A Simple Second-Order Implicit-Explicit Asymptotic Preserving Scheme for the Hyperbolic Heat Equations

Louis Reboul, Teddy Pichard, Marc Massot

A Simple Second-Order Implicit-Explicit Asymptotic Preserving Scheme for the Hyperbolic Heat Equations

Louis Reboul,[a] Teddy Pichard,[b] and Marc Massot[c]

CMAP, CNRS, École polytechnique, Institut Polytechnique de Paris, Route de Saclay, 91128 Palaiseau Cedex, France

[a] *Corresponding author: louis.reboul@polytechnique.edu*
[b] *Electronic mail: teddy.pichard@polytechnique.edu*
[c] *Electronic mail: marc.massot@polytechnique.edu*

Abstract. We propose an asymptotic preserving (AP) Implicit-Explicit (ImEx) scheme for the hyperbolic heat equation in the diffusive regime. This scheme is second order in time and space and l^∞-stabile under relatively low constraints on the time step, and without requiring the use slope limiters. The construction exploits a formalism developed in a previous work and, compared to it, leads to a simpler implementation but it loses uniform accuracy on paper. Stability and accuracy are verified on numerical examples, and even uniform accuracy is obtained on those. Finally, we discuss extensions of this method to the non linear case of isothermal Euler equations with friction.

INTRODUCTION

We devise and study the properties of a numerical method for the hyperbolic heat equations (HHE) [1, 2, 3, 4]

$$\partial_t E + \frac{1}{\varepsilon}\partial_x F = 0, \tag{1a}$$

$$\partial_t F + \frac{1}{\varepsilon}\partial_x E = -\frac{\sigma}{\varepsilon^2}F. \tag{1b}$$

The parameter ε embodies the long-term behavior and stiff collisional source term. This set of equations degenerates in the regime $\varepsilon \to 0$: rewriting (1b) under a relaxation formalism

$$\partial_t F = -\frac{1}{\varepsilon}\left(\partial_x E + \frac{\sigma}{\varepsilon}F\right), \tag{2}$$

one formally obtains when $\varepsilon \to 0$ the equivalence $\frac{1}{\varepsilon}F \sim -\frac{1}{\sigma}\partial_x E$ and consequently (1a) turns into

$$\partial_t E - \partial_x\left(\frac{1}{\sigma}\partial_x E\right) = 0. \tag{3}$$

Classic numerical approaches, such as those fully explicit or splitting techniques coupled with HLL, Roe or Rusanov approximate Riemann solvers, are known both to require very restrictive time steps and for being poorly accurate in the diffusive regimes (see for instance [5, 6, 7, 8]). Higher order methods, such as Strang splitting coupled with a MUSCL-Hancock scheme, do not cirvumvent this issue. Asymptotic preserving methods are designed to preserve uniform stability and accuracy properties in all regimes. This has been abundantly documented in the literature [7, 9, 10, 11, 12, 13, 14, 15, 16].

In a previous work [17], we developed a method based on coupled time-space approach that is uniformly accurate and stable in all regimes for more general, including nonlinear, sets of equations. In the present work, we present specifically for the HHE a much simpler version of this scheme in the sense that it is easier to design, to implement and to generalize. The cost of this simplification is, at the theoretical level, a loss of accuracy in intermediate regimes, which is not observed in our numerical experiments.

The present approach aims at extending the construction of [18] (see also [17]) in a simplified framework to analyze it. As in [18], we aim at extending this approach to Euler-Poisson system to simulate plasma in hall thrusters including sheaths, and the reader is referred to [19] for more details on this application.

In the next section, we detail the steps and the formalism that allow to derive the method. We analyze its numerical properties in the following section. Its theoretical performances are verified by numerical simulations in the third section. Finally, we discuss possible extensions of the scheme in the final section.

CONSTRUCTION OF THE METHOD

We consider the domain $\Omega = [0,1]$ and uniform mesh of $N \geq 0$ cells C_j of width $\Delta x = 1/N$, that is $C_j = [x_{j-1/2}, x_{j+1/2}]$ with $x_{j+1/2} = j\Delta x$. We aim to approximate the cell averages $w_j^n \approx (1/\Delta x)\int_{C_j} w(t^n,x)\,\mathrm{d}x$ of quantities w at the discrete times $t^n = n\Delta t$, $n \geq 0$, $\Delta t > 0$. Since only second order applications are considered in this work, we will use indiscriminately finite-difference and finite-volume formalisms as, under sufficient regularity, we have $w(t^n,x_j) = (1/\Delta x)\int_{C_j} w(t^n,x)\,\mathrm{d}x + \mathscr{O}\left(\Delta x^2\right)$. We use the framework detailed in [17] to design our method:

$$\frac{E_j^{n+1}-E_j^n}{\Delta t} + \frac{1}{\varepsilon}\left[\partial_x F\right]_j^{n+1/2} = 0, \tag{4a}$$

$$\frac{F_j^{n+1}-F_j^n}{\Delta t} + \frac{1}{\varepsilon}\left[\partial_x E\right]_j^{n+1/2} = -\frac{\sigma}{\varepsilon^2}F_j^{n+1/2}. \tag{4b}$$

The fluxes $\left[\partial_x w\right]_j^{n+1/2}$ are not yet defined, but they need to be consistent with $\partial_x w(t^{n+1/2},x_j)$ at time $t^{n+1/2}$ and position x_j. The main idea of the ImEx approach consists in isolating the part corresponding to the fast dynamics hidden in the term $\left[\partial_x F\right]_j^{n+1/2}$ and choosing it implicit in the scheme. To do so, we formally differentiate (1b) and (1a) with respect to space and construct another scheme on the fluxes:

$$\frac{\left[\partial_x F\right]_j^{n+1/2} - \left[\partial_x F\right]_j^n}{\Delta t/2} = -\frac{1}{\varepsilon}\left[\partial_{xx} E\right]_j^n - \frac{\sigma}{\varepsilon^2}\left[\partial_x F\right]_j^{n+1/2}, \tag{5a}$$

$$\frac{\left[\partial_x E\right]_j^{n+1/2} - \left[\partial_x E\right]_j^n}{\Delta t/2} = -\frac{1}{\varepsilon}\left[\partial_{xx} F\right]_j^n, \tag{5b}$$

