An asymptotic-preserving active flux scheme for the hyperbolic heat equation in the diffusive scaling

Junming Duan[1], Wasilij Barsukow[2], Christian Klingenberg[3]

Abstract

The Active Flux (AF) method is a compact, high-order finite volume scheme that enhances flexibility by introducing point values at cell interfaces as additional degrees of freedom alongside cell averages. The method of lines is employed here for temporal discretization. A common approach for updating point values relies on the Jacobian Splitting (JS) method, which incorporates upwinding. A key advantage of the AF method over standard finite volume schemes is its structure-preserving property, motivating the investigation of its asymptotic-preserving (AP) behavior in the diffusive scaling. We show that the JS-based AF method without any modification is AP for solving the hyperbolic heat equation, in the sense that the limit scheme is a discretization of the limit heat equation. We use formal asymptotic analysis, discrete Fourier analysis, and numerical experiments to illustrate our findings.

Keywords: hyperbolic heat equation, finite volume method, active flux, asymptotic-preserving, asymptotic analysis, Fourier analysis

Mathematics Subject Classification (2020): 65M08, 65M12, 65M20, 35L65

Contents

1 Introduction **2**

2 Hyperbolic heat equation and its limit in the diffusive scaling **3**

3 1D active flux scheme **4**
3.1 Update of cell average 4
3.2 Update of point value 5

4 Formal asymptotic analysis **5**

5 An alternative point value update **7**

6 Fourier analysis **8**
6.1 Fourier analysis of the PDEs 8
6.2 Fourier analysis of the numerical scheme 9
6.2.1 Point value update using Jacobian splitting 9
6.2.2 Point value update using the alternating flux 12
6.2.3 Discussion 13

[1]Institute of Mathematics, University of Würzburg, Emil-Fischer-Straße 40, 97074 Würzburg, Germany, junming.duan@uni-wuerzburg.de

[2]Institut de Mathématiques de Bordeaux (IMB), CNRS UMR 5251, University of Bordeaux, 33405 Talence, France, wasilij.barsukow@math.u-bordeaux.fr

[3]Institute of Mathematics, University of Würzburg, Emil-Fischer-Straße 40, 97074 Würzburg, Germany, christian.klingenberg@uni-wuerzburg.de

7 2D active flux scheme **14**
7.1 Formal asymptotic analysis 15

8 Numerical results **18**

9 Conclusion **21**

A Example 8.1 with non-well-prepared data **26**

1 Introduction

Particle systems in physics, such as rarefied gases and neutron transport, can be modeled at different scales. At a microscopic level, particles move and collide individually, and the system is often described by kinetic models, which provide a statistical description using probability distributions. At a macroscopic level, the mean free path, or the average distance between collisions, is small compared to a macroscopic length, and the particles behave like fluids. In this case, macroscopic models such as fluid or diffusion equations offer accurate and computationally efficient approximations. The passage from microscopic to macroscopic descriptions can be formally derived via asymptotic analysis.

Numerically, multiscale problems where the mean free path varies across several orders of magnitude pose significant challenges. Traditional kinetic solvers require grid resolutions smaller than the mean free path, making computations prohibitively expensive. To address this, asymptotic preserving (AP) schemes have been developed to seamlessly transition between kinetic and macroscopic models. Their stability and convergence are independent of the mean free path, and allow coarse grids, which are crucial for applications such as inertial confinement fusion. The AP schemes were first proposed in [33, 32] for the steady-state solutions to neutron transport equation in the diffusive regime and have been extended to a broad range of kinetic models. Some AP finite volume schemes were constructed based on careful modification of numerical flux [26] and/or upwinding treatment of source term [20]. Different reformulations were proposed for designing AP schemes, e.g., even-odd decomposition by rewriting linear transport equations as parity equations [30, 27]. Another popular way is micro-macro decomposition by decomposing the density distribution function into the local Maxwellian plus the deviation, e.g., for the radiative heat transfer equations [31], linear kinetic equations [34], nonlinear Boltzmann equation [10], and so on. High-order AP discontinuous Galerkin (DG) schemes were developed using micro-macro decomposition and several different numerical fluxes [24, 35]. For a more comprehensive review, we refer readers to [25] and references therein.

The active flux (AF) method is a new compact finite volume method [16, 15, 17, 36], inspired by [37]. It simultaneously evolves cell averages and additional degrees of freedom, chosen as point values located at cell interfaces. Thanks to these point values, the AF method does not need Riemann solvers (unlike Godunov methods), as the numerical solution is continuous across the cell interface. The AF methods can be roughly divided into two classes based on the evolution of the point value. The original ones evolve the cell average through Simpson's rule for flux quadrature in time, and employ exact or approximate evolution operators to evolve the point values and to obtain the numerical solutions at the flux quadrature points. Examples are exact evolution operators for linear equations [9, 19, 17, 37], p-system [18], and approximate evolution operators for Burgers' equation

[16, 15, 36, 6], the compressible Euler equations in one spatial dimension [16, 23, 6], multidimensional Euler equations [18], and hyperbolic balance laws [8, 7], etc. The method of bicharacteristics was used for the derivation of truly multidimensional approximative evolution operators [12]. The other so-called generalized, or semi-discrete AF methods adopt a method of lines, where the evolution of the cell average and point value is written in semi-discrete form and integrated in time by using Runge-Kutta methods. Examples of this approach are [1, 2, 3, 4] based on Jacobian splitting (JS) and [13, 14] based on flux vector splitting.

The AF method is superior to standard finite volume methods due to its structure-preserving property. For multi-dimensional acoustic equations, it preserves the vorticity and stationary states [9], and for acoustics with gravity, it is naturally well-balanced [8]. These encouraging results lead us to the central question of this paper: Does the AF method also possess the AP property in the diffusive scaling? To start, we examine the hyperbolic heat equation, the lowest-order angular discretization of the transport equation. This equation is also known as the P_1 model, the first-order formulation of the telegraph equation, or the damped wave equation [11]. We present formal asymptotic analysis and discrete Fourier analysis for the JS-based AF scheme for solving the hyperbolic heat equation, and show that it is AP. Several 1D and 2D numerical results verify our theoretical findings. The key feature of such JS-based AF method is that it is automatically AP without any modification.

The remainder of this paper is structured as follows. Section 2 introduces the hyperbolic heat equation and its limit heat equation by using formal asymptotic analysis. Section 3 gives the 1D AF scheme based on the JS for the point value update, and Section 4 derives the limit scheme via formal asymptotic analysis. As we observe order degradation for the JS-based AF scheme in the limit, an alternative point value update is discussed in Section 5. A discrete Fourier analysis is adopted to study the convergence order of the scheme and its limit in Section 6. Section 7 discusses the 2D case. Numerical tests are conducted in Section 8 to experimentally demonstrate the accuracy and AP property. Section 9 concludes the paper with final remarks.

2 Hyperbolic heat equation and its limit in the diffusive scaling

The hyperbolic heat equation in the diffusive scaling reads

$$\begin{cases} p_t + \dfrac{1}{\epsilon}\nabla \cdot \boldsymbol{u} = 0, \\ \boldsymbol{u}_t + \dfrac{1}{\epsilon}\nabla p + \dfrac{\sigma}{\epsilon^2}\boldsymbol{u} = \boldsymbol{0}, \end{cases} \tag{1}$$

where $0 < \epsilon < 1$ and $\sigma > 0$.

To study the asymptotic limit of (1) as $\epsilon \to 0$, we assume sufficient regularity of the solutions and perform a formal analysis by expanding them as a power series in ϵ,

$$p = p^{(0)} + \epsilon p^{(1)} + \epsilon^2 p^{(2)} + \dots, \qquad \boldsymbol{u} = \boldsymbol{u}^{(0)} + \epsilon \boldsymbol{u}^{(1)} + \epsilon^2 \boldsymbol{u}^{(2)} + \dots. \tag{2}$$

The asymptotic behavior is determined by inserting (2) into (1). Matching the terms of

the same order, we have

$$\begin{aligned}\epsilon^{-2} &: \boldsymbol{u}^{(0)} = \mathbf{0},\\ \epsilon^{-1} &: \nabla p^{(0)} + \sigma \boldsymbol{u}^{(1)} = \mathbf{0},\\ \epsilon^{0} &: p_t^{(0)} + \nabla \cdot \boldsymbol{u}^{(1)} = 0.\end{aligned}$$

Thus, the leading order solutions satisfy

$$p_t^{(0)} = \nabla \cdot \left(\frac{1}{\sigma} \nabla p^{(0)} \right), \quad \boldsymbol{u}^{(0)} = \mathbf{0},$$

i.e., $p^{(0)}$ is the solution to the limit heat equation.

3 1D active flux scheme

This section presents the 1D semi-discrete AF methods for the 1D hyperbolic heat equation,

$$\begin{cases} p_t + \dfrac{1}{\epsilon} u_x = 0, \\ u_t + \dfrac{1}{\epsilon} p_x + \dfrac{\sigma}{\epsilon^2} u = 0. \end{cases} \tag{3}$$

As the advection and source term are both stiff as $\epsilon \ll 1$, the fully-discrete methods are obtained using the 3rd-order four-stage Diagonally Implicit Runge-Kutta (DIRK) method from [28, 29], which is stiffly accurate.

3.1 Update of cell average

Assume that a 1D computational domain is divided into N uniform cells $I_i = [x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}]$ with the cell size $\Delta x = x_{i+\frac{1}{2}} - x_{i-\frac{1}{2}}$ and cell centers $x_i = (x_{i-\frac{1}{2}} + x_{i+\frac{1}{2}})/2$. The degrees of freedom consist of the cell averages and point values at the cell interfaces,

$$\begin{aligned}&\bar{p}_i(t) = \frac{1}{\Delta x} \int_{I_i} p_h(x,t) \mathrm{d}x, \quad \bar{u}_i(t) = \frac{1}{\Delta x} \int_{I_i} u_h(x,t) \mathrm{d}x,\\ &p_{i+\frac{1}{2}}(t) = p_h(x_{i+\frac{1}{2}}, t), \quad u_{i+\frac{1}{2}}(t) = u_h(x_{i+\frac{1}{2}}, t),\end{aligned}$$

where $p_h(x,t)$ and $u_h(x,t)$ are numerical approximations. As for finite volume methods, the update of the cell average is found upon integrating (3) over I_i,

$$\frac{\mathrm{d}}{\mathrm{d}t} \bar{p}_i = -\frac{1}{\epsilon \Delta x} \left(u_{i+\frac{1}{2}} - u_{i-\frac{1}{2}} \right), \tag{4}$$

$$\frac{\mathrm{d}}{\mathrm{d}t} \bar{u}_i = -\frac{1}{\epsilon \Delta x} \left(p_{i+\frac{1}{2}} - p_{i-\frac{1}{2}} \right) - \frac{\sigma}{\epsilon^2} \bar{u}_i. \tag{5}$$

Note that the flux at the cell interface is available directly, without the need for a Riemann solver.

3.2 Update of point value

For the point value evolution, the Jacobian splitting (JS) [2] is adopted. The Jacobian matrix of (3) can be diagonalized,

$$\boldsymbol{J} = \frac{1}{\epsilon}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} \frac{1}{\epsilon} & 0 \\ 0 & -\frac{1}{\epsilon} \end{pmatrix}\begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{pmatrix} =: \boldsymbol{R}\boldsymbol{\Lambda}\boldsymbol{R}^{-1},$$

and split into positive and negative parts as

$$\boldsymbol{J}^{+} = \boldsymbol{R}\boldsymbol{\Lambda}^{+}\boldsymbol{R}^{-1} = \frac{1}{\epsilon}\begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}, \quad \boldsymbol{J}^{-} = \boldsymbol{R}\boldsymbol{\Lambda}^{-}\boldsymbol{R}^{-1} = \frac{1}{\epsilon}\begin{pmatrix} -\frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{pmatrix},$$

such that $\boldsymbol{\Lambda}^{\pm} = \frac{1}{2}\left(\boldsymbol{\Lambda} \pm |\boldsymbol{\Lambda}|\right)$. Then the point value update can be written as

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix} p_{i+\frac{1}{2}} \\ u_{i+\frac{1}{2}} \end{pmatrix} = -\left[\boldsymbol{J}^{+} D^{+}_{i+\frac{1}{2}}\begin{pmatrix} p \\ u \end{pmatrix} + \boldsymbol{J}^{-} D^{-}_{i+\frac{1}{2}}\begin{pmatrix} p \\ u \end{pmatrix}\right] - \frac{\sigma}{\epsilon^2}\begin{pmatrix} 0 \\ u_{i+\frac{1}{2}} \end{pmatrix},$$

where $D^{+}_{i+\frac{1}{2}}$ and $D^{-}_{i+\frac{1}{2}}$ are component-wise upwind finite difference operators. These are obtained by first reconstructing a parabola $u_{\mathtt{para}}(x)$ in the cell I_i satisfying

$$\frac{1}{\Delta x}\int_{I_i} u_{\mathtt{para}}(x)\mathrm{d}x = \bar{u}_i, \quad u_{\mathtt{para}}(x_{i\pm\frac{1}{2}}) = u_{i\pm\frac{1}{2}},$$

whose derivative at $x_{i+\frac{1}{2}}$ yields a high-order accurate upwind finite difference

$$D^{+}_{i+\frac{1}{2}}(u) = u'_{\mathtt{para}}(x_{i+\frac{1}{2}}) = \frac{1}{\Delta x}\left(2u_{i-\frac{1}{2}} - 6\bar{u}_i + 4u_{i+\frac{1}{2}}\right).$$

For negative eigenvalues, the reconstruction in cell I_{i+1} is differentiated at $x_{i+\frac{1}{2}}$ instead, yielding

$$D^{-}_{i+\frac{1}{2}}(u) = \frac{1}{\Delta x}\left(-4u_{i+\frac{1}{2}} + 6\bar{u}_{i+1} - 2u_{i+\frac{3}{2}}\right).$$

