ET-S72, ET-PL07 and ET-R96 schemes may be observed in Figure 4.9 and Figure 4.10, for the PPM-PL07 and PPM-L04 reconstruction schemes, respectively. We notice that the cube edges appear on the error graph when we use the ET-S72 and ET-PL07 schemes. This is an example of grid imprinting. Although all schemes are second-order, the scheme ET-R96 do not seem to produce grid imprinting in the reconstruction problem.



(a) *ET-S72.*

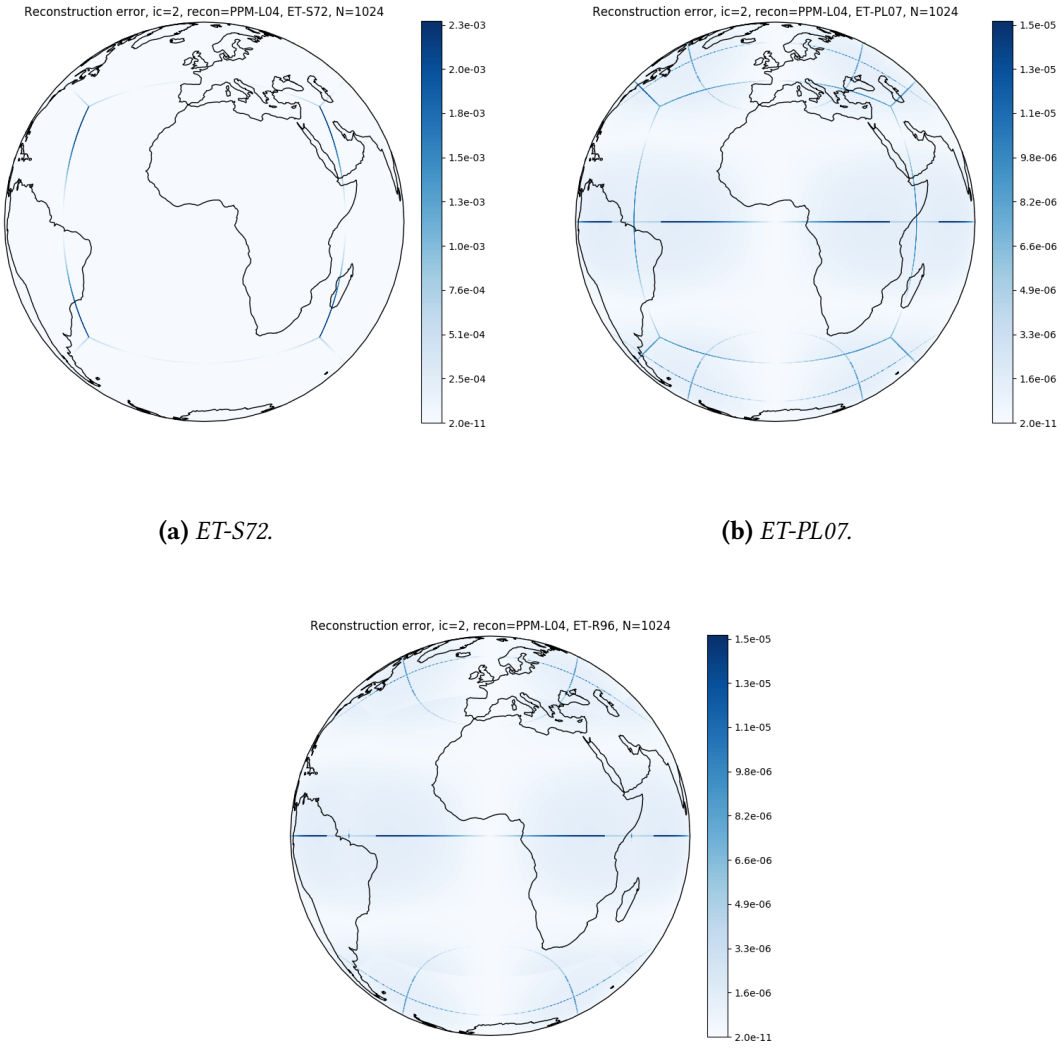


(b) *ET-PL07.*



(c) *ET-R96.*

Figure 4.9: *Error for the edge midpoint values considering the trigonometric function (Equation (4.22)) with edges treatment schemes ET-S72 (a), ET-PL07 (b) and ET-R96 for the cubed-sphere with* $N = 1024$ *using the reconstruction PPM-PL07.*



(a) *ET-S72.*

(b) *ET-PL07.*

(c) *ET-R96.*

Figure 4.10: *As Figure in 4.9 but using the PPM-L04 scheme.*

Chapter 5

Cubed-sphere finite-volume methods

5.1 Advection finite-volume scheme

Consistent Treatment of Boundaries: Boundary conditions play a crucial role in maintaining mass conservation. The treatment of boundaries should ensure that there is no spurious mass exchange between the model domain and the exterior. Various approaches, such as using ghost cells or extrapolation methods, can be employed to enforce mass conservation at the boundaries.

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. 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.

Appendix A

Numerical Analysis

A.1 Finite-difference estimates

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

Lemma A.1. *Let $F \in C^5(\mathbb{R})$, $x_0 \in \mathbb{R}$ and $h > 0$. 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 $F \in C^4(\mathbb{R})$, $x_0 \in \mathbb{R}$ and $h > 0$. 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_2 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 $F \in C^4(\mathbb{R})$, $x_0 \in \mathbb{R}$ and $h > 0$. 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:

$$F(x_0-2h) - 7F(x_0-h) + 16F(x_0) - 16F(x_0+h) + 7F(x_0+2h) - F(x_0+3h) = 2F^{(3)}(x_0)h^3 - 2C_3h^4,$$

we get the desired Formula (A.12) with C_3 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. □

A.2 Lagrange interpolation

Given real numbers, called nodes, $x_0 < x_1 < \ldots < x_m$, we define the k-th Lagrange polynomial by

$$L_k(x) = \prod_{j=0, j\neq k}^{m} \frac{x - x_j}{x_k - x_j}.$$

They satisfy $L_k(x_j) = \delta_{kj}$, where δ_{kj} is the Kronecker delta. Given a function f defined at the nodes x_j, its interpolating polynomial of degree m is given by:

$$P_m(x) = \sum_{k=0}^{m} f(x_k)L_k(x).$$

Indeed, this polynomial interpolates f since $P_m(x_j) = f(x_j)$. It is well known that P_m always exists and is unique. Besides that, we have the following error formula for Lagrange interpolation.

Theorem A.1. *Let $f \in C^{m+1}(\mathbb{R})$. Then, then there is ξ in the smallest interval containing $x_0, \ldots, x_m, x$ such that:*

$$f(x) - P_m(x) = \omega(x)\frac{f^{(m+1)}(\xi)}{(m+1)!}, \tag{A.15}$$

where $\omega(x) = (x - x_0)(x - x_1)\ldots(x - x_m)$.

Proof. See Stoer and Bulirsch (2002, Theorem 2.1.4.1. on p. 49). □

A.3 Numerical integration

The following mean value theorem for integrals is a very useful tool when working with numerical integration errors.