which provides the flux terms $\left[\partial_x w\right]_j^{n+1/2}$ as functions of the explicit terms in brackets $[w]_j^n$ which still need to be defined too. A first order scheme based on this approach was shown to provide a control on the fast dynamics in equation (4a), see [17]. The term $F_j^{n+1/2}$ still remain to be chosen. It was shown in [17] that the midpoint rule does not ensure the desired stability in the resulting scheme. Instead a reverse Runge-Kutta method (see [17, 20, 21, 22, 23]) is used:

$$\begin{aligned} F_j^{n+1/2} &= F_j^{n+1} - \frac{\Delta t}{2}\left[\partial_t F\right]_j^{n+1} \\ &= F_j^{n+1} + \frac{\Delta t}{2\varepsilon}\left(\left[\partial_x E\right]_j^{n+1/2} + \frac{\sigma}{\varepsilon}F_j^{n+1}\right) \\ &= \left(1 + \frac{\sigma\Delta t}{2\varepsilon^2}\right)F_j^{n+1} + \frac{\Delta t}{2\varepsilon}\left[\partial_x E\right]_j^{n+1/2}. \end{aligned} \tag{6}$$

The main difference in the construction compared to [17] is the choice of fixing the flux term $\left[\partial_x E\right]_j^{n+1/2}$ at time $t^{n+1/2}$ in the definition of $F_j^{n+1/2}$. As illustrated below, this choice also leads to a second order accuracy but it impacts the stability property and the implementation of the scheme. Injecting the solution of (5a) into (4a) and injecting both (5b) and (6) into (4b) provides:

$$\frac{E_j^{n+1}-E_j^n}{\Delta t} + \frac{M_1}{\varepsilon}\left[\partial_x F\right]_j^n - \frac{M_1\Delta t}{2\varepsilon}\left[\partial_{xx} E\right]_j^n = 0, \qquad M_1 = \frac{1}{1+\frac{\sigma\Delta t}{2\varepsilon^2}}, \tag{7a}$$

$$\frac{F_j^{n+1}-F_j^n}{\Delta t} + \frac{M_2}{\varepsilon}\left[\partial_x E\right]_j^n - \frac{M_2\Delta t}{2\varepsilon}\left[\partial_{xx} F\right]_j^n = -\frac{\sigma M_2}{\varepsilon^2}F_j^n, \qquad M_2 = \frac{1+\frac{\sigma\Delta t}{2\varepsilon^2}}{1+\frac{\sigma\Delta t}{\varepsilon^2}\left(1+\frac{\sigma\Delta t}{2\varepsilon^2}\right)}. \tag{7b}$$

Finally, we choose the explicit spatial derivative terms using centered differences to obtain the scheme:

$$\frac{E_j^{n+1}-E_j^n}{\Delta t} + \frac{M_1}{\varepsilon}\frac{F_{j+1}^n - F_{j-1}^n}{2\Delta x} - \frac{M_1\Delta t}{2\varepsilon^2}\frac{E_{j+1}^n - 2E_j^n + E_{j-1}^n}{\Delta x^2} = 0, \tag{8a}$$

$$\frac{F_j^{n+1}-F_j^n}{\Delta t} + \frac{M_2}{\varepsilon}\frac{E_{j+1}^n - E_{j-1}^n}{2\Delta x} - \frac{M_2\Delta t}{2\varepsilon^2}\frac{F_{j+1}^n - 2F_j^n + F_{j-1}^n}{\Delta x^2} = -M_2\frac{\sigma}{\varepsilon^2}F_j^n. \tag{8b}$$

NUMERICAL ANALYSIS OF THE METHOD

We now proceed to the numerical analysis of the method, that is we study the stability and the accuracy in each regime.

Stability Properties

In order to study the stability of the method, we diagonalize the convective part of equations (8), yielding the variables $u = \sqrt{M_2}E + \sqrt{M_1}F$ and $v = \sqrt{M_2}E - \sqrt{M_1}F$. These variables tend to the Riemann invariant of the HHE (1), $\tilde{u} = E + F$ and $\tilde{v} = E - F$, in the limit $\Delta t \ll \varepsilon$. We can then rewrite the scheme in terms of this two variables to obtain:

$$u_j^{n+1} = \lambda_1 u_j^n + \lambda_2 u_{j+1}^n + \lambda_3 u_{j-1}^n + \lambda_4 v_{j+1}^n + \lambda_5 v_j^n + \lambda_4 v_{j-1}^n$$
$$v_j^{n+1} = \lambda_1 v_j^n + \lambda_2 v_{j+1}^n + \lambda_3 v_{j-1}^n + \lambda_4 u_{j+1}^n + \lambda_5 u_j^n + \lambda_4 u_{j-1}^n$$

with $\lambda_1 = \left(1 - \frac{M_+\Delta t^2}{\varepsilon^2\Delta x^2} - \frac{M_2\sigma\Delta t}{2\varepsilon^2}\right)$, $\lambda_2 = \left(\frac{M_+\Delta t^2}{2\varepsilon^2\Delta x^2} - \frac{\tilde{M}\Delta t}{2\varepsilon\Delta x}\right)$, $\lambda_3 = \left(\frac{M_+\Delta t^2}{2\varepsilon^2\Delta x^2} + \frac{\tilde{M}\Delta t}{2\varepsilon\Delta x}\right)$, $\lambda_4 = \frac{M_-\Delta t^2}{2\varepsilon^2\Delta x^2}$, $\lambda_5 = \left(\frac{M_2\sigma\Delta t}{2\varepsilon^2} - \frac{M_-\Delta t^2}{\varepsilon^2\Delta x^2}\right)$ where $\tilde{M} = \sqrt{M_1 M_2}$ and $M_\pm = (M_1 \pm M_2)/2$.

Proposition 1 *The scheme* (8) *is l^∞-stable on the variables (u,v) under the constraint*

$$\varepsilon\Delta x =: \Delta t_{min} \le \Delta t \le \Delta t_{max} := \frac{\sigma\Delta x^2}{4}\frac{1+\sqrt{1+2\left(\frac{4\varepsilon}{\Delta x}\right)^2}}{2}.$$

The scheme (8) *is l^2-stable under the constraint $\Delta t \le \Delta t_{max}$.*

Proof: Remark that $\lambda_1 + \lambda_2 + \lambda_3 + 2\lambda_4 + \lambda_5 = 1$ and therefore if $\lambda_i \ge 0$ for all $i \in \{1,2,3,4,5\}$, then the quantities u_j^{n+1} and v_j^{n+1} at time t^{n+1} are convex combinations of those at time t^n, and we obtain a discrete maximum principle and l^∞ stability.