Therefore, the point value update can be rewritten as

$$\frac{\mathrm{d}}{\mathrm{d}t}p_{i+\frac{1}{2}} = -\frac{1}{2\epsilon}\Big[(D^{+}_{i+\frac{1}{2}} - D^{-}_{i+\frac{1}{2}})p + (D^{+}_{i+\frac{1}{2}} + D^{-}_{i+\frac{1}{2}})u\Big], \tag{6}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}u_{i+\frac{1}{2}} = -\frac{1}{2\epsilon}\Big[(D^{+}_{i+\frac{1}{2}} + D^{-}_{i+\frac{1}{2}})p + (D^{+}_{i+\frac{1}{2}} - D^{-}_{i+\frac{1}{2}})u\Big] - \frac{\sigma}{\epsilon^2}u_{i+\frac{1}{2}}. \tag{7}$$

4 Formal asymptotic analysis

Expand sufficiently smooth numerical solutions as a power series in ϵ,

$$\bar{p}_i = \bar{p}_i^{(0)} + \epsilon\bar{p}_i^{(1)} + \epsilon^2\bar{p}_i^{(2)} + \ldots, \qquad \bar{u}_i = \bar{u}_i^{(0)} + \epsilon\bar{u}_i^{(1)} + \epsilon^2\bar{u}_i^{(2)} + \ldots,$$

$$p_{i+\frac{1}{2}} = p_{i+\frac{1}{2}}^{(0)} + \epsilon p_{i+\frac{1}{2}}^{(1)} + \epsilon^2 p_{i+\frac{1}{2}}^{(2)} + \ldots, \qquad u_{i+\frac{1}{2}} = u_{i+\frac{1}{2}}^{(0)} + \epsilon u_{i+\frac{1}{2}}^{(1)} + \epsilon^2 u_{i+\frac{1}{2}}^{(2)} + \ldots,$$

and substitute them into the schemes (4)-(7). After equating the coefficients of equal powers of ϵ, one has

$$
\begin{aligned}
\epsilon^{-2}: \quad & \bar{u}_i^{(0)} = u_{i+\frac{1}{2}}^{(0)} = 0, \\
\epsilon^{-1}: \quad & \bar{u}_i^{(1)} = -\frac{1}{\sigma\Delta x}\left(p_{i+\frac{1}{2}}^{(0)} - p_{i-\frac{1}{2}}^{(0)}\right), \\
& u_{i+\frac{1}{2}}^{(1)} = -\frac{1}{2\sigma}\Big[(D_{i+\frac{1}{2}}^{+} + D_{i+\frac{1}{2}}^{-})p^{(0)} + (D_{i+\frac{1}{2}}^{+} - D_{i+\frac{1}{2}}^{-})u^{(0)}\Big] \\
& \qquad = -\frac{1}{2\sigma}\left(D_{i+\frac{1}{2}}^{+} + D_{i+\frac{1}{2}}^{-}\right)p^{(0)}, \\
\epsilon^{0}: \quad & \frac{\mathrm{d}}{\mathrm{d}t}\bar{p}_i^{(0)} = -\frac{1}{\Delta x}\left(u_{i+\frac{1}{2}}^{(1)} - u_{i-\frac{1}{2}}^{(1)}\right), \\
& \frac{\mathrm{d}}{\mathrm{d}t}p_{i+\frac{1}{2}}^{(0)} = -\frac{1}{2}\Big[(D_{i+\frac{1}{2}}^{+} - D_{i+\frac{1}{2}}^{-})p^{(1)} + (D_{i+\frac{1}{2}}^{+} + D_{i+\frac{1}{2}}^{-})u^{(1)}\Big].
\end{aligned}
$$

The second equation at order ϵ^{-1} implies

$$
\begin{aligned}
-u_{i+\frac{1}{2}}^{(1)} = \frac{1}{2\sigma}(D_{i+\frac{1}{2}}^{+} + D_{i+\frac{1}{2}}^{-})p^{(0)} = & \frac{1}{\sigma\Delta x}\left(p_{i-\frac{1}{2}}^{(0)} - 3\bar{p}_i^{(0)} + 3\bar{p}_{i+1}^{(0)} - p_{i+\frac{3}{2}}^{(0)}\right) \\
= \frac{1}{\sigma}\left(p_x^{(0)}\Big|_{x_{i+\frac{1}{2}}} - \frac{\Delta x^2}{12}p_{xxx}^{(0)}\Big|_{x_{i+\frac{1}{2}}} + \mathcal{O}(\Delta x^4)\right), &
\end{aligned}
$$

thus the first equation at order ϵ^0 can be simplified as

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\bar{p}_i^{(0)} &= \frac{1}{\sigma\Delta x}p_x^{(0)}\Big|_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} - \frac{\Delta x}{12\sigma}p_{xxx}^{(0)}\Big|_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} + \mathcal{O}(\Delta x^3) \\
&= \frac{1}{\sigma\Delta x}p_x^{(0)}\Big|_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} + \mathcal{O}(\Delta x^2),
\end{aligned}
$$

which is a finite volume approximation of $p_t^{(0)} = \frac{1}{\sigma}p_{xx}^{(0)}$ with the truncation error $\mathcal{O}(\Delta x^2)$. Note that the last equality uses the Lipschitz continuity of $p_{xxx}^{(0)}$ if $p^{(0)}$ is sufficiently smooth. The second equation of the order ϵ^0 can be simplified as

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}p_{i+\frac{1}{2}}^{(0)} =& -\frac{1}{2}\Big[(D_{i+\frac{1}{2}}^{+} - D_{i+\frac{1}{2}}^{-})p^{(1)} + (D_{i+\frac{1}{2}}^{+} + D_{i+\frac{1}{2}}^{-})u^{(1)}\Big] \\
=& -\frac{1}{\sigma\Delta x}\Big[u_{i-\frac{1}{2}}^{(1)} - 3\bar{u}_i^{(1)} + 3\bar{u}_{i+1}^{(1)} - u_{i+\frac{3}{2}}^{(1)}\Big] + \frac{1}{2}\Big[(D_{i+\frac{1}{2}}^{-} - D_{i+\frac{1}{2}}^{+})p^{(1)}\Big] \\
=& \frac{1}{\sigma\Delta x}\Big[\frac{1}{2}(D_{i-\frac{1}{2}}^{+} + D_{i-\frac{1}{2}}^{-})p^{(0)} - 3\left(p_{i+\frac{1}{2}}^{(0)} - p_{i-\frac{1}{2}}^{(0)}\right) + 3\left(p_{i+\frac{3}{2}}^{(0)} - p_{i+\frac{1}{2}}^{(0)}\right) \\
& - \frac{1}{2}(D_{i+\frac{3}{2}}^{+} + D_{i+\frac{3}{2}}^{-})p^{(0)}\Big] + \frac{1}{2}\Big[(D_{i+\frac{1}{2}}^{-} - D_{i+\frac{1}{2}}^{+})p^{(1)}\Big] \\
=& \frac{1}{\sigma\Delta x^2}\Big(p_{i-\frac{3}{2}}^{(0)} - 3\bar{p}_{i-1}^{(0)} + 3p_{i-\frac{1}{2}}^{(0)} + 3\bar{p}_i^{(0)} - 8p_{i+\frac{1}{2}}^{(0)} + 3\bar{p}_{i+1}^{(0)} + 3p_{i+\frac{3}{2}}^{(0)} - 3\bar{p}_{i+2}^{(0)} + p_{i+\frac{5}{2}}^{(0)}\Big) \\
& + \frac{1}{\Delta x}\left(-p_{i-\frac{1}{2}}^{(1)} + 3\bar{p}_i^{(1)} - 4p_{i+\frac{1}{2}}^{(1)} + 3\bar{p}_{i+1}^{(1)} - p_{i+\frac{3}{2}}^{(1)}\right) \\
=& \frac{1}{\sigma}\left(p_{xx}^{(0)}\Big|_{x_{i+\frac{1}{2}}} + \frac{1}{12}p_{xxxx}^{(0)}\Big|_{x_{i+\frac{1}{2}}}\Delta x^2\right) - \frac{1}{30}p_{xxxx}^{(1)}\Big|_{x_{i+\frac{1}{2}}}\Delta x^3 + \mathcal{O}(\Delta x^4),
\end{aligned}
$$

whose right-hand side recovers an approximation of $\frac{1}{\sigma}p_{xx}^{(0)}$ with the truncation error $\mathcal{O}(\Delta x^2)$.

Proposition 4.1. Assuming sufficient regularity of the solutions, as $\epsilon \to 0$, the leading order solutions of the 1D JS-based AF schemes (4)-(7) satisfy

$$\begin{aligned}
\bar{u}_i^{(0)} &= u_{i+\frac{1}{2}}^{(0)} = 0, \\
\frac{\mathrm{d}}{\mathrm{d}t}\bar{p}_i^{(0)} &= \frac{1}{2\sigma\Delta x}\left[(D_{i+\frac{1}{2}}^{+} + D_{i+\frac{1}{2}}^{-})p^{(0)} - (D_{i-\frac{1}{2}}^{+} + D_{i-\frac{1}{2}}^{-})p^{(0)}\right], \qquad (8) \\
\frac{\mathrm{d}}{\mathrm{d}t}p_{i+\frac{1}{2}}^{(0)} &= \frac{1}{\sigma\Delta x}\Big[\frac{1}{2}(D_{i-\frac{1}{2}}^{+} + D_{i-\frac{1}{2}}^{-})p^{(0)} - 3\left(p_{i+\frac{1}{2}}^{(0)} - p_{i-\frac{1}{2}}^{(0)}\right) + 3\left(p_{i+\frac{3}{2}}^{(0)} - p_{i+\frac{1}{2}}^{(0)}\right) \\
&\quad - \frac{1}{2}(D_{i+\frac{3}{2}}^{+} + D_{i+\frac{3}{2}}^{-})p^{(0)}\Big] + \mathcal{O}(\Delta x^3), \qquad (9)
\end{aligned}$$

where (8)-(9) are approximations of $p_t^{(0)} = \frac{1}{\sigma}p_{xx}^{(0)}$ with the truncation errors $\mathcal{O}(\Delta x^2)$. To be specific,

$$\frac{\mathrm{d}}{\mathrm{d}t}\bar{p}_i^{(0)} = \frac{1}{\sigma\Delta x}p_x^{(0)}\Big|_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} + \mathcal{O}(\Delta x^2), \quad \frac{\mathrm{d}}{\mathrm{d}t}p_{i+\frac{1}{2}}^{(0)} = \frac{1}{\sigma}p_{xx}^{(0)}\Big|_{x_{i+\frac{1}{2}}} + \mathcal{O}(\Delta x^2).$$

The truncation error may not fully explain the convergence order of the AF scheme (or other compact schemes, such as Discontinuous Galerkin, see [39]). Indeed, if one considers the most classical AF scheme for the linear advection equation, the truncation error apparently is 2nd order while the overall accuracy is actually 3rd order (see e.g. [2, 22, 38]). To examine more precisely the accuracy of the limit schemes, a Fourier analysis is performed in Section 6, after discussing an alternative point value update next.

5 An alternative point value update

The following is another way to update the point values:

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}p_{i+\frac{1}{2}} &= -\frac{1}{\epsilon}D_{i+\frac{1}{2}}^{+}u, \\
\frac{\mathrm{d}}{\mathrm{d}t}u_{i+\frac{1}{2}} &= -\frac{1}{\epsilon}D_{i+\frac{1}{2}}^{-}p - \frac{\sigma}{\epsilon^2}u_{i+\frac{1}{2}}.
\end{aligned}$$

It is inspired by the "alternating flux" in [24]. This choice of upwinding is very different from how stable methods are usually achieved for hyperbolic systems of PDEs: Characteristics are not considered and instead, a left bias is chosen on one variable and a right bias on another. We do not find this prescription to yield a stable method in two space dimensions. The reason for discussing it are its theoretical properties that will be of importance in the subsequent analysis of Section 6.