Theorem A.2 (Mean value theorem for integrals)**.** *If $f \in C([a, b])$, and g is a integrable function in $[a, b]$ whose sign does not change in $[a, b]$, then there exists $c \in]a, b[$ such that*

$$\int_a^b f(x)g(x)\,dx = f(c)\int_a^b g(x)\,dx.$$

Proof. See Courant and John (1999, p. 143). □

A.3.1 Midpoint rule

When considering finite-volume schemes, it is useful to compare the average value on a control volume of a function with its value at the control volume centroid. In the following theorems, for the one and two dimensional cases, respectively, we show that the value of a function at the centroid of a control volume given a second-order approximation to its average value on the control volume.

Theorem A.3. *If $f \in C^2([x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}])$, then*

$$\frac{1}{\Delta x}\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} f(x)\,dx - f(x_i) = C_1\Delta x^2, \tag{A.16}$$

where C_1 is a constant that depends only on f, and $x_i = \frac{x_{i+\frac{1}{2}}+x_{i-\frac{1}{2}}}{2}$, $\Delta x = x_{i+\frac{1}{2}} - x_{i-\frac{1}{2}}$.

Proof. From Taylor's expansion, it follows that, for $x \in [x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}]$, we have:

$$f(x) = f(x_i) + f(x_i)(x - x_i) + f''(\xi)\frac{(x-x_i)^2}{2}, \tag{A.17}$$

for some ξ between x and x_i. Therefore:

$$\begin{aligned}\frac{1}{\Delta x}\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} f(x)\,dx - f(x_i) &= \frac{1}{\Delta x}\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} \left(f'(x_i)(x-x_i) + f''(\xi)\frac{(x-x_i)^2}{2}\right) dx \\ &= \frac{1}{\Delta x}\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} f''(\xi)\frac{(x-x_i)^2}{2}\,dx.\end{aligned}$$

Using the mean value theorem for integrals (see Theorem A.2), we have:

$$\frac{1}{\Delta x}\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} f(x)\,dx - f(x_i) = f''(\eta)\frac{1}{\Delta x}\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} \frac{(x-x_i)^2}{2}\,dx = f''(\eta)\frac{\Delta x^2}{24}$$

for some $\eta \in [x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}]$, from which the proposition follows with

$$C_1 = \frac{1}{24}f''(\eta). \tag{A.18}$$

□

Theorem A.4. *If $f \in C^2([x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}] \times [y_{j-\frac{1}{2}}, y_{j+\frac{1}{2}}])$, then*

$$\left|\frac{1}{\Delta x\Delta y}\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}}\int_{y_{j-\frac{1}{2}}}^{y_{j+\frac{1}{2}}} f(x,y)\,dx\,dy - f(x_i,y_j)\right| \le C_1\Delta x^2 + C_2\Delta x\Delta y + C_3\Delta y^2, \tag{A.19}$$

where C_1, C_2 and C_3 are constants that depend only on f, and $x_i = \frac{x_{i+\frac{1}{2}}+x_{i-\frac{1}{2}}}{2}$, $y_i = \frac{y_{j+\frac{1}{2}}+y_{j-\frac{1}{2}}}{2}$, $\Delta x = x_{i+\frac{1}{2}} - x_{i-\frac{1}{2}}$, $\Delta y = y_{j+\frac{1}{2}} - y_{j-\frac{1}{2}}$.

Proof. From Taylor's expansion, it follows that, for $x \in [x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}]$, $y \in [y_{j-\frac{1}{2}}, y_{j+\frac{1}{2}}]$, we have:

$$f(x,y) = f(x_i,y_j) + \frac{\partial f}{\partial x}(x_i,y_j)(x-x_i) + \frac{\partial f}{\partial y}(x_i,y_j)(y-y_j) + \frac{1}{2}\left(\frac{\partial^2 f}{\partial x^2}(\xi,\theta)(x-x_i)^2 + 2\frac{\partial^2 f}{\partial x \partial y}(\xi,\theta)(x-x_i)(y-y_j)\frac{\partial^2 f}{\partial y^2}(\xi,\theta)(y-y_j)^2\right)$$

for some ξ between x and x_i, and θ between y and y_j. Therefore:

$$\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}}\int_{y_{j-\frac{1}{2}}}^{y_{j+\frac{1}{2}}} f(x,y)\,dy\,dx - \Delta x\Delta y f(x_i,y_j) = \frac{1}{2}\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}}\int_{y_{j-\frac{1}{2}}}^{y_{j+\frac{1}{2}}} \frac{\partial^2 f}{\partial x^2}(\xi,\theta)(x-x_i)^2\,dy\,dx + \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}}\int_{y_{j-\frac{1}{2}}}^{y_{j+\frac{1}{2}}} \frac{\partial^2 f}{\partial x \partial y}(\xi,\theta)(x-x_i)(y-y_j)\,dy\,dx + \frac{1}{2}\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}}\int_{y_{j-\frac{1}{2}}}^{y_{j+\frac{1}{2}}} \frac{\partial^2 f}{\partial y^2}(\xi,\theta)(y-y_j)^2\,dy\,dx$$

Using the mean value theorem for integrals (see Theorem A.2), we have:

$$\frac{1}{\Delta x\Delta y}\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}}\int_{y_{j-\frac{1}{2}}}^{y_{j+\frac{1}{2}}} f(x,y)\,dx\,dy - f(x_i,y_j) = \frac{\partial^2 f}{\partial x^2}(\eta_1,\lambda_1)\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}}\int_{y_{j-\frac{1}{2}}}^{y_{j+\frac{1}{2}}} \frac{(x-x_i)^2}{2\Delta x\Delta y}\,dx\,dy$$
$$+ \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}}\int_{y_{j-\frac{1}{2}}}^{y_{j+\frac{1}{2}}} \frac{\partial^2 f}{\partial x\partial y}(\xi,\theta)\frac{(x-x_i)(y-y_j)}{\Delta x\Delta y}\,dx\,dy + \frac{\partial^2 f}{\partial x^2}(\eta_2,\lambda_2)\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}}\int_{y_{j-\frac{1}{2}}}^{y_{j+\frac{1}{2}}} \frac{(y-y_j)^2}{2\Delta x\Delta y}\,dx\,dy$$
$$= \frac{\partial^2 f}{\partial x^2}(\eta_1,\lambda_1)\frac{\Delta x^2}{24} + \frac{\partial^2 f}{\partial y^2}(\eta_2,\lambda_2)\frac{\Delta y^2}{24} + \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}}\int_{y_{j-\frac{1}{2}}}^{y_{j+\frac{1}{2}}} \frac{\partial^2 f}{\partial x\partial y}(\xi,\theta)\frac{(x-x_i)(y-y_j)}{\Delta x\Delta y}\,dx\,dy$$

for $\eta_1, \eta_2 \in [x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}]$, $\lambda_1, \lambda_2 \in [y_{j-\frac{1}{2}}, y_{j+\frac{1}{2}}]$, from which the proposition follows. □

A.3.2 Multi-step schemes

Let us consider the following problem: given a function $f \in C^{m+1}([0,T])$, a discretization of $[0,T]$ given by $t^n = n\Delta t$, $\Delta t = \frac{T}{N_T}$, for some $N_T \in \mathbb{N}$, we wish to estimate $\int_{t^n}^{t^{n+1}} f(t)\,dt$ using the values $f(t_{n-k})$, for $k = 0, \cdots, m$. This kind of problem arises, for instance, when we are interested in computing departure points as in Equation 2.11. We can estimate the desired integral by computing the interpolating polynomial of $f(t_{n-k})$, for $k = 0, \cdots, m$ and then integrating this polynomial. This approach is exactly what is used in multi-step Adams-Bashforth methods. On the next theorem, we give an expression the error of this approach.

