An *a posteriori* strategy for adaptive schemes in time and space

Maria T. Malheiro*1, Gaspar J. Machado†2, and Stéphane Clain‡3

1Centre of Mathematics and Department of Mathematics, University of Minho, Campus of
Azurém, 4800-058 Guimarães, Portugal
2Centre of Physics and Department of Mathematics, University of Minho, Campus of
Azurém, 4800-058 Guimarães, Portugal
3Centre of Physics and Department of Mathematics, University of Minho, Campus of
Azurém, 4800-058 Guimarães, Portugal

October, 2020

Abstract

A nonlinear adaptive procedure for optimising both the schemes in time and space is proposed in view
of increasing the numerical efficiency and reducing the computational time. The method is based on a
four-parameter family of schemes we shall tune in function of the physical data (velocity, diffusion), the
characteristic size in time and space, and the local regularity of the function leading to a nonlinear procedure.
The *a posteriori* strategy we adopt consists in, given the solution at time $t^n$, computing a candidate solution
with the highest accurate schemes in time and space for all the nodes. Then, for the nodes that present some
instabilities, both the schemes in time and space are modified and adapted in order to preserve the stability
with a large time step. The updated solution is computed with node-dependent schemes both in time and
space. For the sake of simplicity, only convection-diffusion problems are addressed as a prototype with a
two-parameters five-points finite difference method for the spatial discretisation together with an explicit
time two-parameters four-stages Runge-Kutta method. We prove that we manage to obtain an optimal
time-step algorithm that produces accurate numerical approximations exempt of non-physical oscillations.

1 Introduction

High order discrete schemes for equations involving hyperbolic operators are likely to produce non physical
oscillations. Even for linear problems, a nonlinear routine is mandatory to control the over- and under-shooting
in the vicinity of points where the solution presents large gradients. Technologies such as MUSCL or WENO [19,
20], among the most popular, have been developed for half a century and manage to efficiently reduce or eliminate
the numerical instabilities, namely to avoid the oscillations near the discontinuities and the extrema points. More
specifically, for the convection-diffusion problems, different approaches have been considered: variable-order non-
oscillatory scheme (VONOS), hybrid-linear parabolic approximation (HLPA), sharp and monotonic algorithm
for realistic transport (SMART), weighted-average coefficient ensuring boundedness (WACEB), convergent and
universally bounded interpolation scheme for the treatment of advection (CUBISTA) and an adaptive bounded
version of the QUICK with estimated streaming terms (QUICKEST) called ADBQUICKEST (see references
for these methods in [12]). The numerical solution obtained with these schemes is at least second-order accurate
in regions where the solution is smooth enough but retains the first-order approximation in regions where the
solution presents large gradients for the sake of stability.

All the stabilisation procedures mainly address the scheme in space whereas the scheme in time is merely
discretised with a Runge-Kutta (RK) method or its Strong Stability Preserving (SSP) version [26, 27]. Very little
attention has been paid on the time discretisation and a nonlinear dynamical procedure for optimising both
the schemes in time and space is desirable to increase the numerical efficiency. Several traditional numerical
methods were revisited in order to articulate space and time schemes together, aiming to increase the allowable
time step. Bourchtein [5] constructed an explicit central difference method of second order applied to the
one- and two-dimensional advection equations based on the generalised leap-frog type method with the main
goal of increasing the allowable time step, with some deterioration in the accuracy of the solution. Chadha
and Malden [7] consider the numerical solution of a linear time dependent advection–diffusion problem by an
implicit two-weight, three-point central finite difference scheme. They extend the scheme proposed by them in [5], to incorporate an optimal time step selection algorithm for the method. The resulting method, based on optimal values of weights and optimal time-stepping, is of fifth-order in space, and third-order in time.

Other numerical schemes are based on prediction-correction techniques [25]. In [3], the authors present an adaptive finite element scheme for the advection-reaction-diffusion equation based on a stabilized finite element method combined with a residual error estimator. The adaptive process corrects the meshes allowing to capture boundary and inner layers very sharply and without significant oscillations. In [17], the authors have developed some a posteriori error estimates using a stabilized scheme combined with a shock-capturing technique to control the local oscillations in the crosswind direction. In [28], the author has introduced an error estimate for the advection-diffusion equation based on the solution of local problems on each element of the triangulation. Artificial compression method (ACM) based filter scheme is also investigated in [29]. The spatially fourth-order (or higher non-dissipative) scheme in space is used at all times but additional numerical dissipation is made at the shock layers to control the stability.

In the last decade, the a posteriori paradigm has been developed to prevent the numerical solution from creating non physical oscillations: the so-called Multidimensional Optimal Order Detection Method (MOOD) method [9, 10]. The principle consists in building a candidate solution with the highest order scheme in space. Then the guess solution admissibility is analysed using detectors to check some physical properties of non physical oscillations. The nodes that present non-compliance’s solution are tagged and the numerical scheme is only altered for that points by reducing the scheme order (basically adding more viscosity or reducing the polynomial reconstruction degree). The numerical approximation for the cured nodes and their neighbours is computed again to eliminate the oscillations. The goal of this the paper is the design of an adaptive technique, based on the a posteriori paradigm, to provide the optimal choice of the time and space schemes, that enables to compute a stable solution with the largest time step.

To present our strategy, we deal with the one-dimension linear convection-diffusion problem since all the important ingredients are already addressed in this simple scalar equation and enable to highlight the connections between the scheme in time and the scheme in space. Notice that, even a linear problem requires a nonlinear routine for stabilisation when dealing with rough solutions and the convection-diffusion equation involves the two main operator in simulations. For the sake of simplicity, we only consider two schemes in space and two schemes in time with very different characteristics and we aim to optimise the space accuracy and the time step to produce physically admissible numerical solutions with no spurious oscillations.

The crucial point lies on the confrontation between the discretisations in space and time: the eigenvalues associated to the space scheme have to fit into the stability domain of the RK scheme. Several parameters play a major role in the trade-off between accuracy and stability. On the one hand, the scheme in space together with the Péclet number determine the eigenvalues distribution in the complex plane. On the other hand, the RK scheme stability region and the accuracy is characterised by the Butcher tableau [6]. At last, the stability condition between the two schemes is controlled by the Courant–Friedrichs–Lewy (CFL) condition depending on the time discretisation parameter \( \Delta t \).

In this study, we consider a two-parameter family of finite difference schemes in space that characterise the spectrum and the accuracy. Similarly, the time discretization is a two-parameter family of schemes by considering a four-stage RK method where we impose to be at least second-order. The global method is a four-parameter family of schemes we shall tune in function of the physical data (velocity, diffusion), the characteristic size in time and space, and the local regularity of the function leading to a nonlinear procedure since the scheme depends on the approximation itself.

The design of the schemes in space has to respect some basic principles in order to produce an eligible blending. First we only consider five-point finite difference methods involving the same centred stencil but the coefficients are node dependent. Similarly the four-stage RK scheme is also node dependent and it is mandatory that the four sub-steps are the same for all the nodes for the sake of compatibility leading to some additional constraints in the design of the Butcher’s tableau. The a posteriori strategy we adopt consists in, given the solution at time \( t^n \), computing a candidate solution with the highest accurate scheme in time and space for all the nodes. Then applying the detectors, we determine the nodes to be corrected and switch the scheme in space. Unfortunately, it usually leads to a reduction of the local time step for the sake of stability. To overcome such a problem and preserve a large time step, we may have to switch the time scheme. We then obtain, the optimal combination of time and space schemes for each node and iteration while we use the common time step to update the solution.

The organisation of the paper is the following. We briefly present in Section 2 the spectral analysis of the five-point finite difference method focusing on the spectral curves description with respect to the two free parameters. Section 3 is dedicated to the four-stage RK method where we analyse the impact of the two free parameters on the stability region. The stability of the time and space schemes combination is carried out in Section 4 where we determine the optimal time step for each scenario. Finally, Section 5 presents our a posteriori method to design a node by node optimal scheme both in time and space. At last in Section 6 conclusions and perspectives are drawn.
2 Discrete convection-diffusion operator analysis

This section is dedicated to the spectrum of the discrete convection-diffusion operator we shall use in our stability analysis. Let $\phi = \phi(x)$ be a smooth 1-periodic function defined in $\mathbb{R}$, i.e., $\phi(x + 1) = \phi(x)$. We define the convection-diffusion operator

$$E[\phi] = -u\phi' + \kappa\phi'' , \quad (1)$$

where $u \geq 0$ and $\kappa \geq 0$ are the convective and diffusive coefficients, respectively. We restrict the study case to the bounded domain $[0, 1]$ by applying a periodic condition. Let $I \in \mathbb{N}$ and $\Delta x = 1/I$. We denote $x_i = i\Delta x$, $i \in \mathbb{Z}$, a uniform discretisation of the real axis and set $\Psi = (\psi_i)_{i \in \mathbb{Z}}$ a vector with an infinite number of real value entries. Periodicity yields $\psi_{i+I} = \psi_i$, for all $i \in \mathbb{Z}$, hence the relevant data is only given by components $\psi_i$, $i = 1, \ldots, I$. We shall use the same notation $\Psi = (\psi_i)_{i=1}^I$ to denote both the whole vector and its finite representation, being the other components given by periodicity.