Using the formal asymptotic analysis, we have

$$\begin{aligned}
\epsilon^{-2}: &\quad \bar{u}_i^{(0)} = u_{i+\frac{1}{2}}^{(0)} = 0, \\
\epsilon^{-1}: &\quad \bar{u}_i^{(1)} = -\frac{1}{\sigma\Delta x}\left(p_{i+\frac{1}{2}}^{(0)} - p_{i-\frac{1}{2}}^{(0)}\right), \\
&\quad u_{i+\frac{1}{2}}^{(1)} = -\frac{1}{\sigma}D_{i+\frac{1}{2}}^{-}p^{(0)}, \\
\epsilon^{0}: &\quad \frac{\mathrm{d}}{\mathrm{d}t}\bar{p}_i^{(0)} = -\frac{1}{\Delta x}\left(u_{i+\frac{1}{2}}^{(1)} - u_{i-\frac{1}{2}}^{(1)}\right), \\
&\quad \frac{\mathrm{d}}{\mathrm{d}t}p_{i+\frac{1}{2}}^{(0)} = -D_{i+\frac{1}{2}}^{+}u^{(1)}.
\end{aligned}$$

Thus the limit scheme is

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\bar{p}_i^{(0)} &= \frac{1}{\sigma\Delta x^2}\left(D^-_{i+\frac{1}{2}}p^{(0)} - D^-_{i-\frac{1}{2}}p^{(0)}\right),\\ \frac{\mathrm{d}}{\mathrm{d}t}p^{(0)}_{i+\frac{1}{2}} &= -\frac{1}{\Delta x}\left(2u^{(1)}_{i-\frac{1}{2}} - 6\bar{u}^{(1)}_i + 4u^{(1)}_{i+\frac{1}{2}}\right)\\ &= \frac{1}{\sigma\Delta x^2}\left[2D^-_{i-\frac{1}{2}}p^{(0)} - 6\left(p^{(0)}_{i+\frac{1}{2}} - p^{(0)}_{i-\frac{1}{2}}\right) + 4D^-_{i+\frac{1}{2}}p^{(0)}\right].\end{aligned}$$

6 Fourier analysis

6.1 Fourier analysis of the PDEs

We first perform a Fourier analysis for the hyperbolic heat equation (3). Make the following Fourier ansatz of the solutions

$$p(x,t) = \hat{p}(t)\exp(\mathbb{i}\omega x), \quad u(x,t) = \hat{u}(t)\exp(\mathbb{i}\omega x),$$

and substitute them into (3), then we have the following evolution equation

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix}\hat{p}\\ \hat{u}\end{bmatrix} = \underbrace{-\frac{1}{\epsilon}\begin{bmatrix}0 & \mathbb{i}\omega\\ \mathbb{i}\omega & \epsilon^{-1}\sigma\end{bmatrix}}_{=:\mathcal{E}}\begin{bmatrix}\hat{p}\\ \hat{u}\end{bmatrix}. \tag{10}$$

The eigenvalues of the evolution matrix $\mathcal{E}$ are

$$\lambda_{1,2} = \frac{-\sigma \pm s}{2\epsilon^2}$$

with $s := \sqrt{\sigma^2 - 4\epsilon^2\omega^2}$, which leads to

$$\lambda_1 = -\frac{\omega^2}{\sigma} - \frac{\omega^4\epsilon^2}{\sigma^3} + \mathcal{O}(\epsilon^4), \qquad \lambda_2 = -\frac{\sigma}{\epsilon^2} + \frac{\omega^2}{\sigma} + \frac{\omega^4\epsilon^2}{\sigma^3} + \mathcal{O}(\epsilon^4), \tag{11}$$

Given the initial condition

$$p_0 = \hat{p}_0\exp(\mathbb{i}\omega x), \quad u_0 = \hat{u}_0\exp(\mathbb{i}\omega x), \tag{12}$$

one can decompose $[\hat{p}_0, \hat{u}_0]^\top = V_1 + V_2$ in the eigenvectors of $\mathcal{E}$ as follows

$$V_1 = [\hat{p}_0 + \mathcal{O}(\epsilon),\ \mathcal{O}(\epsilon)]^\top, \quad V_2 = [\mathcal{O}(\epsilon),\ \hat{u}_0 + \mathcal{O}(\epsilon)]^\top.$$

The solution of (10) is

$$[\hat{p}, \hat{u}]^\top = \exp(\lambda_1 t)V_1 + \exp(\lambda_2 t)V_2.$$

In the limit $\epsilon \to 0$, the first eigenvalue λ_1 tends to $-\frac{1}{\sigma}\omega^2$, which corresponds to the solution of the limit heat equation for p, as can be seen from V_1. The second eigenvalue diverges as $\epsilon \to 0$ and corresponds to the initial layer, occurring if the initial data are not well-prepared, i.e., if $\hat{u}_0 \neq 0$ as is obvious from the nature of V_2. The corresponding evolution $\exp(\lambda_2 t) = \exp\left(-\frac{\sigma t}{\epsilon^2}\right)$ decays rapidly with time if ϵ is small. Seen as a function of ϵ, this

term has a Taylor series uniformly zero, and hence escapes the formal analysis presented in Section 2. For later reference, we state here the eigenvectors

$$V_1 = \frac{1}{2s}\Big[\hat{p}_0(s+\sigma) - 2\mathbb{i}\epsilon\omega\hat{u}_0,\ \hat{u}_0(s-\sigma) - 2\mathbb{i}\epsilon\omega\hat{p}_0\Big], \tag{13}$$

$$V_2 = \frac{1}{2s}\Big[\hat{p}_0(s-\sigma) + 2\mathbb{i}\epsilon\omega\hat{u}_0,\ \hat{u}_0(s+\sigma) + 2\mathbb{i}\epsilon\omega\hat{p}_0\Big], \tag{14}$$

where $s - \sigma = \mathcal{O}(\epsilon^2)$.

Remark 6.1. Note that s may be complex. In fact, for $\epsilon = 0$ it is real corresponding to the parabolic limit, while for $\sigma = 0$, one has $s = \pm 2\mathbb{i}\epsilon\omega$ and $\lambda_{1,2} = \pm\mathbb{i}\omega/\epsilon$, i.e., the imaginary eigenvalues expected in the hyperbolic regime.

6.2 Fourier analysis of the numerical scheme

Inspired by [39, 21], we now perform a similar[4] Fourier analysis for the AF scheme by taking the following Fourier ansatz of the numerical solutions

$$\bar{p}_i(t) = \hat{p}_1(t)\exp(\mathbb{i}\omega x_i), \qquad p_{i+\frac{1}{2}}(t) = \hat{p}_2(t)\exp(\mathbb{i}\omega x_i), \tag{15}$$

$$\bar{u}_i(t) = \hat{u}_1(t)\exp(\mathbb{i}\omega x_i), \qquad u_{i+\frac{1}{2}}(t) = \hat{u}_2(t)\exp(\mathbb{i}\omega x_i), \tag{16}$$

where ω is the wave number[5]. Let also $t_x = \exp(\mathbb{i}\omega\Delta x)$.

The coefficient $\hat{w}_0 = [\hat{p}_1(0), \hat{u}_1(0), \hat{p}_2(0), \hat{u}_2(0)]^\top$ of the initial data $\hat{w}_0\exp(\mathbb{i}\omega x_i)$ for the numerical solution in the cell $I_i = [x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}]$ is obtained based on (12),

$$\begin{aligned}\hat{w}_0 &= \exp(-\mathbb{i}\omega x_i)\left[\frac{1}{\Delta x}\int_{I_i} p_0(x)\mathrm{d}x,\ \frac{1}{\Delta x}\int_{I_i} u_0(x)\mathrm{d}x,\ p_0(x_{i+\frac{1}{2}}),\ u_0(x_{i+\frac{1}{2}})\right]^\top \\ &= \left[\hat{p}_0\frac{2\sin\left(\frac{\omega\Delta x}{2}\right)}{\omega\Delta x},\ \hat{u}_0\frac{2\sin\left(\frac{\omega\Delta x}{2}\right)}{\omega\Delta x},\ \hat{p}_0\exp\left(\mathbb{i}\omega\tfrac{\Delta x}{2}\right),\ \hat{u}_0\exp\left(\mathbb{i}\omega\tfrac{\Delta x}{2}\right)\right]^\top.\end{aligned} \tag{17}$$

6.2.1 Point value update using Jacobian splitting

Substitute the ansatz (15)–(16) into the AF schemes (4)-(7):

$$\frac{\mathrm{d}}{\mathrm{d}t}[\hat{p}_1, \hat{u}_1, \hat{p}_2, \hat{u}_2]^\top = \boldsymbol{G}[\hat{p}_1, \hat{u}_1, \hat{p}_2, \hat{u}_2]^\top,$$

where the evolution matrix is

$$\boldsymbol{G} = \frac{1}{\epsilon\Delta x}\begin{bmatrix} 0 & 0 & 0 & t_x^{-1} - 1 \\ 0 & -\epsilon^{-1}\sigma\Delta x & t_x^{-1} - 1 & 0 \\ 3(1+t_x) & 3(1-t_x) & -(t_x^{-1} + 4 + t_x) & t_x - t_x^{-1} \\ 3(1-t_x) & 3(1+t_x) & t_x - t_x^{-1} & -\epsilon^{-1}\sigma\Delta x - (t_x^{-1} + 4 + t_x) \end{bmatrix}.$$

[4]The calculations are performed using Mathematica.

[5]It is of no relevance whether we associate the point values with x_i or $x_{i+\frac{1}{2}}$, as the latter would amount to an extra overall factor $\exp(\mathbb{i}\omega\Delta x/2)$ and hence only a redefinition of $\hat{p}_2$ or $\hat{u}_2$.

It is difficult to calculate the eigenvalues of $\boldsymbol{G}$, and we thus construct them as a power series in Δx

$$\tilde{\lambda} = \Delta x^{-1}\tilde{\lambda}^{(-1)} + \tilde{\lambda}^{(0)} + \Delta x\tilde{\lambda}^{(1)} + \Delta x^2\tilde{\lambda}^{(2)} + \Delta x^3\tilde{\lambda}^{(3)} + \text{h.o.t.},$$

and solve the characteristic equation

$$\det(\tilde{\lambda}\boldsymbol{I}_4 - \boldsymbol{G}) = 0$$

by matching the coefficients of the same order in Δx. The eigenvalues are

$$\begin{aligned}
\tilde{\lambda}_1 &= \lambda_1 - \frac{\Delta x^3\omega^4}{72\epsilon} + \mathcal{O}(\Delta x^4),\\
\tilde{\lambda}_2 &= \lambda_2 - \frac{\Delta x^3\omega^4}{72\epsilon} + \mathcal{O}(\Delta x^4),\\
\tilde{\lambda}_3 &= -\frac{6}{\epsilon\Delta x} - \frac{\tilde{s}+\sigma}{2\epsilon^2} + \frac{\omega^2\Delta x}{\epsilon} - \frac{2\Delta x^2\omega^4}{\tilde{s}} - \frac{5\Delta x^3\omega^4}{72\epsilon} + \mathcal{O}(\Delta x^4),\\
\tilde{\lambda}_4 &= -\frac{6}{\epsilon\Delta x} + \frac{\tilde{s}-\sigma}{2\epsilon^2} + \frac{\omega^2\Delta x}{\epsilon} + \frac{2\Delta x^2\omega^4}{\tilde{s}} - \frac{5\Delta x^3\omega^4}{72\epsilon} + \mathcal{O}(\Delta x^4),
\end{aligned}$$

where $\tilde{s} := \sqrt{\sigma^2 - 36\epsilon^2\omega^2}$. The first and second eigenvalues are physical (i.e., converging to (11) as $\Delta x \to 0$), with the other two spurious modes decaying exponentially as $\Delta x \to 0$. Note that the second physical eigenvalue corresponds to the initial layer. Observe also the presence of a singularity in the physical eigenvalues as $\epsilon \to 0$; we will come back to this aspect in Section 6.2.3 after presenting the rest of the analysis.

Similarly, we calculate a set of eigenvectors $\tilde{v}_1, \dots, \tilde{v}_4$ of $\boldsymbol{G}$ by assuming that they are power series of Δx. Then the discrete initial data $\hat{w}_0$ can be decomposed in the basis of the eigenvectors of $\boldsymbol{G}$ as $\hat{w}_0 = \hat{w}_0^{(1)}\tilde{v}_1 + \hat{w}_0^{(2)}\tilde{v}_2 + \hat{w}_0^{(3)}\tilde{v}_3 + \hat{w}_0^{(4)}\tilde{v}_4 =: \tilde{V}_1 + \tilde{V}_2 + \tilde{V}_3 + \tilde{V}_4$ with, ignoring terms $\mathcal{O}(\Delta x^4)$,

$$\begin{aligned}
\tilde{V}_1 = \Big[&- \frac{\omega\epsilon(2p_0\omega\epsilon + \mathbb{i}u_0(s-\sigma))}{s(s-\sigma)} + \frac{\Delta x^2\omega^3\epsilon(2p_0\omega\epsilon + \mathbb{i}u_0(s-\sigma))}{24s(s-\sigma)},\\
&- \frac{(s-\sigma)(\mathbb{i}p_0\omega\epsilon + \sigma u_0) + 2u_0\omega^2\epsilon^2}{s(s-\sigma)} + \frac{\Delta x^2\omega^2((s-\sigma)(\mathbb{i}p_0\omega\epsilon + \sigma u_0) + 2u_0\omega^2\epsilon^2)}{24s(s-\sigma)},\\
&- \frac{\omega(288p_0\omega\epsilon^2 + 144\mathbb{i}u_0\epsilon(s-\sigma))}{144s(s-\sigma)} - \frac{\Delta x\omega(-72u_0\omega\epsilon(s-\sigma) + 144\mathbb{i}p_0\omega^2\epsilon^2)}{144s(s-\sigma)}\\
&- \frac{\Delta x^2\omega(-36p_0\omega^3\epsilon^2 - 18\mathbb{i}u_0\omega^2\epsilon(s-\sigma))}{144s(s-\sigma)}\\
&- \frac{1}{144s(s-\sigma)}\Big(\Delta x^3\omega(2p_0\omega^3\epsilon(s-\sigma) - 6\mathbb{i}p_0\omega^4\epsilon^2 - 2\mathbb{i}\sigma u_0\omega^2(s-\sigma)\\
&+ 3u_0\omega^3\epsilon(s-\sigma) - 4\mathbb{i}u_0\omega^4\epsilon^2)\Big),\\
&\frac{-288\mathbb{i}p_0\omega\epsilon + 144(s-\sigma)u_0}{288s} + \frac{\Delta x(144p_0\omega^2\epsilon + 72\mathbb{i}(s-\sigma)u_0\omega)}{288s}\\
&+ \frac{\Delta x^2(36\mathbb{i}p_0\omega^3\epsilon - 18(s-\sigma)u_0\omega^2)}{288s}\\
&+ \frac{\Delta x^3(-2\mathbb{i}p_0\omega^3(s+\sigma) - 6p_0\omega^4\epsilon - u_0\omega^3(4\omega\epsilon + 3\mathbb{i}s) + 3\mathbb{i}\sigma u_0\omega^3)}{288s}\Big]^\top,
\end{aligned} \tag{18}$$