Theorem A.5. *If $f \in C^{m+1}([0,T])$, $t^n = n\Delta t$, $n = 0, \cdots, N_T$, $\Delta t = \frac{T}{N_T}$ for some $N_T \in \mathbb{N}$, then:*

$$\int_{t^n}^{t^{n+1}} f(t)\,dt = \Delta t \sum_{k=0}^{m}\left(\int_0^1 L_k(s)\,ds\right) f(t_{n-k}) + \frac{(\Delta t)^{k+1}}{(m+1)!} f^{(m+1)}(\eta)\int_0^1 \omega(s)\,ds, \qquad \text{(A.20)}$$

where $w(s) = s(s+1)\cdots(s+m)$, $\eta \in [t^{n-m}, t^n]$.

Proof. We introduce auxiliary functions $\theta(s) = s\Delta t + t_n$, $s \in [-m, 1]$ and $g(s) = f(\theta(s))$. It

is clear that $f(t_{n-k}) = g(-k)$, for $k = -1, 0, \cdots, m$. Hence, we can write:

$$\int_{t^n}^{t^{n+1}} f(t)\,dt = \Delta t \int_0^1 f(\theta(s))\,ds = \Delta t \int_0^1 g(s)\,ds. \tag{A.21}$$

Defining the nodes $s_k = -k$ for $k = 0, \cdots, m$, it follows from Theorem A.1 that the interpolating polynomial P_m of $g(s_k)$ satisfies:

$$g(s) - P_m(s) = \omega(s)\frac{g^{(m+1)}(\xi)}{(m+1)!}, \tag{A.22}$$

where $\xi \in [-m, 1]$. Substituting Equation (A.22) in Equation (A.21), we obtain

$$\int_{t^n}^{t^{n+1}} f(t)\,dt = \Delta t \sum_{k=0}^{m} \left(\int_0^1 L_k(s)\,ds \right) g(-k) + \frac{\Delta t}{(m+1)!} \int_0^1 g^{(m+1)}(\xi)\omega(s)\,ds. \tag{A.23}$$

Since $w(s)$ does not change its sign in $[0, 1]$ it follows from Theorem A.2 that:

$$\int_{t^n}^{t^{n+1}} f(t)\,dt = \Delta t \sum_{k=0}^{m} \left(\int_0^1 L_k(s)\,ds \right) g(-k) + \frac{\Delta t}{(m+1)!} g^{(m+1)}(\bar{\xi}) \int_0^1 \omega(s)\,ds, \tag{A.24}$$

for some $\bar{\xi} \in [-m, 1]$. Notice that by the chain rule we get $g^{(m+1)}(s) = (\Delta t)^k f^{(m+1)}(\theta(s))$, therefore Equation (A.24) in terms of f reads:

$$\int_{t^n}^{t^{n+1}} f(t)\,dt = \Delta t \sum_{k=0}^{m} \left(\int_0^1 L_k(s)\,ds \right) f(t_{n-k}) + \frac{(\Delta t)^{k+1}}{(m+1)!} f^{(m+1)}(\eta) \int_0^1 \omega(s)\,ds, \tag{A.25}$$

where $\eta \in [t^{n-m}, t^n]$, which is the desired identity. □

In the following corollaries, we give the explicit formulas for Equation (A.25) for $m = 0, m = 1, m = 2$. This is achieved by computing the terms $\int_0^1 L_k(s)\,ds$ and $\int_0^1 \omega(s)\,ds$, which are trivial to be computed.

Corollary A.1. *If $f \in C^1([0, T])$, $t^n = n\Delta t$, $n = 0, \cdots, N_T$, $\Delta t = \frac{T}{N_T}$ for some $N_T \in \mathbb{N}$, then:*

$$\int_{t^n}^{t^{n+1}} f(t)\,dt = \Delta t f(t_n) + \frac{\Delta t^2}{2} f'(\bar{t}), \tag{A.26}$$

for some $\bar{t} \in [t^n, t^{n+1}]$.

Corollary A.2. *If $f \in C^2([0, T])$, $t^n = n\Delta t$, $n = 0, \cdots, N_T$, $\Delta t = \frac{T}{N_T}$ for some $N_T \in \mathbb{N}$, then:*

$$\int_{t^n}^{t^{n+1}} f(t)\,dt = \frac{\Delta t}{2}(3f(t_n) - f(t_{n-1})) + \frac{5\Delta t^3}{12} f^{(2)}(\bar{t}), \tag{A.27}$$

for some $\bar{t} \in [t^{n-1}, t^{n+1}]$.

Corollary A.3. *If $f \in C^3([0,T])$, $t^n = n\Delta t$, $n = 0, \cdots, N_T$, $\Delta t = \frac{T}{N_T}$ for some $N_T \in \mathbb{N}$, then:*

$$\int_{t^n}^{t^{n+1}} f(t)\,dt = \frac{\Delta t}{12}(23f(t_n) - 16f(t_{n-1}) + 5f(t_{n-2})) + \frac{3\Delta t^4}{8} f^{(3)}(\bar{t}), \tag{A.28}$$

for some $\bar{t} \in [t^{n-2}, t^{n+1}]$.

When using these schemes for and ODE written in its integral form, $m = 0$ gives the classical Euler method; for $m = 1$ we get the second-order Adams-Bashforth scheme and for $m = 2$ we have the third-order Adams-Bashforth scheme.

A.4 PPM reconstruction accuracy analysis

In this Section, we are going to investigate the accuracy of the PPM reconstruction process. As we pointed out in Section 2.4.1, the approximation of q at the control volumes edges given by Equation (2.46) is fourth-order accurate when $q \in C^4(\mathbb{R})$. This is proved as a Corollary of the following Proposition A.1.

Proposition A.1. *Let $q \in C^4(\mathbb{R})$, $\overline{x} \in \mathbb{R}$ and $h > 0$. Then, the following identity holds:*

$$q(\overline{x}) = \frac{7}{12}\left(\frac{1}{h}\int_{\overline{x}}^{\overline{x}+h} q(x)\,dx + \frac{1}{h}\int_{\overline{x}-h}^{\overline{x}} q(x)\,dx\right) - \frac{1}{12}\left(\frac{1}{h}\int_{\overline{x}+h}^{\overline{x}+2h} q(x)\,dx + \frac{1}{h}\int_{\overline{x}-2h}^{\overline{x}-h} q(x)\,dx\right) + C_1 h^4, \tag{A.29}$$

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 fixed $a \in \mathbb{R}$ 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 (A.29) may be rewritten as:

$$q(\overline{x}) = \frac{4}{3}\left(\frac{Q(\overline{x}+h) - Q(\overline{x}-h)}{2h}\right) - \frac{1}{3}\left(\frac{Q(\overline{x}+2h) - Q(\overline{x}-2h)}{4h}\right) + C_1 h^4, \tag{A.30}$$

which consists of finite-difference approximations. Thus, Equation (A.29) follows from Lemma A.1 with:

$$C_1 = C_1(\mu_1, \mu_2) = \frac{1}{720}\left(6q^{(4)}(\mu 1) - 32q^{(4)}(\mu_2)\right), \tag{A.31}$$

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

Corollary A.4. *It follows from Proposition A.1 with $\overline{x} = x_{i+\frac{1}{2}}$ and $h = \Delta x$ that $q_{i+\frac{1}{2}}$ given by*

Equation (2.46) *satisfies:*

$$q(x_{i+\frac{1}{2}}) - q_{i+\frac{1}{2}} = C_1 \Delta x^4, \tag{A.32}$$

with C_1 *given by Equation* (A.31)*, whenever* $q \in C^4(\mathbb{R})$.