2.1 Five-point discrete schemes

A generic conservative five-point numerical scheme is defined by an ordered list of 5 coefficients which we indicate with

$$E = (a_{-2}, a_{-1}, a_0, a_1, a_2), \quad a_j \in \mathbb{R}, j = -2, \ldots, 2, \quad (2)$$

where the coefficients satisfy the null summation constraint for the sake of conservation

$$a_{-2} + a_{-1} + a_0 + a_1 + a_2 = 0.$$

For any $I$-periodic vector $\Psi = (\psi_i)_{i \in \mathbb{Z}}$, the E-scheme applied to $\Psi$ provides the vector $E\Psi$ given component-wise by

$$(E\Psi)_i = a_{-2}\psi_{i-2} + a_{-1}\psi_{i-1} + a_0\psi_i + a_1\psi_{i+1} + a_2\psi_{i+2}, \quad i \in \mathbb{Z}.$$

Periodicity of $\Psi$ yields that $E\Psi$ also satisfies the periodicity property hence the finite vector representation $E\Psi = \{(E\Psi)_i\}_{i=1}^I$ completely describes the whole vector. We highlight the four particular cases

$$E_1 = \left(\frac{1}{12}, -\frac{2}{3}, 0, 2, -\frac{1}{12}\right), \quad E_2 = \left(-\frac{1}{12}, \frac{4}{3}, -\frac{5}{2}, -\frac{3}{4}, -\frac{1}{12}\right),$$

$$E_3 = \left(-\frac{1}{2}, 1, 0, -1, \frac{1}{2}\right), \quad E_4 = (1, -4, 6, -4, 1),$$

that provide approximations for the first-, second-, third-, and fourth-order derivatives, respectively. Denoting $\Psi = (\phi(x_i))_{i \in \mathbb{Z}}$ for any regular 1-periodic function $\phi = \phi(x)$, consistency errors read

$$E_1\Psi_i = \phi^{(1)}(x_i) + \Theta(\Delta x^4), \quad (E_2\Psi)_i = \phi^{(2)}(x_i) + \Theta(\Delta x^4), \quad (E_3\Psi)_i = \phi^{(3)}(x_i) + \Theta(\Delta x^2), \quad (E_4\Psi)_i = \phi^{(4)}(x_i) + \Theta(\Delta x^2).$$

The fourth-order five-point discretisation of operator $[\mathbb{I}]$ is given by the optimal combination

$$E = -\frac{u}{\Delta x} \left( E_1 + \frac{E_2}{Pe} \right),$$

where $Pe = \frac{u\Delta x}{\kappa}$ represents the cell Péclet number. Instabilities may appear and one has to damp the oscillations by using third- and fourth-derivative approximations. To this end, we consider more general five-point conservative schemes of the form

$$E = E(\theta_3, \theta_4, Pe) = -\frac{u}{\Delta x} \left( E_1 + \frac{E_2}{Pe} + \theta_3 E_3 + \theta_4 E_4 \right), \quad (3)$$

parameterised by $\theta_3, \theta_4 \in \mathbb{R}$, and $Pe$. Substituting the expressions of $E_1$, $E_2$, $E_3$, and $E_4$, scheme $E$ reads

$$E(\theta_3, \theta_4, Pe) = -\frac{u}{\Delta x} \left( \frac{(12\theta_4 - 3\theta_3 + 1)Pe + 1}{12Pe}, \frac{(12\theta_4 - 3\theta_3 + 2)Pe + 4}{3Pe}, \frac{12\theta_4 Pe + 5}{2Pe}, \frac{-(12\theta_4 + 3\theta_3 - 2)Pe + 4}{3Pe}, \frac{(12\theta_4 + 6\theta_3 - 1)Pe + 1}{12Pe} \right). \quad (4)$$

Remark 1 If $u = 0$ and $\kappa \neq 0$ (pure diffusive problem), $Pe = 0$ and expression $[\mathbb{I}]$ should be rewritten specifically for $u = 0$ taking into account $\theta_3$ and $\theta_4$. If $\kappa = 0$ and $u \neq 0$ (pure convective problem) we will write
Letting \( S \) schematically, we have

\[
\text{Remark 2}
\]

It is worth noting that the spectral curve shape is characterised by function \( \rho \).

2.2 Spectra

Due to the periodicity assumption, scheme (\( \mathbf{1} \)) results in a circulant square matrix \( A \) of order \( I \) with entries

\[
a_{ij} = \begin{cases} 0 & \text{if } j = i, \text{ otherwise.} \end{cases}
\]

Schematically, we have

\[
A = \begin{bmatrix}
    a_0 & a_1 & a_2 & 0 & 0 & 0 & \cdots & 0 & a_{-2} & a_{-1} \\
    a_{-1} & a_0 & a_1 & a_2 & 0 & 0 & \cdots & 0 & 0 & a_{-2} \\
    a_{-2} & a_{-1} & a_0 & a_1 & a_2 & 0 & \cdots & 0 & 0 & 0 \\
    0 & a_{-2} & a_{-1} & a_0 & a_1 & a_2 & \cdots & 0 & 0 & 0 \\
    \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
    a_2 & 0 & 0 & 0 & 0 & 0 & \cdots & a_{-1} & a_0 & a_1 \\
    a_1 & a_2 & 0 & 0 & 0 & 0 & \cdots & a_{-2} & a_{-1} & a_0
\end{bmatrix}
\]

The eigenvectors of the circulant matrices associated to schemes \( E_1, E_2, E_3, \) and \( E_4 \) are the same, hence they also are the eigenvectors of the circulant matrix associated to scheme \( E \) and independent of \( \theta_3, \theta_4, \) and \( Pe \). They are given by

\[ \lambda_i(E_{\theta_3, \theta_4, Pe}) = \frac{\lambda_i(E_1) - \lambda_i(E_2)}{Pe} + \theta_3 \lambda_i(E_3) + \theta_4 \lambda_i(E_4) \]

The eigenvalues for the four particular operators read:

- \( \lambda_i(E_1) = i \sin(2\pi i \Delta x) \left( 1 - \frac{1}{3} \left\{ \cos(2\pi i \Delta x) - 1 \right\} \right) \in \left[ -\frac{4}{3}, \frac{4}{3} \right] \)
- \( \lambda_i(E_2) = (\cos(2\pi i \Delta x) - 1) \left( 2 - \frac{1}{3} \left\{ \cos(2\pi i \Delta x) - 1 \right\} \right) \in \left[ -\frac{16}{3}, 0 \right] \)
- \( \lambda_i(E_3) = 2i \sin(2\pi i \Delta x) (\cos(2\pi i \Delta x) - 1) \in [-2i, 2i] \)
- \( \lambda_i(E_4) = 4 \left( \cos(2\pi i \Delta x) - 1 \right)^2 \in [0, 16] \)

Letting \( I \to +\infty \), we obtain the continuous parameterised spectral curve

\[
\lambda(s; \theta_3, \theta_4, Pe) = \frac{u}{\Delta x} \rho(s; \theta_3, \theta_4, Pe), \quad s \in [0, 1],
\]

with

\[
\rho(s; \theta_3, \theta_4, Pe) = x(s; \theta_3, \theta_4, Pe) + iy(s; \theta_3, \theta_4, Pe)
\]

where

\[
x(s; \theta_3, \theta_4, Pe) = \frac{1}{Pe} \left\{ \cos(2\pi s) - 1 \right\} \left[ 2 - \left\{ \cos(2\pi s) - 1 \right\} \left( \frac{1}{3} + 4Pe\theta_4 \right) \right],
\]

\[
y(s; \theta_3, \theta_4, Pe) = -\sin(2\pi s) \left[ 1 - \left\{ \cos(2\pi s) - 1 \right\} \left( \frac{1}{3} - 2\theta_3 \right) \right].
\]

Remark 2 It is worth noting that the spectral curve shape is characterised by function \( \rho(s; \theta_3, \theta_4, Pe) \) while \( \frac{u}{\Delta x} \) is a scaling factor we shall blend with the time step parameter to produce a CFL-like coefficient.
2.3 Centered and upwind schemes

The centered scheme corresponds to $\theta_3 = 0$ and $\theta_4 = 0$ and provides the optimal fourth-order of accuracy. The corresponding scheme for finite and positive $Pe$ reads

$$E_c = -\frac{u}{\Delta x} \left( \frac{1}{12 Pe}, \frac{4 + 2 Pe}{3 Pe}, \frac{5}{2 Pe}, \frac{4 - 2 Pe}{3 Pe}, \frac{1 - Pe}{12 Pe} \right).$$

We present in Table 1 the spectral curves for the centered scheme when $Pe \in [0, +\infty[$ and also for $Pe = 0$ and $Pe = +\infty$.

Table 1: Spectral curves — centered scheme: $\theta_3 = 0$ and $\theta_4 = 0$.

| Pe       | spectrum curve $\lambda$                                                                 | $\rho_c$ |
|----------|------------------------------------------------------------------------------------------|----------|
|          | $\lambda(s; Pe) = \frac{u}{\Delta x}(x_c(s; Pe) + i y_c(s; Pe))$                        |          |
| $]0, +\infty[ $ | $x_c(s; Pe) = \frac{1}{Pe}(\cos(2\pi s) - 1) \left(2 - \frac{1}{3}(\cos(2\pi s) - 1)\right)$ | $0$      |
|          | $y_c(s; Pe) = -\sin(2\pi s) \left(1 - \frac{1}{3}(\cos(2\pi s) - 1)\right)$              |          |
| $0$      | $\lambda(s) = \frac{\kappa}{\Delta x^2}(x_c(s) + i y_c(s))$                            |          |
|          | $x_c(s) = (\cos(2\pi s) - 1) \left(2 - \frac{1}{3}(\cos(2\pi s) - 1)\right)$             |          |
|          | $y_c(s) = 0$                                                                              |          |
| $+\infty$| $\lambda(s) = \frac{u}{\Delta x}(x_c(s) + i y_c(s))$                                    |          |
|          | $x_c(s) = 0$                                                                              |          |
|          | $y_c(s) = -\sin(2\pi s) \left(1 - \frac{1}{3}(\cos(2\pi s) - 1)\right)$                 |          |

We define the weak upwind scheme by cancelling coefficient $a_2$ since we have assumed $u \geq 0$. Hence the following relation has to be satisfied

$$\theta_4 + \frac{\theta_3}{2} = \frac{Pe - 1}{12 Pe}.$$  \hfill (10)

In order to get the optimal accuracy, we set $\theta_3 = 0$ leading to a third-order and deduce by relation (10) that $\theta_4 = \frac{Pe - 1}{12 Pe}$. The scheme for finite and positive $Pe$ then reads

$$E_w = -\frac{u}{\Delta x} \left( \frac{1}{6}, \frac{Pe + 1}{Pe}, \frac{Pe + 4}{2 Pe}, \frac{Pe - 3}{3 Pe}, 0 \right).$$

We present in Table 2 the spectral curves for the weak-upwind scheme for $Pe \in [0, +\infty[$ and also for $Pe = 0$ and $Pe = +\infty$.