$$
\begin{aligned}
\tilde{V}_2 = \Big[& \frac{\mathbb{i}\epsilon\omega(2\mathbb{i}p_0\epsilon\omega + u_0(s+\sigma))}{s(s+\sigma)} - \frac{\mathbb{i}\Delta x^2\epsilon\omega^3(2\mathbb{i}p_0\epsilon\omega + u_0(s+\sigma))}{24s(s+\sigma)}, \\
& \frac{u_0\sigma + \mathbb{i}p_0\epsilon\omega + \frac{1}{2}u_0(s-\sigma)}{s} - \frac{\Delta x^2\omega^2(u_0\sigma + \mathbb{i}p_0\epsilon\omega + \frac{1}{2}u_0(s-\sigma))}{24s}, \\
& \frac{\omega\left(-2p_0s\omega\epsilon^2 + \mathbb{i}\sigma u_0\epsilon(s+\sigma) - 4\mathbb{i}u_0\omega^2\epsilon^3\right)}{s^2(s+\sigma)} + \frac{\Delta x\omega\left(-2\mathbb{i}p_0s\omega^2\epsilon^2 - \sigma u_0\omega\epsilon(s+\sigma) + 4u_0\omega^3\epsilon^3\right)}{2s^2(s+\sigma)} \\
& + \frac{\Delta x^2\omega\left(2p_0s\omega^3\epsilon^2 - \mathbb{i}\sigma u_0\omega^2\epsilon(s+\sigma) + 4\mathbb{i}u_0\omega^4\epsilon^3\right)}{8s^2(s+\sigma)} \\
& + \frac{1}{144s^2(s+\sigma)}\Big(\Delta x^3\omega(2p_0\omega^3\epsilon(\sigma^2 + s\sigma + \omega\epsilon(-4\omega\epsilon + 3\mathbb{i}s)) - 2\mathbb{i}\sigma^2u_0\omega^2(s+\sigma) \\
& + 4\mathbb{i}u_0\omega^4\epsilon^2(s+2\sigma) + 3\sigma u_0\omega^3\epsilon(s+\sigma) - 12u_0\omega^5\epsilon^3)\Big), \\
& \frac{2\mathbb{i}p_0\omega\epsilon + (s+\sigma)u_0}{2s} + \frac{\Delta x(-2p_0\omega^2\epsilon + \mathbb{i}(s+\sigma)u_0\omega)}{4s} \\
& - \frac{\Delta x^2(2\mathbb{i}p_0\omega^3\epsilon + (s+\sigma)u_0\omega^2)}{16s} \\
& + \frac{\Delta x^3(-2\mathbb{i}p_0\omega^3(s-\sigma) + 6p_0\omega^4\epsilon + u_0\omega^3(4\omega\epsilon - 3\mathbb{i}s) - 3\mathbb{i}\sigma u_0\omega^3)}{288s}\Big]^\top,
\end{aligned}
\tag{19}
$$

$$
\tilde{V}_3 = \left[0,\ 0,\ \frac{(6\epsilon\omega\hat{p}_0 + \mathbb{i}(\tilde{s}-\sigma)\hat{u}_0)\omega^3\Delta x^3}{144\tilde{s}},\ \frac{(6\epsilon\omega\hat{u}_0 + \mathbb{i}(\tilde{s}+\sigma)\hat{p}_0)\omega^3\Delta x^3}{144\tilde{s}}\right]^\top,
$$

$$
\tilde{V}_4 = \left[0,\ 0,\ \frac{(-6\epsilon\omega\hat{p}_0 + \mathbb{i}(\tilde{s}+\sigma)\hat{u}_0)\omega^3\Delta x^3}{144\tilde{s}},\ \frac{(-6\epsilon\omega\hat{u}_0 + \mathbb{i}(\tilde{s}-\sigma)\hat{p}_0)\omega^3\Delta x^3}{144\tilde{s}}\right]^\top.
$$

In this decomposition, consider first the contribution of the spurious eigenvectors $\tilde{V}_3, \tilde{V}_4$. It is of order $\mathcal{O}(\Delta x^3)$, i.e., of the order of accuracy of the method, uniformly in ϵ: expanding the stated terms in $\tilde{V}_3$ and $\tilde{V}_4$ as a power series in ϵ, one finds

$$
\left[0,\ 0,\ \mathcal{O}(\epsilon),\ \frac{1}{72}\mathbb{i}\Delta x^3\omega^3\hat{p}_0 + \mathcal{O}(\epsilon)\right], \quad \left[0,\ 0,\ \frac{1}{72}\mathbb{i}\Delta x^3\omega^3\hat{u}_0 + \mathcal{O}(\epsilon),\ \mathcal{O}(\epsilon)\right],
$$

up to terms $\mathcal{O}(\Delta x^4)$. The decomposition along the two physical modes $\tilde{V}_1, \tilde{V}_2$ quantifies which part of the data is evolving according to the first (physical) eigenvalue $\tilde{\lambda}_1 \simeq \lambda_1$, and which - according to the second. In order to determine the error, one needs to compare (18)-(19) to the averages and point values of the same decomposition at the PDE level. These latter are obtained by replacing $\hat{p}_0$ and $\hat{u}_0$ in (17) by the respective components of V_1 and V_2 from (13)–(14). After subtraction, the error of $\tilde{V}_1$ (the mode corresponding to the limit heat equation) reads

$$
\left[0,\ 0,\ \frac{(2\epsilon\omega\hat{p}_0 + \mathbb{i}(s-\sigma)\hat{u}_0)\omega^3\Delta x^3}{144s},\ \frac{(2\epsilon\omega\hat{u}_0 + \mathbb{i}(s+\sigma)\hat{p}_0)\omega^3\Delta x^3}{144s}\right],
$$

where $\mathcal{O}(\Delta x^4)$ terms are omitted. The third entry is a term $O(\epsilon)$, while the fourth is $\frac{1}{72}\mathbb{i}\Delta x^3\omega^3\hat{p}_0 + \mathcal{O}(\epsilon)$. The error of $\tilde{V}_2$ is

$$
\left[0,\ 0,\ \frac{(-2\epsilon\omega\hat{p}_0 + \mathbb{i}(s+\sigma)\hat{u}_0)\omega^3\Delta x^3}{144s},\ \frac{(-2\epsilon\omega\hat{u}_0 + \mathbb{i}(s-\sigma)\hat{p}_0)\omega^3\Delta x^3}{144s}\right],
$$

where the third entry is $\frac{1}{72}\mathbb{i}\Delta x^3\omega^3\hat{u}_0 + \mathcal{O}(\epsilon)$, and the fourth entry is $\mathcal{O}(\epsilon)$. Thus, these errors are of third order uniformly in ϵ.

The analysis thus shows that, in principle, the relevant errors of the Fourier decomposition are $\mathcal{O}(\Delta x^3)$. For the eigenvectors, this is uniformly in ϵ. Before condensing these results into a statement on the error of the method, we perform analogous computations for the point value update based on the alternating flux. With the two sets of results in place, in Section 6.2.3 we present the remainder of the error analysis based on the Fourier decomposition.

6.2.2 Point value update using the alternating flux

The same computations are now performed for the AF method with the point update based on the prescription from Section 5. As mentioned there, this kind of upwinding does not correspond to the usual way of achieving stability. However, it is stable in one spatial dimension and the comparison between the two kinds of point value updates helps elucidating the behavior of the numerical methods in the limit $\epsilon \to 0$.

When the point values are updated with the "alternating flux" prescription, the scheme is

$$\frac{\mathrm{d}}{\mathrm{d}t}[\hat{p}_1, \hat{u}_1, \hat{p}_2, \hat{u}_2]^\top = \boldsymbol{G}[\hat{p}_1, \hat{u}_1, \hat{p}_2, \hat{u}_2]^\top,$$

with the evolution matrix

$$\boldsymbol{G} = \frac{1}{\epsilon\Delta x}\begin{bmatrix} 0 & 0 & 0 & t_x^{-1}-1 \\ 0 & -\epsilon^{-1}\sigma\Delta x & t_x^{-1}-1 & 0 \\ 0 & 6 & 0 & -4-2t_x^{-1} \\ -6t_x & 0 & 4+2t_x & -\epsilon^{-1}\sigma\Delta x \end{bmatrix}.$$

It can be diagonalized explicitly. The eigenvalues are

$$\begin{aligned}
\tilde{\lambda}_1 &= \lambda_1 + \frac{\omega^6\Delta x^4}{540s} + \mathcal{O}(\Delta x^5), \\
\tilde{\lambda}_2 &= \lambda_2 - \frac{\omega^6\Delta x^4}{540s} + \mathcal{O}(\Delta x^5), \\
\tilde{\lambda}_3 &= -\frac{6\mathbb{i}}{\epsilon\Delta x} - \frac{\sigma}{2\epsilon^2} + \mathbb{i}\Delta x\frac{12\epsilon^2\omega^2+\sigma^2}{48\epsilon^3} + \mathcal{O}(\Delta x^3), \\
\tilde{\lambda}_4 &= \frac{6\mathbb{i}}{\epsilon\Delta x} - \frac{\sigma}{2\epsilon^2} - \mathbb{i}\Delta x\frac{12\epsilon^2\omega^2+\sigma^2}{48\epsilon^3} + \mathcal{O}(\Delta x^3).
\end{aligned}$$

Observe first of all that this time, the physical eigenvalues $\tilde{\lambda}_1, \tilde{\lambda}_2$ are 4th-order accurate, and that the error is uniform in ϵ. This is a difference from the results with the JS obtained above. We refrain from stating the entire expansion of the spurious eigenvalues $\tilde{\lambda}_3, \tilde{\lambda}_4$, since the contribution of the spurious modes is again negligible: Upon decomposing the initial data $\hat{w}_0$ in the eigenvectors, the terms $\tilde{V}_3$ and $\tilde{V}_4$ are, neglecting contributions $\mathcal{O}(\Delta x^4)$,

$$\begin{aligned}
\tilde{V}_3 &= \left[0,\ 0,\ \frac{1}{144}\Delta x^3\omega^3(\hat{u}_0 - \mathbb{i}\hat{p}_0),\ \frac{1}{144}\Delta x^3\omega^3(\hat{p}_0 + \mathbb{i}\hat{u}_0)\right], \\
\tilde{V}_4 &= \left[0,\ 0, -\frac{1}{144}\mathbb{i}\Delta x^3\omega^3(\hat{p}_0 - \mathbb{i}\hat{u}_0), -\frac{1}{144}\Delta x^3\omega^3(\hat{p}_0 - \mathbb{i}\hat{u}_0)\right].
\end{aligned}$$

Observe that these are independent of ϵ. Finally, we compute the difference between the projection of the exact mode corresponding to λ_1 onto the degrees of freedom and the corresponding discrete mode $\tilde{V}_1$, finding

$$\left[\frac{\epsilon\omega^4\hat{u}_0\Delta x^3}{72s}, -\frac{\epsilon\omega^4\hat{p}_0\Delta x^3}{72s}, -\frac{1}{144}\mathbb{i}\omega^3\left(1+\frac{\sigma}{s}\right)\hat{p}_0\Delta x^3, \frac{1}{144}\mathbb{i}\omega^3\left(1-\frac{\sigma}{s}\right)\hat{u}_0\Delta x^3\right]$$

up to terms $\mathcal{O}(\Delta x^4)$. The error of the mode corresponding to the initial layer is

$$\left[-\frac{\epsilon\omega^4\hat{u}_0\Delta x^3}{72s}, \frac{\epsilon\omega^4\hat{p}_0\Delta x^3}{72s}, -\frac{1}{144}\mathbb{i}\omega^3\left(1-\frac{\sigma}{s}\right)\hat{p}_0\Delta x^3, \frac{1}{144}\mathbb{i}\omega^3\left(1+\frac{\sigma}{s}\right)\hat{u}_0\Delta x^3\right]$$

up to terms $\mathcal{O}(\Delta x^4)$. All terms are of the order of accuracy of the method and also $\mathcal{O}(\epsilon)$.

6.2.3 Discussion

Following [21] we observe that the error of the method can be related to the errors of the eigendecomposition of $\boldsymbol{G}$ as performed in the previous Sections in the following way:

$$\begin{aligned}\|\pi\hat{q}(t)-\hat{q}_h(t)\|_{L^2} &= \left\|\sum_{k=1}^{2}(\pi V_k\exp(\lambda_k t)-\tilde{V}_k\exp(\tilde{\lambda}_k t))+\sum_{k=3}^{4}\tilde{V}_k\exp(\tilde{\lambda}_k t))\right\|_{L^2}\\ &\leq \sum_{k=1}^{2}\|\pi V_k-\tilde{V}_k\||\exp(\lambda_k t)|+\sum_{k=1}^{2}\|\tilde{V}_k\||\exp(\lambda_k t)-\exp(\tilde{\lambda}_k t)|\\ &\quad+\sum_{k=3}^{4}\|\tilde{V}_k\||\exp(\tilde{\lambda}_k t)|.\end{aligned}$$

Here π denotes the projection (17) of a mode $[\hat{p}_0,\hat{u}_0]$ associated to the PDE (a vector in $\mathbb{C}^2$) onto a Fourier mode associated to the degrees of freedom of the scheme (a vector in $\mathbb{C}^4$).

For both the JS and the alternating flux, we have seen that the contribution of the spurious modes is negligible, i.e., $\|\tilde{V}_k\|\in\mathcal{O}(\Delta x^3)$ for $k=3,4$. Similarly, the error $\|\pi V_k-\tilde{V}_k\|$, $k=1,2$ of the physical modes is $\mathcal{O}(\Delta x^3)$ as well.

Finally,

$$|\exp(\lambda_k t)-\exp(\tilde{\lambda}_k t)|\leq Ct|\lambda_k-\tilde{\lambda}_k|.$$

This value is $\mathcal{O}(\Delta x^4)$ for the alternating flux, and $\mathcal{O}\left(\frac{\Delta x^3}{\epsilon}\right)$ for the JS. We conclude that in the limit, the scheme based on JS suffers a loss of accuracy, while the one based on the "alternating flux" prescription does not. The loss of accuracy can be alleviated by choosing Δx sufficiently small in relation to ϵ. This is confirmed in Example 8.1. The formal analysis presented here, however, is insufficient to obtain theoretical convergence rates, which is beyond the scope of the paper. Below, we present detailed numerical studies of the order of accuracy for various values of ϵ.