Remark A.1. *Similarly, one can show that the formulas are given by Equation* (2.49) *and Equation* (2.49) *are fifth-order accurate.*

The parabolic function from (2.40) given with coefficients specified before approximates q with order 3 when $q \in C^4(\mathbb{R})$. In order to check this, for $x \in X_i$ we rewrite Equation (2.40) as:

$$q_i(x; Q) = q_{L,i} + \frac{(\Delta q_i + q_{6,i})}{\Delta x}(x - x_{i-\frac{1}{2}}) - \frac{q_{6,i}}{\Delta x^2}(x - x_{i-\frac{1}{2}})^2 \tag{A.33}$$

and we write q using its Taylor expansion assuming $q \in C^4(\mathbb{R})$:

$$q(x) = q(x_{i-\frac{1}{2}}) + q'(x_{i-\frac{1}{2}})(x - x_{i-\frac{1}{2}}) + \frac{q''(x_{i-\frac{1}{2}})}{2}(x - x_{i-\frac{1}{2}})^2 + \frac{q^{(3)}(\theta_i)}{6}(x - x_{i-\frac{1}{2}})^3, \tag{A.34}$$

where $\theta_i \in X_i$. Comparing Equation (A.33) with Equation (A.34), it is reasonable to seek to some bound to the expressions:

$$q'(x_{i-\frac{1}{2}}) - \frac{(\Delta q_i + q_{6,i})}{\Delta x}, \tag{A.35}$$

and:

$$\frac{q''(x_{i-\frac{1}{2}})}{2} - \left(- \frac{q_{6,i}}{\Delta x^2} \right). \tag{A.36}$$

We have seen that term $q_{L,i}$ gives a fourth-order approximation to $q(x_{i-\frac{1}{2}})$. The Corollary A.5 shall prove that the term (A.35) has a bound proportional to Δx^2, and the Corollary A.6 shall prove that the term (A.36) 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 of the Δx-grid function Q. Combining Equation (2.43) with Equations (2.47) and (2.48), we may write $q_{6,i}$ as:

$$q_{6,i} = \frac{1}{4}\left(Q_{i-2} - 6Q_{i-1} + 10Q_i - 6Q_{i+1} + Q_{i+2} \right). \tag{A.37}$$

Recalling the definition of Δq_i from Equation (2.41), and applying Equations (2.47) and (2.48), we may express Δq_i as:

$$\Delta q_i = \frac{1}{12}\left(Q_{i-2} - 8Q_{i-1} + 8Q_{i+1} - Q_{i+2} \right). \tag{A.38}$$

Finally, we combine Equations (A.37) and (A.38) and write their sum as:

$$\frac{(\Delta q_i + q_{6,i})}{\Delta x} = \frac{2Q_{i-2} - 13Q_{i-1} + 15Q_i - 5Q_{i+1} + Q_{i+2}}{6\Delta x}. \tag{A.39}$$

The next Proposition A.2 proves that Equation (A.39) approximates $q'(x_{i-\frac{1}{2}})$ with order

2.

Proposition A.2. *Let $q \in C^3(\mathbb{R})$, $\overline{x} \in \mathbb{R}$ and $h > 0$. Then, the following identity holds:*

$$q'(\overline{x}) = \frac{1}{6h}\left(\frac{2}{h}\int_{\overline{x}-2h}^{\overline{x}-h} q(x)\,dx - \frac{13}{h}\int_{\overline{x}-h}^{\overline{x}} q(x)\,dx + \frac{15}{h}\int_{\overline{x}}^{\overline{x}+h} q(x)\,dx - \frac{5}{h}\int_{\overline{x}+h}^{\overline{x}+2h} q(x)\,dx + \frac{1}{h}\int_{\overline{x}+2h}^{\overline{x}+3h} q(x)\,dx\right) + C_2 h^2, \tag{A.40}$$

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 $a \in \mathbb{R}$ fixed as in Equation (2.37). Like in Proposition A.2, we have:

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

which consists of the finite-difference scheme from Lemma A.2. Therefore, Equation (A.40) follows from Lemma A.2 with:

$$C_2 = C_2(\mu_1, \mu_2) = \frac{1}{24}\Big(128 q^{(3)}(\mu_1) - 116 q^{(3)}(\mu_2)\Big), \tag{A.41}$$

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

Corollary A.5. *It follows from Proposition A.2 with $\overline{x} = x_{i-\frac{1}{2}}$ and $h = \Delta x$ that Δq_i given by Equation (A.38) and $q_{6,i}$ given by Equation (A.37) satisfy:*

$$q'(x_{i-\frac{1}{2}}) - \frac{(\Delta q_i + q_{6,i})}{\Delta x} = C_2 \Delta x^2, \tag{A.42}$$

with C_2 given by Equation (A.41), whenever $q \in C^3(\mathbb{R})$.

Now, we analyse the following expression:

$$-\frac{2q_{6,i}}{\Delta x^2} = -\frac{1}{2\Delta x^2}\Big(Q_{i-2} - 6Q_{i-1} + 10Q_i - 6Q_{i+1} + Q_{i+2}\Big). \tag{A.43}$$

deduced from Equation (A.37) and we prove in Proposition A.3 that Equation (A.43) approximates $q''(x_{i-\frac{1}{2}})$ with order 1.

Proposition A.3. *Let $q \in C^3(\mathbb{R})$, $\overline{x} \in \mathbb{R}$ and $h > 0$. Then, the following identity holds:*

$$q''(\overline{x}) = \frac{1}{2h^2}\left(-\frac{1}{h}\int_{\overline{x}-2h}^{\overline{x}-h} q(x)\,dx + \frac{6}{h}\int_{\overline{x}-h}^{\overline{x}} q(x)\,dx - \frac{10}{h}\int_{\overline{x}}^{\overline{x}+h} q(x)\,dx + \frac{6}{h}\int_{\overline{x}+h}^{\overline{x}+2h} q(x)\,dx - \frac{1}{h}\int_{\overline{x}+2h}^{\overline{x}+3h} q(x)\,dx\right) + C_3 h, \tag{A.44}$$

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

Proof. Similarly to Proposition A.2 using the same function Q, we have:

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

which consists of the finite-difference scheme from Lemma A.3. Therefore, Equation (A.44) follows from Lemma A.3 with:

$$C_3 = C_3(\mu_1, \mu_2) = \frac{1}{48}\Big(104 q^{(3)}(\mu_1) - 128 q^{(3)}(\mu_2)\Big), \tag{A.45}$$

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

Corollary A.6. *It follows from Proposition A.3 with $\overline{x} = x_{i-\frac{1}{2}}$ and $h = \Delta x$ that $q_{6,i}$ given by Equation (2.46) satisfies:*

$$q''(x_{i-\frac{1}{2}}) - \left(-\frac{2q_{6,i}}{\Delta x^2}\right) = C_3 \Delta x, \tag{A.46}$$

with C_3 given by Equation (A.45), whenever $q \in C^3(\mathbb{R})$.