The strong upwind scheme consists in cancelling both coefficients $a_1$ and $a_2$ leading to the second-order full upwind three-point scheme with

$$\theta_3 = \frac{3 - Pe}{3 Pe} \quad \text{and} \quad \theta_4 = \frac{3 Pe - 7}{12 Pe}.$$  

The corresponding scheme for finite and positive $Pe$ reads

$$E_s = -\frac{u}{\Delta x} \left( \frac{Pe - 2}{2 Pe}, -\frac{2 Pe - 2}{Pe}, \frac{3 Pe - 2}{2 Pe}, 0, 0 \right).$$

We present in Table 3 the spectral curves for the strong upwind scheme when $Pe \in [0, +\infty[$ and also for $Pe = 0$ and $Pe = +\infty$. 5
Table 2: Spectral curves — weak upwind scheme: $\theta_3 = 0$, $\theta_4 = P_e - \frac{1}{12} P_e$.

| $P_e$ | spectrum curve $\lambda$ | $\rho_w$ |
|-------|---------------------------|--------|
| $\lambda(x; P_e) = \frac{u}{\Delta x}(x_w(x; P_e) + iy_w(x; P_e))$ | ![Diagram](image1) |
| $[0, +\infty[ x_w(x; P_e) = \frac{1}{P_e} \left( \cos(2\pi x) - 1 \right) \left( 2 - \frac{P_e}{3} \left( \cos(2\pi x) - 1 \right) \right)$ | ![Diagram](image2) |
| $y_w(x; P_e) = -\sin(2\pi x) \left( 1 - \frac{1}{3} \left( \cos(2\pi x) - 1 \right) \right)$ | ![Diagram](image3) |

$\lambda(s) = \frac{\kappa}{\Delta x} (x_w(s) + iy_w(s))$

$0$

$x_w(s) = 2 \left( \cos(2\pi s) - 1 \right)$

$y_w(s) = 0$

$+\infty$

$x_w(s) = -\frac{1}{3} \left( \cos(2\pi s) - 1 \right)^2$

$y_w(s) = -\sin(2\pi s) \left( 1 - \frac{1}{3} \left( \cos(2\pi s) - 1 \right) \right)$

Table 3 reports the relevant points of the spectral curves of the centered, weak upwind and strong upwind schemes marked on Tables 1, 2 and 3 respectively.

We compare in Table 4 the spectral curves for the centered, weak upwind, and strong upwind schemes when $P_e = 1, 5, 30$. We observe the similarity between centered and weak upwind space discretisations.

To highlight the interest of considering upwind schemes in order to improve the stability for large Péclet number, we present the extreme situation of a pure steady-state convective problem $\mathcal{C}[\phi] = 0$ with $u = 1$ and $\kappa = 0$ ($P_e = +\infty$) — benchmark 1. The manufactured solution is given by

$$
\phi(x; \delta) = \frac{1}{\pi} \left( 1 - \frac{2}{\pi} \arccos \left( 1 - \delta \sin \left( \frac{\pi}{2} (2x - 1) \right) \right) \right) \left( \arctan \left( \frac{1}{\delta} \sin(\pi x) \right) \right),
$$

where parameter $\delta$ controls the roughness of the function. In the present case, we take $\delta = 0.1$ and carry out the numerical simulation with $I = 25$. We display in Figure 1 the shape of the solution and report oscillations for the centered scheme approximation while the weak and strong upwind schemes eliminate the numerical artefact.

![Figure 1](image4)
Table 3: Spectral curves — strong upwind scheme: $\theta_3 = \frac{3-P_c}{3P_c}$, $\theta_4 = \frac{3P_c-7}{12P_c}$.

| $Pe$ | spectrum curve $\lambda$ | $\rho_3$ |
|------|--------------------------|---------|
|      |                         | [0, 1[  |
|      | $\lambda(s; Pe) = \frac{u}{\Delta x} (x_s(s; Pe) + iy_y(s; Pe))$ |         |
| $\in [0, +\infty[$ | $x_s(s; Pe) = \frac{1}{Pe} (\cos(2\pi s) - 1) \left( 2 - (Pe - 2) \left\{ \cos(2\pi s) - 1 \right\} \right)$ |         |
|      | $y_y(s; Pe) = -\sin(2\pi s) \left( 1 - \frac{Pe - 2}{Pe} \left\{ \cos(2\pi s) - 1 \right\} \right)$ |         |
|      | $\lambda(s) = \frac{\kappa}{\Delta x^2} \rho(s) = \frac{\kappa}{\Delta x^2} (x(s) + iy(s))$ |         |
| $0$ | $x_s(s) = 2 \cos(2\pi s) \left( \cos(2\pi s) - 1 \right)$ |         |
|      | $y_y(s) = -2 \sin(2\pi s) \left( \cos(2\pi s) - 1 \right)$ |         |
| $+\infty$ | $x_s(s) = -(\cos(2\pi s) - 1)^2$ |         |
|      | $y_y(s) = -\sin(2\pi s) [2 - \cos(2\pi s)]$ |         |

3 Time-dependent convection-diffusion equation

We now turn to the time-dependent problem, considering the one-dimensional, 1-periodic in space, convection-diffusion equation. We seek function $\phi = \phi(x, t)$ solution of

$$\phi_t = \mathcal{C}[\phi] + f, \quad \text{in } \Omega \times (0, t_f],$$

where $f = f(x, t)$ is a regular, 1-periodic in space, source term, and $t_f > 0$ is the final time. Initial condition is prescribed with $\phi(x, 0) = \phi^0(x), x \in \Omega$, while the periodic condition reads $\phi(0, t) = \phi(1, t), t > 0$.

Applying the method of lines, we seek for an approximation of the solution of the ordinary differential system of equations

$$\frac{d\Phi}{dt} = A\Phi + F, \quad \text{in } (0, t_f],$$

where vector $\Phi = \Phi(t) = (\phi_i(t))_{i=1}^N$, $\phi_i(t) \approx \phi(x_i, t)$, and $A = A(\theta_3, \theta_4, Pe)$ is the circulant matrix associated to the scheme parameterised with $\theta_3$, $\theta_4$, and $Pe$. On the other hand, $F = F(t) = (f_i(t))_{i=1}^N$, $f_i(t) = f(x_i, t)$, while the initial condition is given by $\phi_i(0) = \phi(x_i, 0)$.

3.1 Time discretisation

We aim at designing a multi-stage Runge-Kutta (RK) method to compute numerical approximations in time that provides the better trade-off between stability, accuracy, and positivity taking into account the cell Péclet number $Pe$ and the parameterisation of the spatial scheme with respect to $\theta_3$ and $\theta_4$.

Let $N$ be a positive integer and $(t^n)_{n=0}^N$ the discrete times. We consider the uniform subdivision $t^n = n\Delta t$, 

Table 4: Relevant points of the spectral curves

| $Pc$ | scheme | $A$ | $B$ | $C$ |
|------|--------|-----|-----|-----|
| $[0, +\infty[$ | centered weak upw | $0, 0$ | $(-2\sqrt{\frac{Pc-1}{2Pc}}, 1.37)$ | $(-\frac{16}{3Pc}, 0)$ |
|        |        |     | $(-0.5 - \frac{2.45}{Pc}, 1.37)$ | $(-\frac{4(3 + Pc)}{3Pc}, 0)$ |
| $[0, 1]$ | strong upw | $0, 0$ | $\left(\frac{4(1 - Pc)}{Pc}, 0\right)$ | $\left(\frac{1}{Pc(Pc-2)}, (Pc-1)\sqrt{\frac{3-2Pc}{Pc(Pc-2)}}\right)$ |
| $[1, +\infty[$ | centered | $0, 0$ | $(0, 0)$ | $(-5.33, 0)$ |
|        | weak upw | $0, 0$ | $(-4.0, 0)$ | $(-5.33, 0)$ |
|        | strong upw |     |     | $(-5.33, 0)$ |

where $\xi = 3Pc^3 - 10Pc + 9$.

Table 5: Spectral curves for different values of $Pc$ for the three pairs $\theta_3, \theta_4$ cases.

| scheme | $\theta_3$ | $\theta_4$ | spectra |
|--------|------------|------------|---------|
| centered (fourth-order) | 0 | 0 | |

weak upwind (third-order) | 0 | $\frac{Pc-1}{12Pc}$ | |

strong upwind (second-order) | $\frac{3-Pc}{4Pc}$ | $\frac{3Pc-7}{12Pc}$ | |
\[ n = 0, \ldots, N, \text{ with the time step } \Delta t = \frac{n}{N}. \] The generic s-stage Runge-Kutta method to solve the initial value ODE system \( \text{[13]} \) is given by
\[
\Phi_i^{n+1} = \Phi_i^n + \Delta t \sum_{j=1}^s b_j R_i^{n,j},
\]
\[
R_i^{n,j} = A\Phi_i^{n,j} + F(t^n), \quad j = 1, \ldots, s,
\]
where \( \Phi_i^n = (\phi_i^n)_{j=1}^s, \) \( \phi_i^n \approx \phi(x_i, t^n) \), and \( \Phi_i^{n,j} = (\phi_i^{n,j})_{j=1}^s, \) \( \phi_i^{n,j} \approx \phi(x_i, t^{n,j}) \), with \( t^{n,j} = t^n + \Delta t c_j \), are the intermediate time sub-steps.

We store the entries \((a_{ij}), (b_j), \) and \((c_j)\) in matrix \( A_{BT} \in \mathbb{R}^{s \times s}\), vectors \( b_{BT} \in \mathbb{R}^s \) and \( c_{BT} \in \mathbb{R}^s \) respectively, presented in a table called Butcher tableau:
\[
\begin{array}{c|cccc}
\text{c}_{BT} & A_{BT} & b_{BT} & c_{BT} \\
\hline
\end{array}
\]
\[
\begin{array}{cccc}
c_1 & a_{11} & \cdots & a_{1s} \\
\vdots & \vdots & & \vdots \\
c_s & a_{s1} & \cdots & a_{ss} \\
\end{array}
\]
\[
\begin{array}{c}
b_1 \\
\vdots \\
b_s \\
\end{array}
\]
Notice that the explicit Runge-Kutta is achieved if \( a_{ij} = 0 \) for \( i \leq j \).

Equation \( \text{[13]} \) leads to the uncoupled linear differential system
\[
\frac{d\phi}{dt} = \lambda(\theta_3, \theta_4, Pe)\tilde{\phi} + \tilde{f}, \quad \tilde{\phi}(0) = \tilde{\phi}_0, \quad i = 1, \ldots, I,
\]
(14)
taking into account the circulant matrix \( A \) eigenvalues given by equation \( \text{[5]} \), where \( \tilde{\phi}_i = \tilde{\phi}_i(t) \) and \( \tilde{f}_i = \tilde{f}_i(t) \) are the projections of \( \phi_i \) and \( f_i \) respectively, on the eigenbasis.