7 2D active flux scheme

This section presents the 2D semi-discrete AF methods for the hyperbolic heat equation

$$\begin{cases} p_t + \dfrac{1}{\epsilon} u_x + \dfrac{1}{\epsilon} v_y = 0, \\ u_t + \dfrac{1}{\epsilon} p_x + \dfrac{\sigma}{\epsilon^2} u = 0, \\ v_t + \dfrac{1}{\epsilon} p_y + \dfrac{\sigma}{\epsilon^2} v = 0. \end{cases}$$

Consider a 2D uniform Cartesian mesh with $N_1 \times N_2$ cells $I_{ij} = [x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}] \times [y_{j-\frac{1}{2}}, y_{j+\frac{1}{2}}]$ with cell centers $((x_{i-\frac{1}{2}} + x_{i+\frac{1}{2}})/2, (y_{j-\frac{1}{2}} + y_{j+\frac{1}{2}})/2)$ and cell sizes $\Delta x, \Delta y$. The degrees of freedom contain the cell averages, face-centered point values, and nodal values at corners. The cell average is updated following the finite volume manner

$$\begin{aligned} \frac{\mathrm{d}\bar{p}_{ij}}{\mathrm{d}t} &= -\frac{1}{\epsilon \Delta x}\left(\hat{u}_{i+\frac{1}{2},j} - \hat{u}_{i-\frac{1}{2},j}\right) - \frac{1}{\epsilon \Delta y}\left(\hat{v}_{i,j+\frac{1}{2}} - \hat{v}_{i,j-\frac{1}{2}}\right), \\ \frac{\mathrm{d}\bar{u}_{ij}}{\mathrm{d}t} &= -\frac{1}{\epsilon \Delta x}\left(\hat{p}_{i+\frac{1}{2},j} - \hat{p}_{i-\frac{1}{2},j}\right) - \frac{\sigma}{\epsilon^2}\bar{u}_{ij}, \\ \frac{\mathrm{d}\bar{v}_{ij}}{\mathrm{d}t} &= -\frac{1}{\epsilon \Delta y}\left(\hat{p}_{i,j+\frac{1}{2}} - \hat{p}_{i,j-\frac{1}{2}}\right) - \frac{\sigma}{\epsilon^2}\bar{v}_{ij}, \end{aligned} \tag{20}$$

where the numerical fluxes are

$$\begin{aligned} \hat{q}_{i+\frac{1}{2},j} &= \frac{1}{6}\left(q_{i+\frac{1}{2},j-\frac{1}{2}} + 4q_{i+\frac{1}{2},j} + q_{i+\frac{1}{2},j+\frac{1}{2}}\right), \\ \hat{q}_{i,j+\frac{1}{2}} &= \frac{1}{6}\left(q_{i-\frac{1}{2},j+\frac{1}{2}} + 4q_{i,j+\frac{1}{2}} + q_{i+\frac{1}{2},j+\frac{1}{2}}\right), \quad q = p, u, v. \end{aligned}$$

For the point value update, the Jacobian in the x- and y-directions are split as

$$\boldsymbol{J}_x^+ = \frac{1}{\epsilon}\begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \boldsymbol{J}_x^- = \frac{1}{\epsilon}\begin{pmatrix} -\frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & -\frac{1}{2} & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

$$\boldsymbol{J}_y^+ = \frac{1}{\epsilon}\begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix}, \quad \boldsymbol{J}_y^- = \frac{1}{\epsilon}\begin{pmatrix} -\frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 \\ \frac{1}{2} & 0 & -\frac{1}{2} \end{pmatrix}.$$

Then using the JS-based discretizations in [13, 3], the point value update at the corner is

$$\begin{aligned} \frac{\mathrm{d}p_{i+\frac{1}{2},j+\frac{1}{2}}}{\mathrm{d}t} &= -\frac{1}{2\epsilon}\Big[(D_x^+)_{i+\frac{1}{2},j+\frac{1}{2}}(p+u) - (D_x^-)_{i+\frac{1}{2},j+\frac{1}{2}}(p-u)\Big] \\ &\quad -\frac{1}{2\epsilon}\Big[(D_y^+)_{i+\frac{1}{2},j+\frac{1}{2}}(p+v) - (D_y^-)_{i+\frac{1}{2},j+\frac{1}{2}}(p-v)\Big], \\ \frac{\mathrm{d}u_{i+\frac{1}{2},j+\frac{1}{2}}}{\mathrm{d}t} &= -\frac{1}{2\epsilon}\Big[(D_x^+)_{i+\frac{1}{2},j+\frac{1}{2}}(p+u) + (D_x^-)_{i+\frac{1}{2},j+\frac{1}{2}}(p-u)\Big] - \frac{\sigma}{\epsilon^2}u_{i+\frac{1}{2},j+\frac{1}{2}}, \\ \frac{\mathrm{d}v_{i+\frac{1}{2},j+\frac{1}{2}}}{\mathrm{d}t} &= -\frac{1}{2\epsilon}\Big[(D_y^+)_{i+\frac{1}{2},j+\frac{1}{2}}(p+v) + (D_y^-)_{i+\frac{1}{2},j+\frac{1}{2}}(p-v)\Big] - \frac{\sigma}{\epsilon^2}v_{i+\frac{1}{2},j+\frac{1}{2}}, \end{aligned} \tag{21}$$

where

$$
\begin{aligned}
(D_x^+)_{i+\frac{1}{2},j+\frac{1}{2}}(a) &= a_{i-\frac{1}{2},j+\frac{1}{2}} - 4a_{i,j+\frac{1}{2}} + 3a_{i+\frac{1}{2},j+\frac{1}{2}},\\
(D_x^-)_{i+\frac{1}{2},j+\frac{1}{2}}(a) &= -3a_{i+\frac{1}{2},j+\frac{1}{2}} + 4a_{i+1,j+\frac{1}{2}} - a_{i+\frac{3}{2},j+\frac{1}{2}}.
\end{aligned}
$$

The point value update at the face center $(x_{i+\frac{1}{2}}, y_j)$ is

$$
\begin{aligned}
\frac{\mathrm{d}p_{i+\frac{1}{2},j}}{\mathrm{d}t} &= -\frac{1}{2\epsilon}\Big[(D_x^+)_{i+\frac{1}{2},j}(p+u) - (D_x^-)_{i+\frac{1}{2},j}(p-u)\Big] - \frac{1}{\epsilon}(D_y)_{i+\frac{1}{2},j}v,\\
\frac{\mathrm{d}u_{i+\frac{1}{2},j}}{\mathrm{d}t} &= -\frac{1}{2\epsilon}\Big[(D_x^+)_{i+\frac{1}{2},j}(p+u) + (D_x^-)_{i+\frac{1}{2},j}(p-u)\Big] - \frac{\sigma}{\epsilon^2}u_{i+\frac{1}{2},j},\\
\frac{\mathrm{d}v_{i+\frac{1}{2},j}}{\mathrm{d}t} &= -\frac{1}{\epsilon}(D_y)_{i+\frac{1}{2},j}p - \frac{\sigma}{\epsilon^2}v_{i+\frac{1}{2},j},
\end{aligned}
\tag{22}
$$

where

$$
\begin{aligned}
(D_x^+)_{i+\frac{1}{2},j}(a) &= a_{i-\frac{1}{2},j} - 4a_{i,j} + 3a_{i+\frac{1}{2},j},\\
(D_x^-)_{i+\frac{1}{2},j}(a) &= -3a_{i+\frac{1}{2},j} + 4a_{i+1,j} - a_{i+\frac{3}{2},j},\\
(D_y)_{i+\frac{1}{2},j}(a) &= a_{i+\frac{1}{2},j+\frac{1}{2}} - a_{i+\frac{1}{2},j-\frac{1}{2}},
\end{aligned}
$$

and the cell center value is obtained by

$$
\begin{aligned}
a_{i,j} = \frac{1}{16}\Big[36\bar{a}_{i,j} - 4\Big(&a_{i-\frac{1}{2},j} + a_{i+\frac{1}{2},j} + a_{i,j-\frac{1}{2}} + a_{i,j+\frac{1}{2}}\Big)\\
&- \Big(a_{i-\frac{1}{2},j-\frac{1}{2}} + a_{i+\frac{1}{2},j-\frac{1}{2}} + a_{i-\frac{1}{2},j+\frac{1}{2}} + a_{i+\frac{1}{2},j+\frac{1}{2}}\Big)\Big].
\end{aligned}
\tag{23}
$$

The point value update at the face center $(x_i, y_{j+\frac{1}{2}})$ is

$$
\begin{aligned}
\frac{\mathrm{d}p_{i,j+\frac{1}{2}}}{\mathrm{d}t} &= -\frac{1}{\epsilon}(D_x)_{i,j+\frac{1}{2}}u - \frac{1}{2\epsilon}\Big[(D_y^+)_{i,j+\frac{1}{2}}(p+v) - (D_y^-)_{i,j+\frac{1}{2}}(p-v)\Big],\\
\frac{\mathrm{d}u_{i,j+\frac{1}{2}}}{\mathrm{d}t} &= -\frac{1}{\epsilon}(D_x)_{i,j+\frac{1}{2}}p - \frac{\sigma}{\epsilon^2}u_{i,j+\frac{1}{2}},\\
\frac{\mathrm{d}v_{i,j+\frac{1}{2}}}{\mathrm{d}t} &= -\frac{1}{2\epsilon}\Big[(D_y^+)_{i,j+\frac{1}{2}}(p+v) + (D_y^-)_{i,j+\frac{1}{2}}(p-v)\Big] - \frac{\sigma}{\epsilon^2}v_{i,j+\frac{1}{2}}.
\end{aligned}
\tag{24}
$$

7.1 Formal asymptotic analysis

Similar to the analysis in Section 4, under the assumption of sufficient regularity of numerical solutions, one can first match the terms of order ϵ^{-2} in (20)-(24) and obtain

$$
\bar{u}_{ij}^{(0)} = \bar{v}_{ij}^{(0)} = u_\zeta^{(0)} = v_\zeta^{(0)} = 0, \tag{25}
$$

where ζ denotes the point values at face centers or corners.

For the terms of order ϵ^{-1}, one has

$$
\begin{aligned}
&\bar{u}_{ij}^{(1)} = -\frac{1}{\sigma\Delta x}\left(\hat{p}_{i+\frac{1}{2},j}^{(0)} - \hat{p}_{i-\frac{1}{2},j}^{(0)}\right),\ \bar{v}_{ij}^{(1)} = -\frac{1}{\sigma\Delta y}\left(\hat{p}_{i,j+\frac{1}{2}}^{(0)} - \hat{p}_{i,j-\frac{1}{2}}^{(0)}\right),\\
&u_{i+\frac{1}{2},j+\frac{1}{2}}^{(1)} = -\frac{1}{2\sigma}\Big[(D_x^+)_{i+\frac{1}{2},j+\frac{1}{2}} + (D_x^-)_{i+\frac{1}{2},j+\frac{1}{2}}\Big]p^{(0)},\\
&v_{i+\frac{1}{2},j+\frac{1}{2}}^{(1)} = -\frac{1}{2\sigma}\Big[(D_y^+)_{i+\frac{1}{2},j+\frac{1}{2}} + (D_y^-)_{i+\frac{1}{2},j+\frac{1}{2}}\Big]p^{(0)},\\
&u_{i+\frac{1}{2},j}^{(1)} = -\frac{1}{2\sigma}\Big[(D_x^+)_{i+\frac{1}{2},j} + (D_x^-)_{i+\frac{1}{2},j}\Big]p^{(0)},\\
&v_{i+\frac{1}{2},j}^{(1)} = -\frac{1}{\sigma}(D_y)_{i+\frac{1}{2},j}p^{(0)},\\
&u_{i,j+\frac{1}{2}}^{(1)} = -\frac{1}{\sigma}(D_x)_{i,j+\frac{1}{2}}p^{(0)},\\
&v_{i,j+\frac{1}{2}}^{(1)} = -\frac{1}{2\sigma}\Big[(D_y^+)_{i,j+\frac{1}{2}} + (D_y^-)_{i,j+\frac{1}{2}}\Big]p^{(0)},
\end{aligned}
\tag{26}
$$

which are generally some approximations of $u^{(1)} = -\frac{1}{\sigma}p_x^{(0)}, v^{(1)} = -\frac{1}{\sigma}p_y^{(0)}$. Take the terms related to $u^{(1)}$ as examples while those of $v^{(1)}$ can be treated similarly. The Simpson's rule is 4th order,

$$
\hat{p}_{i+\frac{1}{2},j}^{(0)} = \frac{1}{6}\left(p_{i+\frac{1}{2},j-\frac{1}{2}}^{(0)} + 4p_{i+\frac{1}{2},j}^{(0)} + p_{i+\frac{1}{2},j+\frac{1}{2}}^{(0)}\right) = \left(\frac{1}{\Delta y}\int_{y_{j-\frac{1}{2}}}^{y_{j+\frac{1}{2}}} p^{(0)}\mathrm{d}y\right)\Bigg|_{x_{i+\frac{1}{2}}} + \mathcal{O}(\Delta y^4),
$$

thus

$$
\bar{u}_{ij}^{(1)} = -\frac{1}{\sigma\Delta x}\left(\frac{1}{\Delta y}\int_{y_{j-\frac{1}{2}}}^{y_{j+\frac{1}{2}}} p^{(0)}\mathrm{d}y\right)\Bigg|_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} + \mathcal{O}(\Delta y^4),
$$

which is a finite volume approximation of $u^{(1)} = -\frac{1}{\sigma}p_x^{(0)}$. By using the Taylor expansion,

$$
\begin{aligned}
u_{i+\frac{1}{2},j+\frac{1}{2}}^{(1)} &= -\frac{1}{2\sigma}\Big[(D_x^+)_{i+\frac{1}{2},j+\frac{1}{2}} + (D_x^-)_{i+\frac{1}{2},j+\frac{1}{2}}\Big]p^{(0)} = -\frac{1}{\sigma}p_x^{(0)}\Big|_{x_{i+\frac{1}{2}},y_{j+\frac{1}{2}}} + \mathcal{O}(\Delta x^2),\\
u_{i+\frac{1}{2},j}^{(1)} &= -\frac{1}{2\sigma}\Big[(D_x^+)_{i+\frac{1}{2},j} + (D_x^-)_{i+\frac{1}{2},j}\Big]p^{(0)} = -\frac{1}{\sigma}p_x^{(0)}\Big|_{x_{i+\frac{1}{2}},y_j} + \mathcal{O}(\Delta x^2),\\
u_{i,j+\frac{1}{2}}^{(1)} &= -\frac{1}{\sigma}(D_x)_{i,j+\frac{1}{2}}p^{(0)} = -\frac{1}{\sigma}p_x^{(0)}\Big|_{x_i,y_{j+\frac{1}{2}}} + \mathcal{O}(\Delta x^2),
\end{aligned}
$$

so that they are approximations of $u^{(1)} = -\frac{1}{\sigma}p_x^{(0)}$ at corresponding points.