With the aid of Corollaries A.4, A.5, and A.6, we are able to prove that the PPM reconstruction approximates q with order 3. Indeed, we prove this on the follow up Proposition A.4.

Proposition A.4. *Let $q \in C^4([a,b])$. Then, the Piecewise-Parabolic function given by Equation (2.40) with the parameters $q_{R,i}$ and $q_{L,i}$ obeying Equations (2.47) and (2.48) gives a third-order approximation to q on the control volume X_i. Namely, there exist constants M_1 and M_2 such that*

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

Proof. For $x \in X_i$, from Equations (A.34) and (A.33), we have:

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

Using this fact with Corollaries A.4, A.5, and A.6, we have:

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

where C_1, C_2 and C_3 are given by Equations (A.31), (A.41) and (A.45), respectively, and

$$C_4 = C_4(\theta_i) = \frac{q^{(3)}(\theta_i)}{6}. \tag{A.47}$$

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

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

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 A.2. *Replacing the formulas for $q_{R,i}$ and $q_{L,i}$ given by Equations (2.47)and (2.48) by the formulas given by Equations (2.49) and (2.50), does not change the order of convergence of the parabolic approximation.*

A.5 Convergence of 1D FV-SL schemes

A.5.1 Consistency and convergence

Hereafter, we are going to use the notations introduced in Section 2.1.1. To move towards the convergence of 1D-FV schemes, for Problem 2.4 we introduce the local truncation error (LTE hereafter) τ_i^n following LeVeque (2002):

$$Q_i(t^{n+1}) = Q_i(t^n) - \lambda\left(\mathcal{F}(Q(t^n)(\mathcal{S}_{i+\frac{1}{2}}), \tilde{u}^n_{i+\frac{1}{2}}) - \mathcal{F}(Q(t^n)(\mathcal{S}_{i-\frac{1}{2}}), \tilde{u}^n_{i-\frac{1}{2}})\right) + \Delta t\tau_i^n. \tag{A.48}$$

Notice the LTE is obtained by replacing the exact solution in Equation (2.20). Since $Q_i(t^n)$ is the exact solution of Equation (2.9), the LTE may be rewritten as

$$\tau_i^n = \frac{1}{\Delta x}\left[\left(\frac{1}{\Delta t}\int_{t^n}^{t^{n+1}} (uq)(x_{i+\frac{1}{2}}, t)\,dt - \mathcal{F}(Q(t^n)(\mathcal{S}_{i+\frac{1}{2}}), \tilde{u}^n_{i+\frac{1}{2}})\right) + \left(\frac{1}{\Delta t}\int_{t^n}^{t^{n+1}} (uq)(x_{i-\frac{1}{2}}, t)\,dt - \mathcal{F}(Q(t^n)(\mathcal{S}_{i-\frac{1}{2}}), \tilde{u}^n_{i-\frac{1}{2}})\right)\right]. \tag{A.49}$$

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.

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

Definition A.1 (Consistency). *Let us consider the framework of Problem 2.4. A 1D-FV scheme is said to be consistency in the p-norm if for any sequence of $(\Delta x^{(k)}, \Delta t^{(k)}, \lambda)$-discretizations, $k \in \mathbb{N}$, with $\lim_{k\to\infty} \Delta x^{(k)} = \lim_{k\to\infty} \Delta t^{(k)} = 0$, we have:*

$$\lim_{k\to\infty}\left[\max_{1\le n\le N_T^{(k)}} \|\tau^n\|_{p,\Delta x^{(k)}}\right] = 0,$$

and it is said to be consistent with order P in the $p-$norm if

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

From Equation (A.49), 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 2.5 we shall address how the numerical flux from PPM approximates the time-averaged flux at edges.

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

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

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

Definition A.2 (Convergence). *Let us consider the framework of Problem 2.4. A 1D-FV scheme is said to be convergent in the p-norm if for any sequence of $(\Delta x^{(k)}, \Delta t^{(k)}, \lambda)$-discretizations, $k \in \mathbb{N}$, with $\lim_{k\to\infty} \Delta x^{(k)} = \lim_{k\to\infty} \Delta t^{(k)} = 0$, we have:*

$$\lim_{k\to\infty}\left[\max_{1\le n\le N_T^{(k)}} \|E^n\|_{p,\Delta x^{(k)}}\right] = 0,$$

and it is said to converge with order P in the $p-$norm if

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

Subtracting Equation (2.20) from Equation (A.48) we get the following equation for the error:

$$E_i^{n+1} = E_i^n - \lambda\left[\left(\mathcal{F}(Q(t^n)(\mathcal{S}_{i+\frac{1}{2}}), \tilde{u}^n_{i+\frac{1}{2}}) - \mathcal{F}(Q^n(\mathcal{S}_{i+\frac{1}{2}}), \tilde{u}^n_{i+\frac{1}{2}})\right) - \left(\mathcal{F}(Q(t^n)(\mathcal{S}_{i-\frac{1}{2}}), \tilde{u}^n_{i-\frac{1}{2}}) - \mathcal{F}(Q^n(\mathcal{S}_{i-\frac{1}{2}}), \tilde{u}^n_{i-\frac{1}{2}})\right)\right] + \tau_i^n \Delta t. \tag{A.50}$$

Notice that if $q, u \in C^3$, we can rewrite Equation A.49 as:

$$\tau_i^n = \left[\frac{1}{\Delta x \Delta t}\int_{t^n}^{t^{n+1}}\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} \frac{\partial(uq)}{\partial x}(x,t)\,dx\,dt - \left(\frac{\mathcal{F}(Q(t^n)(\mathcal{S}_{i+\frac{1}{2}}), \tilde{u}^n_{i-\frac{1}{2}})) - \mathcal{F}(Q(t^n)(\mathcal{S}_{i+\frac{1}{2}}), \tilde{u}^n_{i-\frac{1}{2}})}{\Delta x}\right)\right].$$