To deal with the stability of the Runge-Kutta scheme, we consider the homogeneous problem deriving from \( \text{[14]} \), by cancelling the source term \( \tilde{f} \). Let \( z_i = \Delta t \lambda_i(\theta_3, \theta_4, Pe) \). The s-stage order \( p \) explicit Runge-Kutta scheme for the homogeneous equation associated with equation \( \text{[14]} \) reads
\[
\tilde{\phi}^{n+1}_i = R_{ps}(z_i; w_{p+1}, \ldots, w_s)\tilde{\phi}^n_i, \quad \text{with} \quad R_{ps}(z; w_{p+1}, \ldots, w_s) = 1 + \sum_{k=1}^p \frac{z^k}{k!} + \sum_{k=p+1}^s w_k z^k, \quad z \in \mathbb{C},
\]
(15)
where \( R_{ps} \) is the polynomial transfer function and \( w_k \in \mathbb{R}, \) \( k = p + 1, \ldots, s, \) stand for the free parameters when \( p < s \). No free parameters are available if \( s = p \) and we just denote \( R_p(z) \equiv R_{pp}(z) \) for the sake of simplicity. We recall that stability scheme is achieved when the complex values \( z \) are such that the absolute value of polynomial \( R_{ps} \) is lower than one and the set
\[
\text{SR}_{ps}(w_{p+1}, \ldots, w_s) = \{ z \in \mathbb{C} : |R_{ps}(z; w_{p+1}, \ldots, w_s)| < 1 \},
\]
(16)
characterises the absolute stability region \( \text{[15]} \). In addition, for \( s = p \), we denote \( \text{SR}_p \equiv \text{SR}_{pp} \).

As an example, we plot in Figure \( \text{[2]} \) the stability regions for three classical schemes: (i) \( \text{SR}_2 \) — order 2 with \( w_3 = w_4 = 0 \), (ii) \( \text{SR}_3 \) — order 3 with \( w_3 = \frac{1}{6}, w_4 = 0 \), and (iii) \( \text{SR}_4 \) — order 4 with \( w_3 = \frac{1}{24}, w_4 = \frac{1}{24} \).

### 3.2 Design of Optimal Absolute stability regions

The goal is to design an efficient explicit Runge-Kutta method allowing step sizes as large as possible, still preserving the linear stability. The key idea of the optimisation of Runge-Kutta methods for temporal integration is to determine the parameters that yield both the largest stability limit and the highest accuracy (expressed in terms of dissipation and dispersion). In \( \text{[19]} \), the authors focus on constructing a stability polynomial which allows the largest absolutely stable step size and corresponding Runge-Kutta method (number of stages and Butcher tableau) for a given problem when the spectrum of the initial value problem is known. They formulated a stability optimisation problem and constructed an algorithm based on convex optimisation techniques. They considered a global convergence in the case that the order of approximation is one and the spectrum encloses a star-like region. Their optimality criterion is the stability of the method. Ait-Haddou in \( \text{[3]} \) used the theory of polar forms of polynomials to obtain sharp bounds on the radius of the largest disc (absolute stability radius), and on the length of the largest possible real interval (parabolic stability radius) to be inscribed in the stability region associated to the stability polynomial of an explicit Runge-Kutta method. Schlegel et. al. \( \text{[22]} \) constructed a multirate time-step integration method for the convection equation. The method decouples
different physical regions so that the time step size constraint becomes a local instead of a global restriction. Moreover, Schlegel introduced a generic recursive multirate Runge-Kutta scheme of third order accuracy. In [18], the author developed an optimization of the explicit two-derivative sixth-order Runge–Kutta method in order to obtain low dissipation and dispersion errors. The method depends on two free parameters, used for the optimisation and the spatial derivatives are discretized by finite differences and Petrov–Galerkin approximations.

In this study, we focus on 4-stage RK schemes ($s = 4$) at least second-order ($p \geq 2$) as a guideline for a more general situation. As indicated by Table 5, the shape of spectral curve $s \rightarrow \lambda(s; \theta_3, \theta_4, Pe)$ highly depends on the cell Péclet number and the parameters values. On the other hand, stability regions are controlled by $w_3$ and $w_4$ parameters and should be adapted in function of the spectrum curve to optimally embedded the curve into $\text{SR}_{ps}(w_3, w_4)$. Such an optimisation problem is almost intractable due to the high non-linearity involved in the construction of the functional to minimise and we observe there exist three major scenarios: (A) the spectral curve is almost vertical, (B) almost horizontal, and (C) an intermediate case (see Table 5). Consequently, we aim to determine two absolute stability regions corresponding to the two extreme scenarios.

### 3.2.1 The imaginary axis

Scenario (A) takes place for low diffusion schemes where the spectral curve is getting closer to the vertical axis as long as the Péclet number increases. We also deal with the fourth-order centered scheme where the spectrum lies “near” the left of the imaginary axis. Consequently, one has to design a RK scheme by seeking real constants $w_3$ and $w_4$ such that the stability region includes the largest segment of the imaginary axis centred at the origin.

The solution is given by the following optimisation problem

$$\max_{w_3, w_4 \in \mathbb{R}} \{ \eta; [-\eta, \eta] \subset \text{SR}_{24}(w_3, w_4) \},$$

but no analytical solution can be exhibited. Nevertheless, in [15], the authors present a solution of the optimal problem when one maximises $\eta$ for the three-parameters functional $\text{SR}_{14}(w_2, w_3, w_4)$ (which contains the particular case $\text{SR}_{24}(w_3, w_4)$, since we have one more free parameter). It is shown that polynomial

$$P_4(z) = 1 + z + \frac{5}{9} z^2 + \frac{4}{27} z^3 + \frac{4}{81} z^4$$

provides the best largest segment of the imaginary axis with $\eta_{\text{max}} = 3$.

We plot in Figure 3 the two stability regions associated to the popular RK4 scheme and the optimal solution proposed in [15]. We observe that the RK4 scheme provides an excellent approximation with $\eta_{\text{max}} = 2\sqrt{2}$, [15]. Of course, the optimal case would provide a slightly bigger $\eta_{\text{max}}$, but we consider that, for our application, the RK4 is an excellent candidate for the first scenario.

### 3.2.2 The real axis

Scenario (B) concerns numerical schemes with large diffusion characterised by low Péclet numbers. Therefore, we seek real constants $w_3$ and $w_4$ such that the stability region includes the largest segment of the negative real axis starting at the origin. An additional difficulty is that the stability region may not be a simply connected
Figure 3: Stability region associated to polynomials $P_4$ and $R_4$.

region as exemplified in Figure 4. Consequently, we only consider the first connected region which contains the origin as the effective stability region. There are also stabilized explicit Runge-Kutta methods as, for example, Runge-Kutta-Chebyshev methods (RKC) dedicated to extended real stability intervals and useful for semi-discrete parabolic problems. A second-order RKC method was initially proposed by van der Houwen and Sommeijer [14] and a family of second- and fourth-order Orthogonal-Runge-Kutta-Chebyshev methods (ROCK) were proposed by Abdulle and Abdulle and Medovikov, [11, 2] but as we said these methods are based on Runge-Kutta and for our purposes, we only deal with original Runge-Kutta.

Figure 4: Stability regions for different values of $w_3$ and $w_4$.

Riha, in [21], proved that the optimal stability polynomial $\bar{R}_{ps}(z)$ of order $p$ and degree $s$

$$\bar{R}_{ps}(z; \bar{w}_1, \ldots, \bar{w}_s) = 1 + \sum_{k=1}^{p} \frac{z^k}{k!} + \sum_{k=p+1}^{s} \bar{w}_k z^k$$

such that

$$|\bar{R}_{ps}(z; \bar{w}_1, \ldots, \bar{w}_s)| \leq 1, \quad z \in [-\zeta, 0],$$

is unique and satisfies the so-called ripple property:

**Property 1 (ripple property)** Polynomial $R_{ps}(z; w_1, \ldots, w_s)$ satisfies the ripple property, if and only if, there exist $s - p + 1$ points $x_0 < x_1 < \ldots < x_{s-p} < 0$, with $x_0 = -\zeta_{\text{max}}$, such that

$$R_{ps}(x_i; w_1, \ldots, w_s) = -R_{ps}(x_{i+1}; w_1, \ldots, w_s), \quad i = 0, 1, \ldots, s - p - 1,$$

$$|R_{ps}(x_i; w_1, \ldots, w_s)| = 1, \quad i = 0, 1, \ldots, s - p.$$

There are no explicit analytic expressions for the optimal $s - p + 1$ coefficients $w_k$, $k = p + 1, \ldots, s$, but the ripple property has been used to construct approximations to the optimal stability polynomial $R_{ps}(z)$ as proposed for example in [15, 23], where the authors show that the optimal bound $\zeta_{\text{max}}$ depends on $s$ and satisfies $\zeta_{\text{max}} = c_p s^2$, $c_p \in \mathbb{R}$, asymptotically with $s \to \infty$.

For $p = 2$, there is a suitable approximate polynomial $B_s$ based on Chebyshev polynomials, given by M.
Bakker in 1971, that generates about 80% of the optimal interval and for \( s = 4 \), the Bakker polynomial reads

\[
B_4(z) = 1 + z + \frac{z^2}{2} + \frac{2z^3}{25} + \frac{z^4}{250},
\]

where \( \zeta_{\text{max}} = 10 \), \( w_3 = \frac{2}{25} \), and \( w_4 = \frac{1}{250} \). We plot in Figure 5 the Baker polynomial representation and the corresponding stability region.

![Figure 5: Baker polynomial \( B_4 = R_{24}(x; \frac{2}{25}, \frac{1}{250}) \) graph (left) and the corresponding stability region (right).](image)

Such a stability region is suitable for \( Pe = 0 \) but not for small values of \( Pe \) since it is not radial at the origin. Consequently, we have considered a small perturbation of parameters \( w_3 \) and \( w_4 \) in order to produce a radial region still preserving a large interval on the real negative axis. We found a good trade-off with \( w_3 = 0.0834 \) and \( w_4 = 0.0042 \) providing \( \zeta_{\text{max}} \approx 11 \). We plot in Figure 6 the polynomial curve and the corresponding region.