One verifies from their definition that the coefficients λ_3 and λ_4 are always non-negative.

Following [17], after a few simplifications, the condition $\lambda_1 \ge 0$ is satisfied if $\Delta t \le \Delta t_{max}$. This bound is sharp in the regime $\varepsilon \ll \Delta t$ but not in the regime $\Delta t \ll \varepsilon$.

Remarking that $M_+ \ge \tilde{M}$, the condition $\lambda_2 \ge 0$ is satisfied whenever $\Delta t \ge \Delta t_{min}$.

Finally, straightforwards considerations show the condition $\lambda_5 \ge 0$ is always met as long as $\Delta t_{min} \le \Delta t \le \Delta t_{max}$.

The l^2-stability follows for $\Delta t_{min} \le \Delta t \le \Delta t_{max}$ and can be straightforwardly derived when $\Delta t < \Delta t_{min}$. ■

In the regime $\Delta t \ll \varepsilon$, the two bounds Δt_{min} and Δt_{max} may collide. However, one verifies that $M_+ \le M_1$ and $M_2 \le M_1$, which provides

$$\lambda_1 \ge 1 - \frac{M_1\Delta t^2}{\varepsilon^2\Delta x^2} - \frac{M_1\sigma\Delta t}{2\varepsilon^2},$$

which is non-negative if $\Delta t = \varepsilon\Delta x$. As a result, choosing $\Delta t = \max(\Delta t_{min}, \Delta t_{max})$ always ensures l^∞ stability for the u and v variables, which translate into strong stability properties for E and F as we will numerically demonstrate in the third section.

Accuracy Properties

The accuracy study is conducted via consistency error computations and we denote the consistency errors $c_j^n(w)$ as the error obtained by replacing the quantities w_j^n by the exact value $w(t^n, x_j)$ in (8).

Proposition 2 *The consistency errors of the scheme* (8) *satisfy*

$$c_j^n(E) = \mathcal{O}\left(\Delta t^2\right) + \mathcal{O}\left(\frac{M_1}{\varepsilon}\Delta t^2\right) + \mathcal{O}\left(\frac{M_1}{\varepsilon^2}\Delta t^2\right) + \mathcal{O}\left(\frac{M_1}{\varepsilon}\Delta x^2\right) + \mathcal{O}\left(M_1\Delta t\frac{\Delta x^2}{\varepsilon^2}\right), \tag{9a}$$

$$c_j^n(F) = \mathcal{O}\left(\Delta t^2\right) + \mathcal{O}\left(\frac{M_2\Delta x^2}{\varepsilon}\right) + \mathcal{O}\left(\frac{M_2\Delta t}{2\varepsilon^2}\Delta x^2\right). \tag{9b}$$

Proof: Assuming $F_j^n = F(t^n, x_j)$, $E_j^n = E(t^n, x_j)$ and using Taylor expansions provides

$$\begin{aligned}\partial_x F(t^{n+1/2}, x_j) - [\partial_x F]_j^{n+1/2} &= M_1 \left(\partial_x F(t^{n+1/2}, x_j) - [\partial_x F]_j^n + \frac{\Delta t}{2\varepsilon} [\partial_{xx} E]_j^n + \frac{\sigma \Delta t}{2\varepsilon^2} \partial_x F(t^{n+1/2}, x_j) \right) \\ &= \frac{M_1 \Delta t}{2} \left(\partial_{tx} F(t^{n+1/2}, x_j) + \frac{1}{\varepsilon} \partial_{xx} E(t^{n+1/2}, x_j) + \frac{\sigma}{\varepsilon^2} \partial_x F(t^{n+1/2}, x_j) \right) \\ &\quad + M_1 \left(\mathscr{O}\left(\Delta t^2\right) + \mathscr{O}\left(\frac{\Delta t^2}{\varepsilon}\right) + \mathscr{O}\left(\Delta x^2\right) + \mathscr{O}\left(\Delta t \frac{\Delta x^2}{\varepsilon}\right) \right).\end{aligned}$$

Eventually, we obtain (9a). Similarily, Taylor expansions provide

$$\begin{aligned}c_j^n(F) &= \frac{F_j^{n+1} - F_j^n}{\Delta t} + \frac{M_2}{\varepsilon} \frac{E_{j+1}^n - E_{j-1}^n}{2\Delta x} - \frac{M_2 \Delta t}{2\varepsilon^2} \frac{F_{j+1}^n - 2F_j^n + F_{j-1}^n}{\Delta x^2} + M_2 \frac{\sigma}{\varepsilon^2} F_j^n \\ &= \partial_t F_j^{n+1/2} + \frac{M_2}{\varepsilon} \partial_x E_j^n - \frac{M_2 \Delta t}{2\varepsilon^2} \partial_{xx} F_j^n + \frac{\sigma M_2}{\varepsilon^2} F_j^n + \mathscr{O}\left(\Delta t^2\right) + \mathscr{O}\left(\frac{M_2 \Delta x^2}{\varepsilon}\right) + \mathscr{O}\left(\frac{M_2 \Delta t}{2\varepsilon^2} \Delta x^2\right),\end{aligned}$$

and eventually (9b). ■

Concerning these errors, we note that:

- The terms $\mathscr{O}\left(\Delta t^2\right)$ are independent of ε and is second order as desired.
- The second term $\mathscr{O}\left(\frac{M_1}{\varepsilon} \Delta t^2\right)$ in (9a) is always less restrictive than the third one $\mathscr{O}\left(\frac{M_1}{\varepsilon^2} \Delta t^2\right)$.
- In the regime $\Delta t \leq \varepsilon \Delta x$ then $\frac{M_1}{\varepsilon^2} \Delta t^2 \leq \Delta x^2$ and we obtain the desired second order accuracy. In the regime $\Delta t \geq \varepsilon \Delta x$, then $M_1 \leq 1/(1 + \Delta x/\varepsilon)$ and under the condition $\Delta t \leq \Delta t_{max}$ we have

  $$\mathscr{O}\left(\frac{M_1}{\varepsilon^2} \Delta t^2\right) = \mathscr{O}\left(\frac{1}{\varepsilon^2} \frac{1}{1 + \frac{\Delta x \sqrt{\Delta x^2 + \varepsilon^2}}{\varepsilon^2}} \Delta x^2 \left(\Delta x^2 + \varepsilon^2\right)\right) = \mathscr{O}\left(\frac{\Delta x^2 + \varepsilon^2}{\varepsilon^2 + \Delta x \sqrt{\Delta x^2 + \varepsilon^2}} \Delta x^2\right) = \mathscr{O}\left(\Delta x^2\right).$$