Using the midpoint rule for integration (Theorem A.3) and the mean value theorem for integrals (Theorem A.2), we have:

$$\begin{aligned}\tau_i^n &= \left[\frac{1}{\Delta t}\int_{t^n}^{t^{n+1}} \left(\frac{\partial(uq)}{\partial x}(x_i,t) + \frac{\Delta x^2}{24}\frac{\partial(uq)}{\partial x}(\xi,t)\right)dt - \left(\frac{\mathcal{F}(Q(t^n)(\mathcal{S}_{i+\frac{1}{2}}), \tilde{u}^n_{i+\frac{1}{2}})) - \mathcal{F}(Q(t^n)(\mathcal{S}_{i-\frac{1}{2}}), \tilde{u}^n_{i-\frac{1}{2}})}{\Delta x}\right)\right] \\ &= \left[\frac{1}{\Delta t}\int_{t^n}^{t^{n+1}} \frac{\partial(uq)}{\partial x}(x_i,t)\,dt - \left(\frac{\mathcal{F}(Q(t^n)(\mathcal{S}_{i+\frac{1}{2}}), \tilde{u}^n_{i+\frac{1}{2}})) - \mathcal{F}(Q(t^n)(\mathcal{S}_{i-\frac{1}{2}}), \tilde{u}^n_{i-\frac{1}{2}})}{\Delta x}\right)\right] + \frac{\Delta x^2}{24}\frac{\partial^3(uq)}{\partial x^3}(\xi,\bar{t}),\end{aligned} \tag{A.51}$$

for $\xi \in X_i$ and $\bar{t} \in [t^n, t^{n+1}]$. Therefore, if $q, u \in C^3$ the scheme is consistent, if and only if, $\frac{1}{\Delta t}\int_{t^n}^{t^{n+1}} \frac{\partial(uq)}{\partial x}(x_i,t)\,dt$ is approximated by $\frac{\mathcal{F}(Q(t^n)(\mathcal{S}_{i+\frac{1}{2}}), \tilde{u}^n_{i+\frac{1}{2}})) - \mathcal{F}(Q(t^n)(\mathcal{S}_{i-\frac{1}{2}}), \tilde{u}^n_{i-\frac{1}{2}})}{\Delta x}$. This shall be very useful when we consider two-dimensional schemes, where we are going to use the discrete operators to estimate the divergence of velocity fields.

A.5.2 Stability

In order to define the concept of stability, it is useful to introduce an operator representation of 1D-FV schemes. In the context of Problem 2.4, we define the operators $\mathcal{H}_{\Delta x,n} : \mathbb{R}^{\Delta x} \to \mathbb{R}^{\Delta x}$ whose i-th entry is given by:

$$[\mathcal{H}_{\Delta x,n}(Q)]_i = Q_i - \lambda\left(\mathcal{F}(Q(\mathcal{S}_{i+\frac{1}{2}}), \tilde{u}^n_{i+\frac{1}{2}}) - \mathcal{F}(Q(\mathcal{S}_{i-\frac{1}{2}}), \tilde{u}^n_{i-\frac{1}{2}})\right), \tag{A.52}$$

for $i = 1, \dots, N$, $n = 0, \dots, N_T - 1$. 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 $\mathcal{H}_{\Delta x,n}$. Thus, Equation (2.20) may be rewritten in a vector form by

$$Q^{n+1} = \mathcal{H}_{\Delta x,n}(Q^n),$$

and Equation (A.48) in a vector form reads

$$Q(t^{n+1}) = \mathcal{H}_{\Delta x,n}(Q(t^n)) + \Delta t \tau^n,$$

and the error equation (A.50) is given by

$$E^{n+1} = \mathcal{H}_{\Delta x,n}(Q(t^n)) - \mathcal{H}_{\Delta x,n}(Q^n) + \Delta t \tau^n. \tag{A.53}$$

The stability theory focus on uniformly bounding the norm of $\mathcal{H}_{\Delta x,n}(Q(t^n)) - \mathcal{H}_{\Delta x,n}(Q^n)$ (LeVeque, 2002). We define stability as follows.

Definition A.3 (Stability)**.** *In the context of Problem 2.4, a 1D-FV scheme is stable in the $p-$norm if for any $(\Delta x, \Delta t, \lambda)$-discretization of $[a,b] \times [0,T]$ we have:*

$$\|\mathcal{H}_{\Delta x,n}(Q) - \mathcal{H}_{\Delta x,n}(P)\|_{p,\Delta x} \leq (1 + \alpha \Delta t)\|Q - P\|_{p,\Delta x}, \tag{A.54}$$

for all $Q, P \in \mathbb{R}^{\Delta x}$ and α is a constant that does not depend neither on Δx nor on Δt.

Assuming that the scheme is stable in the $p-$norm, then it follows from Equation (A.53) that:

$$\begin{aligned}
\|E^{n+1}\|_{p,\Delta x} &\leq \|\mathcal{H}_{\Delta x,n}(Q(t^n)) - \mathcal{H}_{\Delta x,n}(Q^n)\|_{p,\Delta x} + \Delta t \max_{n=1,\dots,N_T} \|\tau^n\|_{p,\Delta x} \\
&\leq (1 + \alpha \Delta t)\|E^n\|_{p,\Delta x} + \Delta t \max_{n=1,\dots,N_T} \|\tau^n\|_{p,\Delta x} \\
&\leq (1 + \alpha \Delta t)^n \|E^0\|_{p,\Delta x} + \Delta t \max_{n=1,\dots,N_T} \|\tau^n\|_{p,\Delta x} \sum_{k=0}^{n-1} (1 + \alpha \Delta t)^k \\
&\leq e^{\alpha T}(\|E^0\|_{p,\Delta x} + T \max_{n=1,\dots,N_T} \|\tau^n\|_{p,\Delta x}),
\end{aligned} \tag{A.55}$$

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 2.2. 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 $\mathcal{H}_{\Delta x,n}$ 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 2.5, the operator $\mathcal{H}_{\Delta x,n}$ may become non-linear when monotonicity constraints are activated.

Notice that, if $\mathcal{H}_{\Delta x,n}$ is linear, then stability is equivalent to require that

$$\|\mathcal{H}_{\Delta x,n}\|_{p,\Delta x} \leq 1 + \alpha \Delta t,$$

where

$$\|\mathcal{H}_{\Delta x,n}\|_{p,\Delta x} = \sup_{Q \in \mathbb{R}^{\Delta x}} \frac{\|\mathcal{H}_{\Delta x,n}(Q)\|_{p,\Delta x}}{\|Q\|_{p,\Delta x}},$$

is the operator p-norm.

For linear operators, we may use the discrete Fourier transform (Trefethen, 2000) to estimate the 2-norm of $\mathcal{H}_{\Delta x,n}$. This approach is known as Von Neumann stability analysis. We define the nodes $\theta_i = i\frac{2\pi}{N}$, $i = 1, \dots, N$, $\Delta\theta = \frac{2\pi}{N}$, $\theta = (\theta_1, \theta_2, \dots, \theta_N)$. The imaginary unit

is denoted by $\imath$. The Fourier modes are given by:

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

for $k = 1, \dots, N$. 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 $Q \in \mathbb{R}^{\Delta x}$, 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:

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

The idea of Von Neumann stability analysis is to apply the operator $\mathcal{H}_{\Delta x,n}$ 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 $\mathcal{H}_{\Delta x,n}$ 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:

$$\mathcal{H}_{\Delta x,n}(e^{\imath k\theta}) = \rho(k) e^{\imath k\theta},$$

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

$$\|\mathcal{H}_{\Delta x,n}(Q)\|_{2,\Delta x}^2 = N \sum_{k=1}^{N} |a_k|^2 |\rho(k)|^2 \leq \max_{k=1,\dots,N} |\rho(k)|^2 \|Q\|_{2,\Delta x}^2.$$

Therefore:

$$\|\mathcal{H}_{\Delta x,n}\|_{2,\Delta x} \leq \max_{k=1,\dots,N} |\rho(k)|.$$

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

$$\mathcal{F}(Q(\mathcal{S}_{i+\frac{1}{2}}), \tilde{u}^n_{i+\frac{1}{2}}) = \sum_{l \in \mathcal{S}_{i+\frac{1}{2}}} \alpha_{l,i} Q_l,$$