Collecting the terms of order ϵ^0 in (20)-(24) gives

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\bar{p}_{ij}^{(0)} =& -\frac{1}{\Delta x}\left(\hat{u}^{(1)}_{i+\frac{1}{2},j}-\hat{u}^{(1)}_{i-\frac{1}{2},j}\right)-\frac{1}{\Delta y}\left(\hat{v}^{(1)}_{i,j+\frac{1}{2}}-\hat{v}^{(1)}_{i,j-\frac{1}{2}}\right), \\
\frac{\mathrm{d}p^{(0)}_{i+\frac{1}{2},j+\frac{1}{2}}}{\mathrm{d}t} =& -\frac{1}{2}\Big[(D_x^+)_{i+\frac{1}{2},j+\frac{1}{2}}(p^{(1)}+u^{(1)})-(D_x^-)_{i+\frac{1}{2},j+\frac{1}{2}}(p^{(1)}-u^{(1)})\Big] \\
& -\frac{1}{2}\Big[(D_y^+)_{i+\frac{1}{2},j+\frac{1}{2}}(p^{(1)}+v^{(1)})-(D_y^-)_{i+\frac{1}{2},j+\frac{1}{2}}(p^{(1)}-v^{(1)})\Big], \\
\frac{\mathrm{d}p^{(0)}_{i+\frac{1}{2},j}}{\mathrm{d}t} =& -\frac{1}{2}\Big[(D_x^+)_{i+\frac{1}{2},j}(p^{(1)}+u^{(1)})-(D_x^-)_{i+\frac{1}{2},j}(p^{(1)}-u^{(1)})\Big]-(D_y)_{i+\frac{1}{2},j}(v^{(1)}), \\
\frac{\mathrm{d}p^{(0)}_{i,j+\frac{1}{2}}}{\mathrm{d}t} =& -(D_x)_{i,j+\frac{1}{2}}u^{(1)}-\frac{1}{2}\Big[(D_y^+)_{i,j+\frac{1}{2}}(p^{(1)}+v^{(1)})-(D_y^-)_{i,j+\frac{1}{2}}(p^{(1)}-v^{(1)})\Big],
\end{aligned}
\tag{27}
$$

where the last three can be simplified based on the smoothness of $p^{(1)}$ and the Taylor expansion as

$$
\begin{aligned}
\frac{\mathrm{d}p^{(0)}_{i+\frac{1}{2},j+\frac{1}{2}}}{\mathrm{d}t} =& -\frac{1}{2}\Big[(D_x^+)_{i+\frac{1}{2},j+\frac{1}{2}}+(D_x^-)_{i+\frac{1}{2},j+\frac{1}{2}}\Big]u^{(1)}-\frac{1}{2}\Big[(D_y^+)_{i+\frac{1}{2},j+\frac{1}{2}}+(D_y^-)_{i+\frac{1}{2},j+\frac{1}{2}}\Big]v^{(1)} \\
& +\mathcal{O}(\Delta x^3+\Delta y^3), \\
\frac{\mathrm{d}p^{(0)}_{i+\frac{1}{2},j}}{\mathrm{d}t} =& -\frac{1}{2}\Big[(D_x^+)_{i+\frac{1}{2},j}+(D_x^-)_{i+\frac{1}{2},j}\Big]u^{(1)}-(D_y)_{i+\frac{1}{2},j}v^{(1)}+\mathcal{O}(\Delta x^3), \\
\frac{\mathrm{d}p^{(0)}_{i,j+\frac{1}{2}}}{\mathrm{d}t} =& -(D_x)_{i,j+\frac{1}{2}}u^{(1)}-\frac{1}{2}\Big[(D_y^+)_{i,j+\frac{1}{2}}+(D_y^-)_{i,j+\frac{1}{2}}\Big]v^{(1)}+\mathcal{O}(\Delta y^3).
\end{aligned}
\tag{28}
$$

Note that $\bar{u}^{(1)}_{ij}, u^{(1)}_{\zeta}, \bar{v}^{(1)}_{ij}, v^{(1)}_{\zeta}$ appeared in (28) are given in (26), and the cell center value is given by (23).

One can further examine the accuracy of (27). Since

$$
\begin{aligned}
-\hat{u}^{(1)}_{i+\frac{1}{2},j} =& -\frac{1}{6}\left(u^{(1)}_{i+\frac{1}{2},j-\frac{1}{2}}+4u^{(1)}_{i+\frac{1}{2},j}+u^{(1)}_{i+\frac{1}{2},j+\frac{1}{2}}\right) \\
=& \frac{1}{12\sigma}\Big[(D_x^+)_{i+\frac{1}{2},j-\frac{1}{2}}+(D_x^-)_{i+\frac{1}{2},j-\frac{1}{2}}+4(D_x^+)_{i+\frac{1}{2},j} \\
& +4(D_x^-)_{i+\frac{1}{2},j}+(D_x^+)_{i+\frac{1}{2},j+\frac{1}{2}}+(D_x^-)_{i+\frac{1}{2},j+\frac{1}{2}}\Big]p^{(0)} \\
=& \frac{1}{\sigma}\left(\frac{1}{\Delta y}\int_{y_{j-\frac{1}{2}}}^{y_{j+\frac{1}{2}}}p_x^{(0)}\mathrm{d}y\right)\Bigg|_{x_{i+\frac{1}{2}}}+\mathcal{O}(\Delta x^2+\Delta y^4),
\end{aligned}
$$

treating other numerical fluxes similarly, then (27) becomes

$$
\frac{\mathrm{d}}{\mathrm{d}t}\bar{p}_{ij}^{(0)}=\frac{1}{\sigma\Delta x}\left(\frac{1}{\Delta y}\int_{y_{j-\frac{1}{2}}}^{y_{j+\frac{1}{2}}}p_x^{(0)}\mathrm{d}y\right)\Bigg|_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}}+\frac{1}{\sigma\Delta y}\left(\frac{1}{\Delta x}\int_{x_{i-\frac{1}{2}}}^{x_{i-\frac{1}{2}}}p_y^{(0)}\mathrm{d}x\right)\Bigg|_{y_{j-\frac{1}{2}}}^{y_{j+\frac{1}{2}}}+\mathcal{O}(\Delta x^2+\Delta y^2).
\tag{29}
$$

Thus it is a finite volume approximation of $p_t^{(0)} = \frac{1}{\sigma}p_{xx}^{(0)} + \frac{1}{\sigma}p_{yy}^{(0)}$ with the truncation error $\mathcal{O}(\Delta x^2 + \Delta y^2)$. By the Taylor expansion, the other equations in (28) can be written as

$$\frac{\mathrm{d}p_\zeta^{(0)}}{\mathrm{d}t} = \left(\frac{1}{\sigma}p_{xx}^{(0)} + \frac{1}{\sigma}p_{yy}^{(0)}\right)\Bigg|_\zeta + \mathcal{O}(\Delta x^2 + \Delta y^2), \tag{30}$$

for all the point values at face centers or corners.

Proposition 7.1. Assuming sufficient regularity of the solutions, as $\epsilon \to 0$, the leading order solutions of the 2D AF schemes (20)-(24) satisfy (25)-(28), which are approximations of $p_t^{(0)} = \frac{1}{\sigma}p_{xx}^{(0)} + \frac{1}{\sigma}p_{yy}^{(0)}$ with truncation errors $\mathcal{O}(\Delta x^2 + \Delta y^2)$ in the sense of (29)-(30).

8 Numerical results

This section conducts some numerical experiments to verify the convergence rates and AP property of the AF schemes. Unless otherwise stated, periodic boundary conditions and the 3rd-order DIRK [5] are used and the time step size is chosen as

$$\Delta t = C_{\mathtt{CFL}} \min\{\Delta x, \Delta y\},$$

with the CFL number $C_{\mathtt{CFL}} = 1.0$. The linear system is solved by using GMRES with an incomplete LU factorization preconditioner [5].

Example 8.1 (1D accuracy test)**.** Consider the following exact solution [24],

$$p = \frac{1}{r}\exp(rt)\sin(x),\ u = \epsilon\exp(rt)\cos(x),\ r = \frac{-2}{1+\sqrt{1-4\epsilon^2}},$$

on the domain $[0, 2\pi]$ with periodic boundary conditions and $\sigma = 1$. The test is solved until $T = 1$ and the time step size is $0.2\Delta x^{4/3}$ to make the spatial error dominant.

The errors and convergence rates are plotted in Figure 1 for $\epsilon = 0.5, 10^{-2}, 10^{-6}$, which shows that the JS-based AF scheme is 3rd-order accurate for $\epsilon = 0.5$, while as $\epsilon \to 0$ the convergence rates reduce to 2 except for the 4th-order convergence of the point value in u. When $\epsilon = 10^{-2}$, it is verified that the scheme recovers the 3rd-order accuracy when the mesh size decreases to the magnitude of ϵ. The errors and convergence rates with the alternating flux are shown in Figure 2. For $\epsilon = 0.5$, one observes 3rd-order accuracy, while for $\epsilon = 10^{-2}$ and $\epsilon = 10^{-6}$, the convergence order is 4 for the cell average of p and point value of u, and 3 for the other degrees of freedom. The appearance of the super-convergence (4th order instead of 3rd order) needs further study.

Example 8.2 (Square wave)**.** The initial p of this test is a square wave, adapted from [24],

$$p = \begin{cases} 2, & \text{if } |x| < 0.5, \\ 1, & \text{otherwise}, \end{cases}$$

and $u = 0$ in the whole domain $[-1, 1]$. Two cases are considered: the transport regime $\epsilon = 0.7$ and the diffusive regime $\epsilon = 10^{-6}$, with $\sigma = 1$, and the final time is 0.25 and 0.04, respectively.

The results obtained by the JS-based AF scheme on a coarse mesh of 40 cells are shown in Figure 3. It is observed that the scheme can capture shock waves for $\epsilon = 0.7$ correctly, and can also capture the diffusive solution for $\epsilon = 10^{-6}$ well.

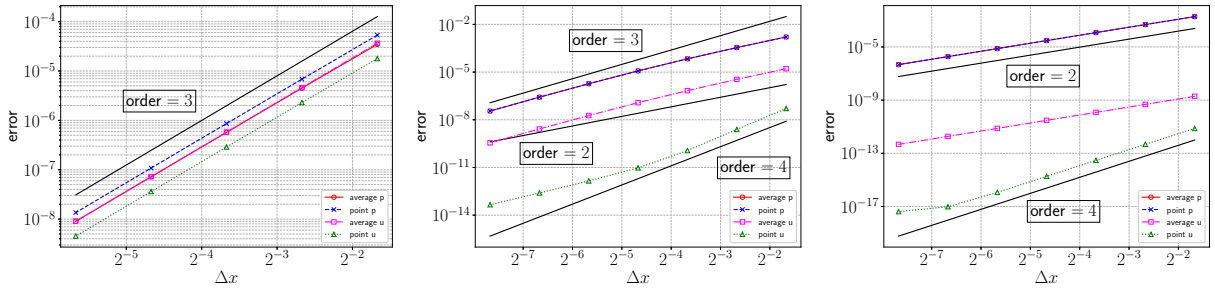

Figure 1: Example 8.1. The errors and convergence rates obtained by the JS-based AF scheme and $\epsilon = 0.5, 10^{-2}, 10^{-6}$ from left to right.

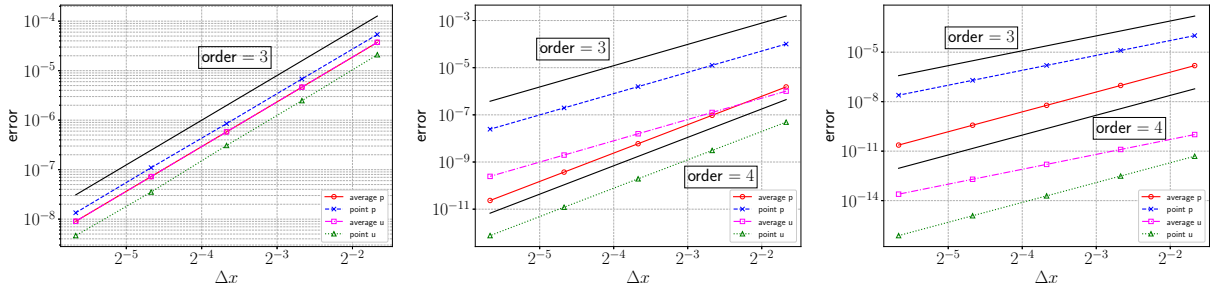

Figure 2: Example 8.1. The errors and convergence rates obtained with the alternating flux and $\epsilon = 0.5, 10^{-2}, 10^{-6}$ from left to right.