  This argument relies on the fact that the function $x \mapsto x^2/(1+x)$ is strictly increasing on $\mathbb{R}_+$.
- The last term $\mathscr{O}\left(M_1 \Delta t \frac{\Delta x^2}{\varepsilon^2}\right)$ in (9a) is also second order since we have $\frac{M_1 \Delta t}{\varepsilon^2} \leq 1/\sigma$. Idem for the third term in (9b).
- The term $\mathscr{O}\left(\frac{M_1}{\varepsilon} \Delta x^2\right)$ in (9a) is the most problematic. Indeed, if $\Delta t \to 0$ then $M_1 \to 1$ and we obtain an error of the form $\mathscr{O}\left(\Delta x^2/\varepsilon\right)$ which is not consistent in the limit $\varepsilon \to 0$. The only way to guarantee the consistency is to assume that the numerical solution stays close to the equilibrium manifold where F and all its derivatives are of order $\mathscr{O}(\varepsilon)$, then the term is of order $\mathscr{O}\left(\Delta x^2\right)$.
- The second term in (9b) is of the form $\mathscr{O}\left(\frac{M_2 \Delta x^2}{\varepsilon}\right)$. It will not be compensated in the regime where $\varepsilon \leq \Delta x$ and $\Delta t \ll \varepsilon$, leading to accuracy loss. A typical example would be $\Delta x = \varepsilon$. Nonetheless, this notation conceals any constant that is not explicitly dependent on Δt, Δx or ε and we will show in the next section, that second order accuracy is observed in practice and maintained throughout all regimes, albeit at a lower overall accuracy than the scheme developed in [17].

NUMERICAL VALIDATION OF THE METHOD

In this section we consider three test cases: one with an analytical solution from which we can perform a convergence study, one Riemann problem to test the robustness of the method with respect to shocks in the hyperbolic regime $\Delta t \ll \varepsilon$, and finally one with a steady state solution to assess the ability of the method to correctly capture such states.

Convergence Study

For the exact test case, we consider the HHE complemented with Dirichlet conditions E_L and E_F imposed on E respectively on the left and right boundaries. One verifies that the following functions are a solution of this problem

$$E(t,x) = f(t)\,g(x) + \frac{E_R - E_L}{x_R - x_L}(x - x_L) + E_L,$$
$$F(t,x) = \varepsilon f'(t) G(x) - \frac{\varepsilon}{\sigma}(E_R - E_L),$$

for $t \geq 0$, $x \in [x_L, x_R]$, where:

$$f(t) = \alpha \frac{\lambda_+ e^{\lambda_- t} - \lambda_- e^{\lambda_+ t}}{\lambda_+ - \lambda_-} + \beta \frac{e^{\lambda_+ t} - e^{\lambda_- t}}{\lambda_+ - \lambda_-}, \quad \lambda_\pm = -\frac{\sigma}{2\varepsilon^2}\left(1 \mp \sqrt{1 - \left(\frac{2\pi\varepsilon}{\sigma}\right)^2}\right),$$
$$g(x) = \sin(\pi(x - x_L)), \quad G(x) = \frac{\cos(\pi(x - x_L))}{\pi}.$$

We fix $\beta = -\frac{\pi^2}{\sigma}\alpha$ such that $\partial_t F(t=0,x) = \mathcal{O}(\varepsilon)$, we use $\alpha = \sigma = 1$ for simplicity. Writing $e_n(w) = w - \bar{w}$ the vector of errors between the exact solution and the numerical one obtained with (4) at the points (t^n, x_j), Figure 1 presents the l^∞-global error $\|e_n(w)\|_\infty$ as a function of Δx for three values of ε. The time step is chosen to be $\Delta t = 0.9\Delta t_{\max}$.

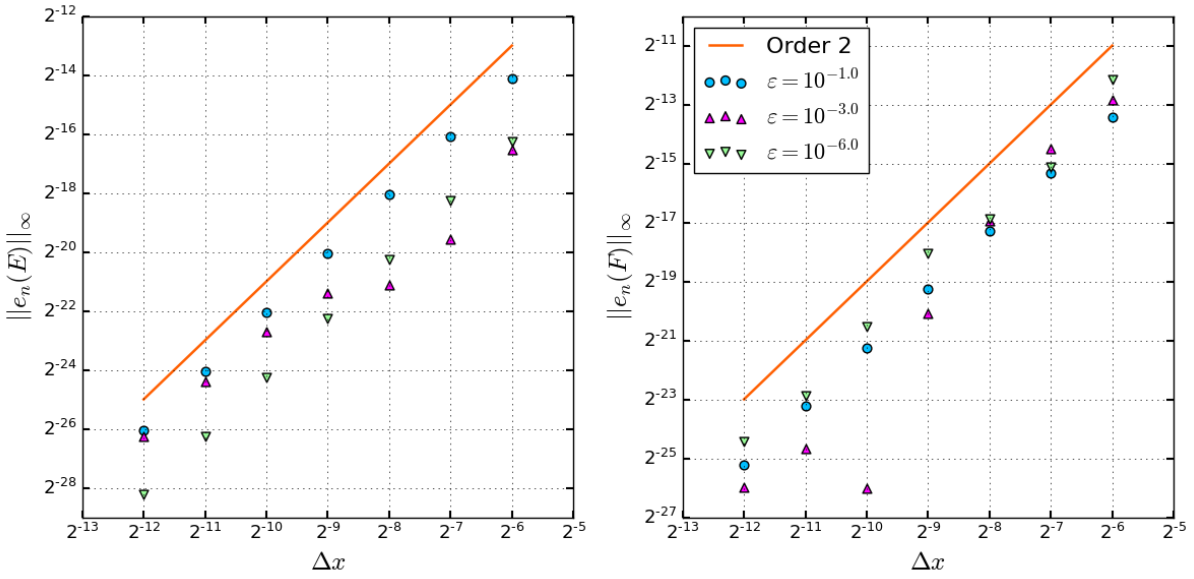

FIGURE 1. l^∞-global errors on E and F obtained with the scheme (8) with $\varepsilon = 10^{-1}$, 10^{-3} and 10^{-6} as functions of Δx.