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

$$
\begin{aligned}
[\mathcal{H}_{\Delta x,n}(e^{\imath k\theta})]_i &= e^{\imath k\theta_i} - \lambda\left(\sum_{l\in S_{i+\frac{1}{2}}} \alpha_{l,i} e^{\imath k\theta_l} - \sum_{l\in S_{i-\frac{1}{2}}} \alpha_{l,i-1} e^{\imath k\theta_{l-1}}\right) \\
&= e^{\imath k\theta_i}\left(1 - \lambda\left(\sum_{l\in S_{i+\frac{1}{2}}} \alpha_{l,i} e^{\imath k\theta_{l-i}} - \sum_{l\in S_{i-\frac{1}{2}}} \alpha_{l,i-1} e^{\imath k\theta_{l-1-i}}\right)\right).
\end{aligned}
$$

Hence, the amplification factor has the form

$$
\rho(k) = 1 - \lambda\left(\sum_{l\in S_{i+\frac{1}{2}}} \alpha_{l,i} e^{\imath k\theta_{l-i}} - \sum_{l\in S_{i-\frac{1}{2}}} \alpha_{l,i-1} e^{\imath k\theta_{l-1-i}}\right). \tag{A.56}
$$

In Section 2.5 we shall analyse $|\rho(k)|$ in terms of the PPM coefficients.

A.5.3 Flux accuracy analysis

With the stencil coefficients, we can compute the amplification factor (Equation (A.56)) 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 A.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.

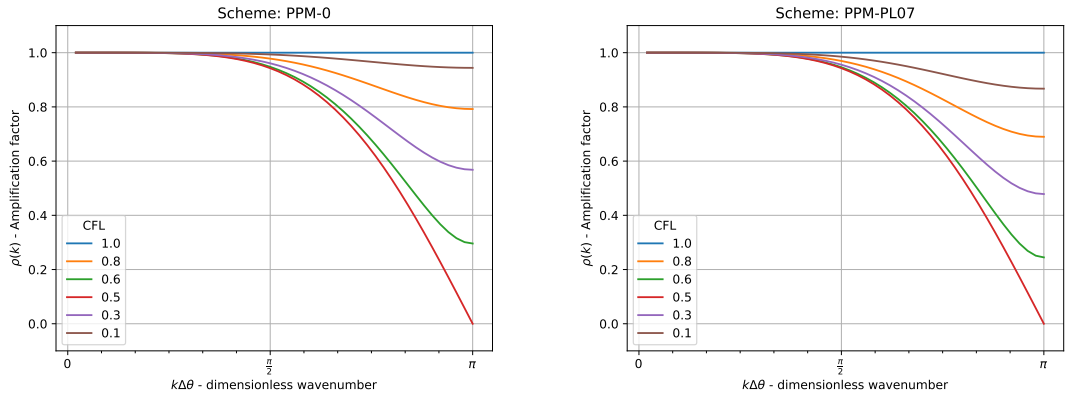


Figure A.1: *Amplification factor for the PPM (left) and hybrid PPM (right) schemes for different CFL numbers.*

A.6 Convergence, consistency and stability of 2D-FV schemes

The notions of convergence, consistency and stability for a 2D-FV schemes are straightforward from these notions for 1D-FV schemes (see Subsections A.5.1 and A.5.2). Indeed, in the context of Problem 3.3, we define the operators $\mathcal{H}_{\Delta x,\Delta y,n} : \mathbb{R}^{N\times N} \to \mathbb{R}^{N\times M}$ whose (i, j) entry is given by:

$$[\mathcal{H}_{\Delta x,\Delta y,n}(Q)]_{ij} = Q_{ij} - \Delta t \mathbb{D}^n_{ij}$$

for $i = 1, \dots, N$, $j = 1, \dots, M$, $n = 0, \dots, N_T - 1$. The 2D-FV is then expressed as

$$Q^{n+1} = \mathcal{H}_{\Delta x,\Delta y,n}(Q^n).$$

The local error truncation $\tau^n \in \mathbb{R}^{N\times M}$ is given by

$$Q(t^{n+1}) = \mathcal{H}_{\Delta x,\Delta y,n}(Q(t^n)) + \Delta t \tau^n.$$

The error equation is given by

$$E^{n+1} = \mathcal{H}_{\Delta x,\Delta y,n}(Q(t^n)) - \mathcal{H}_{\Delta x,\Delta y,n}(Q^n) + \Delta t \tau^n. \tag{A.57}$$

The stability in the p-norm is defined as in the 1D case.

Definition A.4. *A 2D-FV scheme is stable in the $p-$norm if*

$$\|\mathcal{H}_{\Delta x,\Delta y,n}(Q) - \mathcal{H}_{\Delta x,\Delta y,n}(P)\|_{p,\Delta x\times\Delta y} \leq (1 + \alpha\Delta t)\|Q - P\|_{p,\Delta x\times\Delta y}, \tag{A.58}$$

for all $Q, P \in \mathbb{R}^{N\times M}$ and α is a constant that does not depend neither on Δx, Δy, Δt nor on n.

If a 2D-FV scheme is stable in the $p-$norm, similarly to Equation (A.55) we have:

$$\|E^{n+1}\|_{p,\Delta x\times\Delta y} \leq e^{\alpha T}(\|E^0\|_{p,\Delta x\times\Delta y} + T \max_{n=1,\dots,N_T} \|\tau^n\|_{p,\Delta x\times\Delta y}).$$

Again, we point out that from Proposition 3.1, we have that the initial error E^0 shall be second-order accurate. Consistency is defined as in Definition A.1 and convergence is defined as in Definition A.2.

The Von Neumann analysis can be applied when $\mathcal{H}_{\Delta x,\Delta y,n}$ is linear, since we are considering periodic boundary conditions. The idea is the same as in the one-dimensional case, we just apply the operator $\mathcal{H}_{\Delta x,\Delta y,n}$ on the Fourier modes to obtain the amplification factor. We introduce the nodes $\theta_i = i\frac{2\pi}{N}$, $i = 1, \dots, N$, $\Delta\theta = \frac{2\pi}{N}$, $\theta_i = (\theta_1, \theta_2, \dots, \theta_N)$, $\phi_j = j\frac{2\pi}{M}$, $j = 1, \dots, M$, $\Delta\phi = \frac{2\pi}{M}$, $\phi = (\phi_1, \phi_2, \dots, \phi_M)$. For $k_1 = 1, \dots, N$, $k_2 = 1, \dots, M$, the two-dimensional Fourier mode $\boldsymbol{k} = (k_1, k_2)$ from $\mathbb{C}^{N\times M}$ has its (i, j) entry given by $[e^{\imath \boldsymbol{k}\theta}]_{ij} = e^{\imath k_1\theta_i} e^{\imath k_2\phi_j}$.

Notice that if $q, u, v \in C^3$, we can rewrite the LTE as:

$$\tau_{ij}^n = \left[\frac{1}{\Delta x \Delta y \Delta t} \int_{t^n}^{t^{n+1}} \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} \int_{y_{j-\frac{1}{2}}}^{y_{j+\frac{1}{2}}} \nabla \cdot (\boldsymbol{u}q)(x, y, t)\, dy\, dx\, dt - \mathbb{D}_{ij}^n \right].$$

Using the midpoint rule for integration (Theorem A.4), the mean value theorem for integrals (Theorem A.2) and recalling the discrete divergence (Definition 3.5), we have:

$$\tau_{ij}^n = \frac{1}{\Delta t} \int_{t^n}^{t^{n+1}} \nabla \cdot (\boldsymbol{u}q)(x_i, y_j, t)\, dt - \mathbb{D}_{ij}^n + O(\Delta x^2) + O(\Delta y^2). \tag{A.59}$$

Therefore, in order to investigate the consistency, we may compare how well the discrete divergence approximates the divergence.