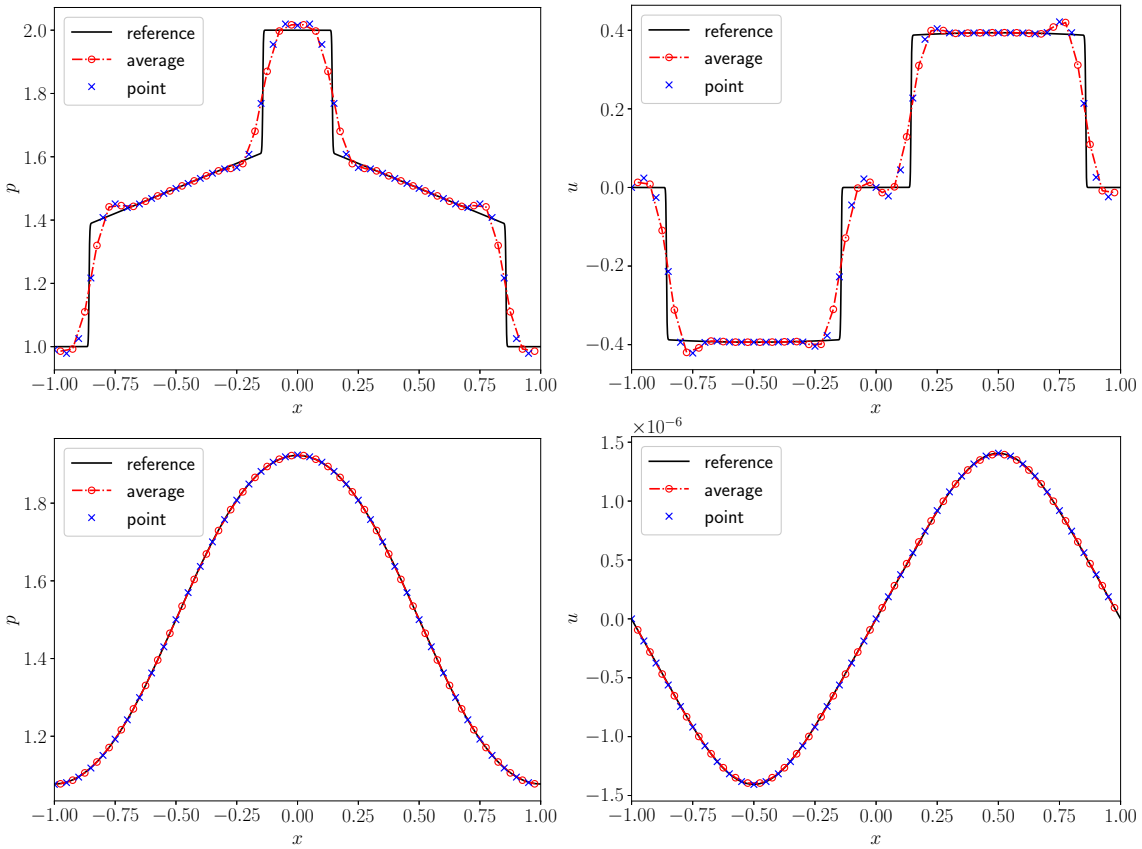

Figure 3: Example 8.2. The results obtained by the JS-based AF scheme. Top row: $\epsilon = 0.7$, bottom row: $\epsilon = 10^{-6}$. Left: p, right: u.

Example 8.3 (Variable σ)**.** This test is used to show that the JS-based AF scheme also works for variable σ. This test problem takes the same initial condition and computational domain as Example 8.2, except for $\epsilon = 1$, and $\sigma = 1 + (10x)^2$. Note that the solution tends to be in the diffusive regime for large σ.

Figure 4 shows the results obtained by the JS-based AF scheme on a coarse mesh of 40 cells. It is seen that the scheme can capture both transport and diffusive regimes well in the domain, which verifies the AP property of the scheme.

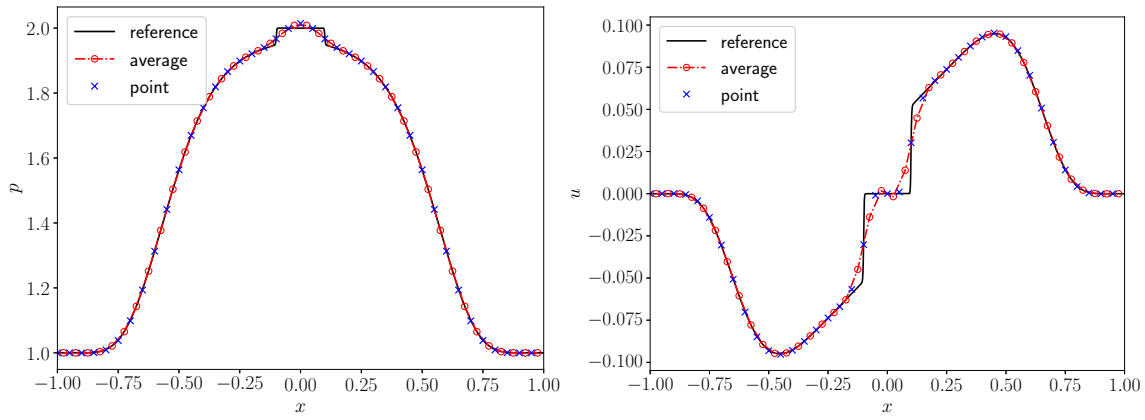

Figure 4: Example 8.3. The results obtained by the JS-based AF scheme. Left: p, right: u.

Example 8.4 (2D accuracy test)**.** We construct the following exact solution similar to Example 8.1,

$$p = \frac{2}{r}\exp(rt)\sin(x)\sin(y),\ \ u = \epsilon\exp(rt)\cos(x)\sin(y),$$
$$v = \epsilon\exp(rt)\sin(x)\cos(y),\ \ r = \frac{-4}{1+\sqrt{1-8\epsilon^2}},$$

in the domain $[0, 2\pi] \times [0, 2\pi]$ with periodic boundary conditions and $\sigma = 1$. The final time is $T = 0.1$ and the CFL number is 0.2 to make the spatial error dominant.

The errors and convergence orders are shown in Figure 5. For $\epsilon = 0.3$, the AF gets the 3rd-order accuracy for all the cell averages, also the 3rd order for the point value in p, while the 2nd order for the point value in u, v, which is due to the mesh alignment issue (using the LLF FVS [13] recovers the 3rd order). For $\varepsilon = 10^{-6}$, all the convergence rates are 2.

Example 8.5 (Fundamental solution in transport regime)**.** This test simulates wave propagation in the domain $[-1, 1] \times [-1, 1]$ with $\sigma = \epsilon = 1$ from an initial Dirac function of p [11], i.e., $p_0 = \frac{1}{\Delta x \Delta y}$ for the centered cell average otherwise $p_0 = 0$, and $u_0 = v_0 = 0$ in the whole domain.

Figure 6 shows the results obtained by using the JS-based AF scheme with 101×101 cells. One can observe that the scheme can capture the transport phenomenon and preserve the circular shape of the wave front well.

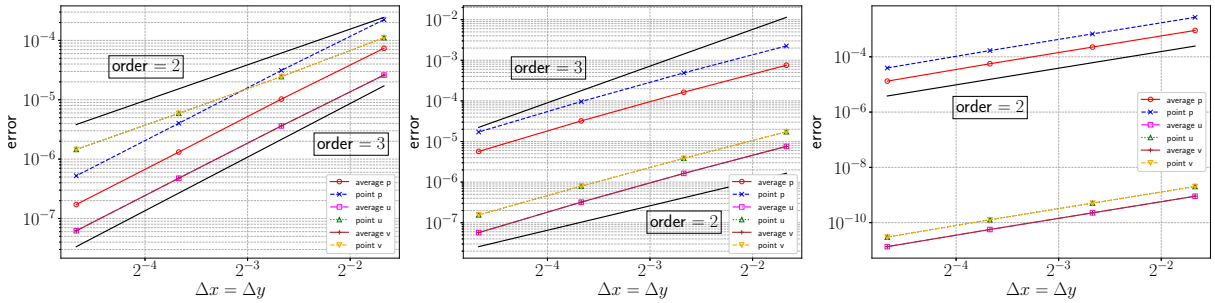

Figure 5: Example 8.4. The errors and convergence rates obtained by the JS-based AF scheme and $\epsilon = 0.3, 10^{-2}, 10^{-6}$ from left to right.

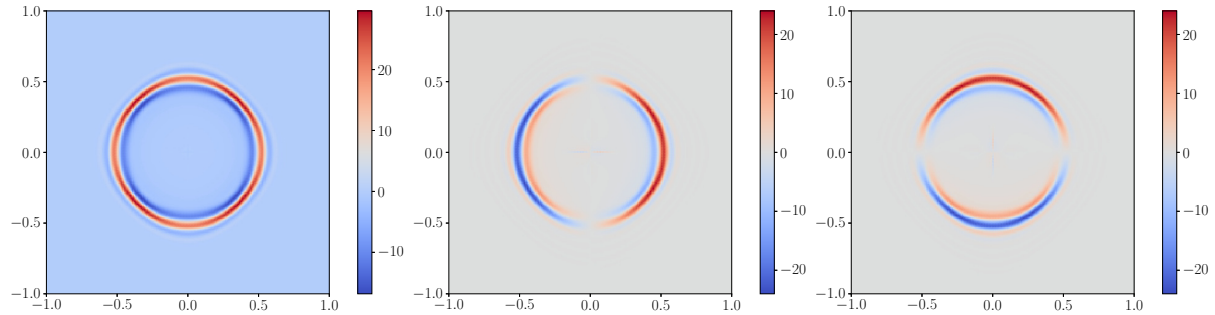

Figure 6: Example 8.5. The results obtained by the JS-based AF scheme. From left to right: p, u, v.

Example 8.6 (Radiation)**.** This is a test problem motivated by radiation simulations, where σ depends on the fluid temperature. We follow the setup in [11]. The domain is a square $[-1, 1] \times [-1, 1]$, and the initial condition is

$$p_0 = 10^{-3} + 100 \exp(-(x^2 + y^2)/0.01), \ u_0 = v_0 = 0.$$

In this test, $\epsilon = 1$, and the coefficient σ is chosen as 1 in the domain except for 10^4 in eight box regions, see Figure 7. For example, in the first quadrant, the two boxes are $[\frac{3}{16}, \frac{7}{16}] \times [\frac{9}{16}, \frac{13}{16}]$, and $[\frac{9}{16}, \frac{13}{16}] \times [\frac{3}{16}, \frac{7}{16}]$, and the other boxes are mirror symmetric with respect to the axes.

The results are obtained by using the 200×200 mesh, shown in Figure 8. The diffusion coefficient $1/\sigma$ is very small in box regions, thus the diffusion behavior is much slower, which can be clearly observed. Our results are comparable to those in [11] (using comparable mesh size), with better resolution near the box regions.

9 Conclusion

In the active flux (AF) methods, the way how point values at cell interfaces are updated is essential to achieve stability, high-order accuracy, and other nice properties. The point value update based on Jacobian splitting (JS) has been used in many existing works. This paper investigates the JS-based AF scheme for solving the hyperbolic heat equation. It is shown that the scheme without any modification is AP in the diffusive scaling by using the formal asymptotic analysis, discrete Fourier analysis, and numerical tests. The convergence

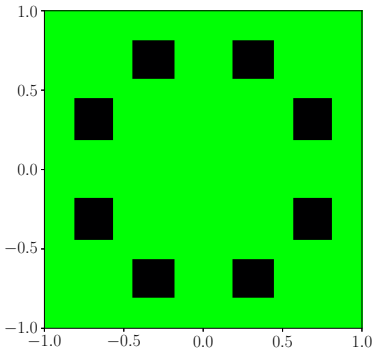

Figure 7: Example 8.6. $\sigma = 1$ in the domain except for $\sigma = 10^4$ in the black box regions.

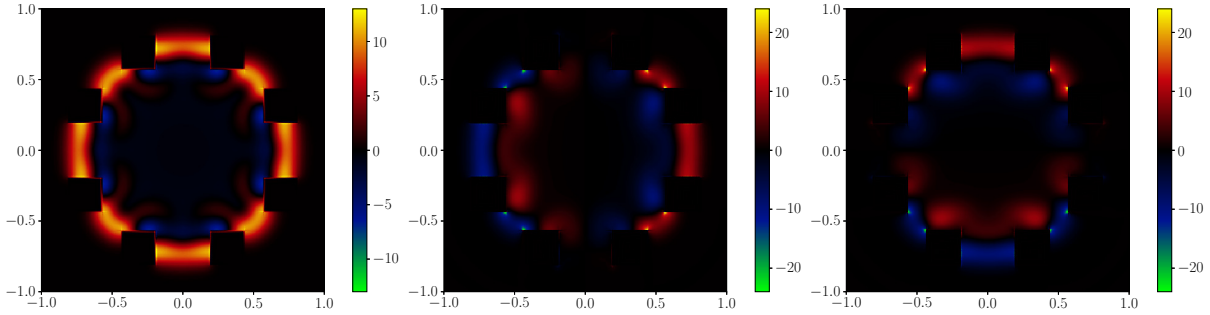

Figure 8: Example 8.6. The results obtained by the JS-based AF scheme. From left to right: p, u, v.

rates are also studied for the scheme and its limit, and the order degradation is observed in the limit.

Acknowledgement

JD was supported by an Alexander von Humboldt Foundation Research Fellowship CHN-1234352-HFST-P. CK and WB acknowledge funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) within *SPP 2410 Hyperbolic Balance Laws in Fluid Mechanics: Complexity, Scales, Randomness (CoScaRa)*, project number 525941602.

References

[1] R. Abgrall, *A combination of residual distribution and the active flux formulations or a new class of schemes that can combine several writings of the same hyperbolic problem: Application to the 1D Euler equationss*, Commun. Appl. Math. Comput., 5 (2023), pp. 370–402, https://doi.org/10.1007/s42967-021-00175-w.

[2] R. Abgrall and W. Barsukow, *Extensions of active flux to arbitrary order of accuracy*, Comput. Methods Appl. Mech. Engrg., 57 (2023), pp. 991–1027, https://doi.org/10.1051/m2an/2023004.

[3] R. Abgrall, W. Barsukow, and C. Klingenberg, *A semi-discrete active flux method for the Euler equations on Cartesian grids*, J. Sci. Comput., 102 (2025), p. 36, https://doi.org/10.1007/s10915-024-02749-1.