The scheme shows second order accuracy independently of the parameter ε. As a comparison, we implemented Strang splitting with Lax-Wendroff for the convective part. The convergence curves are displayed on Figure 2.

The error obtained with this naive discretization is not independent of the parameter ε. In practice, the method is excessively diffusive when $\varepsilon \ll \Delta x$. We attract the attention of the reader on the fact that the considered ε are in the worst case 10^{-3} compared to 10^{-6} for the present method. Simulations with this naive approach with such a low ε is difficult to perform due to its computational cost.

Finally, as a last point of comparison we consider the ImEx2-ctr-DST method developed in [17]. As displayed on Figure 3, the more evolved and intricate ImEx2-ctr-DST method offers better performances in all regimes, although it has the drawback to be more complex to implement.

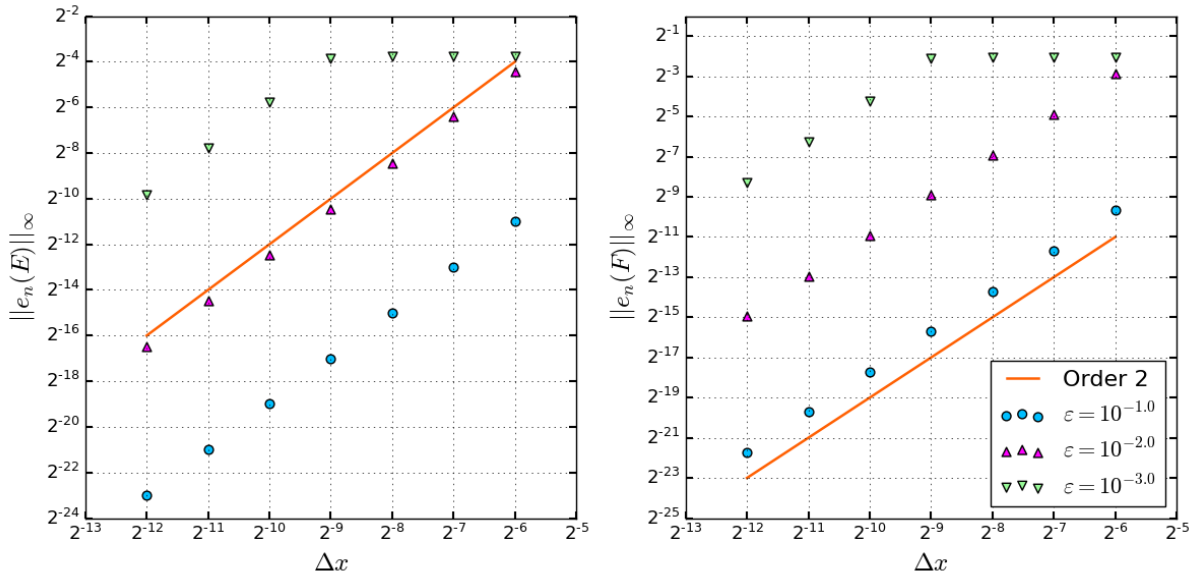

FIGURE 2. l^∞-global errors on E and F obtained using a Strang splitting, with RK2 for the source term and Lax-Wendroff for the convective step with $\varepsilon = 0.1$, 0.01 and 0.001 as functions of Δx.

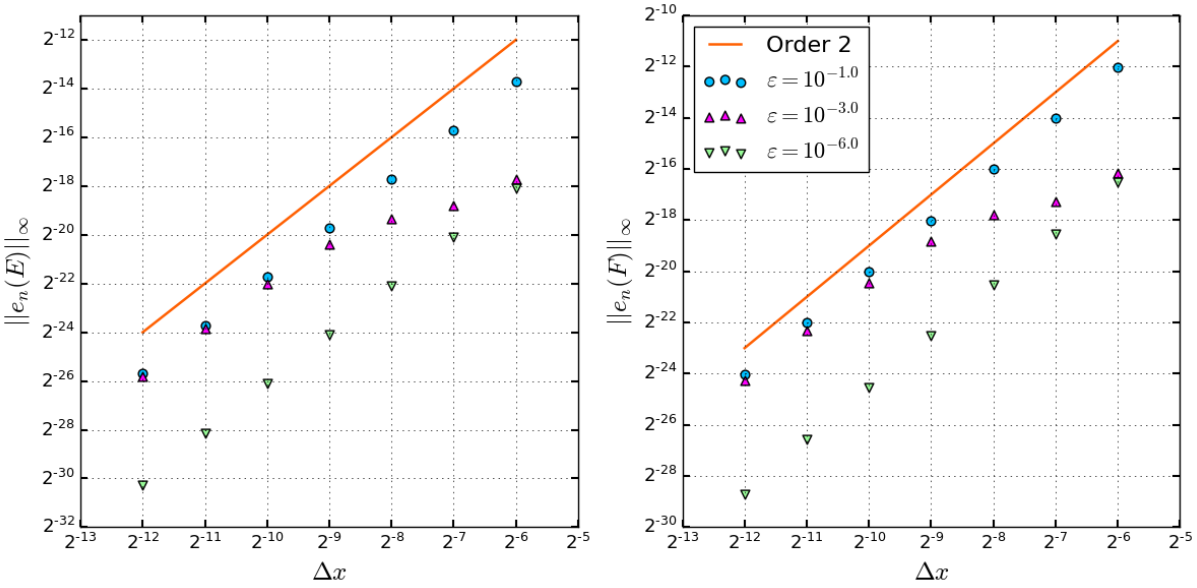

FIGURE 3. l^∞-global errors on E and F obtained using the ImEx2-ctr-DST method from [17] with $\varepsilon = 10^{-1}$, 10^{-3} and 10^{-6} as functions of Δx.

Robustness Study

The second test case is a Riemann problem, with initial condition:

$$E(0,x) = \mathbf{1}_{\left\{x \leq \frac{x_R+x_L}{2}\right\}} E_L + \mathbf{1}_{\left\{x > \frac{x_R+x_L}{2}\right\}} E_R, \quad F(0,x) = 0.$$

This test case is used to validate the stability properties that we have demonstrated in the second section. The present approach is again compared to a MUSCL-Hancock method with minmod limiters coupled with Strang splitting and using RK2 for the source term with the same precisition $N = 64$ and with a fine precision $N = 2048$ as a reference. The numerical approximation of E and F at final time $t_f = 0.15$ with $\sigma = 1$ and $\varepsilon = 0.5$ (hyperbolic regime) are presented on Figure 4.