![Figure 6: Polynomial \( R_{24}(x; 0.083, 0.0042) \) graph (left) and the corresponding stability region (right).](image)

### 3.2.3 Design of a non-negative scheme

A critical issue in some problems is to guarantee the positivity of the solution. Indeed, temperature, concentration, or density are physical quantities that must be non negative at the discrete level. Figure 6 shows that the polynomial does not satisfy this criterion. For instance, if \( x \in \left[ -10.3997, -6.9570 \right] \), the scheme is stable but the polynomial is negative leading to a sign change.

We then consider a more constrained problem and define the new optimisation problem

\[
\max_{w_3, w_4 \in \mathbb{R}} \left\{ \zeta; 0.01 \leq R_{24}(x; w_3, w_4) \leq 0.7, \forall x \in [-\zeta, 0] \right\},
\]

to create the appropriated region. The lower bound 0.01 and upper bound 0.7 are prescribed in order to provide a radial domain. Indeed, extreme bound values 0.0 and 1.0 provide a stability region similar to Bakker polynomial displayed in Figure 5.

Numerical solution of the optimal problem provides \( w_3 = \frac{603}{6998} \) and \( w_4 = \frac{15}{3212} \) with \( \zeta_{\text{max}} \approx 9.43 \) and we plot in Figure 7 the polynomial curve and the corresponding stability region. We obtain a large radial stability region with \( \zeta_{\text{max}} \) close to the one provided by the Bakker polynomial. On the other hand, the domain still preserves an important part of the horizontal axis in comparison to the non positive optimal case given in Figure 6.

To conclude the analysis, scenario (A) concerns the low diffusive situation and the \( D_4 \) polynomial is well adapted for such situation. On the other hand, scenario (B) deals with high diffusive operator and the polynomial

\[
R_D(z) = R_{24} \left( z; \frac{603}{6998}, \frac{15}{3212} \right)
\]

12
Figure 7: Polynomial $R_D(x)$ graph (left) and the corresponding stability region (right).

provides an excellent stability region.

3.3 Butcher tableau construction

Stability issue provides coefficients $w_3$ and $w_4$ for polynomial $R_D$ and one has to compute corresponding entries of Butcher tableau. We briefly outline the method for the general case $R_{24}(z; w_3, w_4)$ and provide the Butcher’s tableau for the two scenarios, taking into account a crucial restriction: the two tableaux must have the same time sub-steps for the sake of compatibility.

An explicit 4-stage Runge-Kutta method is given by the generic Butcher tableau

\[
\begin{array}{c|cccc}
  c_1 & 0 & 0 & 0 & 0 \\
  c_2 & a_{21} & 0 & 0 & 0 \\
  c_3 & a_{31} & a_{32} & 0 & 0 \\
  c_4 & a_{41} & a_{42} & a_{43} & 0 \\
  & b_1 & b_2 & b_3 & b_4 \\
\end{array}
\]

Consistency yields

\[
c_i = \sum_{j=1}^{i-1} a_{ij}, \quad i = 1, \ldots, 4,
\]

while the second-order assumption provides the constraints

\[
\begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix} b = 1, \quad b^T c = \frac{1}{2},
\]

Additional constraints are considered:

- The Runge-Kutta scheme is fourth-order for the non-homogeneous equation $\phi'(t) = f(t)$,

\[
\begin{bmatrix} (c_1)^2 & (c_2)^2 & (c_3)^2 & (c_4)^2 \end{bmatrix} b = \frac{1}{3}, \quad \begin{bmatrix} (c_1)^3 & (c_2)^3 & (c_3)^3 & (c_4)^3 \end{bmatrix} b = \frac{1}{4},
\]

- The four stage method has to satisfy the consistency condition to suit the polynomial $R_{24}(z; w_3, w_4)$,

\[
b^T A c = w_3, \quad b^T A^2 c = w_4.
\]

Conditions (18), (19), and (20) are equivalent to the system

\[
V^T b^T = \begin{bmatrix} 1 & 1 & 1 & 1 \\ c_1 & c_2 & c_3 & c_4 \\ c_1^2 & c_2^2 & c_3^2 & c_4^2 \\ c_1^3 & c_2^3 & c_3^3 & c_4^3 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} = \begin{bmatrix} 1/2 \\ 1/3 \\ 1/4 \end{bmatrix}
\]

together with equations

\[
b_3 a_{32} c_2 + b_4 (a_{42} c_2 + a_{43} c_3) = w_3
\]

\[
b_4 a_{43} a_{32} c_2 = w_4.
\]

Assuming $c_1 = 0$, the determinant of the Vandermonde matrix $V$ is $(c_4 - c_3)(c_3 - c_2)c_2$ and two situations arise:

- If $c_2 \neq 0$, $c_4 \neq c_3$, and $c_3 \neq c_2$, the system has a unique solution $b$. Hence in this case, we choose the free parameters to be $c_2$, $c_3$, $c_4$, and $a_{43}$. From the Vandermonde system, we obtain $b$. 

13
• If \( c_2 = 0 \) or \( c_4 = c_3 \) or \( c_3 = c_2 \), we have a dependent linear system. For instance, assuming that \( c_2 = c_3 \), system (21) turns to be

\[
\begin{bmatrix}
1 & 1 & 1 & 1 \\
0 & c_2 & c_2 & c_4 \\
0 & 0 & 0 & c_4(c_4 - c_2) \\
0 & 0 & 0 & 0
\end{bmatrix}
\begin{bmatrix}
b_1 \\
b_2 \\
b_3 \\
b_4
\end{bmatrix}
= \begin{bmatrix}
1 \\
1/2 \\
1/3 - c_2/2 \\
1/4 - (c_4 + c_2)/3 + c_4c_2/2
\end{bmatrix}.
\tag{24}
\]

The system has several degrees of freedom we have to fix with relations

\[
\frac{1}{4} - \frac{c_4 + c_2}{3} + \frac{c_4c_2}{2} = 0, \quad c_4(c_4 - c_2) \neq 0, \quad c_2 \neq 0.
\]

We get vector \( b \) with

\[
b_4 = \frac{2 - 3c_2}{6c_4(c_4 - c_2)}, \quad b_3 = \frac{3c_4 - 2}{6c_2(c_4 - c_2)} - b_2, \quad b_1 = \frac{c_4 + c_2 - 1}{6c_2c_4}.
\]

Notice that the user has to choose the free parameters \( c_2, a_{43}, \) and \( b_2 \).

From vectors \( c \) and \( b \), we compute the remaining Butcher tableau elements \( a_{21}, a_{31}, a_{32}, a_{41}, \) and \( a_{42} \) and we obtain

\[
a_{21} = c_2, \quad a_{32} = \frac{w_4}{b_4a_{43}c_2}, \quad a_{42} = \frac{w_4 - b_3a_{43}c_3 - b_4c_2a_{43}}{b_4c_2}, \quad a_{31} = c_3 - a_{32}, \quad a_{41} = c_4 - a_{42} - a_{43}.
\]

At last, we present in Table 6 the two Butcher tableaux with 4-stage corresponding to \( R_4 \) and \( R_D \) polynomials. We tag the associated methods as RK4 and RKD, respectively. It is important to notice that the RKD has been designed with the same time sub-steps according to the classical RK4 scheme.

Table 6: Butcher tableaux for the classical scheme RK4 (left) and RKD (right).

| \( \theta \) | \( \theta \) | \( \theta \) | \( \theta \) | \( \theta \) |
|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 |
| 1/2 | 1/2 | 0 | 0 | 0 |
| 1/2 | 0 | 1/2 | 0 | 0 |
| 1 | 0 | 0 | 1 | 0 |
| 1/6 | 1/3 | 1/3 | 1/6 | 1/6 |

| \( \theta \) | \( \theta \) | \( \theta \) | \( \theta \) | \( \theta \) |
|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 |
| 1/2 | 1/2 | 0 | 0 | 0 |
| 1/2 | 0 | 1/2 | 0 | 0 |
| 1/2 | 334/861 | 373/3328 | 0 | 0 |
| 1 | 481/3310 | 587/1655 | 0 | 0 |

| \( \theta \) | \( \theta \) | \( \theta \) |
|---|---|---|
| 1/6 | 0.4 | 4/15 |
| 1/6 | 1/6 | 1/6 |

4 Optimal time step for stability

We now reach the key point of the paper: the confrontation between the RK stability region with the spatial operator spectrum. More precisely, we have, on the one hand, a complex value parametric curve given by relation (6), \( \theta \in [0,1] \rightarrow \lambda(\theta; \theta_3, \theta_4, Pe) \) that contains all the eigenvalues \( \lambda_i \) of the discrete operator. On the other hand, applying the RK scheme with time step \( \Delta t \) and setting \( z_i = \Delta t \lambda_i(\theta_3, \theta_4, Pe) \), one has to choose \( \Delta t \) small enough such that \( z_i \in \text{SR}_{\rho}(u_{w+1}, \ldots, u_w) \) to guarantee the stability.

4.1 The CFL condition

Let

\[
z(\theta; \theta_3, \theta_4, Pe) = \Delta t \lambda(\theta; \theta_3, \theta_4, Pe) = C_{\text{CFL}} \rho(\theta; \theta_3, \theta_4, Pe),
\tag{25}
\]

where \( C_{\text{CFL}} = \frac{u_\Delta}{\Delta x} \) and \( \rho \) be given by relation (2) for \( u \neq 0 \). Stability condition for the discrete problem is achieved if we satisfy the condition

\[
C_{\text{CFL}} \rho(\theta; \theta_3, \theta_4, Pe) \subset \text{SR}_{\Delta t}(w_3, w_4), \quad \theta \in [0,1].
\tag{26}
\]

Note that the scheme stability depend on the six parameters \( \theta_3, \theta_4, w_3, w_4, Pe \), and \( C_{\text{CFL}} \) and the goal of this section is to analyse the stability of the full time-dependent convection-diffusion equation.