[4] R. Abgrall, M. Jiao, Y. Liu, and K. Wu, *Bound preserving Point-Average-Moment PolynomiAl-interpreted (PAMPA) scheme: one-dimensional case*, Oct. 2024, https://doi.org/10.48550/arXiv.2410.14292.

[5] S. Balay, S. Abhyankar, M. F. Adams, S. Benson, J. Brown, P. Brune, K. Buschelman, E. M. Constantinescu, L. Dalcin, A. Dener, V. Eijkhout, J. Faibussowitsch, W. D. Gropp, V. Hapla, T. Isaac, P. Jolivet, D. Karpeev, D. Kaushik, M. G. Knepley, F. Kong, S. Kruger, D. A. May, L. C. McInnes, R. T. Mills, L. Mitchell, T. Munson, J. E. Roman, K. Rupp, P. Sanan, J. Sarich, B. F. Smith, S. Zampini, H. Zhang, H. Zhang, and J. Zhang, *PETSc Web page*, 2024, https://petsc.org/.

[6] W. Barsukow, *The active flux scheme for nonlinear problems*, J. Sci. Comput., 86 (2021), p. 3, https://doi.org/10.1007/s10915-020-01381-z.

[7] W. Barsukow and J. P. Berberich, *A well-balanced active flux method for the shallow water equations with wetting and drying*, Commun. Appl. Math. Comput., (2023), https://doi.org/10.1007/s42967-022-00241-x.

[8] W. Barsukow, J. P. Berberich, and C. Klingenberg, *On the active flux scheme for hyperbolic PDEs with source terms*, SIAM J. Sci. Comput., 43 (2021), pp. A4015–A4042, https://doi.org/10.1137/20M1346675.

[9] W. Barsukow, J. Hohm, C. Klingenberg, and P. L. Roe, *The active flux scheme on Cartesian grids and its low Mach number limit*, J. Sci. Comput., 81 (2019), pp. 594–622, https://doi.org/10.1007/s10915-019-01031-z.

[10] M. Bennoune, M. Lemou, and L. Mieussens, *Uniformly stable numerical schemes for the Boltzmann equation preserving the compressible Navier-Stokes asymptotics*, J. Comput. Phys., 227 (2008), pp. 3781–3803, https://doi.org/10.1016/j.jcp.2007.11.032.

[11] C. Buet, B. Després, and E. Franck, *Design of asymptotic preserving finite volume schemes for the hyperbolic heat equation on unstructured meshes*, Numer. Math., 122 (2012), pp. 227–278, https://doi.org/10.1007/s00211-012-0457-9.

[12] E. Chudzik, C. Helzel, and M. Lukáčová-Medvid'ová, *Active flux methods for hyperbolic systems using the method of bicharacteristics*, J. Sci. Comput., 99 (2024), p. 16, https://doi.org/10.1007/s10915-024-02462-z.

[13] J. Duan, W. Barsukow, and C. Klingenberg, *Active flux methods for hyperbolic conservation laws – flux vector splitting and bound-preservation*, SIAM J. Sci. Comput., 47 (2025), pp. A811–A837, https://doi.org/10.1137/24M1658887.

[14] J. Duan, P. Chandrashekar, and C. Klingenberg, *Active flux for ideal magnetohydrodynamics: A positivity-preserving scheme with the godunov-powell source term*, 2025, https://doi.org/10.48550/ARXIV.2506.04857.

[15] T. Eymann and P. Roe, *Active flux schemes*, in 49th AIAA Aerospace Sciences Meeting including the New Horizons Forum and Aerospace Exposition, Orlando, Florida, Jan. 2011, American Institute of Aeronautics and Astronautics, https://doi.org/10.2514/6.2011-382.

[16] T. Eymann and P. Roe, *Active flux schemes for systems*, in 20th AIAA Computational Fluid Dynamics Conference, Fluid Dynamics and Co-located Conferences, American Institute of Aeronautics and Astronautics, June 2011, https://doi.org/10.2514/6.2011-3840.

[17] T. A. Eymann and P. L. Roe, *Multidimensional active flux schemes*, in 21st AIAA Computational Fluid Dynamics Conference, Fluid Dynamics and Co-located Conferences, American Institute of Aeronautics and Astronautics, June 2013, https://doi.org/10.2514/6.2013-2940.

[18] D. Fan, *On the Acoustic Component of Active Flux Schemes for Nonlinear Hyperbolic Conservation Laws*, thesis, University of Michigan, 2017, http://deepblue.lib.umich.edu/handle/2027.42/140800.

[19] D. Fan and P. L. Roe, *Investigations of a new scheme for wave propagation*, in 22nd AIAA Computational Fluid Dynamics Conference, American Institute of Aeronautics and Astronautics, 2015, https://doi.org/10.2514/6.2015-2449.

[20] L. Gosse and G. Toscani, *An asymptotic-preserving well-balanced scheme for the hyperbolic heat equations*, Cr. Math., 334 (2002), pp. 337–342, https://doi.org/10.1016/S1631-073X(02)02257-4.

[21] W. Guo, X. Zhong, and J.-M. Qiu, *Superconvergence of discontinuous Galerkin and local discontinuous Galerkin methods: Eigen-structure analysis based on Fourier approach*, J. Comput. Phys., 235 (2013), pp. 458–485, https://doi.org/10.1016/j.jcp.2012.10.020.

[22] M. M. Hasan and X. Zeng, *A central compact hybrid-variable method with spectral-like resolution: One-dimensional case*, J. Comput. Appl. Math., 421 (2023), p. 114894, https://doi.org/10.1016/j.cam.2022.114894.

[23] C. Helzel, D. Kerkmann, and L. Scandurra, *A new ADER method inspired by the active flux method*, J. Sci. Comput., 80 (2019), pp. 1463–1497, https://doi.org/10.1007/s10915-019-00988-1.

[24] J. Jang, F. Li, J. M. Qiu, and T. Xiong, *High order asymptotic preserving DG-IMEX schemes for discrete-velocity kinetic equations in a diffusive scaling*, J. Comput. Phys., 281 (2015), pp. 199–224, https://doi.org/10.1016/j.jcp.2014.10.025.

[25] S. Jin, *Asymptotic-preserving schemes for multiscale physical problems*, Acta Numer., 31 (2022), pp. 415–489, https://doi.org/10.1017/S0962492922000010.

[26] S. Jin and C. D. Levermore, *Numerical schemes for hyperbolic conservation laws with stiff relaxation terms*, J. Comput. Phys., 126 (1996), pp. 449–467, https://doi.org/10.1006/jcph.1996.0149.

[27] S. Jin, L. Pareschi, and G. Toscani, *Uniformly accurate diffusive relaxation schemes for multiscale transport equations*, SIAM J. Numer. Anal., 38 (2000), pp. 913–936, https://doi.org/10.1137/S0036142998347978.

[28] C. A. Kennedy and M. H. Carpenter, *Diagonally implicit Runge-Kutta methods for ordinary differential equations. A review*, Tech. Report NF1676L-19716, NASA Langley Research Center, Mar. 2016, https://ntrs.nasa.gov/citations/20160005923.

[29] C. A. Kennedy and M. H. Carpenter, *Diagonally implicit Runge–Kutta methods for stiff ODEs*, Appl. Numer. Math., 146 (2019), pp. 221–244, https://doi.org/10.1016/j.apnum.2019.07.008.

[30] A. Klar, *An asymptotic-induced scheme for nonstationary transport equations in the diffusive limit*, SIAM J. Numer. Anal., 35 (1998), pp. 1073–1094, https://doi.org/10.1137/S0036142996305558.

[31] A. Klar and C. Schmeiser, *Numerical passage from radiative heat transfer to nonlinear diffusion models*, Math. Models Methods Appl. Sci., 11 (2001), pp. 749–767, https://doi.org/10.1142/S0218202501001082.

[32] E. W. Larsen and J. E. Morel, *Asymptotic solutions of numerical transport problems in optically thick, diffusive regimes II*, J. Comput. Phys., 83 (1989), pp. 212–236, https://doi.org/10.1016/0021-9991(89)90229-5.

[33] E. W. Larsen, J. E. Morel, and W. F. Miller, *Asymptotic solutions of numerical transport problems in optically thick, diffusive regimes*, J. Comput. Phys., 69 (1987), pp. 283–324, https://doi.org/10.1016/0021-9991(87)90170-7.

[34] M. Lemou and L. Mieussens, *A new asymptotic preserving scheme based on micro-macro formulation for linear kinetic equations in the diffusion limit*, SIAM J. Sci. Comput., 31 (2008), pp. 334–368, https://doi.org/10.1137/07069479X.

[35] Z. Peng, Y. Cheng, J.-M. Qiu, and F. Li, *Stability-enhanced AP IMEX-LDG schemes for linear kinetic transport equations under a diffusive scaling*, J. Comput. Phys., 415 (2020), p. 109485, https://doi.org/10.1016/j.jcp.2020.109485.

[36] P. Roe, *Is discontinuous reconstruction really a good idea?*, J. Sci. Comput., 73 (2017), pp. 1094–1114, https://doi.org/10.1007/s10915-017-0555-z.

[37] B. Van Leer, *Towards the ultimate conservative difference scheme. IV. A new approach to numerical convection*, J. Comput. Phys., 23 (1977), pp. 276–299, https://doi.org/10.1016/0021-9991(77)90095-X.

[38] X. Zeng, *A high-order hybrid finite difference–finite volume approach with application to inviscid compressible flow problems: A preliminary study*, Comput. Fluids, 98 (2014), pp. 91–110, https://doi.org/10.1016/j.compfluid.2014.02.007.

[39] M. Zhang and C.-W. Shu, *An analysis of three different formulations of the discontinuous Galerkin method for diffusion equations*, Math. Models Methods Appl. Sci., 13 (2003), pp. 395–413, https://doi.org/10.1142/S0218202503002568.

A Example 8.1 with non-well-prepared data

We also compute Example 8.1 with non-well-prepared data $u = 0.1$, and p is kept the same. The reference solution is obtained by the same scheme on a fine mesh. Figure 9 shows the errors and convergence rates. The 3rd-order convergence is observed for $\epsilon = 10^{-2}$ when Δx is small enough, while for $\epsilon = 10^{-6}$, the convergence rate is the 1st order except for the 2nd-order accuracy in the cell average of u.

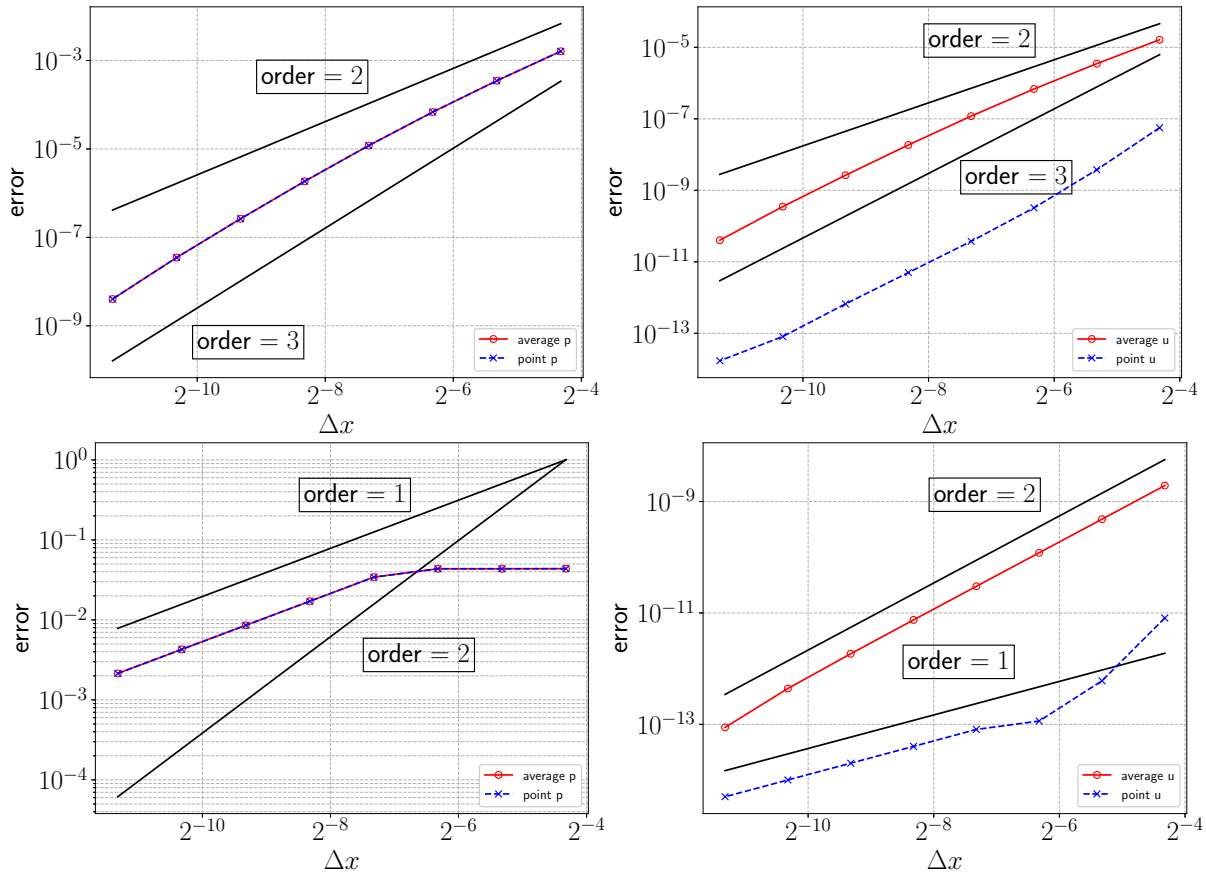

Figure 9: Example 8.1. The errors and convergence rates obtained with the non-well-prepared initial data $u(0) = 0.1$ and the same $p(0)$, where $\epsilon = 10^{-2}$ (upper) and $\epsilon = 10^{-6}$ (lower).