The present method is not only stable, without requiring the use of limiters, but it is also much more accurate than the MUSCL-Hancock approach. However, the range of acceptable Δt in order not to trigger spurious oscillations is extremely small in this regime, leading in practice to only one possible value for Δt.

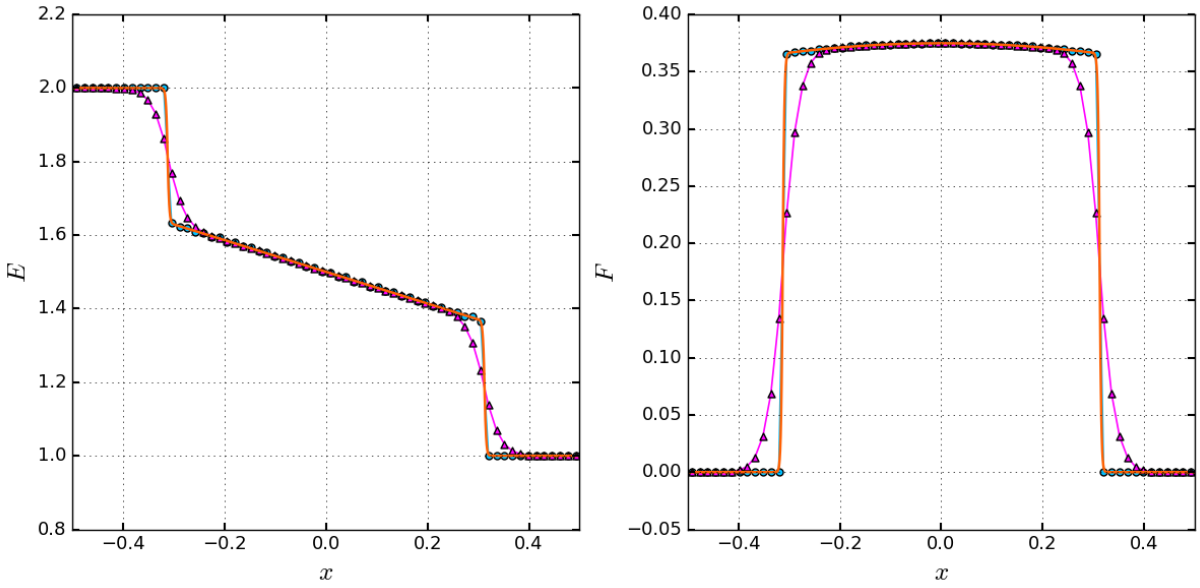

FIGURE 4. Approximated solutions at time $t = 0.15$, with $\sigma = 1$ and $\varepsilon = 0.5$ (hyperbolic regime), computed respectively with the scheme (8) with $\Delta t_{M_2} = \Delta t_{min} = \varepsilon \Delta x$ ($\circ$), with the MUSCL-Hancock method with Strang splitting and RK2 for the source term with $\Delta t_{MH} = 0.9 \min\left(2\varepsilon^2/\sigma, \varepsilon\Delta x/(u_{max}+c)\right)$ using $N = 64$ cells ($\triangle$) and $N = 2048$ cells (solid line) as reference.

Unlike the method of lines, the present approach does not rely on decoupling time and space. This allows to take full advantage of the stability that the source term brings to the system and to achieve l^∞-stability without having recourse to slope limiter which is uncommon for second-order methods. However this is specific to the HHE (1) which has essentially one velocity of propagation of information, and this is not expected to hold for system featuring several speeds of propagation, typically for nonlinear systems. For such system, some form of additional viscosity, for instance via limiters, will be necessary.

With a Non-Constant σ

The last example is a stationary test case with a parameter σ that varies with x. We choose for instance:

$$\sigma(x) = (\sigma_{max} - \sigma_{min}) \frac{1 - \frac{2}{\pi}\arctan\left(\frac{x - \frac{x_c}{2}}{\tau}\right)}{1 - \frac{2}{\pi}\arctan\left(\frac{x_L - \frac{x_c}{2}}{\tau}\right)} + \sigma_{min}.$$

The analytical steady state corresponding to this choice of non constant σ can be computed analytically ([17]). On Figure 5, we compare the accuracy of the present method on this steady state to that of classical approaches.

Although the present approach was not designed specifically to be well-balance, that is to preserve with high accuracy steady states, it appears that is does demonstrate such a property, unlike the classical approach that is more and more inaccurate as $\varepsilon \to 0$ even in a stationary setting.

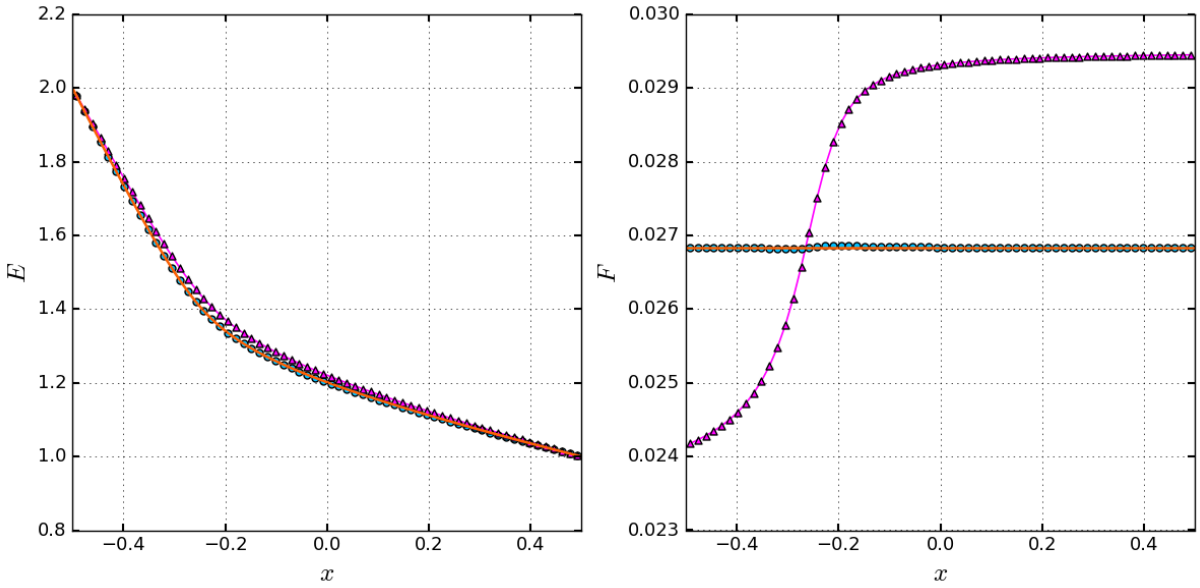

FIGURE 5. Variable E and F obtained at time $t_f = 2$, with non-constant σ and $\varepsilon = 10^{-2}$, respectively with the ImEx2-ctr method (○), $\Delta t_{M_2} = 0.9\Delta t_{max}$ and the MUSCL-Hancock method with Strang splitting, Reverse RK2 (△), $\Delta t_{MH} = 0.9\min\left(2\varepsilon^2/\sigma_{max}, \varepsilon\Delta x/(u_{max}+c)\right)$, using $N = 64$ cells, with the steady state solution used as reference (solid line).