Remark 3 The stability condition combines two main ingredients. On the one hand, the shape of the stability region is characterised by function \( \rho \) that provides a complex value curve where all the eigenvalues lie, independently of the space parameter \( \Delta x \). On the other hand, the CFL value \( C_{\text{CFL}} \) scales the previous curve to fit inside the stability region.
Assuming that scheme in space is given (parameters $\theta_3$, $\theta_4$ are prescribed) and the scheme in time is given (parameters $w_3$, $w_4$ are prescribed), we define the optimal CFL curve as a function of the Péclet number

$$Pe \to \hat{C}_{CFL}(Pe) = \limsup \left\{ C_{CFL} \geq 0, \text{ such that (26) holds} \right\}.$$ 

We compute $\hat{C}_{CFL}(Pe) = \hat{C}_{CFL}(Pe; \theta_3, \theta_4, w_3, w_4)$ with the following algorithm (see Figure 8):

1. Compute eigenvalues $\rho_i$, $i = 1, \ldots, I$, from the space discretisation operator spectrum;
2. Compute intersection points $\tau_i$ of the segment line $O\rho_i$ with the stability region boundary;
3. Compute $\hat{C}_{CFL} = \min_{i=1,\ldots,I} |\tau_i|/|\rho_i|$.

Afterwards the maximum stable time step is given by

$$\Delta t_{\max} = \hat{C}_{CFL} \frac{\Delta x}{u}.$$ (27)

**Remark 4** Two extreme situations require a specific treatment.

- If $Pe = 0$, i.e., $u = 0$, a CFL constant based on the velocity is no longer available. In that case, the spectrum is given in Table 1 for the centered scheme, Table 2 for the weak upwind, Table 3 for the strong upwind and reads

$$\Delta t \lambda(s) = C_{CFL} \rho(s) = \frac{\kappa \Delta t}{\Delta x^2} 2(\cos(2\pi s) - 1) \begin{cases} \left( \frac{\pi}{6} - \frac{i}{6} \cos(2\pi s) \right) & \text{centered,} \\ 1 & \text{weak,} \\ (\cos(2\pi s) - i \sin(2\pi s)) & \text{strong.} \end{cases}$$

In this case the maximum stable time step is given by

$$\Delta t_{\max} = \frac{\hat{C}_{CFL} \Delta x^2}{\kappa}.$$ (27)

- If $Pe = +\infty$, i.e., $\kappa = 0$, one has to pass to the limit to determine the spectrum curve, remaining (27) valid.

As an example, we present in Figure 9 two situations with $Pe = 5$ and $Pe = 10$ where we adjust the CFL constant to fit the spectrum into the stability region: centered scheme in space with the RK$_4$ scheme in time (top left) and the RK$_D$ (down left) while the pictures in second column present the weak upwind case.

### 4.1.1 Analysis of spatial schemes

A full discretised scheme consists in fixing the time and spatial scheme parameters. We have identified two schemes in time — RK$_D$ and RK$_4$ — for low and high Péclet situations and three schemes in space — centered, weak upwind, and strong upwind. We shall discard the strong upwind scheme due to the high dispersion effect. Indeed, let us consider the advection problem $\mathcal{E}[\phi] = 0$ with $u = 1$ and $\kappa = 0$ ($Pe = +\infty$). Function

$$\phi(x, t; \omega) = \sin(2\pi \omega (x - ut))$$
with \( \omega = 3 \), is a periodic solution. We carry out the numerical simulation with the RK4 scheme in time and the three schemes in space with \( \Delta t = \Delta t_{\text{max}} \) (see expression (27)) until the final time \( t_f = 1 \). We plot the numerical solutions in Figure 10 computed with \( I = 25 \) and \( I = 50 \) nodes — benchmark 2. We observe the large phase errors produced by the strong upwind scheme which justify the choice to discard it.

We turn to a situation where a rough function is convected \((u = 1, \kappa = 0, Pe = +\infty)\) and compare the centered and weak upwind scheme. The manufactured solution is given by

\[
\phi(x, t; \delta) = \frac{1}{\pi} \left( 1 - \frac{2}{\pi} \arccos((1 - \delta) \sin(\pi(x - t - \frac{1}{2}))) \right) \left( \arctan\left( \frac{1}{\delta} \sin(\pi(x - t)) \right) \right)
\]

with \( \delta = 0.01 \) that corresponds to a rough solution due to the sharp transition. The numerical solutions are computed with the RK4 scheme in time taking \( \Delta t = \Delta t_{\text{max}} \), \( I = 25 \), and the centered and weak upwind schemes in space — benchmark 3.

In Figure 11 we present the numerical solutions after two time steps. We observe the centered scheme is more unstable with a larger number of over- and undershoots. Of course, a definitive method would consist in employing a high viscous scheme (the simple two-points upwind one) but with a dramatic cut of the accuracy. Hence, the weak upwind scheme is regarded as an alternative to the centered one when large oscillations appear.

### 4.1.2 Optimal CFL curves

The curves \( Pe \rightarrow \hat{C}_{\text{CFL}}(Pe; \theta_3, \theta_4, w_3, w_4) \) are determined for the four situations we want to highlight: RK4 and RKD for time; centered and weak upwind for space. We plot in Figure 12 the optimal CFL curves for the fourth-order centered scheme and the third order weak upwind scheme with \( Pe \in [0.001, 20] \) (note that the oscillations we observe for the centered case are a consequence of the discrete consideration of the spectral curve). We complement the figure with Table 7 indicating the optimal CFL values, \( \hat{C}_{\text{CFL}} \), for \( Pe = 0 \) (an expression depending on \( \Delta x \)) and large Péclet values.

From Figure 12 and Table 7 we draw the following conclusions:
| $I$ | centered | weak upwind | strong upwind |
|-----|-----------|-------------|---------------|
| 25  | ![centered weak upwind strong upwind](image1) | ![centered weak upwind strong upwind](image2) | ![centered weak upwind strong upwind](image3) |
| 50  | ![centered weak upwind strong upwind](image4) | ![centered weak upwind strong upwind](image5) | ![centered weak upwind strong upwind](image6) |

Figure 10: Exact (---) and approximate (→) solutions considering RK$_4$ time scheme for benchmark 2.

Figure 11: Exact (---) and approximate (→) solutions for $t_f = 2\Delta t_{max}$ considering RK$_4$ time scheme for benchmark 3: centered (left) and weak upwind (right).

Figure 12: $Pe \to \hat{C}_{CFL}$ curves for the four scenarios that we are considering: centered (left) and weak upwind (right), with solid — RK$_4$, dashed — RK$_D$.

- **centered scheme** RK$_D$ provides the largest time-steps for $Pe < 5$ while RK$_4$ is more efficient for $Pe > 5$. In particular, the limit for $Pe \to +\infty$ of $\hat{C}_{CFL}$ is 2.55 for RK$_4$ while $\hat{C}_{CFL}$ converges to 0 with RK$_D$.

- **weak upwind scheme** For $Pe < 15$, $\hat{C}_{CFL}$ value is larger with RK$_D$ scheme than RK$_4$. The particular
Table 7: Comparison of $\tilde{C}_{CFL}$ for RK$_D$ and RK$_4$.

| $\text{Pe}$ | RK$_D$ | RK$_4$ |
|-----------|--------|--------|
|           | centered weak upwind | centered weak upwind |
| 0         | $1.77\Delta x^2$ | $2.36\Delta x^2$ | $0.53\Delta x^2$ | $0.70\Delta x^2$ |
| 20        | 1.04    | 1.60   | 1.62   | 1.66   |
| 200       | 0.45    | 1.34   | 2.04   | 1.74   |
| 20000     | 0.09    | 1.25   | 2.06   | 1.75   |
| 200000    | 0.04    | 1.24   | 2.06   | 1.75   |
| $+\infty$ | 0       | 1.24   | 2.55   | 1.75   |

case $Pe = 0$ shows that RK$_D$ case allows a time parameter about three times larger than the RK$_4$ case. For $Pe > 15$, RK$_4$ method turns out to be more efficient.

4.1.3 Convergence order and stability

To compare the convergence error between the centered and the weak upwind discretisations, we consider the advection ($u = 1$, $\kappa = 0$, $Pe = +\infty$) of the manufactured solution \((28)\) with $\delta = 0.1$, corresponding to a smooth solution since the transition takes place in more than 5 nodes.

— benchmark 4. We carry out the RK$_4$ scheme in time with a time step $\Delta t = 0.8\Delta t_{\text{max}}$ until $t_f = 1$. We present in Table 8 the time step, error $E_\infty$, and the respective convergence order $O_\infty$ between two solutions/grids $(\Phi_k^N, I_k)$, for $k = 1, 2$ where $I_1 < I_2$ as

$$O_\infty \equiv O_\infty((\Phi_1^N, I_1), (\Phi_2^N, I_2)) = \frac{\log E_\infty(\Phi_1^N, I_1)/E_\infty(\Phi_2^N, I_2)}{\log I_1/I_2},$$

while we display in Figure 13 the numerical approximations for the centered and weak upwind schemes. We reach the fourth-order in space for the centered method and the expected third-order in space for the upwind case (the fourth-order method in time turns to be insignificant for large values of $I$, since small values of $\Delta t$ are needed due to the CFL condition).

Table 8: Time steps and errors using 80% of the maximal time step given by $\tilde{C}_{CFL}$ considering RK$_4$ time scheme for benchmark 4.

| $I$     | centered | weak upwind |
|---------|-----------|-------------|
| $\Delta t$ | $E_\infty$ | $O_\infty$ | $\Delta t$ | $E_\infty$ | $O_\infty$ |
| 100     | $1.65E-2$ | $1.68E-2$ | 1.40E-2 | 2.20E-2 |
| 200     | $8.25E-3$ | $3.88E-3$ | 2.1 | 6.98E-3 | 6.29E-3 | 1.8 |
| 400     | $4.12E-3$ | $4.88E-4$ | 3.0 | 3.49E-3 | 1.20E-3 | 2.4 |
| 800     | $2.06E-3$ | $3.83E-5$ | 3.7 | 1.75E-3 | 1.70E-4 | 2.8 |
| 1600    | $1.03E-3$ | $2.46E-6$ | 4.4 | 8.73E-4 | 2.15E-5 | 3.0 |

To check the stability condition, we have performed the computation with the RK$_4$ time scheme with three different time steps: (1) the time step $\Delta t_{\text{max}}$ corresponding to the optimal CFL value given by $\tilde{C}_{CFL}(Pe)$, (2) a 20% smaller time step, and (3) a 10% larger time step. We present in Table 9 the errors $E_\infty$ together with the number of iterations $n_{TS}$ needed to reach the final time $t_f = 1$. For the latter case, stability is no longer preserved and the error blows up. With the critical time step $\Delta t_{\text{max}}$, we manage to compute the solution until the final time but with an error slightly larger than the one obtained with the smaller $\Delta t$. Notice that, as expected, the number of steps linearly increases with the number of nodes.