Appendix B

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-dimension-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.

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 pp. 4, 9)

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, 8, 18)

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. 49)

Chen, Y., Weller, H., Pring, S., & Shaw, J. (2017). Comparison of dimensionally split and multi-dimensional atmospheric transport schemes for long time steps. *Quarterly Journal of the Royal Meteorological Society*, *143*(708), 2764–2779. https://doi.org/https://doi.org/10.1002/qj.3125 (cit. on pp. 23, 29, 44, 45)

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, 17, 18, 20–23, 28)

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)

Courant, R., & John, F. (1999). In *Introduction to calculus and analysis i*. Springer Berlin, Heidelberg. https://doi.org/https://doi.org/10.1007/978-3-642-58604-0. (Cit. on p. 71)

Croisille, J.-P. (2013). Hermitian compact interpolation on the cubed-sphere grid. *Journal of Scientific Computing*, *57*. https://doi.org/10.1007/s10915-013-9702-3 (cit. on p. 49)

Csomós, P., Faragó, I., & Havasi, Á. (2005). Weighted sequential splittings and their analysis [Numerical Methods and Computational Mechanics]. *Computers and Mathematics with Applications*, *50*(7), 1017–1031. https://doi.org/https://doi.org/10.1016/j.camwa.2005.08.004 (cit. on p. 38)

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)

Durran, D. (2011). Time discretization: Some basic approaches. In *Numerical techniques for global atmospheric models* (pp. 75–104). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-11640-7_5. (Cit. on p. 18)

Durran, D. R. (2010). Semi-lagrangian methods. In *Numerical methods for fluid dynamics: With applications to geophysics* (pp. 357–391). Springer New York. https://doi.org/10.1007/978-1-4419-6412-0_7. (Cit. on p. 18)

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 pp. 8, 10, 20)

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 pp. 7, 20).

Guo, W., Nair, R. D., & Qiu, J.-M. (2014). A conservative semi-lagrangian discontinuous galerkin scheme on the cubed sphere. *Monthly Weather Review*, *142*(1), 457–475. https://doi.org/10.1175/MWR-D-13-00048.1 (cit. on p. 18)

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. 8, 22)

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)

Holden, H., Karlsen, K., Lie, K.-A., & Risebro, H. (2010). *Splitting methods for partial differential equations with rough solutions: Analysis and matlab programs*. https://doi.org/10.4171/078. (Cit. on p. 38)

Jia, H., & Li, K. (2011). A third accurate operator splitting method. *Mathematical and Computer Modelling*, *53*(1), 387–396. https://doi.org/https://doi.org/10.1016/j.mcm.2010.09.005 (cit. on p. 38)

Jung, J.-H., Konor, C. S., & Randall, D. (2019). Implementation of the vector vorticity dynamical core on cubed sphere for use in the quasi-3-d multiscale modeling framework. *Journal of Advances in Modeling Earth Systems*, *11*(3), 560–577. https://doi.org/https://doi.org/10.1029/2018MS001517 (cit. on p. 52)

Katta, K. K., Nair, R. D., & Kumar, V. (2015a). High-order finite volume shallow water model on the cubed-sphere: 1d reconstruction scheme. *Applied Mathematics and Computation*, *266*, 316–327. https://doi.org/https://doi.org/10.1016/j.amc.2015.04.053 (cit. on p. 49)

Katta, K. K., Nair, R. D., & Kumar, V. (2015b). High-order finite-volume transport on the cubed sphere: Comparison between 1d and 2d reconstruction schemes. *Monthly Weather Review*, *143*(7), 2937–2954. https://doi.org/https://doi.org/10.1175/MWR-D-13-00176.1 (cit. on p. 49)

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. 8)

Leonard, B. P., Lock, A. P., & MacVean, M. K. (1996). Conservative explicit unrestricted-time-step multidimensional constancy-preserving advection schemes. *Monthly Weather Review*, *124*(11), 2588–2606. https://doi.org/https://doi.org/10.1175/1520-0493(1996)124<2588:CEUTSM>2.0.CO;2 (cit. on p. 7)

LeVeque, R. J. (1985). A large time step generalization of godunov's method for systems of conservation laws. *SIAM Journal on Numerical Analysis*, *22*(6), 1051–1073. https://doi.org/10.1137/0722063 (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. 11, 43)

LeVeque, R. J. (2002). *Finite volume methods for hyperbolic problems*. Cambridge University Press. https://doi.org/10.1017/CBO9780511791253. (Cit. on pp. 11, 19, 79, 82)

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, 8, 22–24, 28, 42)

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, 7, 18, 23, 29, 37, 38, 41, 42, 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)

Lu, F., Zhang, F., Wang, T., Tian, G., & Wu, F. (2022). High-order semi-lagrangian schemes for the transport equation on icosahedron spherical grids. *Atmosphere*, *13*(11). https://doi.org/10.3390/atmos13111807 (cit. on p. 18)

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. 26, 44)

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 pp. 5, 58)

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, 8, 21, 28, 42, 49, 60, 65)

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 pp. 49, 56)

Rančić, M. (1992). Semi-lagrangian piecewise biparabolic scheme for two-dimensional horizontal advection of a passive scalar. *Monthly Weather Review*, *120*(7), 1394–1406. https://doi.org/10.1175/1520-0493(1992)120<1394:SLPBSF>2.0.CO;2 (cit. on p. 29)

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 pp. 4, 50)

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)

Richtmyer, R. D., & Morton, K. W. (1968). Difference methods for initial-value problems. *SIAM Review*, *10*(3), 381–383. https://doi.org/10.1137/1010073 (cit. on p. 37)

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, 49, 51, 60)

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, 49, 50, 60)

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)

Stoer, J., & Bulirsch, R. (2002). In *Introduction to numerical analysis*. Springer New York, NY. https://doi.org/https://doi.org/10.1007/978-0-387-21738-3. (Cit. on pp. 19, 71)

Strang, G. (1968). On the construction and comparison of difference schemes. *SIAM Journal on Numerical Analysis*, *5*(3), 506–517. https://doi.org/10.1137/0705041 (cit. on p. 38)

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. 83)

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 p. 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. 26, 82)

Tumolo, G. (2011). *A semi-implicit, semi-lagrangian, p-adaptative discontinuous galerkin method for the rotating shallow-water equations: Analysis and numerical experiments* (Doctoral dissertation). University of Trieste. https://core.ac.uk/download/pdf/41173373.pdf. (Cit. on p. 18)

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, 18, 20)

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. 8)

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. 8)

Zerroukat, M., & Allen, T. (2022). On the corners of the cubed-sphere grid. *Quarterly Journal of the Royal Meteorological Society*, *148*(743), 778–783. https://doi.org/https://doi.org/10.1002/qj.4230 (cit. on p. 57)

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)