DISCUSSION ON A POSSIBLE EXTENSION TO THE ISOTHERMAL EULER EQUATIONS WITH FRICTION

In this last section, we discuss how to extend the method to more general systems with stiff collisional source terms, and in particular we consider the Euler-friction equation, that reads:

$$\partial_t \boldsymbol{w} + \frac{1}{\varepsilon}\partial_x \boldsymbol{f}(\boldsymbol{w}) = \frac{\sigma}{\varepsilon^2}\boldsymbol{S}(\boldsymbol{w}) \tag{11}$$

$$\boldsymbol{w} = \begin{pmatrix} \rho \\ \rho u \end{pmatrix}, \quad \boldsymbol{f}(\boldsymbol{w}) = \begin{pmatrix} \rho u \\ \rho u^2 + p(\rho) \end{pmatrix}, \quad \boldsymbol{S}(\boldsymbol{w}) = B\boldsymbol{w}, \quad B = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}. \tag{12}$$

Here we have chosen for simplcity the isothermal case so that $p(\rho) = c^2\rho$ with $c > 0$. The crucial properties here, when aiming further at extending the method to other systems, is that the source is linear and the flux homogeneous, i.e. $\boldsymbol{f}(\boldsymbol{w}) = A(\boldsymbol{w})\boldsymbol{w}$ where $A(\boldsymbol{w})$ is the Jacobian of $\boldsymbol{f}$.

Following a recipe similar as in [17], the present second order approach extends into:

$$\frac{\boldsymbol{w}_j^{n+1} - \boldsymbol{w}_j^n}{\Delta t} + \frac{1}{\varepsilon}\left[\partial_x \boldsymbol{f}(\boldsymbol{w})\right]_j^{n+1/2} = -\frac{\sigma_j}{\varepsilon^2} B\boldsymbol{w}_j^{n+1/2}, \tag{13}$$

We first devise the term $\left[\partial_x \boldsymbol{f}(\boldsymbol{w})\right]_j^{n+1/2}$. Assuming sufficient regularity, we have:

$$\partial_t \boldsymbol{f}(\boldsymbol{w}) = A(\boldsymbol{w})\partial_t \boldsymbol{w} = -\frac{1}{\varepsilon}A(\boldsymbol{w})\partial_x \boldsymbol{f}(\boldsymbol{w}) - \frac{\sigma}{\varepsilon^2}A(\boldsymbol{w})B\boldsymbol{w}.$$

Formally, the identity $\left[\boldsymbol{f}(\boldsymbol{w})\right]_j^{n+1/2} = \left[\boldsymbol{f}(\boldsymbol{w})\right]_j^n + \frac{\Delta t}{2}\left[\partial_t \boldsymbol{f}(\boldsymbol{w})\right]_j^n$ is rewritten:

$$\left[\boldsymbol{f}(\boldsymbol{w})\right]_j^{n+1/2} = \left[\boldsymbol{f}(\boldsymbol{w})\right]_j^n - \frac{\Delta t}{2\varepsilon}\left[\left(A(\boldsymbol{w})\partial_x \boldsymbol{f}(\boldsymbol{w})\right)\right]_j^n - \frac{\Delta t\sigma_j}{2\varepsilon^2}\left[\left(A(\boldsymbol{w})B\boldsymbol{w}\right)\right]_j^*, \tag{14}$$

where the flux term is chosen explicit and the source term is chosen of the form:

$$[A(\boldsymbol{w})B\boldsymbol{w}]_j^* = A(\boldsymbol{w}_j^n)B\boldsymbol{w}_j^*, \tag{15}$$
$$\boldsymbol{w}_j^* = \boldsymbol{w}_j^n - \frac{\Delta t}{2\varepsilon}\left[\partial_x \boldsymbol{f}(\boldsymbol{w})\right]_j^n - \frac{\Delta t \sigma_j}{2\varepsilon^2} B\boldsymbol{w}_j^* = I_{M_1}\left(\boldsymbol{w}_j^n - \frac{\Delta t}{2\varepsilon}\left[\partial_x \boldsymbol{f}(\boldsymbol{w})\right]_j^n\right),$$

with $I_{M_1} = Diag(1, M_1)$ and $M_1 = 1/(1+\sigma\Delta t/\left(2\varepsilon^2\right))$ as in (8). By injection we then obtain:

$$\left[\partial_x \boldsymbol{f}(\boldsymbol{w})\right]_j^{n+1/2} = \left[\partial_x \boldsymbol{f}_{M_1}(\boldsymbol{w})\right]_j^n - \frac{\Delta t}{2\varepsilon}\left[\partial_x\left(A_{M_1}(\boldsymbol{w})\partial_x \boldsymbol{f}(\boldsymbol{w})\right)\right]_j^n, \tag{16a}$$
$$A_{M_1}(\boldsymbol{w}) = A(\boldsymbol{w})I_{M_1}, \qquad \boldsymbol{f}_{M_1}(\boldsymbol{w}) = A_{M_1}(\boldsymbol{w})\boldsymbol{w}, \tag{16b}$$

which, applied to (12), provides:

$$\boldsymbol{f}_{M_1}(\boldsymbol{w}) = \begin{pmatrix} M_1 \rho u \\ (2M_1 - 1)\rho u^2 + p(\rho) \end{pmatrix}, \tag{16c}$$
$$A_{M_1}(\boldsymbol{w})\partial_x \boldsymbol{f}(\boldsymbol{w}) = \begin{pmatrix} M_1 \partial_x\left(\rho u^2 + p(\rho)\right) \\ 2uM_1\partial_x\left(\rho u^2 + p(\rho)\right) + \left(\left(p'(\rho)\right)^2 - u^2\right)\partial_x(\rho u) \end{pmatrix}. \tag{16d}$$