4.2 Hybrid scheme

Hybrid time scheme consists in choosing between the RK$_4$ and the RK$_D$ scheme in function of the Péclet number to provide the largest time step. For space (and time) dependent velocity and diffusion coefficient, the scheme
then would be different from one node to another. For example, for the centered scheme in space, the left panel of Figure 12 shows that the highest $\hat{C}_{\text{CFL}}(Pe)$ values are reached with the RK$_D$ scheme when $Pe < 5$ while the $\hat{C}_{\text{CFL}}(Pe)$ is larger with RK$_4$ when $Pe > 5$. We implement an hybrid scheme that switches from one method to the other in function of the velocity and diffusion coefficients to optimise the time step and reduce the computational effort.

We consider the space-dependent parameter equation $\partial_t \phi + \mathcal{E}(x)[\phi] = f(x)$ with

$$
\mathcal{E}(x)[\phi] = -u(x)\phi' + \kappa(x)\phi''
$$

and $f$ a given source term. Since the cell Péclet number depends on the position, we set $Pe_i = \frac{u_i\Delta x}{\kappa_i}$ for node
for a small number of nodes, we have the property

\[ \hat{x} \]

At the numerical level, smooth solutions are functions where the numerical derivative is bounded for \( \Delta x \) small enough and transitions between successive extremes are spread on, at least, five or six nodes. In that case, even with a large Péclet number, the centered scheme is stable and the time step is ruled by the hybrid scheme condition. On the other hand, to handle sharp gradients or even discontinuities, the weak upwind scheme turns out to be the candidate method substituting the centered scheme, possibly leading to a change of the scheme in time. At time \( t^n \) and for each node \( i \), the scheme in space has to be chosen with respect to the local regularity for that particular node. Then the scheme in time is chosen with respect to the scheme in space, the local cell

\[
CFL_{\text{w}} = \max \left( CFL_{\text{c}}, CFL_{\text{u}} \right)
\]

that confirms we have taken the best scheme in time for each node providing the larger global time step \( \Delta t \).

### 5 The a posteriori method for optimal time step scheme

At the numerical level, smooth solutions are functions where the numerical derivative is bounded for \( \Delta x \) small enough and transitions between successive extremes are spread on, at least, five or six nodes. In that case, even with a large Péclet number, the centered scheme is stable and the time step is ruled by the hybrid scheme condition. On the other hand, to handle sharp gradients or even discontinuities, the weak upwind scheme turns out to be the candidate method substituting the centered scheme, possibly leading to a change of the scheme in time. At time \( t^n \) and for each node \( i \), the scheme in space has to be chosen with respect to the local regularity for that particular node. Then the scheme in time is chosen with respect to the scheme in space, the local cell

\[
CFL_{\text{w}} = \max \left( CFL_{\text{c}}, CFL_{\text{u}} \right)
\]

that confirms we have taken the best scheme in time for each node providing the larger global time step \( \Delta t \).
Figure 14: $\hat{C}_{\text{CFL}}$ plots for benchmark 5: centered (left) and weak upwind (right). $\hat{C}_{\text{CFL,hybrid}}$ values ($\bigtriangleup$); $\hat{C}_{\text{CFL,RK}_4}$ curve (−−), and $\hat{C}_{\text{CFL,RK}_D}$ curve (−−).

Péclet, and $\hat{C}_{\text{CFL}}$ to provide the optimal local time step $\Delta t^n$. Then, using the time parameter $\Delta t^n$ given by relation (30), we update the solution at time $t^{n+1}$.

In order to make the optimal choice for the space and time schemes (node by node), we use the a posteriori paradigm also mentioned as MOOD method for Multi-dimensional Optimal Order Detector [9, 10]. In this context, the strategy is based on the choice between a high accurate scheme in space (the centered scheme) and a high stable scheme (the weak upwind scheme) together with the associated optimal time scheme we studied in the hybrid scheme section.

5.1 Basics on a posteriori strategy

We present a brief description of the a posteriori paradigm and introduce the notations we need in this section. Assume that the numerical solution $\Phi^n$ is known at time $t^n$.

1. We compute a candidate solution $\Phi^*$ for time $t^{n+1}$ using the most accurate schemes in space and time, namely the centered scheme and the RK$_4$.

2. We check, node by node, the admissibility of the solution using detectors, i.e. small routines that analyse specific aspects of the approximations such as extrema, oscillations, and physical property violations if any.

3. The nodes detected as non-admissible are computed again but with the weak upwind and time schemes using the hybrid scheme procedure.

Remark 5 The method is tagged a posteriori since we analyse a candidate solution $\Phi^*$ after computing a time step. On the contrary, MUSCL and WENO method are said a priori since we perform the limitation strategy based on solution $\Phi^n$, [19, 20].

5.1.1 Node space and time scheme tables

In practice, we introduce two tables CSS and CTS for Cell Space Scheme and Cell Time Scheme, respectively, with the following rules.

- We set $\text{CSS}[i] = 0$ if we use the centered scheme otherwise $\text{CSS}[i] = 1$ for the weak upwind scheme.
- We set $\text{CTS}[i] = 0$ if we use the RK$_4$ scheme otherwise $\text{CTS}[i] = 1$ for the RK$_D$ scheme.
Given the numerical solution \( \Phi^{*,0} = \Phi^n \) and tables CSS and CTS, we rewrite the RK scheme taking the two tables into account. For each node \( i \) the terms \( \Phi_i^{*,j} \) and \( \mathcal{R}_i^{*,j} \) are computed with

\[
\Phi_i^{*,j} = \Phi_i^{*,0} + \Delta t \sum_{s=1}^{s} a_j(\text{CTS}[i]) \mathcal{R}_i^{*,s}, \\
\Phi_i^{*} = \Phi_i^{*,0} + \Delta t \sum_{j=1}^{s} b_j(\text{CTS}[i]) \mathcal{R}_i^{*,j}, \\
\mathcal{R}_i^{*,j} = E_j(\text{CSS}[i]) \Phi^{*,j} + F(x_i, t^{*,j}), \quad j = 1, \ldots, s.
\]

Expressions \( a_j(\text{CTS}[i]) \) and \( b_j(\text{CTS}[i]) \) indicate that we use RK4 scheme if \( \text{CTS}[i] = 0 \) or RKD scheme if \( \text{CTS}[i] = 1 \). On the other hand, the term \( E_j(\text{CSS}[i]) \) states that we use the centered scheme for node \( i \) if \( \text{CSS}[i] = 0 \) or the weak upwind scheme if \( \text{CSS}[i] = 1 \).

5.1.2 Detectors

Detectors are small routines to check a specific property of the candidate solution. We assemble the detectors in a chain of operations which basically indicate if a node would be cured (change the scheme in space) or not. We list hereafter the detectors we use in the present document and refer to [11] for a detailed presentation of the most useful detectors.

- **ED.** The Extrema Detector intends to localise the extrema of the numerical function by checking

\[ s_i = \text{sign} \left( (\phi_i^* - \phi_{i+1}^*)(\phi_i^* - \phi_{i-1}^*) \right). \]

If \( s_i > 0 \) we have an extremum and the detector is activated and returns \( \text{true} \) otherwise it corresponds to a monotone situation and the detector returns \( \text{false} \).

- **SCD.** The Small Curvature Detector helps to select the very small oscillations that we consider innocuous from the stability point of view. We calculate the variation quantity

\[ v_i = \max \left( \frac{|\phi_i^* - \phi_{i+1}^*|}{\Delta x}, \frac{|\phi_i^* - \phi_{i-1}^*|}{\Delta x} \right) \]

and, for a user parameter \( \theta > 0 \), the detector returns \( \text{true} \) if \( v_i < \theta \Delta x \) and \( \text{false} \) otherwise since a small value of \( v_i \) indicates oscillations with too low magnitude to be considered as an issue.

- **LOD.** Local Oscillation detection aims to detect variations deriving from oscillations. Indeed, a local oscillation is characterised by a variation of the curvature sign. To this end, we compute the second derivative

\[ \chi_i = \frac{\phi_{i+1}^* - 2\phi_i^* + \phi_{i-1}^*}{\Delta x}. \]

Then the detector is deactivated (return \( \text{false} \)) if \( \chi_{i-1}, \chi_i, \chi_{i+1} \) have the same sign and is activated (return \( \text{true} \)) if one of the curvatures has a different sign of the two others.

- **SD.** Smooth Detector consists in assessing the local numerical smoothness to determine if an extremum is physical or if it corresponds to a numerical artefact. To evaluate the smoothness of the solution we use, once again, the curvature and define the minimum and maximum absolute curvature

\[ \chi_{i,m} = \min(\chi_{i-1}, \chi_i, \chi_{i+1}), \quad \chi_{i,M} = \max(\chi_{i-1}, \chi_i, \chi_{i+1}). \]

For a given user parameter \( \theta \in [0, 1] \), the solution is not smooth (detector activated \( \text{true} \)) if \( \chi_{i,m} < \theta \chi_{i,M} \) since we detect large variation of curvature on the three consecutive points. Otherwise, the detector returns \( \text{false} \) that indicates the solution is considered smooth enough at the numerical level.

5.1.3 Detector chain and a posteriori cure

The detectors being defined, we assemble it into an ordered chain that enables to decide whether a node would be corrected or not. The detector chain is carried out for nodes with \( \text{CSS}[i]=0 \), that is the centered scheme. We preserve the high accuracy scheme if there is no extremum (ED \( \text{false} \)) or generated by too small variations (SCD \( \text{false} \)). When a potential oscillation is detected (ED \( \text{true} \)), we try to relax the scheme (and preserve the accuracy) with other detectors that assess if the extremum is a real physical one. The LOD is activated if we observe the sign change of the curvature and then we set \( \text{CSS}[i] = 1 \) to indicate that the node has to be cured. In the same way, a node with too large variations of the curvature is considered as a problematic node and we set \( \text{CSS}[i] = 1 \) if SD is \( \text{true} \). At the end of the chain, we have a new CSS map that indicates the points we
have to compute again with the weak upwind scheme. The time scheme is determined with the hybrid scheme strategy and we flag the $\text{CTS}[i]$ accordingly.