Concerning the source term, we use again the Reverse Runge-Kutta method as in (6):

$$\boldsymbol{w}_j^{n+1/2} = \boldsymbol{w}_j^{n+1} + \frac{\Delta t}{2\varepsilon}\left(\left[\partial_x \boldsymbol{f}(\boldsymbol{w})\right]_j^{n+1/2} + \frac{\sigma_j}{\varepsilon} B\boldsymbol{w}_j^{n+1}\right). \tag{17}$$

Combining (16a) and (17) one finally obtain the scheme:

$$\frac{\boldsymbol{w}_j^{n+1} - \boldsymbol{w}_j^n}{\Delta t} + \frac{1}{\varepsilon}\left[I_{M_2}\partial_x \boldsymbol{f}_{M_1}(\boldsymbol{w})\right]_j^n - \frac{\Delta t}{\varepsilon^2}\left[I_{M_2}\partial_x\left(A_{M_1}\partial_x \boldsymbol{f}(\boldsymbol{w})\right)\right]_j^n = -\frac{\sigma_j M_2}{\varepsilon^2} B\boldsymbol{w}_j^n. \tag{18}$$

Then the spacial terms are discretized using centered difference, with the centered velocity u that appears in (16c) and (16d) computed via Roe average.

This scheme is virtually as easy to derive and implement than the first order scheme presented in [17], a significant improvement as compared to the ImEx2-ctr-DST method that is much more involved.

This scheme was implemented and successfully tested in the diffusive ($\varepsilon \ll \Delta x$) and intermediate ($\varepsilon \sim \Delta x$) regimes. In the hyperbolic regime however, the method used to stabilized ImEx-ctr-DST did not work for the present method. Additional effort is required to make the method viable in this regime, in the sense that it should be robust to shocks. As future work, we plan to derive limiters adapted to this case to solve this difficulty.

CONCLUSION

To conclude, we have derived a second order asymptotic preserving method that is significantly more simple to devise and to implement than methods with similar performances presented in previous work [17]. The stability constraint of the method does not go to 0 when $\varepsilon \to 0$, and for a well chosen range of values for the time step Δt the method even exhibits l^∞ stability properties, which was tested and confirmed numerically. Regarding accuracy, the method can theoretically lose an order of convergence in the intermediate regime $\varepsilon \sim \Delta x$, but was shown in practice to be of uniform second order accuracy in all regimes, making it a competitive method. The method can be straightforwardly extended to more complex sets of equations with stiff linear source terms respecting the same scaling as the one presented in introduction. However, additional work is still necessary in order to make these extensions robust to shocks in the hyperbolic regime. Extension to multidimensional problems are also considered.

ACKNOWLEDGMENTS

We acknowledge the precious help of Loïc Gouarin and the use of the samurai code he develops (`https://github.com/hpc-maths/samurai`) within the framework of Initiative HPC@Maths (PI M. Massot and L. Gouarin - `https://initiative-hpc-maths.gitlab.labos.polytechnique.fr/site/`).

The first author of this proceeging is funded by a joint PhD grant from the French Ministry of Defence (AID - Agence Innovation Défense) and the Région Île-de-France (DIM MathInnov).

REFERENCES

1. P. Vernotte, C. R. Acad. Sci. Paris **246**, 3154–3155 (1958), and 1948 - Volume 227 - pages 43 and 114.
2. C. Cattaneo, Atti del Seminario Matematico e Fisico dell'Università di Modena **3**, 3–21 (1948).
3. C. Cattaneo, C. R. Acad. Sci. Paris **247**, 431–433 (1958).
4. J. C. Maxwell, Philos. Trans. A: Math. Phys. Eng. Sci. **157**, 49–88 (1867).
5. S. Jin, Rivista di Matematica della Universita di Parma **3**, 177–216 (2012).
6. R. E. Caflisch, S. Jin, and G. Russo, SIAM J. Numer. Anal. **34**, 246–281 (1997).
7. L. Gosse and G. Toscani, C. R. Acad. Sci. Paris **334**, 337–342 (2002).
8. C. Buet, B. Després, and E. Franck, Numer. Math. **122**, 227–278 (2012).
9. S. Jin and C. D. Levermore, J. Comput. Phys. **126**, 449–467 (1996).
10. C. Berthon and R. Turpault, Numer. Methods Partial Differ. Equ. **27**, 1396–1422 (2011).
11. C. Chalons and R. Turpault, Numer. Methods Partial Differ. Equ. **35**, 1538–1561 (2019).
12. S. Jin, L. Pareschi, and G. Toscani, SIAM J. Numer. Anal. **35**, 2405–2439 (1998).
13. S. Boscarino, L. Pareschi, and G. Russo, SIAM J. Sci. Comput. **35**, A22–A51 (2013).
14. A. Klar, SIAM J. Numer. Anal. **35**, 1073–1094 (1998).
15. S. Jin, L. Pareschi, and G. Toscani, SIAM J. Numer. Anal. **38**, 913–936 (2000).
16. G. Albi, G. Dimarco, and L. Pareschi, SIAM J. Sci. Comput. **42**, A2402–A2435 (2020).
17. L. Reboul, T. Pichard, and M. Massot, arXiv preprint arXiv:2205.09993 (2022).
18. A. Alvarez-Laguna, T. Pichard, T. Magin, P. Chabert, A. Bourdon, and M. Massot, J. Comput. Phys (2020).
19. L. Reboul, M. Massot, and A. Alvarez-Laguna, https://arxiv.org/pdf/2212.14590.pdf (2023).
20. J. C. Butcher, *Numerical methods for ordinary differential equations*, 3rd ed. (John Wiley & Sons, Ltd., Chichester, 2016) pp. xxiii+513, with a foreword by J. M. Sanz-Serna.
21. E. Hairer, S. P. Nørsett, and G. Wanner, *Solving ordinary differential equations. I*, 2nd ed., Springer Series in Computational Mathematics, Vol. 8 (Springer-Verlag, Berlin, 1993) pp. xvi+528, nonstiff problems.
22. N. N. Kalitkin and I. P. Poshivaylo, Mathematical Models and Computer Simulations **6**, 272–285 (2014).
23. L. M. Skvortsov, Mathematical Models and Computer Simulations **9**, 498–510 (2017).