From the $\text{CSS}$ map and the cell Péclet number, we modified the $\text{CTS}$ map following the rule given by the hybrid scheme. At the end of the day, we get a new $\Delta t^n_i$ for each node $i$ and use the minimum time step for $\Delta t^n$ following (30).

**Remark 6** Computational resources are reduced by only computing again the node $i$ that have been cured together with the neighbour nodes that may be affected by the values of $\Phi^i_\star$ during the four-stage Runge-Kutta procedure. In practice, very few nodes are modified by the a posteriori correction (less than 5%, see [10]) and the additional cost is of the same order.

### 5.2 Numerical tests

We present two numerical tests to examine the efficiency of the a posteriori method. The first sanity check consists in carrying out a simulation with a regular solution. Indeed, for smooth approximations, the limiter strategy has to preserve the higher accuracy and the chain detector has to return $\text{CSS}[i] = 0$ for all nodes $i$.

— benchmark 6. We consider the manufactured solution (11) with $\delta = 0.15$. We take constant physical parameters $u = 1$, $\kappa = 2.7778\times10^{-03}$ and a grid of $I = 60$ nodes to obtain the cell Péclet number $Pe = 6$. The simulation is carried out until the final time $t_f = 0.5$, corresponding to half a revolution.

![Figure 16: Exact ( ) and approximate ( ) solutions for $t_f = 0.5$ the full RK4+centered scheme (left) and the a posteriori strategy (right): the smooth solution case — benchmark 6.](image)

We display in Figure 16 the numerical approximation with the most accurate scheme (RK4+centered) while we reproduce on the right side the approximation computed with the a posteriori strategy. We have checked that the table $\text{CSS}$ has never been altered during the simulation, that is, computation have been achieved with the centered scheme and the RK4 scheme in time due to the Péclet number.

— benchmark 7. The last benchmark deals with a rough function (regarded to the characteristic mesh size) using the manufactured solution (11) but with $\delta = 0.015$ corresponding to a steep variation we assimilate as a shock regarded to the small number of nodes $I = 60$. We take constant physical parameters $u = 1$ and $\kappa = 5.5556\times10^{-03}$ to obtain the cell Péclet number $Pe = 3$. The simulation is again carried out until the final time $t_f = 0.5$, corresponding to half a revolution.

![Figure 17: the solution obtained with the “unlimited” RK4+centered scheme (left) and the a posteriori method (right).](image)

We display in Figure 17 the solution obtained with the “unlimited” RK4+centered scheme (left) and the a posteriori method (right). Clearly, the steep gradient provokes oscillations when employing the low diffusive centered scheme while the introduction of the weak upwind scheme in some nodes (indicated with the red X on the figure) manages to stabilise the solution and strongly reduces the over- and under-shooting. Moreover the scheme in time on the node $i$ such that $\text{CSS}[i] = 1$ (weak upwind) switch to the RK4 once $\text{CTS}[i] = 1$ while the nodes where we maintain the original centered scheme $\text{CSS}[i] = 0$ (centered upwind) use the RK4 once
CTS[i] = 0. Notice that the number of nodes that have been cured is almost 3 or 4, i.e., less than 8% for a 60-nodes grid.

6 Conclusions and further work

We have developed a strategy to analyse and optimise the stability based, on the one hand, on the two-parameter family of continuous spectral curves that characterise the space discretization and, on the other hand, a two-parameter family of Runge-Kutta stability regions. Optimisation results from the inclusion of the spectral curves into the stability region with the help of the additional CFL parameter. We have detailed the procedure with the five-point finite difference method context but extension to other methods such as finite volume or finite elements methods could be considered. A hybrid time scheme as a function of the Péclet number have been proposed and analysed with the objective of providing the largest time step while preserving the stability. At last, we have presented an adaptation of the a posteriori strategy to handle the schemes in space and time to preserve both the accuracy and the stability, even for rough solutions, while we optimise the time step to reduce the computational effort.

Acknowledgements

G.J. Machado and S. Clain acknowledge the financial support by FEDER – Fundo Europeu de Desenvolvimento Regional, through COMPETE 2020 – Programa Operacional Fatores de Competitividade, and the National Funds through FCT – Fundação para a Ciência e a Tecnologia, project no. UID/FIS/04650/2019.

M.T. Malheiro acknowledge the financial support by Portuguese Funds through FCT (Fundação para a Ciência e a Tecnologia) within the Projects UIDB/00013/2020 and UIDP/00013/2020 of CMAT-UM.

M.T. Malheiro, G.J. Machado, and S. Clain acknowledge the financial support by FEDER – Fundo Europeu de Desenvolvimento Regional, through COMPETE 2020 – Programa Operacional Fatores de Competitividade, and the National Funds through FCT – Fundação para a Ciência e a Tecnologia, project no. POCI-01-0145-FEDER-028118.

References

[1] A. Abdulle, Fourth order Chebyshev methods with recurrence relation, SIAM J. Sci. Comput. 23 (2002) 2042–2055.
[2] A. Abdulle, A.A. Medovikov, Second order Chebyshev methods based on orthogonal polynomials, Numer. Math. 90 (2001) 1–18.
[3] R. Ait-Haddou, New stability results for explicit Runge-Kutta methods, BIT Numer. Math. (2019) 585–612.
[4] R. Araya, E. Behrens, R. Rodriguez, An adaptive stabilized finite element scheme for the advection-reaction-diffusion equation, Appl. Numer. Math. 54 (2005) 491–503.
[5] A. Bourchtein, L. Bourchtein, Explicit finite difference schemes with extended stability for advection equations, J. Comput. Appl. Math. 236 (2012) 3591–3604.
[6] J. Butcher, Numerical methods for Ordinary Differential Equations, Wiley, second ed., 2008.
[7] N. M. Chadha, N. Madden, An optimal time-stepping algorithm for unsteady advection-diffusion problems, J. Comput. Appl. Math. 294 (2016) 57–77.
[8] N. M. Chadha, N. Madden, A two-weight scheme for a time-dependent advection–diffusion problem, Lect. Notes Comput. Sci. Eng. 81 (2011) 99–108.

[9] S. Clain, S. Diot, R. Loubère, A high-order finite volume method for hyperbolic systems: multi-dimensional optimal order detection (MOOD), J Comput Phys 230(10) (2011) 4028–50.

[10] S. Diot, S. Clain, R. Loubère, Improved detection criteria for the multi-dimensional optimal order detection (MOOD) on unstructured meshes with very high-order polynomials, Comput Fluids 64 (2012) 43–63.

[11] S. Clain, R. Loubère, G. Machado, a posteriori stabilized sixth-order finite volume scheme for one-dimensional steady-state hyperbolic equations, Adv. Comput. Math. 44 (2018) 571-607.

[12] V. G. Ferreira, F. A. Kurokawa, R. A. B. Queiroz, M. K. Kaibara, C. M. Oishi, J. A. Cuminato, A. Castelo, M. F. Tomé, S. McKee, Assessment of a high-order finite difference upwind scheme for the simulation of convection–diffusion problems, Int. J. Numer. Meth. Fluids 60 (2009) 1–26.

[13] F. R. Gantmacher, The Theory of Matrices, Chelsea Publishing Co., NY 1960.

[14] P. J. van der Houwen, B. P. Sommeijer, On the internal stability of explicit, m-stage Runge-Kutta methods for large m-values, Z. Angew. Math. Mech. 60 (1980) 479–485.

[15] W. Hundsdorfer, J. Verwer, Numerical Solution of Time-Dependent Advection-Diffusion-Reaction Equations, Springer–Verlag Berlin Heidelberg, 2003.

[16] D. I. Ketcheson, A. J. Ahmadia, Optimal Stability Polynomials for Numerical Integration of Initial Value Problems, Comm. App. Math. and Comp. Sci. 7 (10) (2012) 247–271.

[17] T. Knopp, G. Lube, G. Rapin, Stabilized finite element methods with shock capturing for advection–diffusion problems, Comput. Methods Appl. Mech. Engrg. 191 (2002) 2997–3013.

[18] G. V. Krivovichev, Optimized low-dispersion and low-dissipation two-derivative Runge-Kutta method for wave equations, J Appl Math Comput, 63 (2020) 787–811.

[19] B. van Leer, Towards the Ultimate Conservative Difference Scheme, V. A Second Order Sequel to Godunov’s Method, J. Com. Phys., 32 (1979) 101–136.

[20] X.-D. Liu, S. Osher, T. Chan, Weighted Essentially Non-oscillatory Schemes, J. Comput. Phys., 115 (1994) 200–212.

[21] W. Riha, Optimal stability polynomials, Computing 9 (1972) 37–43.

[22] M. Schlegel, O. Knoth, M. Arnold, R. Wolke, Multirate Runge-Kutta schemes for advection equations, J. Comput. Appl. Math. 226 (2009) 345–357.

[23] L. M. Skorvtsov, Explicit stabilized Runge-Kutta methods, Comput. Math. Math. Phys., 51 (2011) 7 1153–1166.

[24] L. N. Trefethen, Finite Difference and Spectral Methods for Ordinary and Partial Differential Equations, Unpublished Text, (1996) http://people.maths.ox.ac.uk/trefethen/pdetext.html.

[25] R. Verfürth, A Review of a Posteriori Error Estimation and Adaptive Mesh-Refinement Techniques, Wiley–Teubner, Stuttgart, 1996.

[26] C.-W. Shu, Total-variation diminishing time discretizations, SIAM J. Sci. Statist. Comput. 9 (1988), 1073–1084.

[27] C.-W. Shu, S. Osher, Efficient implementation of essentially non-oscillatory shock-capturing schemes, J. Comput. Phys. 77 (1988) 439–471.

[28] S. Wang, An a posteriori error estimate for finite element approximations of a singularly perturbed advection–diffusion problem, J. Comput. Appl. Math. 87 (1997) 227–242.

[29] H. C. Yee, N. D. Sandham, M. J. Djomehri, Low dissipative high order shock-capturing methods using characteristic-based filters, J. Comput. Phys. 150 (1999) 199–238.